Cloning and analysis of three new homeobox genes from the nematode *Caenorhabditis elegans*

Dominique Schaller, Claudia Wittmann, Albert Spicher, Fritz Müller and Heinz Tobler
Institute of Zoology, University of Fribourg, Pérolles, CH-1700 Fribourg, Switzerland

Received January 8, 1990; Revised and Accepted February 23, 1990 EMBL accession nos X17075-X17077

ABSTRACT

Three homeobox-containing genes from the nematode *Caenorhabditis elegans* are described. Two of them (*ceh-11* and *ceh-12*) were isolated from a genomic library by hybridization at low stringency with the *Ascaris lumbricoides* homeobox *AHB-1*. The first clone contains a homeobox defining a new class of homeoboxes (*ceh-11*). This gene maps on the third chromosome of *C. elegans*, at the same locus as *egl-5*, a gene already known to be essential for the determination of specific neurons. In the second clone, sequence analysis revealed the existence of the third helix of a putative homeobox (*ceh-12*) which is interrupted by an intron located upstream of the codon for the amino acid 45 of the homeodomain. Using the *ceh-11* homeobox as a probe, a third homeobox (*ceh-13*) was isolated from a cDNA library. As *ceh-13* belongs to the *labial* class of homeoboxes, we conclude that, at the time when the nematode lineage diverged from the myriapod-insect and the vertebrate lineages, the duplication which led to the *Antp* and the *labial* families of homeoboxes had already taken place.

INTRODUCTION

The homeobox was first discovered as a 180 bp long region of homology shared by several homeotic genes in *Drosophila melanogaster* (for review see 1 and 2). Subsequently, homeoboxes were found in many higher eukaryotes including other arthropods, annelids, ascidians, echinoderms, brachiopods, tapeworms, molluscs and chordates, but not at first in coelenterates, nematodes, sponges, flatworms, slime molds, fungi, or bacteria. Only relatively recently have homeobox-containing genes been isolated from the nematodes *Caenorhabditis elegans* (3,4,5,6) and *Ascaris lumbricoides* (Spicher *et al.*, in preparation). Such genes were shown to control the fate of cells during development of *D. melanogaster* and *C. elegans*.

The homeobox codes for a homeodomain which is supposed to bind specifically DNA sequences through an α helix-turn-α helix motif. Recently it has been proven by NMR analysis that the *Antennapedia* homeodomain adopts this conformation in solution (7). Sequence-specific binding has been shown *in vitro* for several *D. melanogaster* homeodomain-containing proteins (8,9,10) and for the murine *Hox-1.5* homeodomain (11).

A complete cell-by-cell description of the development of *C. elegans* has been achieved (12), and now a library of ordered cosmid clones representing the entire genome is brought near to completion (13,14). Therefore, it is of great interest to search for homeobox-containing genes as potential developmental regulatory genes in *C. elegans*. Mapping the putative homeobox loci to the physical contig map of the genome will allow the subsequent identification of potentially corresponding genes from the correlated genetic map.

In this paper we report the cloning and sequence analysis of three putative homeobox-containing genes from *C. elegans*. One of them, *ceh-11*, maps at the same locus as *egl-5*, a gene already known to be essential for the correct determination of certain neurons and other cells.

MATERIAL AND METHODS

Libraries and probes

The genomic clones were isolated from a library prepared with *C. elegans* DNA partially digested by *Eco*RI and ligated into the *Eco*RI site of λEMBL4 (the genomic library was kindly provided by Drs. Karen L. Bennett and Samuel Ward).

For the screening of the genomic library, we used the 2.3 kb insert of the plasmid pC53 containing the *A. lumbricoides* homeobox *AHB-1* and adjacent sequences (Spicher *et al.*, in preparation). The cDNA library (a gift of Drs. Julie Arhinger and Judith E. Kimble) was derived from total RNA of *C. elegans* eggs and the cDNA was inserted into the *Eco*RI cloning site of λgt10.

Hybridization conditions

Low stringency hybridizations for Southern transfer and library screening were performed as described by McGinnis *et al.* (15). We screened approx. 3×10^4 phages from the genomic library, corresponding to 5 genome equivalents. From the cDNA library, 2×10^6 phages were screened.

High stringency hybridizations of Southern blots were carried out by standard procedures (16).

Sequencing of DNA

Fragments from restriction endonuclease digests of the lambda clones were subcloned into the vector pUC18. Smaller fragments from the plasmid clones were then subcloned into the vector M13mp18 for sequencing by the chain termination method (17).

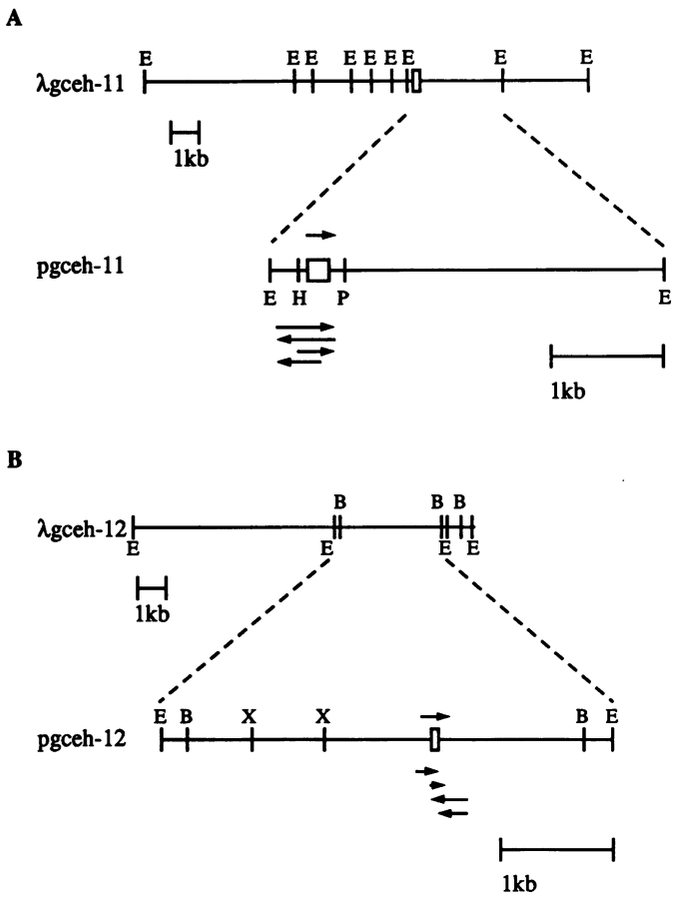

Fig. 1. Restriction maps and sequencing strategy of *ceh-11* (panel A) and *ceh-12* (panel B). Homeoboxes are shown as open boxes. Arrows underneath the homeoboxes represent the sequencing strategy and arrows above the homeoboxes denote the direction of transcription as deduced from the sequence. The *Eco*RI restriction map from λceh-12 is only partial. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pvu*II; X, *Xba*I.

The Klenow fragment as well as Sequenase™ were used for sequencing.

RESULTS AND DISCUSSION

Isolation of two genomic clones containing a homeobox

We had previously shown by Southern hybridization that the *C. elegans* genome contains sequences crosshybridizing at low stringency with an *Antp*-like homeobox from *A. lumbricoides*, *AHB-1* (Spicher *et al.*, in preparation). By screening a *C. elegans* genomic library under the same conditions, we isolated two λ clones (λgceh-11 and λgceh-12). Both phage DNAs were mapped with restriction endonucleases and the position of their homeoboxes was determined by hybridizing the phage DNA to the *A. lumbricoides* homeobox *AHB-1* at low stringency. In each phage, one single *Eco*RI fragment crosshybridized with *AHB-1*. These fragments were subcloned in the plasmid pUC18 (pgceh-11 and pgceh-12). From the insert of each plasmid, the region of homology with *AHB-1* was sequenced (Fig. 1).

λceh-11 contains a homeobox of a new type

In the first clone we found a homeobox (*ceh-11*) of 180 bp, located in an open reading frame of at least 273 bp (cf. Fig. 2). Downstream of the homeobox, the reading frame remains open

ceh-11

```
          tttcagCATCATCAAAGAAAGGCCGTCAAACGTATCAACGCTATCAAACATCA
                S  S  K  K  G  R  Q  T  Y  Q  R  Y  Q  T  S

GTTCTTGAAGCGAAATTCCAACAGTCATCATACGTGTCAAAGAAACAACGTGAAGAGCTC
V  L  E  A  K  F  Q  Q  S  S  Y  V  S  K  K  Q  R  E  E  L

CGCCTGCAGACTCAATTGACAGATCGTCAAATCAAAATCTGGTTCCAAAATCGTCGAATG
R  L  Q  T  Q  L  T  D  R  Q  I  K  I  W  F  Q  N  R  R  M

AAGGCGAAAAAAGAGAAGCAAAGAGTAGATGATCACACGGAACATACTCCACTTCTAC
K  A  K  K  E  K  Q  R  V  D  D  H  T  E  H  T  P  L  L
```

ceh-12

```
tttcagATCAAAATCTGGTTCCAAAATCGTCGAATGAAAAACAAACGTTGCCCATCTTCAACTCCA
      I  K  I  W  F  Q  N  R  R  M  K  N  K  R  C  P  S  S  T  P

ATTCAATCAACTTCCTCTTCTTGATAATCTTTTTATATTTGATCACTTTACTTaataaaT
I  Q  S  T  S  S  S  OpalOchre
```

Fig. 2. Sequence of the *ceh-11* and *ceh-12* homeoboxes and flanking regions. From *ceh-11*, 234 bp are represented, starting from the splice acceptor consensus sequence and including the entire homeobox. From *ceh-12*, 125 bp are shown including the last exon of the homeobox and flanking sequences. The conceptual translation product is represented under the nucleotide sequence. The amino acids of the homeodomain are underlined. Nucleotides from the splice acceptor consensus sequence as well as from the polyadenylation consensus signal are in lower case. Arrows indicate the 5′ end of the putative exons.

for a further 13 codons that have been sequenced. Three basepairs upstream of the homeobox, there is a *C. elegans* splice acceptor consensus sequence, TTTCAG (18). An intron situated just upstream of the homeobox is a feature found in many homeobox-containing genes.

The *ceh-11* homeobox encodes an amino acid sequence which has 34 residues out of 60 identical (57% of identity) with the *Drosophila Antp* homeodomain (19,20). Among more than 100 published sequences, the homeodomain with the highest level of similarity with *ceh-11* is *XlHbox2*, an *Antp* class homeodomain from *X. laevis* (21), with 35 identical residues (58%). As *ceh-11* is only distantly related to the *Antp* class of homeoboxes and does not fit into any other class listed by Scott *et al.* (2), it defines a new family of homeoboxes.

High stringency hybridization of pgceh-11 to a genomic Southern blot shows this sequence to be unique in the *C. elegans* genome. By fingerprinting, J. Sulston and A. Coulson (personal communication) determined the position of the genomic lambda clone on the physical map of the *C. elegans* genome completed to 90−95%. The mapping falls on the third chromosome within the region of *egl-5*. Furthermore, a cosmid containing our entire λ clone has been shown to be able to rescue an *egl-5* mutant (A. Chisholm, personal communication), suggesting the possibility of *ceh-11* being identical with *egl-5*, which is a key gene for the acquisition of the identity of two hermaphrodite-specific motor neurons (HSNs, (22)) and for a variety of other cell fates in the tail. Mutants defective for this gene generate cells that are mispositioned and lack serotonin, a characteristic neurotransmitter of HSNs. Therefore, *egl-5* mutants may be defective in the determination of the HSN fate (23). Interestingly, *mab-5* (4), another homeobox gene, maps very closely to *ceh-11*.

λceh-12 contains the third helix of a putative homeobox

In the sequence of the second clone, we found the third helix of a hypothetical homeobox (*ceh-12*). Out of sixteen amino acids, twelve are identical with *Antp*. Immediately upstream of the codon for the amino acid 45, sequence analysis revealed a *C. elegans* splice acceptor concensus sequence, TTTCAG/R. Further

upstream of this motif, we found an **opal** stop codon. Moreover, the local A+T concentration rises abruptly and no more significant homology with known homeoboxes can be found. We conclude that an intron interrupts the homeobox between the codons for the amino acids 44 and 45. The position of the intron in *ceh-12*, inside the third putative helix, is conserved in six other known homeoboxes, namely one in *C. elegans* (*ceh-2*, (6)) and five in *Drosophila* (*labial* (24), *Distal-less* (25), *proboscipedia* (26), *Abd-B* (27) and *NK-1* (28)). Unlike the situation in other organisms analyzed so far, introns located inside homeoboxes seem to be quite frequent in *C. elegans*. Out of fourteen known homeoboxes from *C. elegans*, eleven have been found to be interrupted by at least one intron (*ceh-2, ceh-3, ceh-4, ceh-5, ceh-6, ceh-7, ceh-8, ceh-14* (6), *mec-3* (3), *unc-86* (5) and *ceh-12*).

The open reading frame containing the homeobox ends with an **opal** stop codon 33 bp downstream of the homeobox. 29 bp further downstream, we found a polyadenylation consensus sequence, AATAAA. We suppose that this position marks the authentic 3′ end of the *ceh-12* coding portion.

High stringency hybridization of pgceh-12 to a genomic Southern blot shows this sequence to be unique in the *C. elegans* genome. The corresponding genomic λ clone maps on the first chromosome (J. Sulston and A. Coulson, personal communication). No genes have yet been found to map at the very same position.

Isolation of a cDNA clone containing a labial-like homeobox and its corresponding genomic clone

Using *ceh-11* as a probe, we screened a cDNA library from eggs of *C. elegans* at low stringency. A positive λ clone was isolated and its insertion subcloned in plasmids. The region crosshybridizing with *ceh-11* was sequenced (Fig. 3), and we found a homeobox (*ceh-13*) coding for a homeodomain which has 33 out of 60 amino acids identical with *Antp* (55%). A computer search shows that *ceh-13* is most similar to homeodomains belonging to the *labial* class, namely to the two murine homeodomains *hox-2.9* (P. Murphy, personal communication) with 44 identical residues (73%) and *hox-1.6* (29) with 43 identical residues (72%), as well as to the *Drosophila* homeodomain *labial* (30) with 41 identical residues (68 %). We conclude therefore that *ceh-13* is evolutionarily related to the *labial* class of homeoboxes.

The similarity with *labial* extends somewhat downstream of the 60th amino acid, the residues 61, 62, 64 and 65 being identical in both homeodomains (Fig. 4). Upstream of the homeobox, no more significant similarity with any known gene could be found.

Using the *ceh-13* homeobox as a probe, we isolated the corresponding genomic clone, λgceh-134. Unfortunately, λgceh-134 maps to a contig which is not yet physically mapped on a *C. elegans* chromosome (J. Sulston and A. Coulson, personal communication). The contig contains a second known homeobox from the *Antp*-class, *ceh-15* (31), which is located very close to *ceh-13*.

CONCLUSIONS

Three loci containing homeoboxes in the genome of *C. elegans* have been analyzed. The three loci coincide with the putative homeobox loci numbers 7 (*ceh-12*), 20 (*ceh-11*) and 38 (*ceh-13*), isolated independently by Bürglin *et al.* ((6) and T. Bürglin, personal communication). Here we demonstrate that these loci indeed contain homeobox sequences.

```
TGCATACAAAACGGTCACAAAGACCAGCTGCCCCAAAAAAGAAGGTGATAGACGAAAATG
M  H  T  K  R  S  Q  R  P  A  A  P  K  K  K  V  I  D  E  N

GAACAAATCGAACCAACTTCACAACACATCAACTGACAGAGCTTGAAAAGGAATTTCACA
G  T  N  R  T  N  F  T  T  H  Q  L  T  E  L  E  K  E  F  H

CAGCAAAATATGTCAATCGAACGAGACGAACTGAAATTGCTTCAAATCTGAAACTGCAGG
T  A  K  Y  V  N  R  T  R  R  T  E  I  A  S  N  L  K  L  Q

AGGCTCAGGTCAAAATCTGGTTTCAAAATCGTCGAATGAAGGAGAAAAAGCGTGAAAAAG
E  A  Q  V  K  I  W  F  Q  N  R  R  M  K  E  K  K  R  E  K

AAAAGGCTATTCCTGGCC
E  K  A  I  P  G
```

Fig. 3. Sequence of the *ceh-13* homeobox and flanking regions. The conceptual translation product is represented below the nucleotide sequence. The amino acids of the homeodomain are underlined.

```
                1.....10.....20.....30.....40.....50.....60
                |     |     |     |     |     |     |     |
Antp            RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKTKGEP
ceh-13          NGTN-TNF-TH-LT-------TAK-VN-T--T---SN-K-Q-A-V----------E--RE-E-AI-
lab             NNS--TNF-NK-LT-------------A------NT-Q-N-T-V----------Q--RV-EGLI-
hox-1.6         PNAV-TNF-TK-LT---------K----A--V---AS-Q-N-T-V----------Q--RE-EGLL-
```

Fig. 4. Alignment of the conceptual translation product of the *ceh-13* homeobox and its flanking sequences with the *lab* and *hox-1.6* homeodomains. Amino acid residues are numbered according to Gehring (1). The *Antp* homeodomain is represented as reference.

All these homeoboxes have a *Gln* at position 9 of their recognition helix. According to the theory proposed by Hanes and Brent (32), our homeoboxes should therefore have an *Antp* type DNA-binding specificity.

The presence in nematodes of *labial*-like (*ceh-13*) as well as *Antp*-like homeodomains (*AHB-1* and *mab-5*) suggests that the common ancestor of the nematode, vertebrate and myriapod-insect lineages, already possessed homeobox-containing genes of both types. Thus, at the time when the nematode lineage diverged from the myriapod-insect and the vertebrate lineages, the gene duplication which led to the *Antp* and the *labial* families of homeoboxes had already taken place. It was suggested that the ancestor of the myriapod-insect lineage contained a *labial*-like and an *Abd-B*-like homeotic gene (33). Until now, an *Abd-B*-like homeobox has not been found in nematodes.

So far, homeoboxes tightly related to *ceh-11* have not been described in any other organism. The question remains whether this type of homeobox is exclusive to nematodes or if it is also present but yet undetected in vertebrates and other invertebrates. We are currently in the process of searching for its homologue in *A. lumbricoides*.

Two loci (*ceh-11* and *ceh-13*) out of three are in close vicinity to already known homeobox-containing genes, suggesting that in *C. elegans* some homeobox-containing genes are organized in 'mini-clusters', comparable to the large clusters of homeoboxes known in vertebrates and insects.

ACKNOWLEDGEMENTS

We thank Drs. John Sulston and Alan Coulson for the mapping of the genomic clones, Drs. Karen L. Bennett and Samuel Ward for the gift of the genomic library and Drs. Julie Ahringer and Judith E. Kimble for the cDNA library. We are also grateful to Drs. Andrew Chisholm, Thomas R. Bürglin and Paula Murphy for communicating results before publication.

We wish to thank Chantal Bulliard and Colette Dedual-

Frassineti for excellent technical assistance. This research was supported by SNF grants Nr 3.113-0.85 and 3.141-0.88.

REFERENCES

1. Gehring,W.J. (1987) Science 236, 1245–1252.
2. Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) Bioch. Bioph. A. 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. Bürglin,T.R.,Finney,M., Coulson,A. and Ruvkun,G. (1989) Nature 341, 239–243.
7. Otting,G., Yan-qui,Q., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1988) EMBO J. 7, 4305–4309.
8. Chen,H.Z., Hoey,T. and Zubay,G. (1988) Mol. Cell. Biochem. 79, 181–189.
9. Desplan,C., Theis,J. and O'Farrell,P.H. (1985) Nature 318, 630–635.
10. Hoey,T. and Levine,M. (1988) Nature 332, 858–861.
11. Fainsod,A., Bogarad,L.D., Ruusala,T., Lubin,M., Crothers,D.M. and Ruddle,F.H. (1988) Proc. Natl. Acad. Sci. U.S.A. 83, 9532–9536.
12. Sulston,J. (1988) in Wood,W.B. (ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp. 123–155.
13. Coulson,A., Sulston,J., Brenner,S. and Karn,J. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 7821–7825.
14. Coulson,A.,Waterston,R., Kiff,J., Sulston,J. and Kohara,Y. (1988) Nature 335, 184–186.
15. McGinnis,W., Garber,R.L., Wirz,J., Kuroiwa,A. and Gehring,W.J. (1984) Cell 37, 403–408.
16. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press. Cold Spring Harbor.
17. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.
18. Emmons,S.W. (1988) in Wood,W.B. (ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp. 47–79.
19. McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) Nature 308, 428–433.
20. Scott,M.P. and Weiner,A.J. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 4115–4119.
21. Müller,M.M., Carrasco,A.E. and De Robertis,E.M. (1984) Cell 39, 157–162.
22. Desai,C., Garriga,G., McIntire,S.L. and Horvitz,H.R. (1988) Nature 336, 638–646.
23. Chalfie,M. and White,J. (1988) in Wood,W.B. (ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp. 337–391.
24. Mlodzik,M., Fjose,A. and Gehring,W.J. (1988) The EMBO Journal 7, 2569–2578.
25. Cohen,S.M., Brönner,G., Küttner,F., Jürgens,G. and Jäckle,H. (1989) Nature 338, 432–434.
26. Diederich,R.J., Merrill,V.K.L., Pultz,M.A. and Kaufman,T.C. (1989) Genes & Development 3, 399–414.
27. Celniker,S.E., Keelan,D.J. and Lewis,E.B. (1989) Genes & Development 3, 1424–1436.
28. Kim,Y. and Nirenberg,M. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 7716–7720.
29. Baron,A., Featherstone,M.S., Hill,R.E., Hall,A., Gaillot,B. and Duboule,D. (1987) EMBO J. 6, 2977–2986.
30. Hoey,T., Doyle,H.J., Harding,K., Wedeen,C. and Levine,M. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 4809–4813.
31. Kamb,A., Weir,M., Rudy,B., Varmus,H. and Kenyon,C. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 4372–4376.
32. Hanes,S.D. and Brent,R. (1989) Cell 57, 1275–1283.
33. Akam,M., Dawson,I. and Tear,G. (1988) Development 104 (suppl.), 123–133.