*4874–4881* *Nucleic Acids Research, 1996, Vol. 24, No. 24* 

# Class III POU genes of zebrafish are predominantly expressed in the central nervous system

**Philipp Spaniol, Caroline Bornmann, Giselbert Hauptmann and Thomas Gerster***

Biozentrum der Universität Basel, Abteilung Zellbiologie, Klingelbergstraße 70, CH-4056 Basel, Switzerland

Received September 18, 1996; Revised and Accepted November 8, 1996 EMBL/GenBank/DDBJ accession nos Y07905–Y07907

## ABSTRACT

**POU genes encode a family of transcription factors involved in a wide variety of cell fate decisions and in the regulation of differentiation pathways. We have searched for POU genes in the zebrafish, a popular model organism for the study of early development of vertebrates. Besides five putative pseudogenes we have identified five POU genes that are expressed during embryogenesis. Probes obtained by PCR were used to isolate full-length cDNAs. Four of the isolated genes encode proteins with class III POU domains. Analysis of genomic clones suggests that the fish genes in general do not contain introns, similar to class III genes of mammals. However, the C-termini of two of the encoded proteins vary due to facultative splicing of a short intervening sequence. These two genes show very strong similarities in their sequence. They have probably arisen by gene duplication, possibly as part of a larger scale duplication of part of the zebrafish genome. Analysis of the expression of the class III genes shows that they are predominantly expressed in the central nervous system and that they may play important roles in patterning the embryonic brain.**

## INTRODUCTION

The zebrafish (*Danio rerio*) is a popular model organism to study early embryonic development of vertebrates (1–3). The completely transparent embryos develop rapidly and are easily accessible. Furthermore, large-scale mutagenesis screens have been performed in several laboratories identifying hundreds of genes involved in controlling early development (4,5). Analysis of these mutants will be instrumental for an increased genetic understanding of developmental pathways in vertebrates. Many genes involved in the regulation of early embryogenesis have been isolated from other species. Molecular characterization of homologs of many such genes in zebrafish is complementing the genetic approaches and will provide possible candidate genes for some of the obtained mutants (see e.g. ref. 6).

POU genes encode a class of DNA-binding proteins interacting with DNA through a bipartite domain of ~150 amino acids. The POU domain consists of a N-terminal POU specific region separated by a short linker from a particular type of homeodomain (reviewed in refs 7,8). Both the POU-specific and the homeodomain contribute to DNA binding each via helix–turn–helix motifs. The two subdomains are also involved in several types of protein–protein interactions (reviewed in ref. 8). POU domain proteins generally act as transcription factors but some of them also have additional functions e.g. in DNA replication. POU genes are found all across the animal kingdom. According to the sequence of their POU domain these genes have been grouped into at least six classes (reviewed in ref. 9).

Some POU genes, like *oct-3/4* in mice or *pou-2* in zebrafish, are expressed very early during embryogenesis and are possibly involved in controlling some of the first steps of development (10–13). Expression of other POU genes is initiated later during development. Many of those are transcribed in the forming central nervous system, especially the genes with a class III POU domain. In mammals each of these genes show a very elaborate expression pattern in the brain (reviewed in ref. 14). Similarly, the zebrafish class III gene *zp-50* shows a very complex and dynamic expression pattern in the embryonic brain (15). Genetic analysis of naturally occurring or genetically engineered mutations demonstrate that POU genes are involved in cell fate decisions and in the control of terminal differentiation. For example, genetic defects in several class III genes have been investigated: naturally occurring mutations of *brn-4* in human patients lead to profound sensorineural deafness (16) whereas knock-outs of the mouse *brn-2* gene cause the loss of several types of neurons in the hypothalamus (17,18).

To investigate POU genes that are potentially involved in the control of zebrafish embryogenesis we isolated several cDNAs with POU domain probes obtained by PCR. We have previously described the sequences of two of the identified genes (13,15). In this communication we report the sequences of three more genes encoding putative transcription factors containing class III POU domains. One of these genes has been independently found by others (19). Some of the genes appear to have arisen by gene duplication events that happened after the separation of the actinopterygian lineage from higher vertebrates. The transcripts of two class III genes are alternatively spliced which leads to the formation of POU domain proteins with variable C-termini. This appears to cause a diversity of class III proteins in zebrafish that is larger than in mammals. All these POU genes are expressed predominantly in the central nervous system and may thereby be involved in patterning the developing brain.

*To whom correspondence should be addressed. Tel: +41 61 267 20 72/64; Fax: +41 61 267 20 78; Email: Gerster@ubaclu.unibas.ch

## MATERIALS AND METHODS

### PCR reactions

PCR reactions on genomic DNA were done first with 35 cycles using the two following primers: 5′-TITWYGGIAAIGTITTY-WSICARACIAC-3′ (64-fold degeneracy) and 5′-TGGTTYTG-YAAYMGIMGICARAAR-3′ (256-fold degeneracy). The annealing temperature was 50°C. An aliquot of amplified material was then further amplified over 38 cycles with the nested primers 5′-CCGGAATTCYTIAAIAAYATGTGYAARYT-3′ (32-fold degeneracy) and 5′-GTIRTIMGIGTITGGTTYTGYAAGGAT-CCGGG-3′ (16-fold degeneracy) and annealing at 52°C. For PCR of cDNA libraries 3–6 × $10^7$ phage particles were heated to 100°C for 10 min. Their DNA was then amplified with either of the two primer pairs mentioned above. PCR products were treated with Geneclean™ (Bio101) glass beads to remove low molecular weight DNA. The ends were polished with T4 DNA polymerase and the DNA fragments were phosphorylated with T4 polynucleotide kinase before subcloning and sequencing.

### Library screening and DNA analysis

Standard procedures were used to isolate cDNAs and genomic clones for the different *zp* genes (20). Probes obtained in the original PCR screens were used to probe 5 × $10^5$ plaques of 9–16 h post fertilization (hpf) (neurula) or 20–28 hpf (postsomitogenesis) cDNA libraries constructed in λZAPII (prepared by R. Riggleman and K. Helde, a kind gift from D.J. Grunwald). Inserts of the phages were subcloned into BSSK– by *in vivo* excision. Nested deletions were generated from both ends of the longest cDNAs by the Erase-a-Base system (Promega). These deletions were used for sequence determination using the Sequenase kit and $^{35}$S-dATP (USB/Amersham). Sequences were analyzed by the GCG program suite of DNA analysis programs. A genomic library prepared in λFIXII (Stratagene) was screened using full-length cDNAs of the class III POU genes. Inserts of positive phages were characterized by restriction analysis.

### RNA analysis

RNA was prepared from embryos or adult tissues. For Northern blot analysis 10 µg total RNA were run on formaldehyde agarose gels (20), blotted to nylon filters and hybridized to $^{32}$P-labelled cDNA probes.

*In situ* hybridization analysis was carried out as described (21). Digoxigenin-labelled RNA probes were generated by T7 RNA polymerase transcription of subcloned restriction fragments. In the case of *zp-12* we used a ~570 nucleotide (nt) probe specific for the 5′ UTR reaching from the 5′-end of clone #20 to the *Xmn*I site at –27. The ~840 nt probe for *zp-23* extended from the *Bsm*I site downstream of the second stop codon through most of the 3′-UTR to a *Taq*I site just before the polyadenylation site. The ~2115 nt long *zp-47* probe was derived from an exonuclease deletion clone removing the poly-A tail. That probe started at the *Asp*718 site 59 nucleotides upstream of the end of the open reading frame and extended through the entire 3′-UTR region.

## RESULTS

### PCR screen for zebrafish POU genes

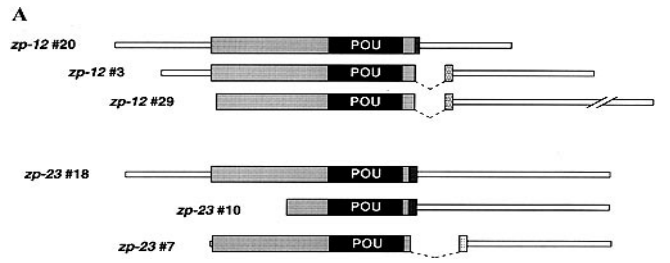
To identify zebrafish POU genes we performed two consecutive rounds of PCR reactions on genomic DNA. We used nested primers targeting conserved regions in POU domains of other species. Sixty randomly picked clones of the subcloned population of PCR products were analyzed by sequencing. Forty-one clones were related to POU domains and could be grouped into 11 different classes (data not shown). Sequence inspection of five classes obviously suggested that they represented either pseudogenes or that they arose only during PCR amplification e.g. by ‘polymerase jumping’ (22).

Since the initial PCR screen was carried out using genomic DNA, intron containing POU genes might have been missed. Therefore, we repeated the PCR reactions on phage DNA from different embryonic cDNA libraries using either of the two primer pairs. The frequency of (mostly artefactual) non-POU sequences was higher in the cDNA screens, probably due to the fact that in these experiments only one round of amplification was performed. Of the 25 subcloned PCR amplificates from a 6–9 hpf library four isolates seemed to encode a novel type of POU domain. This GP-9 group of isolates was subsequently used to clone full-length cDNAs of the zebrafish *pou-2* gene as described elsewhere (12,13). Of the total 37 isolates analyzed from 9–16 hpf neurula or 20–28 hpf postsomitogenesis libraries, three more clones were derived from the *pou-2* gene and nine isolates were virtually identical to sequences present in three groups of PCR products identified previously in the genomic screen. These groups were represented by isolates termed ZP-12, ZP-23 and ZP-47.

To determine whether any other POU gene is expressed at some point during embryogenesis, we $^{32}$P-labelled uncloned PCR products amplified from the cDNA libraries with both sets of amplification primers. These cDNA derived probes were hybridized to dot blots containing DNA from all our zebrafish POU isolates. Using the three different embryonic libraries we found positive hybridization signals in the cases of *pou-2* and of four (ZP-12, ZP-23, ZP-47 and ZP-50) of the eight potential POU genes identified by amplification of genomic DNA (data not shown). The other four genes that were apparently not expressed encoded potential POU domains whose sequence deviated considerable from those found in the databases. Therefore, we consider them to be pseudogenes although we cannot exclude that these genes are expressed at a very low level or at some later point during development. The four genes identified by PCR amplification of genomic DNA that are expressed during embryogenesis all encode class III POU domains. We used representative PCR isolates for each gene (ZP-12, ZP-23, ZP-47 and ZP-50) to isolate full length cDNAs. We report below the molecular characterization of the *zp-12*, *zp-23* and *zp-47* genes, whereas the sequence and expression of *zp-50* has been described earlier (15).

### Cloning of *zp-12* cDNAs

The postsomitogenesis cDNA library was screened with the ZP-12 isolate of the initial PCR screen. A PCR fragment with high similarity to ZP-12 was recently identified also by others (23). From screening 500 000 plaques we identified 36 plaques which repeatedly hybridized to the ZP-12 probe. The DNA sequence of three overlapping clones (#3, #20, #29) was determined (EMBL accession Y07906). Clones #20 and #3 contain an apparently full-length open reading frame (Fig. 1). The frame begins with a motif (MATAA) highly conserved among vertebrate class III POU domain genes. This presumptive start ATG is preceded three codons upstream by a stop codon. cDNA #29 starts 28 bp downstream of the putative initiation codon. In




B

ZP-12

```
  1  MATAASNPYL ASSSILSSAS LVHSESGGGG MQPGSGAVTS VSGGYRGDPT VKMVQSDFMQ   60
 61  GAMAASNGGH MLSHAHQWVT SLPHAAAAAA AAAAAAAAEA GSPWSSSPVG MAGSPQQQDV  120
121  KSSSNREDLH SGTALHHRPS HLGAHQSHQS AWGGTTASHI STITGGQQQS QQSLIYSQPG  180
181  GFTVNGMLNP PGSLVHPGLM RGESPEMDHF HEHHHHQQQH PHHHHQHHA GVHSHDSHSD  240
241  EDTPTSDDLE QFAKQFKQRR IKLGFTQADV GLALGTLYGN VFSQTTICRF EALQLSFKNM  300
301  CKLKPLLNKW LEEADSTTGS PTSIDKIAAQ GRKRKKRTSI EVSVKGALES HFLKCPKPSA  360
361  QEITSLADNL QLEKEVVRVW FCNRRQKEKR MTPPGVPQTP EDVYTHAGNV SADTPPPSMD  420
421  CKR

C-terminus spliced      424  EVCGRLLKRC KFER
C-terminus unspliced    424' MFTET
```

ZP-23

```
  1  MATAASNPYL ASSSILSSGS IVHSDSGGGM QQGSAAVTSV SGGYRGDPTV KMVQSDFMQG   60
 61  AMAASNGGHM LSHAHQWVTS LPHAAAAAAA AAVAAAAECS PWSSSPVGMT GSPGQQDVKN  120
121  NSGRDDLHSG TALHHRAPHL GPHQTHAGAM GSTTAAHIPS LTQSQQQQQS LIYSQPGGFT  180
181  VNGMLSPPGS QSLVHPGLVR GDTPELDHSS HHHHHHHQHQ HHQQAHHGVH SHDPHSDEDT  240
241  PTSDDLEHFA KQFKQRRIKL GFTQADVGLA LGTLYGNVFS QTTICRFEAL QLSFKNMCKL  300
301  KPLLNKWLEE ADSSTGSPTS IDKIAAQGRK RKKRTSIEVS VKGALESHFL KCPKPSAQEI  360
361  TSLADNLQLE KEVVRVWFCN RRQKEKRMTP PGVPQTPEDV YSQVGN

C-terminus spliced      407  GHFLVDYLKC ASLTGPSEPG DQRVTTTSSF HQVILAH
C-terminus unspliced    407' VSADTPPPSM DCKRMFSET
```

ZP-47

```
  1  MATTASNHYN IUTSSPSIVH SEPGSHQQAT AYRDAQTLLQ SDYSLQSNSH PLSHAHQNIT   60
 61  ALSHGEGGPW SSSPLGEQDI KPAVQSPRDE MHNSSNLQHQ SRPPHLVHQT HGNHHDSRAW  120
121  RTTTAAHIPS MATSNGQSLI YSQPSFSVNG LIPGSGQGIH HHSMRDAHED HHSPHLSDHG  180
181  HPPSQHQHQS HQSHHGHSDE DTPTSDDLEQ FAKQFKQRRI KLGFTQADVG LALGTLYGNV  240
241  FSQTTICRFE ALQLSFKNMC KLKPLLNKWL EEADSTGGSP TSLDKIAAQG RKRKKRTSIE  300
301  VSVKGALESH FLKCPKPAAS EITSLADSLQ LEKEVVRVWF CNRRQKEKRM TPPGGPLPGT  360
361  EDVYGDTPPH HGVQTPVQ
```

**Figure 1.** (**A**) Schematic representation of the organization of the cDNAs sequenced from *zp-12* and *zp-23*. White narrow rectangles indicate untranslated 5′ and 3′ sequences, broad shaded boxes the coding region including the POU domain, and broken lines indicate the spliced out intron. (**B**) Predicted sequence of the ZP-12, ZP-23 and ZP-47 proteins. The POU domains are indicated by bold type. The alternative C-termini of ZP-12 and ZP-23 are listed separately under the main body of the respective protein shared by either sequence variant. cDNA sequences have been deposited in the EMBL sequence database under accession numbers Y07906, Y07907 and Y07908. Alignments of the cDNA sequences with the corresponding amino acid sequences can be accessed on the NAR web site.

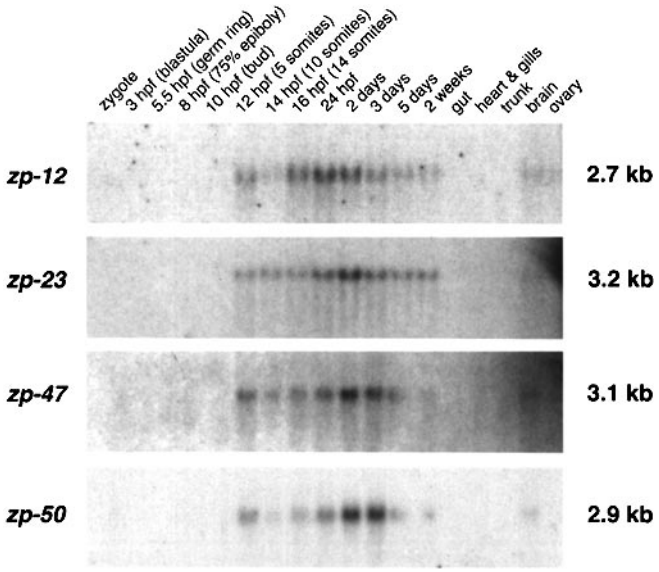



**Figure 2.** Northern blot analysis of *zp* gene expression. Total RNA of the indicated embryonic stages or from adult tissues was size fractionated on agarose gels and blotted. In case of RNA from adult tissues the 'gut' RNA was obtained from all the intestinal organs, 'trunk' RNA was prepared from mostly muscle tissues obtained from the tail, and 'brain' RNA was purified from the brain, the eyes and the attached cranial nerves. Filters were hybridized to radioactive probes specific for either *zp-12*, *zp-23*, *zp-47* or *zp-50*. The estimated size of each transcript species is indicated.

the second half of the reading frame we find a class III POU domain that is 98% conserved with that of mouse *brn-1*. Except for two nucleotide mismatches the three cDNAs are nearly identical over the entire length of the reading frame. Clones #3 and #29 differ however significantly from #20 by the fact that a short intron has been spliced out from their corresponding RNAs (Fig. 1). This 184 bp sequence starts with a canonical GT motif and ends with a consensus AG sequence (24). Most of the open reading frame is unaffected by the splicing event but the resulting proteins have slightly different C-termini: in the unspliced version of the mRNA represented by clone #20 the reading frame is extended by five codons encoded by the facultative intron. If the intron is spliced out, as in cDNAs #3 and #29, the second exon contributes 14 amino acids to the body of the reading frame shared by all three clones. Restriction analysis of additional isolated cDNAs suggests that the intron is spliced out in the majority of the cases (data not shown). Although Northern blot analysis (Fig. 2) suggests a homogenous size of *zp-12* mRNAs we found cDNAs with variable 3′-ends: cDNA #3 ends 14 nucleotides downstream of a perfect AATAAA polyadenylation motif. In contrast, cDNA #20 ends just upstream of an A-rich sequence between nucleotides 1861 and 1880 which may have erroneously served as template for the oligo-dT primers during library preparation. Partial sequencing of additional isolated cDNAs (data not shown) showed 3′-ends with short poly-A stretches at several positions which however were not preceded by obvious AATAAA motifs.

## Cloning of *zp-23* cDNAs

We screened the neurula library prepared from 9–16 hpf embryos with the ZP-23 isolate of the initial PCR screen. Fifteen of 500 000 screened plaques repeatedly hybridized with the ZP-23 probe. The two longest inserts of 2.1 and 2.3 kb were entirely sequenced. Clone #10 is shorter at the 5′-end (Fig. 1) resulting in a truncation of the open reading frame by more than 150 amino acids in comparison to the ORF of clone #7. The predicted N-terminus encoded by clone #7 starts with a MATAA motif conserved among other class III POU domain genes. Besides the 5′-end variation, clone #7 differs from #10 downstream of the POU domain by the removal of an intron sequence of 309 nt. Thereby 37 amino acids of clone #7 replace 19 amino acids at the C-terminus encoded by the unspliced cDNA #10. Restriction analysis of additional cDNAs suggests that the majority of *zp-23* transcripts are unspliced (not shown). Since Northern analysis (Fig. 2) suggested a considerably longer transcript size of around 3 kb we screened with the entire *zp-23* #7 cDNA the cDNA library prepared from 1 day old embryos. We found nearly 100 positive phages in the 500 000 screened plaques. The longest insert (#18) was sequenced (EMBL accession Y07907). This

cDNA was derived from an unspliced mRNA and extended the region of the 5′ untranslated region by 0.5 kb (Fig. 1).

Apart from the spliced out intron, isolate #7 is nearly identical to the sequenced cDNAs #10 and #18, except for the insertion of a CAG trinucleotide at amino acid position 116. An additional five mismatches or deletions both in and outside the coding region have no functional consequences. There is a microheterogeneity at the immediate 3′-end of the three cDNAs resulting in the addition of the poly-A tail either 12 or 16 nucleotides downstream of a canonical AATAAA polyadenylation signal. A RNA (*zfpou1*) with strong similarity to the unspliced *zp-23* #18 has been reported by others (19). However, its open reading frame deviates at 12 nucleotide positions from the sequences of our three *zp-23* cDNA isolates thereby changing eight amino acids in the encoded protein.

### Cloning of *zp-47* cDNAs

The cDNA library prepared from 1 day old embryos was also screened with the ZP-47 isolate of the initial PCR screen. From 500 000 screened plaques eight phages were recovered that repeatedly hybridized to the ZP-47 probe. The clone with the longest insert was subjected to the generation of nested deletions and these were subsequently used for sequence determination (Fig. 1; EMBL accession Y07905). Although the long open reading frame extends up to the 5′-end of the cDNA we believe that translation starts at the first ATG codon found 41 nt downstream of the 5′-end. This start codon is part of a MATTA motif as found in other class III POU genes. The POU domain itself is found as in other class III genes in the second half of the protein. The cDNA ends with a long 3′ untranslated region of over 2 kb in length. This 3′-UTR terminates with an AATAAA polyadenylation signal followed 27 nucleotides downstream by a long poly-A tail.

### Isolation of genomic clones for *zp-12*, *zp-23*, *zp-47* and *zp-50*

We have used the longest cDNA isolates for each of the four class III POU genes to screen a zebrafish genomic library. Inserts of repeatedly hybridizing phages were subcloned and analyzed by restriction digestion (not shown). For this purpose we used restriction enzymes that were cleaving sites close to the 5′- and 3′-ends of the cDNAs. We compared the size of several restriction fragments of genomic clones with fragments derived from the corresponding cDNAs. In all cases we could not detect significant differences in size. This suggests that the genomic sequences are colinear with their respective cDNAs. Therefore, the four genes *zp-12*, *zp-23*, *zp-47* and *zp-50* apparently do not contain intron sequences, with the exception of the facultatively spliced small intervening sequence at the end of the open reading frames of *zp-12* and *zp-23*. Our finding that all four class III POU domain genes isolated in the zebrafish lack introns mirrors the absence of introns in the murine class III genes *brn-1*, *brn-2*, *brn-4* and *oct-6* (25). The non-existence of introns in the class III POU genes contrasts, however, with the presence of multiple introns in other POU genes (see e.g. 12,13,26,27).

### Analysis of *zp* gene expression

Total RNA was prepared from various developmental stages and of several adult tissues. In all cases single major RNA species were detected after hybridization with probes specific for the longest available cDNA sequences (Fig. 2). For *zp-47* a RNA was detected that had an estimated size of ~3.1 kb. For the *zp-50* gene we observed a RNA of ~2.9 kb. In the case of the highly related *zp-12* and *zp-23* genes we estimated transcripts sizes of ~2.7 and ~3.2 kb. The two bands predicted from the alternative splicing of the latter two genes were apparently not resolved in these blots.

Transcripts of all four identified class III genes begin to accumulate between 10 and 12 hpf just after completion of the gastrula period. Maximal RNA levels are observed in all cases after 1 to 2 days of development. *zp* gene expression slightly decreases during the following days but is still clearly detected 2 weeks after fertilization. In adult zebrafish we could detect transcripts of all four genes only in the brain. At the sensitivity of these Northern blots we can, however, not exclude that these genes might also be expressed in other adult tissues, albeit at lower levels.

To investigate the sites of embryonic expression of the *zp* genes we began to study their expression pattern by *in situ* hybridization. In the case of *zp-50* we have found previously that this gene is expressed in a highly dynamic and complex pattern in several regions of the central nervous system (15). To investigate the expression patterns of the other genes we prepared probes specific mostly for untranslated regions thus ensuring that the probes do not crossreact with other class III mRNAs. Similarly to *zp-50*, we found predominant expression of *zp-12*, *zp-23* and *zp-47* in various portions of the CNS (Fig. 3). All *zp* genes are thus expressed in all the major subdivisions of the embryonic CNS, including the fore-, mid- and hindbrain and the spinal cord. Interestingly, the precise arrangement of the expression domains varies between the different genes. Whereas *zp-47* shows some similarities of expression to the *zp-50* pattern, it differs from *zp-12* and *zp-23* expression. For example, *zp-47* and *zp-50* are strongly expressed in the cerebellum whereas the other two genes are not. Although *zp-12* and *zp-23* show in general very similar expression profiles, there are also differences, e.g. in the telencephalon where *zp-23* expression is more widespread. Besides their predominant expression in the CNS *zp-12* and *zp-23* transcripts were detected at notable levels also in the pronephric duct (Fig. 3B and D).

## DISCUSSION

After the discovery of the family of POU genes it was initially attempted to group the family members into four subclasses based on differences in their POU domains (28). The discovery of additional POU genes led to the expansion of this scheme to six different subclasses (7,9). Within any of these classes, genes show high similarity within the POU domain. In addition, a limited degree of homology among class members is also found outside of the DNA binding region, especially among vertebrate genes. While the majority of known POU domain genes can be easily fitted into one of the six subclasses, the sequence comparison in Figure 4 shows that some genes cannot be easily assigned to a particular class. These include the zebrafish *pou-2* gene (12,13), the *Xenopus oct-25*, *oct-60*, *oct-79* and *oct-91* genes (29–31), the rat *sprm-1* gene (32), and the nematode *ceh-18* gene (33). Similarly, the partial sequences of the planarian *dtpou-2* and *djpou-2* POU domains deviate considerably in their homeodomains from other POU genes (34,35) and might thus have to be grouped into a separate class, at least until the evolutionary relationships of invertebrate POU genes are better understood.

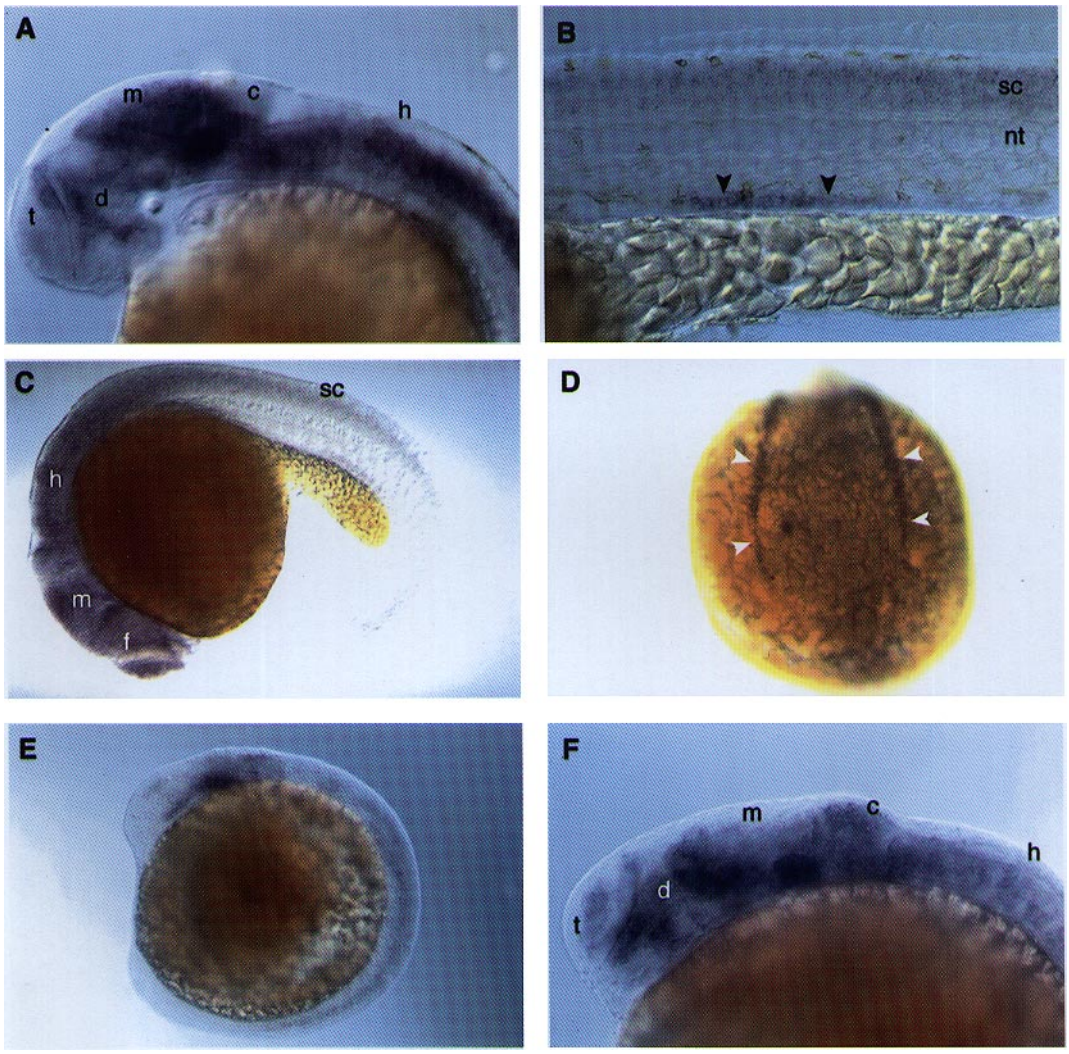



**Figure 3.** *In situ* hybridization of early zebrafish embryos. Probes specific for *zp-12* (**A** and **B**), *zp-23* (**C** and **D**) and *zp-47* (**E** and **F**) have been hybridized to zebrafish embryos and whole-mount preparations were photographed as side views under Nomarski optics (A–C, E–F) or as a dorsal view under a dissection microscope (D). *zp-12* expression is well visible at 27 h post fertilization in several regions of the brain (A) and in the spinal cord and pronephric duct (arrowheads) (B). *zp-23* expression in the pronephric primordium is already visible at the 10 somite stage (arrowheads in D) and expression in all major portion of the CNS is well visible at the 23 somite stage (C). Expression of *zp-47* in distinct domains of the forming CNS is already clearly visible around the 7 somite stage (E) giving then rise to an elaborate pattern of expression at 27 hpf in the fore-, mid- and hindbrain and in the spinal cord (F). Abbreviations: c, cerebellum; d, diencephalon; f, forebrain; h, hindbrain; m, midbrain; nt, notochord; sc, spinal cord; t, telencephalon.

Despite the problem to accommodate all POU genes into one of the proposed six subclasses the four *zp* genes are all obviously of the class III type. The POU domains are >90% identical to those of mammalian class III genes. There are also considerable homologies outside of the POU domain (Fig. 5). Similarly to the intron-lacking mammalian class III POU genes (25), the zebrafish *zp* genes do not contain introns, except for the short sequences at the immediate C-termini of *zp-12* and *zp-23* which are removed in a subset of transcripts. Resembling the mammalian genes also functionally, our analysis shows that the *zp* genes are primarily expressed in the embryonic and adult central nervous system.

Notwithstanding the fact that our screen identified in zebrafish the same number of four class III genes as reported in mammals (25), it is not possible to assign each of the fish gene a direct homolog in the mammalian genome (Fig. 5). For example, no obvious homolog has yet been found in zebrafish for the mammalian *brn-4* and the amphibian *xlpou2* genes. As pointed out previously, *zp-50* is probably the homolog of *xlpou1* in *Xenopus* (15). This gene pair shows some similarity to mammalian *oct-6*. Nevertheless, differences in the expression argue against the assumption that *zp-50/xlpou1* are functional homologs of *oct-6*. On the other hand, the pair of highly related *zp-12* and *zp-23* genes shows very high sequence similarity to mouse *brn-1*. Both zebrafish genes are expressed in very similar domains in the developing central nervous system and in the pronephric Wolffian duct. This expression pattern mirrors the embryonic expression of mammalian *brn-1* detected in several domains of the brain, as well as in the fetal kidney (28). Therefore, we consider both *zp-12* and *zp-23* to be direct homologs of mammalian *brn-1*. The *zp-47* gene is most closely related to *xlpou3* of *Xenopus* and may be a direct homolog of mammalian *brn-2*. Interestingly, a partial sequence of another zebrafish class III POU domain gene, termed *brn-1.2*, has been recently reported (23). Over the known 112 amino acids this sequence differs from the *zp-47* POU domain at

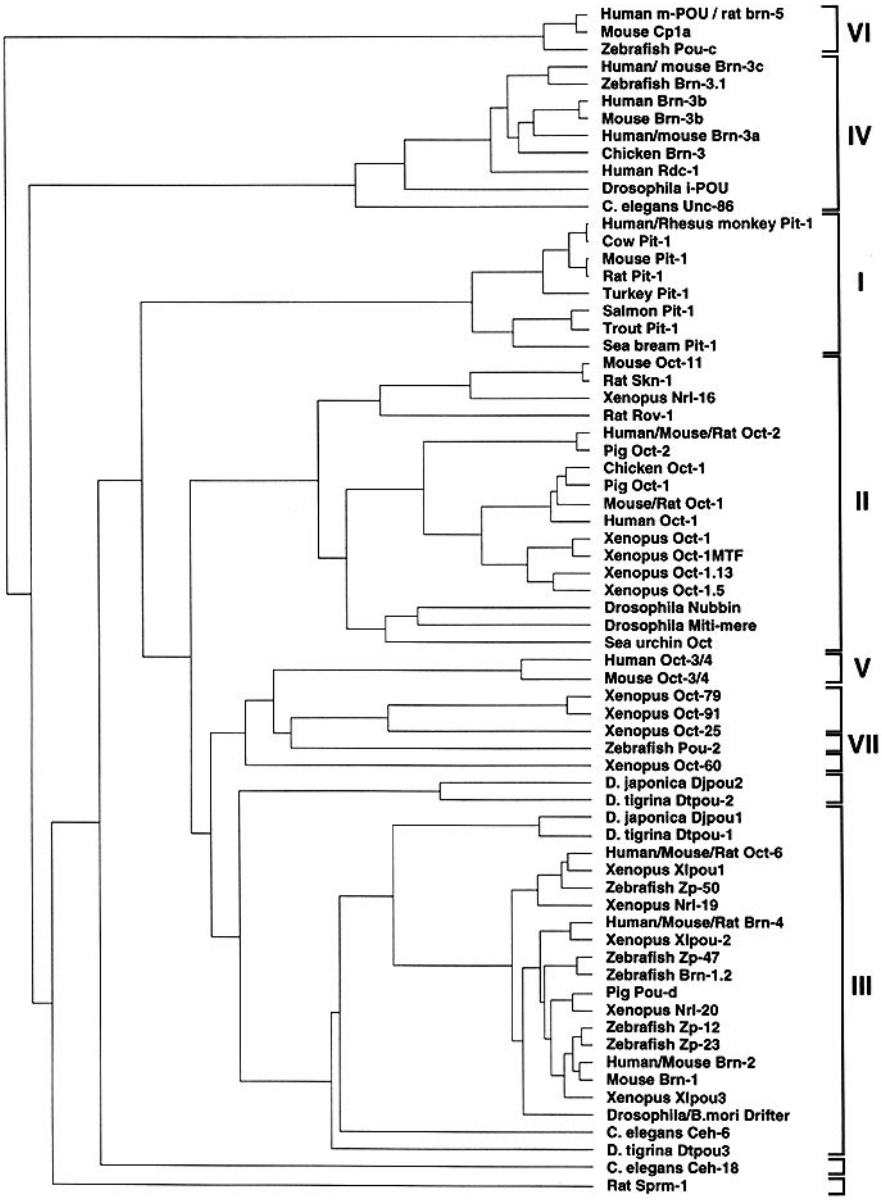



**Figure 4.** Sequence comparison of all POU domains. POU domains deposited in the EMBL and GenBank databases were compared to each other. Their relationship is displayed by the output of the GCG program Pileup. The generally used classification into subgroups is indicated by roman numerals. An alignment of the corresponding sequences can be viewed at the NAR web site.

only two positions. Nevertheless, sequence identity at the nucleic acid level is only ~83%. This clearly demonstrates that *zp-47* and *brn-1.2* are distinct although highly related genes.

In our PCR screen we have identified only class III POU genes and the prototype of a novel class of POU domains (*pou-2*) (13). Comparing the sequences of the PCR primers to POU domain sequences of other species it is clear in retrospect that the homologies were not high enough to amplify members of all the major POU gene classes. In our screen no members of classes II (*oct-1*-like genes) or V (*oct-3/4*) have been amplified. This is unexpected since the primers match well the POU domain sequences of these classes in other species and were capable in control experiments to amplify POU domains of the human *oct-1* and *oct-2* genes, the *Drosophila nubbin* gene and of the murine *oct-3/4* gene (data not shown). Since at least class II genes have been found not only in mammals but also in amphibians and even in invertebrates like sea urchin and *Drosophila*, it would be expected that members of this class were also present in zebrafish. It is currently not clear why we have not identified such genes in the PCR analysis of the cDNA libraries. Possibly, the genes are expressed either at low levels or only at later time points of embryogenesis. Another gene that has certainly been missed in our screen is the *brn-1.2* gene (23) related to *zp-47*. Furthermore, we have not found a direct homolog of the class III genes *xlpou2* (*Xenopus*) and *brn-4* (mammals).

The protein coding sequences of the *zp-12* and *zp-23* genes are highly similar over their entire length. They share, except for their divergent C-termini, a sequence identity of nearly 90%. This sequence similarity suggests that these genes have a common origin and have arisen by gene duplication which must have occurred

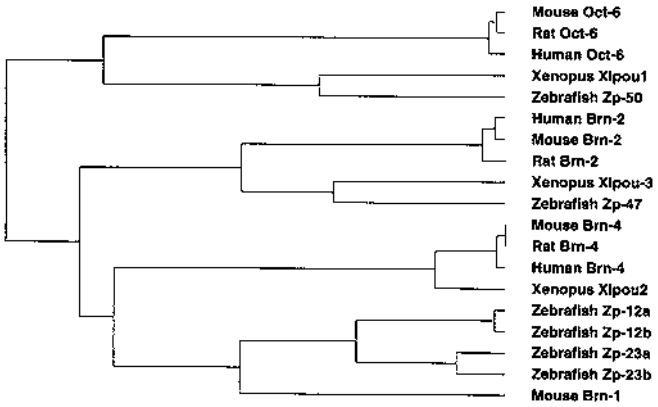



**Figure 5.** Sequence comparison of full-length vertebrate class III POU protein sequences. Those class III proteins of which full-length sequences have been determined were compared. The relationship over the entire length of the proteins has been determined by the GCG program Pileup. An alignment of the corresponding sequences can be viewed at the NAR web site.

after the actinopterygii diverged from the higher vertebrates. The common ancestry of the two genes is also reflected in their splicing pattern: despite the general absence of introns from class III POU genes short intervening sequences are spliced out near the 3′-end of the coding region during processing of a fraction of *zp-12* and *zp-23* transcripts. As a result proteins differing at their C-termini are produced. Since the POU domains are unaltered it can be assumed that these proteins maintain the same DNA-binding potential. However, the amino acid variations downstream of the POU domain might modulate interactions with other regulatory proteins, e.g. analogous to the auxiliary factor P-OTX which specificially interacts with the N-terminus of the POU protein Pit-1 but not with Brn-2 (36). At least in the case of the *zp-23* reading frame the variation of the primary sequence might also affect posttranslational protein modifications since splicing of the transcript removes a potential N-linked glycosylation site (37). Although the *zp-12* and *zp-23* genes resemble each other in their overall splicing pattern, they differ in the details since the position of the 5′ splice site of the intron of *zp-23* is 14 codons more upstream than the 5′ splice site used in two of the *zp-12* cDNAs that we sequenced. An unpublished longer cDNA has been deposited in GenBank (38) that is related to the *brn-1.1* PCR fragment (23) which is derived from the *zp-12* gene. That cDNA appears to lack a sequence corresponding roughly to the intron removed in *zp-23* clone #7. However, it is not clear whether the sequence absent in this *brn-1.1* clone is really an intron since comparison of this segment with our full-length *zp-12* sequence does not show good matches to splicing consensus sequences (24).

Despite the high degree of amino acid sequence conservation over the entire length of the *zp-12* and *zp-23* reading frames the corresponding DNA sequence is only 78% identical, whereas 5′ and 3′ untranslated regions display even less than 65% identity. This indicates that the duplication of these genes has occurred a considerable time ago allowing since the DNA sequences to diverge. Nevertheless, there appears to have been substantial evolutionary pressure to maintain the amino acid sequence both within as well as outside of the POU domain. This is remarkable since in the cases of many transcription factor families the DNA binding motifs often remain highly conserved whereas sequences outside of the DNA binding regions tend to be much less conserved. Similar evolutionary pressure may have applied to maintain the *zp-47* and *brn-1.2* coding sequences relatively homogenous. However, more sequence information about *brn-1.2* will be needed to investigate this point further. Several reasons can be imagined why protein sequences of duplicated genes should remain closely conserved: POU genes with highly similar sequences could be expressed in different cells of the brain. Nevertheless, the genes may still be regulating the transcription of related sets of target genes and could do so only if their sequences are highly preserved. In another scenario, highly related POU genes could be coexpressed in the same cells. Eventual expression of sequence variants could lead to aberrant transcription of some of the target genes which in turn could negatively affect cell functions. Alterations of the protein sequences could therefore not be tolerated.

Our analysis of *zp* gene expression during embryogenesis shows that the different genes are primarily expressed in the brain, although the expression of *zp-12* and *zp-23* in the pronephric duct suggests that some genes may also be involved in patterning of other tissues. The expression patterns of *zp-12* and *zp-23* in the embryonic brain show strong similarities but they are not entirely overlapping. Most importantly, these genes are not expressed homogeneously throughout the brain. They are strongly expressed only in some regions of the brain but not, or at much lower levels, in others. Conversely, the *zp-47* expression pattern in the embryonic brain is distinct and resembles in part that reported for *zp-50* although several differences exist. It therefore appears that each class III POU gene is expressed in a unique pattern that partly overlaps with expression domains of other class III genes. This is somewhat reminiscent of the overlapping ranges of expression of different *Hox* genes in more posterior CNS regions (39). Higher resolution mapping of the individual *zp* gene expression domains may reveal an analogous POU code for the patterning of the anterior brain.

The close relation between the *zp-12* and *zp-23* genes, as well as between *zp-47* and *brn-1.2*, suggests that these genes have arisen by gene duplication. These findings are further evidence that a considerable portion of the zebrafish genome has been duplicated after the branching off of the actinopterygii: several gene families have been found that have more family members in the zebrafish than in mammals. Examples include both the *engrailed*-like (40) and the *otx* genes (41). In both families three gene copies are found in fish but only two occur in mammals. Similarly, four *msx* genes have been found in zebrafish compared to three genes in mice (42). Also the number of zebrafish *dlx* genes exceeds that found in mammals (43). In the case of genes encoding RXR-type nuclear receptors five genes have been found in zebrafish, whereas only three genes are known to exist in mammals (44). Although the precise organization of the family of *ash* helix–loop–helix regulators has not yet been elucidated in zebrafish, it appears that in fish there are two genes that are both most closely related to a single gene in other vertebrates (45). Furthermore, several genes encoding signalling molecules similar to *sonic hedgehog* exist in zebrafish that have no direct mammalian counterparts (46,47). In the face of a possible large-scale duplication of the zebrafish genome and the existence of two pairs of closely related class III POU genes, it is tempting to speculate that a precursor of *zp-12* and *zp-23* might have been duplicated at the same time as a precursor of *zp-47* and *brn-1.2*. It might even be possible that *zp-12* might be physically linked to either *zp-47* or *brn-1.2*, whereas *zp-23* might be located on the same

chromosome as the other member of the *zp-47*/*brn-1.2* gene pair. Whether these genes are arranged in such syntenic relationships will be tested when the position of these genes are established on the genetic map (48).

See supplementary material available in NAR Online.

## ACKNOWLEDGEMENTS

We thank Robert Riggleman, Kathryn Helde and David Grunwald for the cDNA libraries and Andreas Vogel for comments on the manuscript. T.G. is a START fellow of the Swiss National Science Foundation. G.H. was in part supported by a short-term fellowship of the Sandoz-Stiftung zur Förderung der medizinisch-biologischen Wissenschaften. This work was supported by the Swiss National Science Foundation and the cantons of Basel-Stadt and Basel-Land.

## REFERENCES

1 Kimmel,C.B. (1989) *Trends Genet.*, **5**, 293–288.
2 Mullins,M.C. and Nüsslein-Volhard,C. (1993) *Curr. Opin. Genet. Dev.*, **3**, 648–654.
3 Driever,W., Stemple,D., Schier,A.F. and Solnica-Krezel,L. (1994) *Trends Genet.*, **10**, 152–159.
4 Driever,W., Solnica-Krezel,L., Schier,A.F., Neuhauss,S.C.F., Malicki,J., Stemple,D.L., Stainier,D.Y.R., Zwartkruis,D., Abdelilah,S., Rangini,Z., Belak,J. and Boggs,C. (1996) *Development*, **123**, 37–46.
5 Haffter,P., Granato,M., Brand,M., Mullins,M.C., Hammerschmidt,M., Kane,D.A., Odenthal,J., van Eden,F.J.M., Jiang,Y.-J., Heisenberg,C.-P. *et al.* (1996) *Development*, **123**, 1–36.
6 Talbot,W.S., Trevarrow,B., Halpern,M.E., Melby,A.E., Farr,G., Postlethwait,J.H., Jowett,T., Kimmel,C.B. and Kimelman,D. (1995) *Nature*, **378**, 150–157.
7 Verrijzer,C.P. and van der Vliet,P.C. (1993) *Biochim. Biophys. Acta*, **1173**, 1–21.
8 Herr,W. and Cleary,M.A. (1995) *Genes Dev.*, **9**, 1679–1693.
9 Wegner,M., Drolet,D.W. and Rosenfeld,M.G. (1993) *Curr. Opin. Cell. Biol.*, **5**, 488–498.
10 Shimazaki,T., Okazawa,H., Fujii,H., Ikeda,M., Tamai,K., McKay,R.D.G., Muramatsu,M. and Hamada,H. (1993) *EMBO J.*, **12**, 4489–4498.
11 Palmieri,S.L., Peter,W., Hess,H. and Schöler,H.R. (1994) *Dev. Biol.*, **166**, 259–267.
12 Takeda,H., Matsuzaki,T., Oki,T., Miyagawa,T. and Amanuma,H. (1994) *Genes Dev.*, **8**, 45–59.
13 Hauptmann,G. and Gerster,T. (1995) *Mech. Dev.*, **51**, 127–138.
14 Alvarez-Bolado,G., Rosenfeld,M.G. and Swanson,L.W. (1995) *J. Comp. Neurol.*, **355**, 237–295.
15 Hauptmann,G. and Gerster,T. (1996) *Development*, **122**, 1769–1780.
16 de Kok,Y.J.M., van der Maarel,S.M., Bitner-Glindzicz,M., Huber,I., Monaco,A.P., Malcolm,S., Pembrey,M.E., Ropers,H.H. and Cremers,F.P.M. (1995) *Science*, **267**, 685–688.
17 Nakai,S., Kawano,H., Yudate,T., Nishi,M., Kuno,J., Nagata,A., Jishage,K., Hamada,H., Fujii,H., Kawamura,K., Shiba,K. and Noda,T. (1995) *Genes Dev.*, **9**, 3109–3121.
18 Schonemann,M.D., Ryan,A.K., McEvilly,R.J., O'Connell,S.M., Arias,C.A., Kalla,K.A., Li,P., Sawchenko,P.E. and Rosenfeld,M.G. (1995) *Genes Dev.*, **9**, 3122–3135.
19 Matsuzaki,T., Amanuma,H. and Takeda,H. (1992) *Biochem. Biophys. Res. Commun.*, **187**, 1446–1453.
20 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
21 Hauptmann,G. and Gerster,T. (1994) *Trends Genet.*, **10**, 266.
22 Shuldiner,A.R., Nirula,A. and Roth,J. (1989) *Nucleic Acids Res.*, **17**, 4409.
23 Sampath,K. and Stuart,G.W. (1996) *Biochem. Biophys. Res. Commun.*, **219**, 565–571.
24 Stephens,R.M. and Schneider,T.D. (1992) *J. Mol. Biol.*, **228**, 1124–1136.
25 Hara,Y., Rovescalli,A.C., Kim,Y. and Nirenberg,M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3280–3284.
26 Theill,L.E., Hattori,K., Lazzaro,D., Castrillo,J.L. and Karin,M. (1992) *EMBO J.*, **11**, 2261–2269.
27 Matsuo,K., Clay,O., Kunzler,P., Georgiev,O., Urbanek,P. and Schaffner,W. (1994) *Biol. Chem. Hoppe-Seyler*, **375**, 675–683.
28 He,X., Treacy,M.N., Simmons,D.M., Ingraham,H.A., Swanson,L.W. and Rosenfeld,M.G. (1989) *Nature*, **340**, 35–42.
29 Frank,D. and Harland,R.M. (1992) *Development*, **115**, 439–448.
30 Hinkley,C.S., Martin,J.F., Leibham,D. and Perry,M. (1992) *Mol. Cell. Biol.*, **12**, 638–649.
31 Whitfield,T.T., Heasman,J. and Wylie,C.C. (1995) *Dev. Biol.*, **169**, 759–769.
32 Andersen,B., Pearse,R.V., Schlegel,P.N., Cichon,Z., Schonemann,M.D., Bardin,C.W. and Rosenfeld,M.G. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11084–11088.
33 Greenstein,D., Hird,S., Plasterk,R.H.A., Andachi,Y., Kohara,Y., Wang,B., Finney,M. and Ruvkun,G. (1994) *Genes Dev.*, **8**, 1935–1948.
34 Orii,H., Agata,K. and Watanabe,K. (1993) *Biochem. Biophys. Res. Commun.*, **192**, 1395–1402.
35 Stuart,G.W., Zhu,Z., Sampath,K. and King,M.W. (1995) *Gene*, **161**, 299–300.
36 Szeto,D.P., Ryan,A.K., O'Connell,S.M. and Rosenfeld,M.G. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 7706–7710.
37 Schmidt,J.E., von Dassow,G. and Kimelman,D. (1996) *Development*, **122**, 1711–1721.
38 Sampath,K. and Stuart,G.W. (1996) GenBank U43898.
39 Krumlauf,R. (1994) *Cell*, **78**, 191–201.
40 Ekker,M., Wegner,J., Akimenko,M.A. and Westerfield,M. (1992) *Development*, **116**, 1001–1010.
41 Mori,H., Miyazaki,Y., Morita,T., Nitta,H. and Mishina,M. (1994) *Mol. Brain Res.*, **27**, 221–231.
42 Akimenko,M.A., Johnson,S.L., Westerfield,M. and Ekker,M. (1995) *Development*, **121**, 347–357.
43 Stock,D.W., Ellies,D.L., Zhao,Z.Y., Ekker,M., Ruddle,F.H. and Weiss,K.M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 10858–10863.
44 Jones,B.B., Ohno,C.K., Allenby,G., Boffa,M.B., Levin,A.A., Grippo,J.F. and Petkovich,M. (1995) *Mol. Cell. Biol.*, **15**, 5226–5234.
45 Allende,M.L. and Weinberg,E.S. (1994) *Dev. Biol.*, **166**, 509–530.
46 Ekker,S.C., Ungar,A.R., Greenstein,P., von Kessler,D.P., Porter,J.A., Moon,R.T. and Beachy,P.A. (1995) *Curr. Biol.*, **5**, 944–955.
47 Currie,P.D. and Ingham,P.W. (1996) *Nature*, **382**, 452–455.
48 Johnson,S.L., Gates,M.A., Johnson,M., Talbot,W.S., Horne,S., Baik,K., Rude,S., Wong,J.R. and Postlethwait,J.H. (1996) *Genetics*, **142**, 1277–1288.