Volume 277, number 1,2, 250–252 FEBS 09287 December 1990

Conservation of *engrailed*-like homeobox sequences during vertebrate evolution

Peter W.H. Holland and Nicola A. Williams

Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK

Received 1 November 1990

The *Drosophila melanogaster* developmental gene *engrailed (en)* is a member of a distinct subfamily of homeobox genes with a wide phylogenetic distribution. Here we report the use of reduced stringency polymerase chain reaction (PCR) to amplify and clone 8 genes related to *en* from 5 vertebrate species, including representatives of the most ancient vertebrate lineages. Nucleotide and deduced amino acid sequence comparisons between mouse, toad, zebrafish, lamprey and hagfish genes reveal extensive evolutionary conservation, and suggests that 2 *en*-like genes have been retained in most vertebrate lineages.

Engrailed; Homeobox; PCR; Molecular evolution; Vertebrate development

1. INTRODUCTION

The mouse and human genomes each contain two homeobox genes, *En-1* and *En-2*, closely related to the *D. melanogaster* segmentation gene *engrailed (en)*. These genes, which presumably arose by gene duplication, have diverged from each other, and from *D. melanogaster en*, in both sequence and regulation [1–6].

Here we report the results of a comparative approach to investigate the pattern of sequence divergence during the evolution of the vertebrate *en*-like gene. The polymerase chain reaction (PCR) was used to amplify and clone *en*-like homeobox genes from 5 vertebrate species, chosen to include representatives of all 3 extant lineages resultant from the deepest vertebrate radiations. Analysis of multiple recombinant clones from each species allowed insight into homeobox gene family evolution, and sequence conservation, during vertebrate radiation.

2.MATERIALS AND METHODS

Total DNA was purified from each species (mouse, *Mus musculus* strain CBA; clawed toad, *Xenopus laevis*; zebrafish, *Brachydanio rerio*; lamprey, *Lampetra planeri*; hagfish, *Myxine glutinosa*) by standard methods [7,8], and dialysed versus 10 mM Tris-HCl, 1 mM EDTA (pH 8) prior to use in amplification reactions.

Using published sequences of *D. melanogaster* and mouse *en*-like genes, two conserved regions were identified: one within and one flanking the homeodomain. Two alternative positive strand oligonucleotide primers and two negative strand primers were designed to complement the encoding DNA sequences, and synthesized on a Milligen Biosearch 7500 DNA synthesizer. Primers A and C utilized inosine to allow base pairing at variable sites; primers B and D included mixed nucleotide redundancies. The primers were: (primer A, positive strand) 5[] GAIAAGCGICCICGCACIGCCTTCAC 3[]; (primer B, positive strand) 5[] GAVAAGCGGCCGCGCACRGCCTTC 3[]; (primer C, negative strand) 5[] TGGTTGTACAGICCCTGIGCCATGAG 3[]; and (primer D, negative strand) 5[] TGGTTGTACAGNCCCTGNGCCATGAG 3[], where N = A,C,G or T; R = A or G; V = A,C or G; I = inosine.

DNA amplification reactions were performed using a Techne PHC-2 programmable dri-block under conditions recommended by the suppliers of the Taq DNA polymerase (Perkin–Elmer Ltd and Promega Biotec). Following DNA amplification and electrophoresis, the major products were purified and cloned into plasmid vectors [9–11]. Following transformation, multiple recombinant clones from each species were selected via blue/white screening and/or restriction enzyme digestion, before sequencing using T7 DNA polymerase and 7-deaza dGTP sequencing mixes (Pharmacia LKB). Sequences were analyzed using CLUSTAL [12], NIP 1.0 [13] and PHYLIP 3.2 (provided by Dr J. Felsenstein, Seattle, WA).

3. RESULTS AND DISCUSSION

The PCR-based strategy employed in this study was designed to enable amplification of genes related to the *D. melanogaster* segmentation gene *engrailed*. A major product of 233 base pairs was amplified from each of the 5 divergent vertebrates analysed (mouse, toad, zebrafish, lamprey, hagfish); as predicted if the primers had amplified DNA from an uninterrupted gene (or genes) containing an *en*-like homeobox.

Following purification, modification, cloning and transformation of the major amplified product from each of these species, multiple recombinant clones were sequenced, and 8 *en*-like genes identified (two each from zebrafish, hagfish, and toad; one each from lamprey and mouse). Six of these represent novel homeobox genes, whilst two classes of recombinant correspond to the previously reported genes mouse

Correspondence address: P.W.H. Holland, Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

A.

```
                        1                                                                 61
D. melanogaster en      C AGC GAG CAG TTG GCC CGC CTC AAG CGG GAG TTC AAC GAG AAT CGC TAT CTG ACC GAG CGG
Mouse En-1              G GCC GAG CAG CTG CAG AGA CTC AAG GCG GAG TTC CAG GCA AAC CGC TAT ATC ACG GAG CAG
Mouse En-2              T GCT GAG CAG CTC CAG AGG CTC AAG GCT GAG TTT CAG ACC AAC AGG TAC CTG ACA GAG CAG
X. laevis En-1a         T GCT GAG CAG CTC CAG AGA CTG AAG GCT GAG TTC CAA GCC AAC CGC TAC ATC ACA GAG CAG
X. laevis En-1b         T GCT GAG CAG CTC CAG AGA CTG AAG GCT GAG TTC CAG GCC AAT CGC TAC ATC ACA GAG CAG
Zebrafish En-1          A GCG GAG CAA CTA CAG AGA CTC AAG AAT GAA TTC CAG AAT AAT CGT TAC CTG ACG GAG CAA
Zebrafish En-2          G GCG GAG CAG CTT CAG AGA CTC AAG GCC GAG TTC CAG ACC AAC CGC TAC CTG ACC GAG CAG
Lamprey En              G GGC GAG CAG CTG TGC CGC TTG CGC GCC GAG TTC CAG GCG TCG CGC TAC CTC ACG GAG GAG
Hagfish En-A            G GCC GAT CAG CTG GCG CGC CTC CGG GCG GAG TTC CAG GCG AAC CGC TAC CTG ACC GAG GAA
Hagfish En-B            A GTC GAG CAA CTT CAG CGG CTC AAG TCC GAG TTT GGG GCA AGC CGG TAC CTA ACA GAG GCA

                      62                                                                 124
D. melanogaster en    AGA CGC CAG CAG CTG AGC AGC GAG TTG GGC CTG AAC GAG GCG CAG ATC AAG ATC TGG TTC CAG
Mouse En-1            CGG CGA CAG ACC CTC GCC CAG GAG CTC AGC CTG AAT GAG TCC CAG ATC AAG ATC TGG TTC CAA
Mouse En-2            CGG CGC CAG AGT CTG GCA CAG GAG CTC AGC CTG AAC GAG TCT CAG ATC AAG ATT TGG TTC CAG
X. laevis En-1a       AGG AGA CAG AGC TTG GCC CAA GAG CTG AGC CTC AAT GAA TCC CAA ATA AAG ATC TGG TTC CAG
X. laevis En-1b       AGG AGA CAG ACC TTG GCC CAA GAG CTG AGT CTC AAT GAA TCC CAA ATA AAG ATC TGG TTC CAG
Zebrafish En-1        AGG AGA CAA GCG TTG GCC CAG GAA CTC GGC CTG AAC GAG TCT CAA ATC AAA ATC TGG TTT CAA
Zebrafish En-2        CGG CGG CAA AGC CTG GCG CAG GAA CTG GGC CTC AAC GAA TCT CAG ATC AAA ATC TGG TTC CAA
Lamprey En            CGG CGC ACG GCG CTG GCG CGC GAG CTG CGG CTG AAC GAG GCG CAG ATC AAG ATC TGG TTC CAG
Hagfish En-A          CGA CGT CAG AAC CTC GCC CGT GAG CTA AGC TTG AAC GAG GCG CAA ATC AAG ATT TGG TTC CAG
Hagfish En-B          AGG CGA CAA GCG CTC GCC CAG GAA CTG CGA CTC AAC GAG GCT CAG ATC AAG ATC TGG TTC CAG

                     125                                                         181
D. melanogaster en    AAC AAG CGG GCC AAG ATC AAG AAG TCG ACG GGC TCC AAA AAT CCG CTG GCA CTG CAG
Mouse En-1            AAC AAG CGT GCC AAG ATC AAG AAA GCC ACA GGC ATC AAG AAC GGC CTG GCG CTG CAC
Mouse En-2            AAC AAG CGG GCC AAA ATC AAG AAA GCC ACG GGC AAC AAG AAC ACT TTG GCG GTG CAC
X. laevis En-1a       AAC AAA AGG GCC AAG ATC AAA AAG GCT TCG GGG ATG AAG AAT GGC CTG GCT CTC CAT
X. laevis En-1b       AAC AAA AGG GCC AAG ATC AAA AAG GCA TCA GGC ATG AAG AAT GGC CTA GCT CTA CAT
Zebrafish En-1        AAC AAA AGG GCA AAG ATC AAA AAA GCA ACG GGG AAC AAA AAC ACA CTT GCC GTG CAC
Zebrafish En-2        AAC AAG CGG GCC AAA ATC AAA AAG GCC AGC GGC GTC AAG AAC GGT CTG GCA ATA CAC
Lamprey En            AAC AAG CGC GCC AAG ATC AAG AAG GCG AGC GGC GTG AAG AAC GCC CTC GCA CTC TAC
Hagfish En-A          AAC AAA CGC GCC AAG ATC AAG AAA GCG AGC GGC GTT AAG AAC ACC TTG GCC TTG TAC
Hagfish En-B          AAC AAG CGC GCC AAG TTG AAG AAG GCA AAC GGG TTG CGG AAC CCA CTG GCG TTG CAC
```

B.

```
                     1       10        20        30        40        50        60
Drosophila En        SEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKSTGSKNPLALQ
                                                       ..........

Mouse En-1           A---Q---A--QA---I--Q---T-AQ--S---S---------------A--I--G---H
Mouse En-2           A---Q---A--QT------Q---S-AQ--S---S---------------A--N--T--VH
X. laevis En-1a      A---Q---A--QA---I--Q---S-AQ--S---S---------------AS-M--G---H
X. laevis En-1b      A---Q---A--QA---I--Q---T-AQ--S---S---------------AS-M--G---H
Zebrafish En-1       A---Q---N--QN------Q---A-AQ------S---------------A--N--T--VH
Zebrafish En-2       A---Q---A--QT------Q---S-AQ------S---------------AS-V--G--IH
Lamprey En           G---C--RA--QAS-----E--TA-AR--R-------------------AS-V--A---Y
Hagfish En-A         AD-----RA--QA------E---N-AR--S-------------------AS-V--T---V
Hagfish En-B         V---Q---S--GAS-----A---A-AQ--V----------------L--AN-LR-----H
```

Fig. 1. (A) Consensus nucleotide sequences (internal to the primers) from 8 vertebrate *en*-like genes cloned by PCR, aligned with *D. melanogaster en* and mouse *En-1*. (B) Deduced amino acid sequences from (A). Dashes indicate identity with *En*; homeodomain residues are underlined; putative DNA sequence recognition helix.

En-2 [3] and zebrafish *En-2* [14]. To ensure exact sequence determination, multiple independent clones from each gene were sequenced. Fig. 1 shows the consensus nucleotide and deduced amino acid sequences from the 8 *en*-like genes cloned in this study, aligned with *D. melanogaster en* and mouse *En-1*.

Over the 60 amino acid region analysed, 38 residues are conserved between all vertebrate *en*-like genes cloned. The C-terminal portion of the homeodomain is most highly conserved, including a stretch of 12 invariant residues. This extent of conservation, however, is less than in another subfamily of homeobox genes, the *msh*-related genes [10]. We propose that the two *X. laevis* genes we have cloned are both homologues of *En-1*, since both share higher sequence identity with mouse *En-1* than with mouse *En-2*, and since the two *X. laevis* genes are almost identical (9 nucleotides and one amino acid differ over the amplified region). The two genes isolated may represent two *En-1* loci resultant from the recent tetraploidization of the *X. laevis* genome [15]; we therefore designate these genes *En-1a* and *En-1b*.

The sequence of one zebrafish gene isolated in this study is almost identical to that of a gene previously reported [14] and designated zebrafish *En-2*. Although our consensus sequence differs from this gene by two nucleotide differences, we believe the clones derive from the same gene and follow the previous terminology. We suggest the second zebrafish *en*-like gene is the homologue of *En-1*.

Further insight into vertebrate *en*-related gene evolution can be gained by comparing the sequences obtained from the two jawless vertebrates examined, lamprey and hagfish. Two hagfish genes were isolated, both of which are clearly members of the *en*-like homeobox gene subfamily. The genes cannot easily be interpreted as orthologues of *En-1* and *En-2*; hence, to avoid this implicit assumption, we designate them hagfish *En-A* and *En-B*.

In contrast to hagfish, we identified only a single *en*-like gene from the genome of a lamprey, despite extensive sequence determination (23 independent clones derived using primers A and D, and 8 clones using primers B and D, were all found to derive from the same gene). The lamprey *En* sequence shows some unusual amino acid differences from the gnathostome genes, several of which it shares with hagfish *En-A*.

In conclusion, these data reveal that homeodomain sequences encoded by *en*-like genes have been highly conserved during vertebrate radiation, particularly during the radiation of the jawed vertebrates. The isolation of two distinct *en*-like genes from zebrafish parallels the situation in mammals, indicating that duplication of an ancestral *en*-like homeobox gene was an ancient event in vertebrate evolution. However, the more divergent hagfish *en*-like genes may have arisen by an independent gene duplication: a suggestion supported by the isolation of only one lamprey *en*-like gene.

Acknowledgements: We thank Jon Bartlett for technical assistance, John McVey for primer synthesis and advice, Matthew Parkin for help during preliminary experiments, Jonathan Slack (Oxford, UK) for tadpoles, Brigitte Baker (Bath, UK) for lampreys, and Jarl-Ove Stromberg (Kristenberg, Sweden) for hagfish. This work was funded by SERC, the E.P. Abraham Cephalosporin Fund and the Queens College, Oxford.

REFERENCES

[1] Fjose, A., McGinnis, W.J. and Gehring, W.J. (1985) Nature 313, 284–289.
[2] Poole, S.J., Kauver, L.M., Drees, B. and Kornberg, T. (1985) Cell 40, 37–43.
[3] Joyner, A.L. and Martin, G.R. (1987) Genes Dev. 1, 29–38.
[4] Davidson, D., Graham, E., Sime, C. and Hill, R. (1988) Development 104, 305–316.
[5] Davis, C.A. and Joyner, A.L. (1988) Genes Dev. 2, 1736–1744.
[6] Logan, C., Willard, H.F., Rommens, J.M. and Joyner, A.L. (1989) Genomics 4, 206–209.
[7] Hogan, B., Constantini, F. and Lacy, E. (1986) Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratories, New York.
[8] Holland, P.W.H. and Hogan, B.L.M. (1986) Nature 321, 251–253.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratories, New York.
[10] Holland, P.W.H. (1990) Gene (in press).
[11] Lanfear, J. and Holland, P.W.H. (1990) J. Molec. Evol. (in press).
[12] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237–244.
[13] Staden, R. (1986) Nucleic Acid Res. 14, 217–231.
[14] Fjose, A., Eiken, H.G., Njolstad, P.R., Molven, A. and Hordvit, J. (1988) FEBS Lett. 231, 355–360.
[15] Fritz, A.F., Cho, K.W.Y., Wright, C.V.E., Jegalian, B.G. and DeRobertis, E.M. (1989) Dev. Biol. 131, 584–588.