

Evidence for a new family of evolutionarily conserved homeobox genes

Cathy J. Wedeen*, Richard G. Kostriken, Ichiro Matsumura and David A. Weisblat

Department of Molecular and Cellular Biology, 385 Life Sciences Addition, University of California, Berkeley, CA 94720, USA

Submitted February 12, 1990

EMBL accession no. X51752

The homeodomain is a DNA binding domain about 60 amino acids in length that occurs in many developmental regulatory proteins. Based on their degree of relatedness, homeodomain sequences have been grouped into 10 different families plus some unclassified sequences (1). Using a set of degenerate oligonucleotides (hom3 and en5) designed to anneal to homeoboxes, we employed the polymerase chain reaction (PCR) (2), to amplify a portion of a homeobox from genomic DNA of the leech, *Helobdella triserialis* (*Htr*). The amplified homeobox fragment was used as a probe to screen an *Htr* genomic library, from which positive recombinants were plaque purified. From one of these, *Htr*-A2, the homeobox cross-hybridizing region was sequenced. The corresponding homeodomain is unlike those of any of the previously described 10 families but is quite similar to the X1-HBox8 homeodomain (3) from the frog, *Xenopus laevis* (*Xl*) (shown below). Underlined portions of the nucleic acid sequence are homologous to the oligonucleotide primers used in the PCR amplification. Asterisks signify amino acids in the frog sequence that are identical to those in the leech sequence. In the region compared, 37 of 42 amino acids are identical. In addition, X1-HBox8 and *Htr*-A2 homeodomains have in common a histidine (H) residue (amino acid 44) in the putative recognition helix region (4, 5), which is not present in any other homeodomain. For the region compared, this level of sequence identity surpasses that observed between chordates and annelids

for the *engrailed* gene family (unpublished results), one that is considered well conserved throughout many phyla. This high degree of homology identifies a new family of highly conserved homeobox genes.

ACKNOWLEDGEMENTS

We would like to thank Michael Weir for the generous gift of the degenerate oligonucleotides. C.J.W. was funded by Damon Runyon Cancer Fund Grant, DRG 925. This work was funded by N.I.H. Grant HD23328 to D.A.W. The nucleotide sequence data reported here will appear in EMBL, GenBank and DDBJ Nucleotide Sequence Data Bases under accession number X51752.

REFERENCES

1. Scott, M.P., Tamkun, J.W. and Hartzell, G.W. III (1989) *Biochimica et Biophysica Acta* **989**, 25–48.
2. 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.
3. Wright, C.V.E., Schnegelsberg, P. and DeRobertis, E.M. (1988) *Development* **104**, 787–794.
4. Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., DeRobertis, E.M. and Gehring, W.J. (1984) *Nature* **310**, 70–71.
5. Laughon, A. and Scott, M.P. (1984) *Nature* **310**, 25–31.

```

AAC AAA AGA ACT CGC ACA GCT TAC AGC AGA TCA CAA CTT TTC GAA TTA GAG AAA GAG
TTT CAT TTC GAT AAA TAT ATT TCG AGA CCA AGA AGA GTG GAA CTG GCC TCC TCT CTC
AAT TTG ACT GAA CGA CAT ATT AAG ATC TGG TTT CAA AAC CGT CGA ATG AAA TGG AAA
AAA ATG GAA

```

10 20 30 40 50 60

```

Htr NKRTRTAYSRSQLELEKEFHF*FDKYISRPRRVELASSLNLTERHIKIWFQNR*RMKWK*KME*
X.1. *LL*N******VM******E*

```

```

hom3 GAATTCGA (AG) (TC) TCGA (AG) AA (AG) GA (AG) TT (TC) CAT
en5 TT (AG) TT (TC) TG (AG) AACCAGAT (TC) TT (AGT) AT

```

* To whom correspondence should be addressed