Cephalic expression and molecular characterization of *Xenopus En-2*

ALI HEMMATI-BRIVANLOU[1], JOSE R. de la TORRE[1,2], CHRISTINE HOLT[3] and RICHARD M. HARLAND[1,*]

[1]*Department of Molecular and Cellular Biology, Division of Biochemistry and Molecular Biology, University of California, Berkeley, CA 94720, USA*

[2]*Current address: Department of Biochemistry and Biophysics, School of Medicine, University of California, San Francisco, California 94143, USA*

[3]*Department of Biology, University of California, San Diego, La Jolla, California 92093, USA*

* Corresponding author

Summary

We have isolated and characterized cDNAs corresponding to the *Xenopus En-2* gene. Comparison of amino acid sequences between the entire *Xenopus En-2* and the *Drosophila engrailed* proteins confirms conservation of sequences inside as well as proximal to the homeobox and reveals a region of similarity towards the N terminus. Two transcripts encode the *Xenopus* En-2 protein. Both transcripts are regulated temporally in an identical fashion and are likely to be transcribed from two copies of the *En-2* gene.

We have also analyzed the distribution of the protein in the head tissue and in the dissected brain of tailbud stage embryos. In addition to the main band of expression at the midbrain–hindbrain boundary, we show that the protein is expressed in three novel areas: the mandibular arch, the optic tectum and the region of anterior pituitary.

Key words: engrailed, *Xenopus*, mandibular arch, optic tectum, pituitary.

Introduction

The *engrailed* gene defines a subset of homeobox genes that have been highly conserved during evolution between arthropods, echinoderms and vertebrates. Cloned and characterized extensively in the fruit fly, the *engrailed* locus contains two homeobox genes: *engrailed* and *invected* (see Coleman *et al.* 1987). These two genes share regions of extensive similarity inside as well as outside the homeodomain. They are expressed selectively in the same embryonic and larval patterns. *Engrailed* is expressed in the adult central nervous system of both *Drosophila* and grasshopper (N. Patel, personal communication). While the function of *invected* is still unknown, the *engrailed* gene acts first during segmentation of the blastoderm where it is involved in maintaining compartment and segment boundaries (Morata and Lawrence, 1975; Kornberg, 1981; Lawrence and Struhl, 1982). *Engrailed* is subsequently expressed during neurogenesis in a subset of neurons and neuroblasts (DiNardo *et al.* 1985; Brower, 1986; Patel *et al.* 1989). The biological function of *engrailed* during neurogenesis of arthropods has yet to be elucidated. Interestingly, the expression of *engrailed* in the nervous system has been found in all other species where the presence of an *engrailed* gene has

been demonstrated. Based on these findings, it has been speculated that the ancestral function of *engrailed* is to control cell fate during neurogenesis and that the arthropod phylum has co-opted the gene to function during segmentation (Patel *et al.* 1989).

Subsequent to the characterization of the *engrailed* gene in *Drosophila*, the use of nucleic acid probes allowed the detection of *engrailed*-related genes in a variety of organisms including mouse (Joyner *et al.* 1985; Davidson *et al.* 1988), chicken (Darnell *et al.* 1986), sea urchin (Dolecki and Humphreys, 1988), human (Logan *et al.* 1989; Poole *et al.* 1989) and zebrafish (Fjose *et al.* 1988). Perhaps the most spectacular approach to detect the *engrailed* gene in different species was undertaken at the protein level. A monoclonal antibody (mAb 4D9) that recognized the C-terminal part of the homeodomain of the *Drosophila engrailed* and *invected*, allowed the detection of *engrailed related* proteins in a number of invertebrate and vertebrate species including grasshopper, crayfish, lobster, leech, zebrafish, chickens and *Xenopus* (Patel *et al.* 1989; Hemmati-Brivanlou and Harland, 1989; Gardner *et al.* 1988; Weisblat *et al.* 1988). In all chordates, the protein recognized by the 4D9 antibody is expressed during early embryogenesis in a region that corresponds or will correspond to the midbrain–

hindbrain junction (in addition to a subset of mesodermal cells associated with specific muscles and vertebrate components in the case of zebrafish).

In a previous study, we used mAb 4D9 to show that an *engrailed* protein was present in *Xenopus* embryos (Hemmati-Brivanlou and Harland, 1989; Patel *et al.* 1989). From the pattern of expression of the protein, we speculated that it corresponded to the *Xenopus* equivalent of the murine *En-2* transcript. We have since shown that the anterior notochord of the early neurula is a strong source of signal(s) that induce expression of this gene (Hemmati-Brivanlou *et al.* 1990*b*). Here we present the molecular characterization of the *Xenopus En-2* gene. The *En-2* protein provides the earliest available marker of regionalization in the *Xenopus* neural plate, it is therefore important to determine when *En-2* transcripts first appear, since a substantially earlier expression would imply regionalization at even earlier times. We show that *En-2* transcripts are present and expressed at the same time and at the same place as the protein. We present sequences for two independent cDNAs and show that, in addition to the homeobox region, a stretch of amino acids near the N terminus of the protein is similar between the *Drosophila engrailed* protein and the *Xenopus En-2* protein; such a similarity has been noted previously (cited in Patel *et al.* 1989). We have also refined our analysis of expression of the *En-2* protein and find new sites of expression in the head of the tailbud tadpole.

Materials and methods

Embryos

Pigmented as well as albino *Xenopus laevis* embryos were obtained from the Berkeley colony maintained by the laboratory of J. C. Gerhart. Ovulation of females and *in vitro* fertilization were carried out as described by Condie and Harland (1987). Stages according to Nieuwkoop and Faber (1967).

Plasmids, enzymes and antibody

The 4D9 antibody and the zebrafish *engrailed* cDNA clone was a gift from Nipam Patel (University of California Berkeley). The *src* plasmid (Steele *et al.* 1989) was digested with *Eco*RI and *Hind*III and both the 1.8 and the 1.9 kb fragments were purified and labeled by random priming.

cDNA libraries

The anterior neural plate library was generously provided by C. Kintner (Salk Institute). This library was screened using a 2.5 kb *Eco*RI fragment from the zebrafish *engrailed* cDNA clone, under moderately stringent conditions (Condie and Harland, 1987).

The head-specific cDNA expression library was made as follows: 1000 tailbud embryos (stage 28, Nieuwkoop and Faber, 1967) were dissected by a transverse cut, perpendicular to the anterior–posterior axis, between the otic vesicle and the eye. RNA was prepared from the head and tail fractions as described by Condie and Harland, 1987. Polyadenylated RNA was selected by chromatography on oligo(dT)-cellulose. Synthesis of cDNA followed a procedure from Hazel Sive (personal communication): First strand cDNA was primed with oligo(dT), and following synthesis with reverse transcrip-

tase, was tailed with terminal transferase and dGTP. The second strand was primed with a *Not*I–oligo(dC) primer; 5′ CAGACACGTAGCGGCCGC(C)$_{12}$ 3′. To minimize secondary structure in the GC-rich primer, the oligonucleotide was hybridized to a partial complementary strand; 5′ CCGCTACGTGTCTG 3′. Second strands were synthesized with 'sequenase' (USB). Double-stranded cDNA was methylated and *Eco*RI linkers were added. Excess linkers were removed with *Eco*RI and *Not*I and the cDNA was size selected on a 0.6×28 cm column of Sepharose 4BCL. The first fractions from the peak were pooled and ligated into *Not*I–*Eco*RI-digested λZAPII (Stratagene). To reduce background from vector religation, extensively digested λDNA was separated from the released oligonucleotide on a 0.6×13 cm column of Biogel A 5M. After packaging, 3×10^6 independent recombinants were obtained. One million plaques from the unamplified expression library were screened as described by Patel *et al.* (1987), using the method of Huynh *et al.* (1985). The 4D9 antibody was used at a concentration of 1:2 and the secondary antibody, peroxidase-conjugated goat anti-mouse IgG (Biorad) was used at a concentration of 1:500. A single plaque reacted with the antibody.

DNA sequencing

Subclones and deletions of the cDNAs were sequenced on both strands by the dideoxy chain termination method (Sanger *et al.* 1977), using a T7 sequencing kit (from Pharmacia).

Northern blots

For the developmental northern analysis, polyadenylated RNAs from staged embryos were prepared as described by Condie and Harland (1987). For the spatial analysis of *En-2* in embryos, poly(A) RNA from tails and heads of stage 28 embryos were prepared as above. 2 μg of poly(A) RNA from each stage were fractionated by electrophoresis in denaturing formaldehyde agarose gels. The RNA was transfered to nylon membrane and cross-linked by UV irradiation with a Stratalinker (Stratagene). DNA probes were hybridized to the blot at 42°C overnight and RNA probes were hybridized at 65°C. The filter was then washed for 30 min at room temperature with 2×SSC, 0.1% SDS and then for two hours with 50 mM Tris pH 8.0, 2 mM EDTA, 0.5% sodium pyrophosphate, 1×Denhardt's, 5% SDS, 0.05% Lauryl sarcosine for 2 h at 65°C and exposed to X-ray film (Kodak). Filters were exposed for 20 h for the *engrailed* probes and for 4 h for the *src* probe at −80° C with screens.

Immunohistochemistry

Whole-mount embryos were fixed and stained for *En-2* with mAb 4D9 as described by Hemmati-Brivanlou and Harland (1989). Brains were dissected out and immunostained as whole mounts. For staining of the optic fibers, the optic projection was filled with horseradish peroxidase (HRP) by inserting a chip of dried HRP in 2% lysolecithin (Sigma) into the retina after removing the lens. 25 min after HRP application the embryos were fixed in 4% paraformaldehyde in 0.1 M phosphate buffer for two hours, brains were dissected and HRP was visualized as described previously (Hemmati-Brivanlou and Harland, 1989).

Results

Cloning of the Xenopus engrailed *gene*

Two parallel strategies were adopted to isolate the

engrailed gene. The first strategy involved the screening of neurula (stage 17) anterior neural plate library with a zebrafish *En-2* clone. Ten independent clones were isolated after screening a total of 10^6 plaques. The insert from each clone was subcloned into Bluescript KS+. Based on their length and restriction map, these clones were classified into two categories: 2.7 kb *EN* and 1.4 kb *EN* (see Fig. 1B). In parallel, a stage 28 head-specific expression library was screened using the monoclonal antibody 4D9. This antibody has been previously used to indicate the presence of the *engrailed* protein in *Xenopus* (Hemmati-Brivanlou and Harland, 1989; Patel *et al.* 1989). One million plaques were screened and a single positive clone with a 2.7 kb insert was isolated.

One representative clone of each class was sequenced by the dideoxy sequencing method on both strands. The sequence of the 2.7 kb clone is shown in Fig. 1A. The sequence reveals an open reading frame of 265 amino acids containing the homeobox and preceded by a 453 nucleotides of 5' untranslated sequence. The 3' untranslated region is long (1405 nucleotides) and ends with a poly-adenylation signal at nucleotide 2204. The 1.4 kb *engrailed* cDNA also encodes a protein of 265 amino acids. The 5' untranslated sequence of this clone is 453 nucleotides long and the open reading frame is followed by 181 bases of 3' untranslated region. There are large regions of nucleotide sequence similarity between the two clones both in the 5' and 3' untranslated region (Fig. 2) as well as 95% identity in the coding region (data not shown). Moreover, the comparison of *Xenopus* and mouse (Alex Joyner, personal communication) *En-2* 3' untranslated sequences has revealed regions similar between the two species. The significance of this evolutionary conservation is at present not understood.

Amino acid sequence comparison between Drosophila, mouse and Xenopus engrailed proteins

Fig. 3 shows the amino acid comparison between *engrailed* proteins from *Drosophila*, partial mouse *En-1* and *En-2* sequence and the predicted sequence of the two *Xenopus En-2* proteins. The homeodomain of the predicted protein from the *Xenopus* 2.7 kb cDNA shares 97% identity with murine *En-2*, 90% identity with murine *En-1* and 73% identity with *Drosophila engrailed*. Similarly, the homeodomain predicted from the protein encoded by the 1.4 kb cDNA is 98% identical to murine *En-2*, 95% to murine *En-1* and 78% to *Drosophila engrailed*. We therefore conclude first that the *Xenopus engrailed* proteins, as expected, are closer to other chordate *engrailed* proteins than to the arthropod protein (inspection of sequences flanking the homeobox also support this conclusion); second, both *Xenopus* proteins are closer to the mouse *En-2* protein than the *En-1* protein. These facts, in addition to the protein localization in the embryo, allow us to conclude unequivocally that both of these cDNAs encode the *Xenopus* homolog of the murine *En-2* gene.

In addition to the homeobox and the immediate C-terminal flanking sequence, inspection of amino acid sequences upstream of the homeobox reveals two regions of similarity between the frog *En-2* and the *Drosophila engrailed* protein. The first region closest to the homeobox spans 24 amino acids (starting at residue 145) and has also been shown to be conserved in both the *En-1* and *En-2* proteins in mouse (Joyner and Martin, 1987) and the ZF-EN protein in zebrafish (Fjose *et al.* 1988). Through the analysis of full-length cDNA clones, we have found a second region of similarity between frog *En-2* and *Drosophila engrailed* proteins. This region lies near the N terminal domain of the protein (Fig. 3) and spans 12 amino acids, 7 of which are conserved. Interestingly the same motif is also present in the *Drosophila invected* protein (not shown). When this region is compared to murine *En-2*, an extended region of 20 amino acids is found to be identical between *Xenopus* and mouse *En-2* proteins (and is the only extensive region of similarity near the N terminus; A. Joyner personal communication).

Finally, there are 11 amino acid differences between the proteins predicted from the two *Xenopus* cDNAs, two of which are changes at C-terminal part of the homeodomain (Fig. 3). One of these changes is in a region of the homeobox shown to be the epitope recognized by the 4D9 antibody (Patel *et al.* 1989). Patel *et al.* (1989) showed that the glycine residue at position 40 in the homeodomain is crucial for the recognition of the epitope by mAb 4D9, and a glycine to serine substitution prevents recognition. That single amino acid change prevents the antibody from recognizing the mouse *En-2* protein. We therefore conclude that mAb 4D9 is only able to recognise the protein product of one of the *Xenopus En-2* sequences, represented by the 2.7 kb *En-2* cDNA clone.

Two other lines of evidence support this claim: (i) Only the 2.7 kb clone containing the glycine residue was detected in the expression library by the antibody probe even though the 1.4 kb type is eight to ten times more abundant in the anterior neural plate library. (ii) When vectors driving the expression of the 2.7 or the 1.4 kb *En-2* cDNA under the control of inducible or constitutive promoters, are injected in *Xenopus* embryos, only the protein product of the 2.7 kb cDNA, but not the 1.4 kb cDNA, can be detected in whole-mount immunostaining using the 4D9 antibody (Hemmati-Brivanlou unpublished observation). These results are therefore in perfect agreement with the protein recognition epitope defined for the mAb 4D9 (Patel *et al.* 1989).

Expression of Xenopus En-2 mRNAs during early Xenopus development

To address the issue of stage-specific expression of the *En-2* gene, we probed a northern blot of poly(A) RNA isolated from embryos at close time points during neurulation. Using the 1.4 kb *Eco*RI fragment of the small cDNA as probe (Fig. 1), we find that two major transcripts of 3.0 and 1.7 kb are present at stage 14 of embryogenesis (Fig. 4A). This corresponds to the same stage that the protein was initially detected. Both messages peak in expression around stage 20 and decrease around stage 36. The use of 5', 3' and

A

```
-453                                    GCTCTGGAAGCGCAGGGGATTGATGTGCTGCTCCTCCTGCTGTGCAGATGTTACAGTGATATAGACTGTACATGTAATATCTTTCTATTATTT

-360    TATATCTATTATATGTCACACTTGCTGAGCTCCACCACTGGGACTTGGATGGGGCTACAAACATCTCCCTGCTGGCTACTCTGAAGGATTTCTCCCCAGGAGTCTGGATCTGGAAGGAGG

-240    GGGGTAATGAGCCCGTCCTGGGGAAGAAGAAGAAGAAGATGCTGTGGTTGGCAGAAGTGAAACATTAAAAATCTGCAGAGGAACCAACAGAGGAACCAAGAGAACAAGAAGAAGGGGTCAA

-120    AGCCTTTCTGCTGGAACGTCTGTCGGAGCCTTGGGGTTAGAGGCAATTTGAATGTGTAGCTGACTGGAGCCAGTAGGAAACACAAGGGGTTAAAGGCAAGAAGAAGCTTTGGGGAACAGC

   1    ATGGAAGAGAATGAGCAGAATAACAGGGAAGTGGAACCGCAGCAGGAGTCAGGAGAAGAGTCCAACAGGGGGATCCTTCATCAGGCTCCTCCTGGGAACCACCAGCCCCATCACAGGATA
   1    M  E  E  N  E  Q  N  N  R  E  V  E  P  Q  Q  E  S  G  E  E  S  N  R  G  I  L  H  Q  A  P  P  G  N  H  Q  P  H  H  R  I

 121    ACTAACTTCTTCATTGACAACATCCTACGGCCGGAGTTCGGGAGGAGAAAAGAAAGAATAAACCACCAGGATGAGCTCTTTACTGGGAGAGACACTGGGGCTTTGTCTGGGGCAGAGAGC
  41    T  N  F  F  I  D  N  I  L  R  P  E  F  G  R  R  K  E  R  I  N  H  Q  D  E  L  F  T  G  R  D  T  G  A  L  S  G  A  E  S

 241    GGCCACCACCGAGTCAATGTCCCCGAGGGCGCAGGAGGCTCTTCCAAAGTCATCACTGTTACTGGAGAGAAGAAAAGTGATCTAGCGATGGAAGAAACCTTAAAATCCAGAGGACTTAAT
  81    G  H  H  R  V  N  V  P  E  G  A  G  G  S  S  K  V  I  T  V  T  G  E  K  K  S  D  L  A  M  E  E  T  L  K  S  R  G  L  N

 361    GGAGACCATTCCCTCAGCTCAGACTCTGATAGTTCCCAAGCCAGCTCCAAAGCCACTCAAAAGCCTATACTGTGGCCAGCCTGGGTCTACTGCACCAGATATTCAGACAGACCTTCATCA
 121    G  D  H  S  L  S  S  D  S  D  S  S  Q  A  S  S  K  A  T  Q  K  P  I  L  M  P  A  M  V  Y  C  T  R  Y  S  D  R  P  S  S

 481    GGTCCGAGATCCCGAAAACCAAAGAAGAAGAGTGTCAGCAAAGAGGACAAGAGGCCCAGAACTGCCTTCACTGCTGAACAGCTTCAGAGGCTCAAGGCCGAGTTCCAAACTAACAGGTAT
 161    G  P  R  S  R  K  P  K  K  K  S  V  S  K  _E__D__K__R__P__R__T__A__F__T__A__E__Q__L__Q__R__L__K__A__E__F__Q__T__N__R__Y_

 601    TTGACTGAGCAGAGGAGGCAAAGTCTGGCCCCAGGAACTCGGTCTGAACGAATCTCAGATCAAAATCTGGTTTCAGAATAAACGAGCAAAGATTAAAAAATCCACAGGGAACAAAAATTCC
 201    _L__T__E__Q__R__R__Q__S__L__A__Q__E__L__G__L__N__E__S__Q__I__K__I__M__F__Q__N__K__R__A__K__I__K__K_S  T  G  N  K  N  S

 721    CTGGCCCTGCATTTAATGGCCCAAGGACTTTACAACCACGCGACCACTTCAAAGGACGGCAAATCAGACAGCGAATAGTAACGTTTCGGTAGCGCAAAGGCCTCGCAATGCAATAAAAAA
 241    L  A  L  H  L  M  A  Q  G  L  Y  N  H  A  T  T  S  K  D  G  K  S  D  S  E

 841    GCCTGTGTATAAAGAATCTACCAGCATTTGGCTTCGGAGTGTGTTTATTGGGATAAACCCGCAATATTCTATGTAAATATCATTTACAGACAAACAAACAAAAGGAAAAAGTCGAAAATA

 961    TGGACACAACTATTTTTTCTTTGATTTCATTATTATTTTATTTTGTTTAATCCTTTCCAGTCCTTTTTTTTCCTGCCTTTAGATTTTATTTTATACTAAACGATGCTCGTGGATGTCGGA

1081    GAGAAAAAAAAATCAAAAAAAGTCAACCTTTTTTTTTTTTTTTTTGCTAAAGTGGAAAGATGAAATTGCTGCATTTCGCACTACTGTGAATTTAATTAAAAAAAAAAAAAAAGCTAATATTTG

1201    TTGTGTTTTTTTTTGTTTTTGTTAATATTTGAAACTTTTTGTTGTAAAATAAATCAAAGCAGGGGCCATGTCCTGTCGCTGTTGTAATTAAAGGTGTATTTTAATTGATTTAATTTTTTT

1321    GTTTTTATATGCATTTTTTTTTATAGTTGTGCAGAATATGATTTATTTATATACACTGGCGGGGTGATTTATAGTTTTGCACTTAGTGTTTTTTTTTTTAAATGACATATTGGGGCAACTTT

1441    TTTCTTTAAAATGACATATTGGGCAACTTTTTTTTTTTTTTGCTGACAAATTGTATTATTTCCCAGAAATTGAAGATATTGAAGCCTGACAGCTGCATATGTACTAGGAACAAGCAAACCA

1561    AACATTTGCCAATCACTGTCCCAATAAATGGTTAATTTATCCAATCCACTTTCCCCTCCTACTATGACGTAATCCCCCCTTCCCCACTCTGCTGCACCCCCCTTCCCTGTTTATGGATCC

1681    CTCCTTTGTTTCCAACTTCCCTGAATTTTAAAAAAAGACCCTGCAAAAATAATACATGAGCCCAATGAATAATGGAGAGGTTTATAATGAATCAGGTCAATGGACTACAGCTGATTTGGG

1801    CCATTTCTTTTGTTACAGTTCATTTTTTTTTTTAAAGATATTAAAGTTTTTTTTTGTTTTTTAAATCAATCACCGATTTCATCGTAAAGAATGTAATATTTAGCCTTGGACAGAAGTAAAAC

1921    TTAAAAATCATGCAAAAGAACCTCTTTTCTCCAATTCTAGGGGAAACTTACTACATTTATTCTCTTACTGTATATGTACCATTTGCCTTCTTTTAATTGATTTTTTTTTTTTTTAAATTAAA

2041    AAACATAAAAAGGAAAAAAAAAAAATCTGGCAAGACAAAGCCAGAAATCTAATGATACAATTAGTGTAAATACAATGTTTCCAACACTGGCAATGTTTGTTTTCATTTGAATGTACAGAGA

2161    TGCAAATAATAAATATTTCCCCCTTTCTGATTTTAAAAAATAAACCGGAATTCC
```
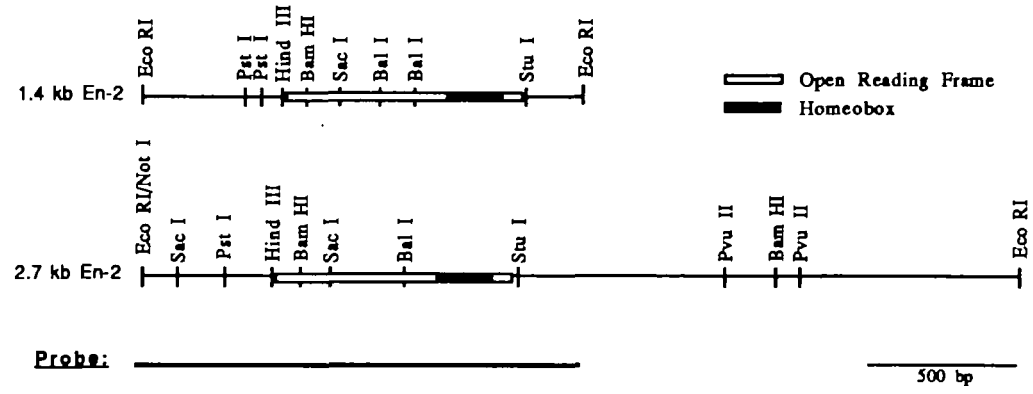
B

Fig. 1. Sequence analysis of the *Xenopus En-2* cDNAs. (A) Nucleotide sequence and the predicted protein sequence of the 2.7 kb *En-2* cDNA. The homeodomain region of the protein is underlined. (B) Restriction map of the two *En-2* cDNAs. The open box represents the open reading frame. Filled boxes represent the homeodomain.

A

```
2.7 kb En-2   -453   --------------------------------GCTCTGGAAGCGCAGGGGATTGATGTGCTGCTCCTCCTGCTGTGCA
1.4 kb En-2   -474   GGAGCGTGTAAAAATTGGTCTCTCTTAGGAGACA::::::::::::::C:T::G::C::::::::::::::::::::::::AAC

2.7 kb En-2   -407   GATGTTACAGTGATATAGACTGTACATGTAATATCTTTCTATTATTTTATATCTATTA------TATGTCACACTTGCTG
1.4 kb En-2   -394   A:GAG:GGG:CAGCTCCT:TA:-::T:TATT:T:AA:AA:TA:::A:CT:::A::CATATATCT:::::::::::CC::C

2.7 kb En-2   -333   AGCTCCACCACTGGGACTTGGATGGGGCTACAAACATCTCCCTGCTGGCTACTCTGAAGGATTTCT-CCCCAGGAGTCTG
1.4 kb En-2   -315   ::::::::::::::::G:::::::::T-:::::::::::::::::::A::::::G::::C:::::::::CC:::::::::AC:::

2.7 kb En-2   -254   GATCTGGAAGGAGGGGGGTAATGAGCCCGTCCTGGGGAAGAAGAAGAAGAAGATGCTGTGGTTGGCAGAAGTGAAACATT
1.4 kb En-2   -236   :::::::::::::-::::::::::::::::::::::::::::::G:::::::::T::C:::::T:::::::::::::::

2.7 kb En-2   -174   AAAAATCTGCAGAGGAACCAACAGAGGAACCAAGAGA----ACAAGAAGAAGGGTCAAAGCCTTT-CTGCTGGAACGTCT
1.4 kb En-2   -154   -:::::::::------------:::::A::::G:::::GAGC:A::A:::G:A:G::G:A::::C:::A:::::::C

2.7 kb En-2    -99   GTCGGAGCCTTGGGGTTAGAGGCAATTTGAATGTGTAGCTGACTGGAGCCAGTAGGAAACACAAGGGGTTAAAGGCAAGA
1.4 kb En-2    -90   C:G::TC-----------:::::::::::::::::::::::::::::::::::::::::::T:::::::::::::::::::::::

2.7 kb En-2    -19   AGAAGCTTTGGGG-AACAGCATG
1.4 kb En-2    -20   :::::::::::AAT:::::::::
```

B

```
2.7 kb En-2    796   TAGTAACGTTTCGGTAGCGCAAAGGCCTCGCAATGCAATAAAAAAGCCT-------------GTGTATAAAGAATCTAC
1.4 kb En-2    796   :::::::::::::::A:::G:::::::::::::::::::::TTTAAGGAAGAATAAAGGGCCA::::::::::T:AA::

2.7 kb En-2    862   CAGCATTTGGCTTCGGAGTGTGTTTATTGGGATAAACCCGCAATATTCTATGTAAATATCATTTACAGAC--AAACAAAC
1.4 kb En-2    876   :::::::::::T:::::::::G:G:::::::::T:::::::::::::::::::::::::A::::::::G:AA:::::::::

2.7 kb En-2    940   AAAAGGAAAAAGTCGAAAATATGGACACAACTA____
1.4 kb En-2    956   ::::AA:::::------::A:C:::ATTC
```

Fig. 2. Comparison of nucleotide sequence in the 5' (A) and 3' (B) untranslated region of the two *Xenopus En-2* cDNAs. Semi colons represent perfect matches, dashes represent gaps and nucleotide differences are indicated. The open reading frame begins at the ATG at the end of the sequences shown. The ATG start codon in (A) as well as the stop codon in (B) are underlined.

homeobox-specific probes from the 2.7 kb clone on stage 20 poly(A) RNA demonstrated that both messages contain the 5' and homeobox sequences, but only the 3 kb transcript contains the most 3' end sequences (data not shown).

Next, we addressed the spatial distribution of the transcripts. Since the antibody used to localize the *En-2* protein only recognizes the protein from one of the transcripts, it was important to know whether both transcripts were also localized in the same region of the embryo. Poly(A) RNAs extracted from the head and tail fractions of the tailbud stage (described above) were subjected to northern blot analysis using the 1.4 kb fragment of the small *engrailed* cDNA as a probe. Fig. 4B shows that both messages are present in the head (anterior to the otic vesicle) and that neither is present in the tail region.

New sites of expression in the craniofacial structures of tailbud tadpole

In our previous study, we showed the distribution of *En-2* protein in neurulae (stage 14) tailbud and tadpoles (stage 28 and above). However, since at stage 20 we find a large increase in the amount of *En-2* mRNAs, we reexamined whether the increase in the mRNA levels correlated with increased protein expression at that stage. A different batch of supernatant also permitted more sensitive whole-mount immunocytochemistry on embryos from stage 20 to stage 28. Fig. 5A shows the staining in the head of a stage 23 embryo with the 4D9 antibody under saturating conditions (see Materials and methods). At this stage, there is nuclear staining at the midbrain and hindbrain junction, the main region of expression of *engrailed*. However, in addition to this staining, we also detect bilaterally symmetrical expression in the branchial region (Fig. 5A and 5C). Frontal sections of a stage 28 embryo (Fig. 5D and 5E) confirmed that this staining is anterior to the first pharyngeal pouch, and is therefore in the mandibular arch. The identity of the cells expressing *En-2* in the mandibular arch is difficult to assess since at this stage they cannot be assigned to a specific anatomical structure. However, since these facial structures arise mostly from the contribution of neural crest cells it is likely that the *En-2*-expressing cells are neural crest derived. The examination of a frontal section of the embryos shown in Fig. 5C reveals that the *En-2* protein is expressed in the presumptive mandibular region before any obvious morphological differentiation of the mandibular processes. This aspect of the expression parallels the expression of *En-2* in the anterior neural plate of the stage 14 neurula where *En-2* is expressed

```
2.7 Kb EN-2    1                                    M-----------------------------------
1.4 Kb En-2    1                                    :-----------------------------------
D.mel en       1                                    MALEDRCSPQSAPSPITLQMQHLHHQQQQQQQQQ


2.7 Kb En-2    2    ---------------------------------------------------------------------
1.4 Kb En-2    2    ---------------------------------------------------------------------
D.mel en      35    QQMQHLHQLQQLQQLHQQQLAAGVFHHPAMAFDAAAAAAAAAAAAAAAHAHAAALQQRLSG


2.7 Kb En-2    2    -------------------------------------------EENEQNNREVEPQQESGEESN
1.4 Kb En-2    2    ---------------------------------------------:::::::::::::::::::::
D.mel en      95    SGSPASCSTPASSTPLTIKEEESDSVIGDMSFHNQTHTTNEEEEAEEDDDIDVDVDDTSA


2.7 Kb En-2   23    RGILHQAPPGNHQPHHRITN FFIDNILRPEFG RRKERINHQDELFTGRDTGALSGAESGH
1.4 Kb En-2   23    :::::::::::::::::::::: :::::::::::: ::::G:S:::::Y:E:::::::::::::
D.mel en     155    G:R:PPPAHQQQSTAKPSLA :S:S:::SDR:: DVQKPGKSIENQASIFRPFEANRSQTAT


2.7 Kb En-2   83    HRVNVPEGAGGSSKVITVTGEKKSDLAMEETLK-------------------------
1.4 Kb En-2   83    :::::::::::::::::::::::::::::::::::-------------------------
D.mel en     215    PSAFTRVDLLEF:RQQQAAAAAATAAM:L:RANFLNCFNPAAYPRIHEEIVQSRLRRSAA


2.7 Kb En-2  116    -------------------------------------------------------------
1.4 Kb En-2  116    -------------------------------------------------------------
D.mel en     275    NAVIPPPMSSKMSDANPEKSALGSLCKAVSQIGQPAAPTMTQPPLSSSASSLASPPPASN


2.7 Kb En-2  116    -------------------------------------------------------------
1.4 Kb En-2  116    -------------------------------------------------------------
D.mel en     335    ASTISSTSSVATSSSSSSSGCSSAASSLNSSPSSRLGASGSGVNASSPQPQPIPPPSAV-


2.7 Kb En-2  116    SRGLNGDHSLSSDSDSSQASSKATQKPIL WPAWVYCTRYSDRPSSGPRSRKPK KKSVSK-
1.4 Kb En-2  116    :::::::::::::::::::::T:::::::Q:M: ::::::::::::::::::::::::: :::::::-
D.mel en     394    ::DSGME-SSDDTRSETGSTTTEGGKNEM ::::::::::::::::::::Y:R:: QPKDKTN
Mouse En-2        ...:M: :::::::::::::::::::::::: ::NPN:-
Mouse En-1        ...:LV ::::::::::::::::::::T::L: ::KNE:-


2.7 Kb En-2  175    EDKRPRTAFTAEQLQRLKAEFQTNRYLTEQRRQSLAQELGLNESQIKIWFQNKRAKIKKS
1.4 Kb En-2  175    :::::::::D:::::::::::::::::::::::::::::::::::S::::::::::::::::::A
D.mel en     453    DE:::::::SS:::A:::R::NE:::::::R:::Q:SS:::::::A:::::::::::::::::
Mouse En-2        ::::::::::::::::::::::::::::::::::::::::::::::S::::::::::::::::::A
Mouse En-1        :::::::::::::::::::::::A:::I:::::::T::::S::::::::::::::::::::::A


2.7 Kb En-2  235    TGNKNSLALHLMAQGLYNHATT SKDGKSDSE ter
1.4 Kb En-2  235    :::::::::::::::::::::S:: :::::::::: ter
D.mel en     513    ::S::P:::Q:::::::::::-:: VPLT:EEE:LEMRMNGQIP ter
Mouse En-2        :::::T::V::::::::::S:: A:E::::::: ter
Mouse En-1        ::I::G:::::::::::::S:: TVQD:DE:: ter
```

Fig. 3. Amino acid comparison between *Xenopus*, mouse (Joyner and Martin, 1987) and *Drosophila engrailed* (Poole *et al.* 1985) proteins. The homeodomain is between residues 175 and 234. Barred boxes represent regions of similarity. Semi colons represent perfect matches, dashes represent gaps and amino acid changes are indicated.
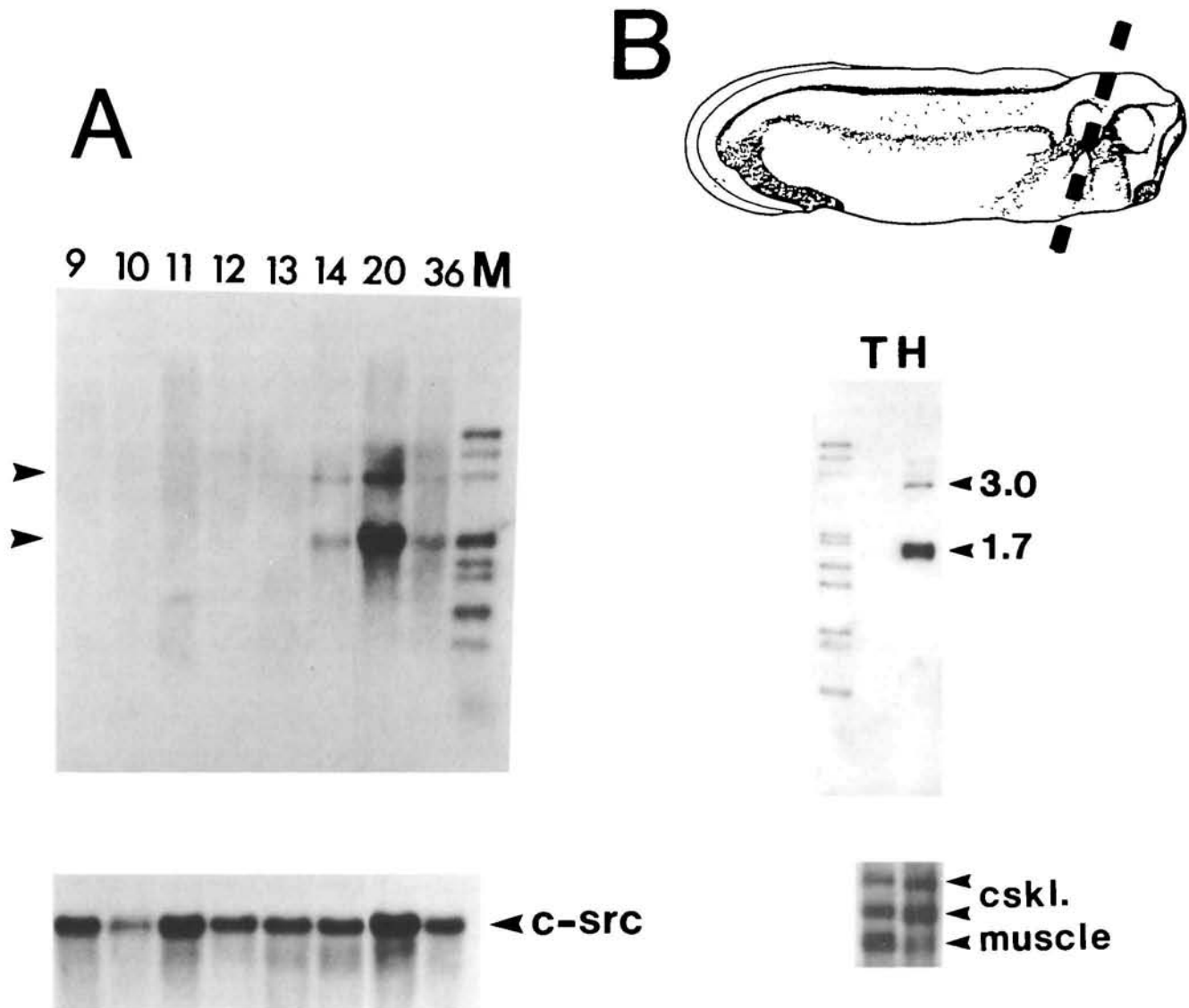
Fig. 4. Characterization of *En-2* transcripts during development. (A) Developmental pattern of expression of *En-2*. The top panel shows *En-2* transcripts in poly(A) RNA from blastula to swimming tadpole. The numbers at the top of the lanes correspond to embryonic stages (Nieuwkoop and Faber, 1967). Stages 9 and 10 are late blastula, stages 10 to 12 are gastrula, stages 12 to 20 are neurula and stage 36 is tadpole. The M lane is the size marker (λ DNA digested with *Eco*RI and *Hind*III and radiolabeled). The bottom panel is the expression of the *Xenopus src* gene during development using the same filter, stripped of the *En-2* probe, to show the amount of poly(A) RNA in each lane. The quantity of *src* transcripts remains fairly constant throughout early *Xenopus* embryogenesis (Condie Ph.D. Thesis, 1989). (B) Spatial distribution of the *En-2* transcripts in a stage 28 tailbud. The top panel represents the dissection (a single cut anterior to the otic vesicle). The middle panel shows that both *En-2* transcripts are present in the head (H) and absent in the tail (T) region. The bottom panel shows the hybridization of the same filter with the actin probe to show abundance of RNA in each lane. cskl=cytoskeletal actin, muscle=muscle actin.

(Hemmati Brivanlou and Harland, 1989) before any morphological boundaries are visible.

To analyze the pattern of expression of *En-2* in the brain of the embryo in more detail, we immunostained whole-mount brain preparations from stage 35/36 and 37/38 embryos with the 4D9 antibody under saturating conditions. Fig. 5F shows that two regions of low-intensity staining occur outside the intensely stained hindbrain–midbrain junction. The first region is in the lateral part of the dorsal midbrain (Fig. 5G) identified

as the optic tectum (Harris, 1989). The expression in this region is 'salt and pepper' like with positive and negative cells intermingled in a bilaterally symmetrical pattern. At this stage of development, the pioneer population of retinal axons are begining to enter the optic tectum (Holt, 1984; Harris, 1986). When the eye is filled with horseradish peroxidase (HRP; Harris, 1986), HRP-stained retinal axons can be seen coursing along the optic tract and terminating in the rostral part of the tectum (Fig. 5H). The *En-2*-positive tectal cells

(Fig. 5H) are located in a position appropriate for receiving direct retinal input; however, further analysis is needed to establish whether the optic fiber terminals actually make direct synaptic contact with the En-2-expressing cells.

A final region of positive staining is in the region of the developing anterior pituitary shown in Fig. 5I. Whereas the cells of the mandibular arch and optic tectum could arise from the stripe of cells that first express the En-2 antigen in the early neurula, the pituitary fate maps to the anterior ridge of the neural plate (Eagleson and Harris, 1990). Thus the pituitary staining appears to result from a distinct population of cells from the rest of the staining.

All the new areas of expression described here are transient, and staining is not detected in more mature tadpoles. This makes it difficult to identify precisely the fate of En-2 expressing cells. Nevertheless, the expression in relatively undifferentiated tissue emphasizes the likely role of the protein in directing cell fate.

Discussion

Molecular characterization of Xenopus En-2

In this work, we describe the molecular characterization of the En-2 gene in Xenopus laevis. We have isolated and sequenced two independent cDNAs encoding the En-2 gene product. These cDNAs predict proteins that differ by 11 amino acids in their coding sequences. Three of these changes are in the homeodomain and one of these changes lies in the epitope recognized by the 4D9 monoclonal antibody so that only one of the two protein products is detected. The En-2 protein can be encoded by at least two transcripts containing very similar 5' end and homeobox sequences. Both of these transcripts are expressed during early neurulation and are localized later in the head of the embryo.

Of the two cDNAs isolated, one, 2666 nucleotides long, corresponds to the large 3.0 kb transcript which encodes the protein recognized by the monoclonal antibody (4D9) originally raised against the Drosophila invected homeodomain. The other cDNA, 1444 nucleotides long encodes the version of the En-2 protein that is not recognized by the 4D9 antibody. Although the smaller cDNA ends with a stretch of poly(A), it does not contain any AAUAAA polyadenylation signal immediately upstream of the poly(A). Although direct proof is not available, two lines of evidence suggest that the small cDNA corresponds to the 1.7 kb transcript. First a DNA probe from the zebrafish En-2 hybridized to six clones from an anterior neural plate library; one of the clones was 2.7 kb long while the other five were 1.4 kb. The latter five shared the same restriction map and, upon analysis of terminal, non-coding sequence (which is most divergent between the cDNAs), were all shown to be the same. This ratio of five to one in favor of the small cDNA, is reminiscent of the ratio of the 1.7 kb to the 3.0 kb transcripts detected in the northern blot analysis. Furthermore, when the mAb 4D9 was

Fig. 5. Localization of En-2 protein in anterior structures and in the whole-mount brain. (A) Expression of En-2 protein in the mandibular arch of a stage 23 embryo. Arrows point to the nuclear staining. (B) Plane of frontal sections of stage 23 and 28 embryos shown in C, D and E. (C) Frontal section of the embryo shown in A; arrows point to the nuclei expressing En-2 in the presumptive mandibular arch. (D) Frontal section of a stage 28 tailbud (differential interference contrast optics) to show pharyngeal structures. MA=Mandibular arch, PP1=Pharyngeal pouch 1 and PP2=Pharyngeal pouch 2. (E) Same section as D (bright-field optics). Arrow indicate the nuclear staining in the mandibular arch. (F) Dissected brain from a stage 35/36 tadpole stained with mAb 4D9. The two boxes delineate the optic tectum shown in G and the pituitary shown in I. (G) Expression of En-2 in the optic tectum. The arrow points to the 'salt and pepper' like expression in the midbrain. (H) Whole-mount brain from a stage 35/36 embryo with HRP-stained retinal axons. The optic fibers are seen to terminate in the optic tectum (t), of the midbrain (mb). (I) Expression of En-2 in the region of the pituitary. The arrow points to the nuclear staining. The major area of the brain are indicated, t=optic tectum, p=Pineal, ot=optic tract, tel=Telencephalon, di=Diencephalon, mb=Midbrain, hb=Hindbrain.

used to screen 10^6 individual plaques from an unamplified expression library, only one positive clone was obtained and, by sequence analysis, corresponded to the larger of the two cDNAs isolated with the DNA probe. Taken together, these two lines of evidence suggest that the small cDNA clone that we have characterized may be a representative of the 1.7 kb mRNA. However, we cannot exclude the possibility that the small mRNA could be generated from either of the genes represented by the two cDNAs clones by differential use of polyadenylation signals. Several potential polyadenylation signals are present in the 3' untranslated region of the 2.7 kb cDNA (Fig. 1).

Three possibilities could explain the pattern of expression that we have shown. The two cDNA classes could result from two alleles of the same gene (using different polyadenylation sites) or they could be the products of the gene duplication event that accompanies polyploidization of Xenopus laevis (reviewed in Kobel and Du Pasquier, 1986), finally they could be two independently evolved genes. Because of a high level of sequence identity in the 5' untranslated sequence, as well as a very high degree of identity in the coding sequence (Figs 2, 3), we believe that the latter possibility is unlikely. The presence of multiple polyadenylation signals in the long 3' untranslated region of the 3.0 kb transcript raises the possibility that these transcripts are alleles of the same gene using different polyadenylation signals. Our attempts to resolve the issue by using different probes along our cDNAs, from translated and untranslated region, in Southern blot analysis are complicated by the fact that Xenopus is a pseudotetraploid organism and all probes hybridize to several bands. We are currently working on characterizing the En-2 genomic locus, which will allow us to resolve this issue definitively.

5A

B

MA
PP1
PP2

C

D

E

t   p

F

t

G

hb   mb   t   ot
di   tel

H

I
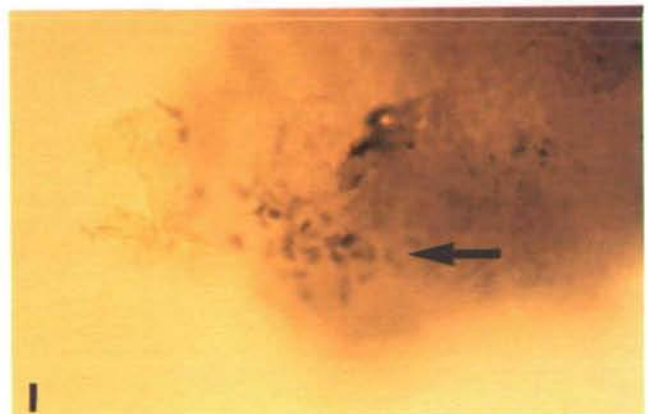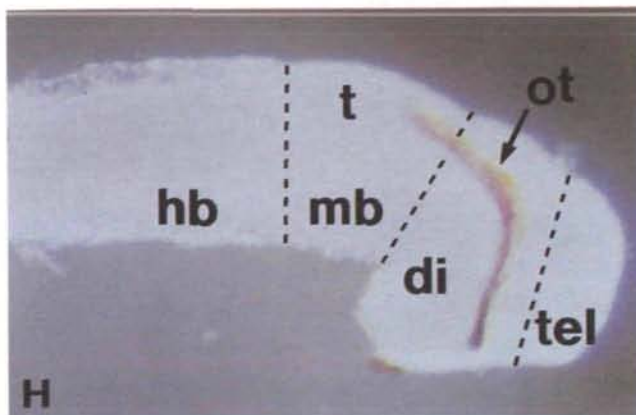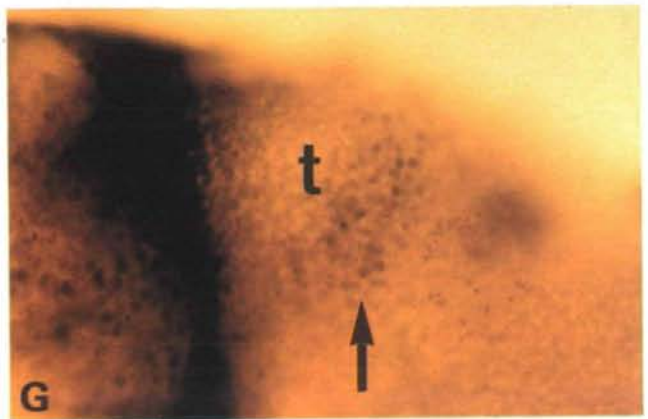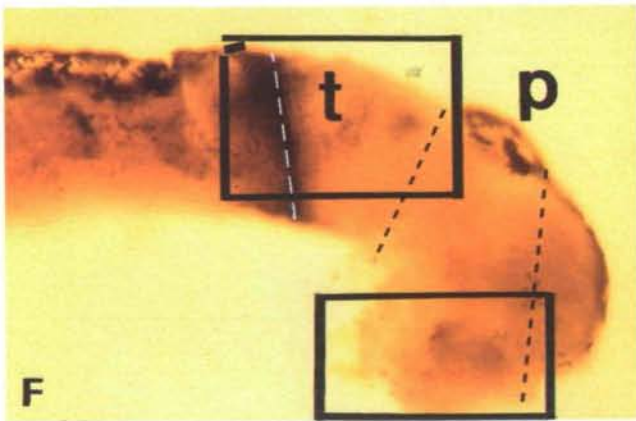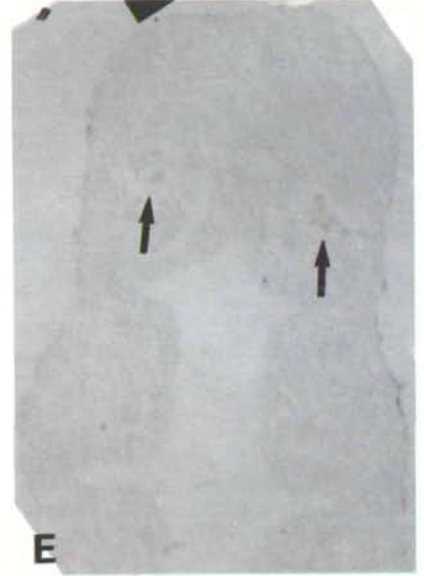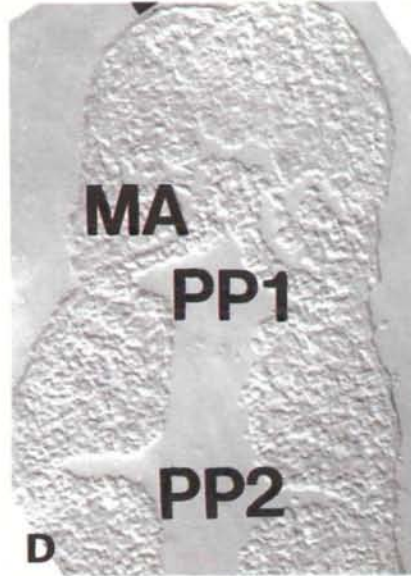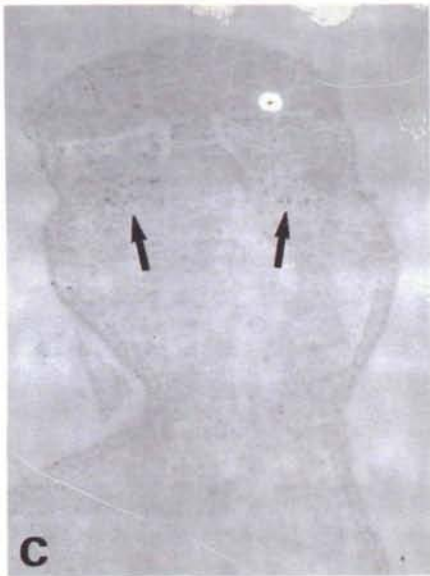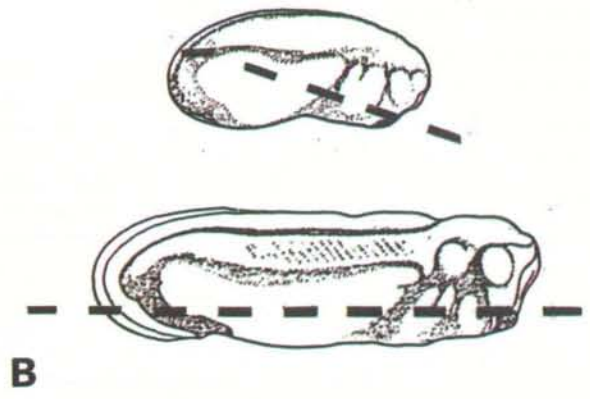
Protein localization in anterior structures
Staining embryos and dissected tissues under saturating conditions, allowed us to define three novel regions of *En-2* expression outside the midbrain–hindbrain boundary (Hemmati-Brivanlou and Harland, 1989). These regions include the mandibular arch, the optic tectum and the region of the pituitary. The expression of the *En-2* protein in most of these regions is in perfect agreement with the fate map study of the *Xenopus* neural plate and ridges done at stage 15 of embryogenesis by Eagleson and Harris (1990). According to the fate map, the stripe where *En-2* is expressed will develop into the following tissues: dorsal and ventral tegmentum (midbrain–hindbrain junction), the optic tectum and the choroid plexus (groups of cells in the dorsolateral side of the hindbrain, see Hemmati-Brivanlou and Harland, 1989). The fate map also predicts that *En-2* will be present in the cerebellum; *En-2* expression in the roof of the metencephalon (anterior hindbrain, also see Hemmati-Brivanlou et al. 1989) coincides with the cerebellar plate that later gives rise to the cerebellum. The latter staining is consistent with the detection of *En-2* RNA in the cerebellum in mouse embryos (Davis et al. 1988).

En-2 is also expressed in the optic tectum, the main target of the optic fibers. The first nerve fibres leave the retina at stage 28, cross the chiasm at stage 32, grow along the optic tract and begin to innervate the rostral optic tectum at stage 37/38 (Holt, 1984). *En-2*-positive cells are located in the lateral part of the mid-tectum. Optic fibres first invade the rostral part of the tectum where they contact the dendrites of cells whose somata are positioned more posteriorly. Thus the *En-2*-positive cells are likely to be the cells with which retinal axons form synapses. The tectum could be involved in the fine-tuned guidance of optic axons to the tectum and/or the selection of their appropriate synaptic partners. *En-2* may be involved in the regulation of expression of such trophic factors. Alternatively, the expression of *En-2* in the tectum may be one of a series of position-specific genes that are turned on in different parts of the brain and govern the expression of local positional guidance factors that are not diffusible (Harris, 1989). Experiments are currently in progress to analyze the behavior of the optic nerve in a region of the midbrain where *En-2* has been ectopically expressed.

The lowest level of expression of *En-2* in the CNS is in the pituitary region. These cells can be visualized by staining dissected brain tissue. In the mouse, *En-2* RNA has also been detected around the precursor of the anterior part of the pituitary, the Rathke's pouch; the RNA has not, however, been detected in the pituitary gland. This transient expression therefore resembles the transient expression of *En-2* protein in the *Xenopus* pituitary.

Between approximately stage 24 to stage 28 of embryogenesis, *En-2* protein is also expressed as a ring around craniofacial structures. This region is colinear with the band of expression in the midbrain–hindbrain junction and is tilted about 45° relative to the anterior–posterior axis. Close examination of the

whole-mount-stained embryos as well as frontal and transverse sections of embryos show that this pattern of expression coincides with the mandibular arch. It has been demonstrated in other vertebrate species that the mandibular arch is mostly, if not entirely, derived from neural crest cells (Couly and Le Douarin, 1988) located at the edge of the neural plate. We have previously shown that the band where the *En-2* protein is expressed at stage 14 extends from the neural groove to the ridges of the neural plate, a region that therefore includes a subset of neural crest cells (Hemmati Brivanlou and Harland, 1989). It is also interesting to note that the presence of an *engrailed* related transcript has been demonstrated in a subset of cranial neural crest cells in embryonic chick cell culture (Gardner et al. 1988). The transient nature of expression in the mandibular arch and the precedent in the chick embryonic cell culture suggests that the stained cells are neural crest derived. However, since the expression is transient, our attempts to find intermediary steps in the movement of these cells have so far been unsuccessful. Alternatively, it has been proposed that the superficial ectoderm covering the branchial arches may be part of a larger developmental unit that defines segmental units called 'ectomeres' in the cephalic region (Couly and Le Douarin, 1990). Ectomeres are arranged in metameric-like fashion in the presumptive facial and hypobranchial ectoderm. This metameric distribution parallels the segmentation of the neural anlage into individual neuromeres. *En-2* could be involved in the determination of fate of the cells of the most anterior ectomere in this stage of neurogenesis.

In a previous study, we found that *En-2* is only expressed in neurectoderm which is underlain by anterior mesoderm (Hemmati-Brivanlou and Harland, 1989). Subsequently, we showed that anterior notochord was the strongest inducer of *En-2* in the late gastrula (Hemmati-Brivanlou et al. 1990b). Interestingly in this study, we find that areas expressing *En-2* (the midbrain–hindbrain junction, the pituitary and the mandibular arch) all lie in the same plane, in a manner similar to cells that express XlHbox 1 (De Robertis et al. 1990). It is tempting to speculate that the same signal radiates from the anterior notochord dorsally to induce *En-2* in the midbrain–hindbrain junction, laterally to induce expression in the mandibular arch and ventrally to induce expression in the pituitary which at stage 36 is in the floor of the forebrain folded underneath the anterior tip of the notochord.

From the pattern of expression, we can conclude that the *En-2* protein, like its counterpart in *Drosophila*, probably possesses multiple functions during development. With the molecular tools in hand, we hope to dissect these biological activities in *Xenopus* embryos.

References

BROWER, D. L. (1986). *Engrailed* gene expression in *Drosophila* imaginal discs. *Embo J.* **10**, 2649–56.

COLEMAN, K. G., POOLE, S. J., WEIR, M. P., SOELLER, W. C. AND KORNBERG, T. (1987). The *invected* gene of *Drosophila*: sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. *Genes and Dev.* **1**, 19–28.

CONDIE, B. C. (1989). Ph.D. Thesis, University of California, Berkeley, USA.

CONDIE, B. C. AND HARLAND, R. M. (1987). Posterior expression of a homeobox gene in early *Xenopus* embryos. *Development* **101**, 93–105.

COULY, G. AND LE DOUARIN, N. M. (1988). The fate map of the cephalic neural primordium at the presomitic to the 3-somite stage in the avian embryo. *Development* **103 Supplement**, 101–113.

COULY, G. AND LE DOUARIN, N. M. (1990). Head morphogenesis in embryonic avian chimeras: *Development* **108**, 543–558.

DARNELL, D. K., KORNBERG, T. AND ORDAHL, C. P. (1986). *ChickEN*: a chick genomic clone with homology to the *Drosophila engrailed* homeobox. *J. Cell Biol.* **103**, 311a.

DAVIDSON, D., GRAHAM, E., SIME, C. AND HILL, R. (1988). A gene with sequence similarity to *Drosophila engrailed* is expressed during the development of the neural tube and vertebrae in the mouse. *Development* **104**, 305–316.

DAVIS, C. A., NOBLE-TOPHAM, S. E., ROSSANT, J. AND JOYNER, A. L. (1988). Expression of the homeo box-containing gene *EN-2* delineates a specific region of the developing mouse brain. *Genes and Dev.* **2**, 361–371.

DE ROBERTIS, E. M., OLIVER, G. AND WRIGHT, C. V. E. (1989). Determination of axial polarity in the vertebrate embryo: Homeodomain proteins and homeogenetic induction. *Cell* **57**, 189–191.

DINARDO, S., KUNER, J. M., THEIS, J. AND O'FARRELL, P. (1985). Development of embryonic pattern in *D. melanogaster* as revealed by accumulation of the nuclear *engrailed* protein. *Cell* **43**, 59–69.

DOLECKI, G. J. AND HUMPHREYS, T. (1988). An *engrailed* class homeo box gene in sea urchins. *Gene* **64**, 21–31.

EAGLESON, G. W. AND HARRIS, W. A. (1990). Mapping of the presumptive brain regions in the neural plate of *Xenopus laevis*. *J. Neurobiol* **21**, 427–440.

FJOSE, A., EIKEN, H. G., NJOLSTAD, P. R., MOLVEN, A. AND HORDVIK, I. (1988). A zebrafish *engrailed*-like homeobox sequence expressed during embryogenesis. *Febs Letters* **231**, 355–60.

GARDNER, C. A., DARNELL, D. K., POOLE, S. J., ORDAHL, C. P. AND BARALD, K. F. (1988). Expression of an *engrailed*-like gene during development of the early embryonic chick nervous system. *J. Neurosci. Res.* **21**, 426–37.

HARRIS, W. A. (1989). Local positional cues in the neuroepithelium guide retinal axons in embryonic *Xenopus* brain. *Nature* **339**, 218–221.

HEMMATI-BRIVANLOU, A., FRANK, D., BOLCE, M., BROWN, B., SIVE, H. AND HARLAND, R. M. (1990a). Localization of specific mRNAs in *Xenopus* embryos by whole mount *in situ* hybridization. *Development* **110**, 325–330.

HEMMATI-BRIVANLOU, A. AND HARLAND, R. M. (1989). Expression of an *engrailed*-related protein is induced in the anterior neural ectoderm of early *Xenopus* embryos. *Development* **106**, 611–617.

HEMMATI-BRIVANLOU, A., STEWART, R. AND HARLAND, R. M. (1990b). Region-specific neural induction of an *engrailed* protein by anterior notochord in *Xenopus*. *Science* **250**, 800–802.

HOLT, C. E. (1984). Does timing of axon outgrowth influence retinotectal topography in *Xenopus*? *J. Neurosci.* **4**, 1130–1152.

HUYNH, T. V., YOUNG, R. A. AND DAVIS, R. W. (1985). Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. In *DNA Cloning*, Vol. 1 (D. M. Glover, ed.), pp. 49–78. Washington, DC: IRL Press.

JOYNER, A. L., KORNBERG, T., COLEMAN, K. G., COX, D. R. AND MARTIN, G. R. (1985). Expression during embryogenesis of a mouse gene with sequence homology to the *Drosophila engrailed* gene. *Cell* **43**, 29–37.

JOYNER, A. L. AND MARTIN, G. R. (1987). *En-1* and *En-2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes and Dev.* **1**, 29–38.

KOBEL, H. R. AND DU PASQUIER, L. (1986). Genetics of polyploid *Xenopus*. *Trends Genet* **2**, 310–315.

KORNBERG, T. (1981). *engrailed*: A gene controlling compartment and segment formation in *Drosophila*. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1095–1099.

LAWRENCE, P. AND STRUHL, G. (1982). Further studies on the *engrailed* phenotype of *Drosophila*. *Embo J.* **1**, 827–33.

LOGAN, C., WILLARD, H. F., ROMMENS, J. M. AND JOYNER, A. L. (1989). Chromosomal localization of the human homeo box-containing genes, EN1 and EN2. *Genomics* **4**, 206–209.

MORATA, G. AND LAWRENCE, P. A. (1975). Control of compartment development by the *engrailed* gene in *Drosophila*. *Nature* **255**, 614–617.

NIEUWKOOP, P. D. AND FABER, J. (1967). *Normal Table of Xenopus laevis (Daudin)*. North Holland, Amsterdam.

PATEL, N. H., MARTIN-BLANCO, E., COLEMAN, K. G., POOL, S. J., ELLIS, M. C., KORNBERG, T. AND GOODMAN, C. S. (1989). Expression of *engrailed* protein in arthropods, annelids, and chordates. *Cell* **58**, 955–968.

PATEL, N. H., SCHAFER, B., GOODMAN, C. S. AND HOLMGREN, R. (1989). The role of segment polarity genes during *Drosophila* neurogenesis. *Genes and Dev.* **3**, 890–904.

PATEL, N. H., SNOW, P. M. AND GOODMAN, C. S. (1987). Characterization and cloning of fasciclin III: a glycoprotein expressed on a subset of neurons and axon pathways in *Drosophila*. *Cell* **48**, 975–988.

POOLE, S. J., KAUVAR, L. M., DREES, B. AND KORNBERG, T. (1985). The *engrailed* locus of *Drosophila*: Structural analysis of an embryonic transcript. *Cell* **40**, 37–43.

POOLE, S. J., LAW, M. L., KAO, F. T. AND LAU, Y. F. (1989). Isolation and chromosomal localization of the human *En-2* gene. *Genomics* **4**, 225–31.

SANGER, F., NICKLEN, S. AND COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

STEELE, R. E., UNGER, T. F., MARDIS, M. J. AND FERO, J. B. (1989). The two *Xenopus laevis src* genes are co-expressed and each produces functional pp60^{src}. *J. biol. Chem.* **264**, 10649–10653.

WEISBLAT, D. A., PRICE, D. J. AND WEDEEN, C. J. (1988). Segmentation in leech development. *Development* **104 Supplement**, 161–168.