Int. J. Dev. Biol. 37: 263-272 (1993)

Original Article

Genomic sequence and embryonic expression of the zebrafish homeobox gene *hox-3.4*

JOHANNA U. ERICSON[1]*, STEFAN KRAUSS[2] and ANDERS FJOSE[3]

[1]*Department of Molecular Genetics, Institute of Medical Biology, University of Tromsø, Tromsø, Norway,*
[2]*Division of Biomedical and Biomolecular Sciences, King's College London, University of London, London, England and*
[3]*Department of Biochemistry, University of Bergen, Bergen, Norway*

ABSTRACT We have analyzed the genomic DNA sequence and embryonic expression pattern of the zebrafish *hox-3.4* gene. Two exons, encoding a protein with a total size of 232 amino acids, were identified within a 3.5 kbp genomic region. Besides the homeodomain, which is identical to that of the human *HOX3D* and mouse *Hox-3.4* genes, the first 58 residues of the N-terminal domain in the predicted Hox-3.4 protein share 48% sequence identity with the gene product of the human cognate. Some of the N-terminal sequence elements are also conserved relative to the two other members of the *Hox-1.3/Hox-2.1/Hox-3.4* paralogy group. In addition, the paralogous genes share a significant degree of sequence identity in non-coding regions. This conservation is particularly evident in the promoter regions of the cognates *hox-3.4*, *Hox-3.4*, and *HOX3D*, where a 180 bp TATA-box-containing element with a 60% identity is located. This is in agreement with the previous finding that the *HOX3D* promoter region contains response elements for other Hox proteins and retinoids. Also with respect to embryonic expression, the zebrafish *hox-3.4* gene is very similar to its mammalian counterparts. Within the central nervous system of 16, 24, and 48 h embryos, *hox-3.4* transcripts were detected throughout most of the spinal cord from a boundary at the posterior end of the hindbrain. In 16 h embryos the *hox-3.4* gene is also active within a restricted region of the tailbud.

KEY WORDS: *zebrafish, homeobox gene, hox-3.4, genomic sequence, embryonic expression*

Introduction

Following the first identification of homeobox sequences in the ANT-C and BX-C homeotic gene clusters of *Drosophila melanogaster* (McGinnis *et al.*, 1984a,c), a large number of related homeobox genes *(Hox)* have been characterized in different vertebrate species, including *Xenopus laevis* (Carrasco *et al.*, 1984), chicken (Rangini *et al.*, 1989), mice (McGinnis *et al.*, 1984b), human (Levine *et al.*, 1984), and zebrafish (Eiken *et al.*, 1987). The strong conservation of both DNA sequence and gene organization among *Hox* genes from different species implies that the clusters had a common ancestor, HOM-C (homologous to the *Drosophila* ANT-C and BX-C clusters in juxtaposition; Beeman, 1987), from which they all have evolved. Relatively high levels of *Hox* gene expression are detected in the central nervous system (CNS) of vertebrates, in which the gene organization reflects the expression pattern, i.e. 3'-anterior-early/5'-posterior-late (Dollé *et al.*, 1989; Izpisúa-Belmonte *et al.*, 1991). Gene paralogues display similar anteroposterior (A-P) but different dorsoventral (D-V) expression patterns, indicating that duplications of the *Hox* cluster (up to four copies in higher vertebrates; Acampora *et al.*, 1989; Graham *et al.*, 1989) might have provided a possibility to direct development of a more complex CNS, not only by the positional information along the A-P axis but also in the D-V orientation (Graham *et al.*, 1991). The correlation between *Hox* expression and the segmental organization of the hindbrain suggests that *Hox* proteins participate in the determination of segmental identities rather than establishing the individual segments (Wilkinson *et al.*, 1989; Hunt and Krumlauf, 1991). Experimental manipulation of *Hox* gene expression in *Xenopus* and mouse embryos have provided more direct evidence for this assumption. Ectopic expression of *Hox-1.1* in transgenic mice leads to posterior transformations in tissues anterior of the regions normally expressing *Hox-1.1* (Puschel *et al.*, 1991). Conversely, loss of function, either by *Hox-3.1* disruption in mouse (Le Mouellic *et al.*, 1992) or anti-*Xlhbox1* antibody injections in *Xenopus* (Wright *et al.*, 1989), displayed anteriorization of a subset of embryonic cells, resulting in transformed vertebrae *(Hox-3.1)* or hindbrain structures *(Xlhbox1)*.

Abbreviations used in this paper: CNS, central nervous system; A-P, anteroposterior; D-V, dorsoventral; RA, retinoic acid; ORF, open reading frame.

***Address for reprints**: Department of Molecular Genetics, Institute of Medical Biology, University of Tromsø, N-9037 Tromsø, Norway. FAX: 47-83-45350.

0214-6282/93/$03.00

Printed in Spain

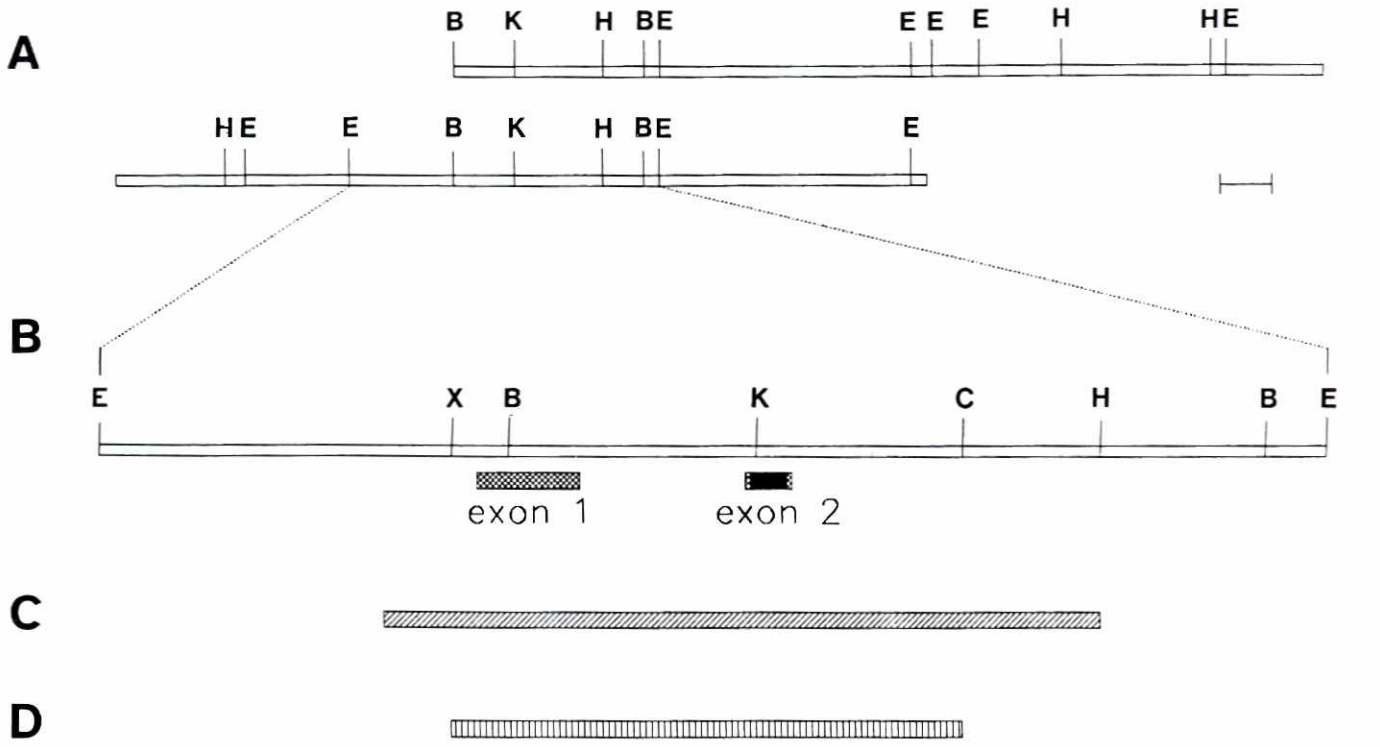


Fig. 1. Genomic organization of the *hox-3.4* gene. (Part A) *shows the restriction enzyme map of the two genomic clones carrying the zebrafish hox-3.4 homeobox.* **(Part B)** *Below, represents the total subcloned region. The size-bar corresponds to 1000 bp in part A and 250 bp in part B. The two predicted exons of the hox-3.4 gene are indicated below the map in B.* **(Part C)** *shows the 3.5 kbp sequenced region, and* **(Part D)** *the restriction fragment used as a probe for in situ hybridization experiments. Abbreviations: B, BamHI; C, ClaI; E, EcoRI; H, HindIII; K, KpnI; X, XbaI.*

Furthermore, these data confirmed that each *Hox* paralogue has a unique function that cannot be replaced by another gene from the same subfamily.

Establishment of the «*Hox* code», i.e. the combination of functionally active *Hox* genes that will specify the identity of a body region, can be influenced by endogenous molecules like retinoic acid (RA) and peptide growth factors. RA, a posteriorizing morphogen, influences the expression of *Hox* genes according to their individual sensitivity, both in whole embryos (Kessel and Gruss, 1991; Morriss-Kay *et al.*, 1991) and in tissue cell culture (Boncinelli *et al.*, 1991; Arcioni *et al.*, 1992). It was suggested by Kessel and Gruss (1991) that during gastrulation, ingressing cells are exposed to a RA signal generated from midline embryonic structures, and will respond with a sequential activation of more and more *Hox* genes, leading to nonidentical, overlapping expression domains of *Hox* genes along the anteroposterior axis.

Some of the *Drosophila* homeoproteins have been shown to work as transcription factors (Biggin and Tjian, 1989; Han *et al.*, 1989; Winslow *et al.*, 1989; Jaynes and O'Farrell, 1991). In vertebrates, little is known about the repressing/derepressing activities of the *Hox* gene products. However, the characterization of the human *HOX3D* promoter sequence revealed potential targets for multiple regulatory mechanisms acting independently (Arcioni *et al.*, 1992), among them trans-activation by three, more posteriorly expressed, *Hox* gene products, i.e. *HOX3C, HOX4D* and *HOX4C* proteins, in addition to the mouse *Hox-4.3* protein (Arcioni *et al.*, 1992). Alignment of the promoter regions from the *HOX3D* gene and its mouse homologue (*Hox-3.4*) showed conservation in all identified regulatory targets (Arcioni *et al.*, 1992). Thus, the vertebrate homeoproteins display evolutionarily conserved regulatory functions mediated by sequence-specific DNA-binding.

In zebrafish *(Brachydanio rerio)*, strong similarities to the mouse *Hox* genes have been observed, both with respect to the gene organization and DNA sequence (Eiken *et al.*, 1987; Njølstad *et al.*, 1988a,b, 1990; Molven *et al.*, 1990, 1992). Evidently there are at least two *Hox* gene clusters in zebrafish, corresponding to the mammalian HOX-2 and -3 complexes (Njølstad *et al.*, 1988b, 1990). In this paper we describe the genomic DNA sequence of the zebrafish *hox-3.4* gene, which encodes a homeoprotein homologous to the human *HOX3D* gene product. Additional sequence conservations were identified in the non-coding regions including the promoter. Results obtained by *in situ* hybridization analysis of the embryonic zebrafish *hox-3.4* expression pattern are also consistent with the spatial distribution of transcripts reported for the murine *Hox-3.4* gene (Gaunt *et al.*, 1990).

Results

Identification of the zebrafish hox-3.4 protein-encoding sequence

Two recombinant lambda-EMBL3 clones, C25 and C26, containing genomic zebrafish DNA, were previously shown to carry the *hox-3.4* homeobox sequence (Eiken, *et al.*, 1987; Njølstad *et al.*, 1990). The unique Kpn I-site in lambda C25, found in the homeobox region of *hox-3.4* (Eiken *et al.*, 1987), was used to identify the *hox-3.4*-containing subclones (Fig. 1). DNA sequencing of a 3.5 kbp region, including 1.7 kbp downstream and 1.8 kbp upstream of the Kpn I

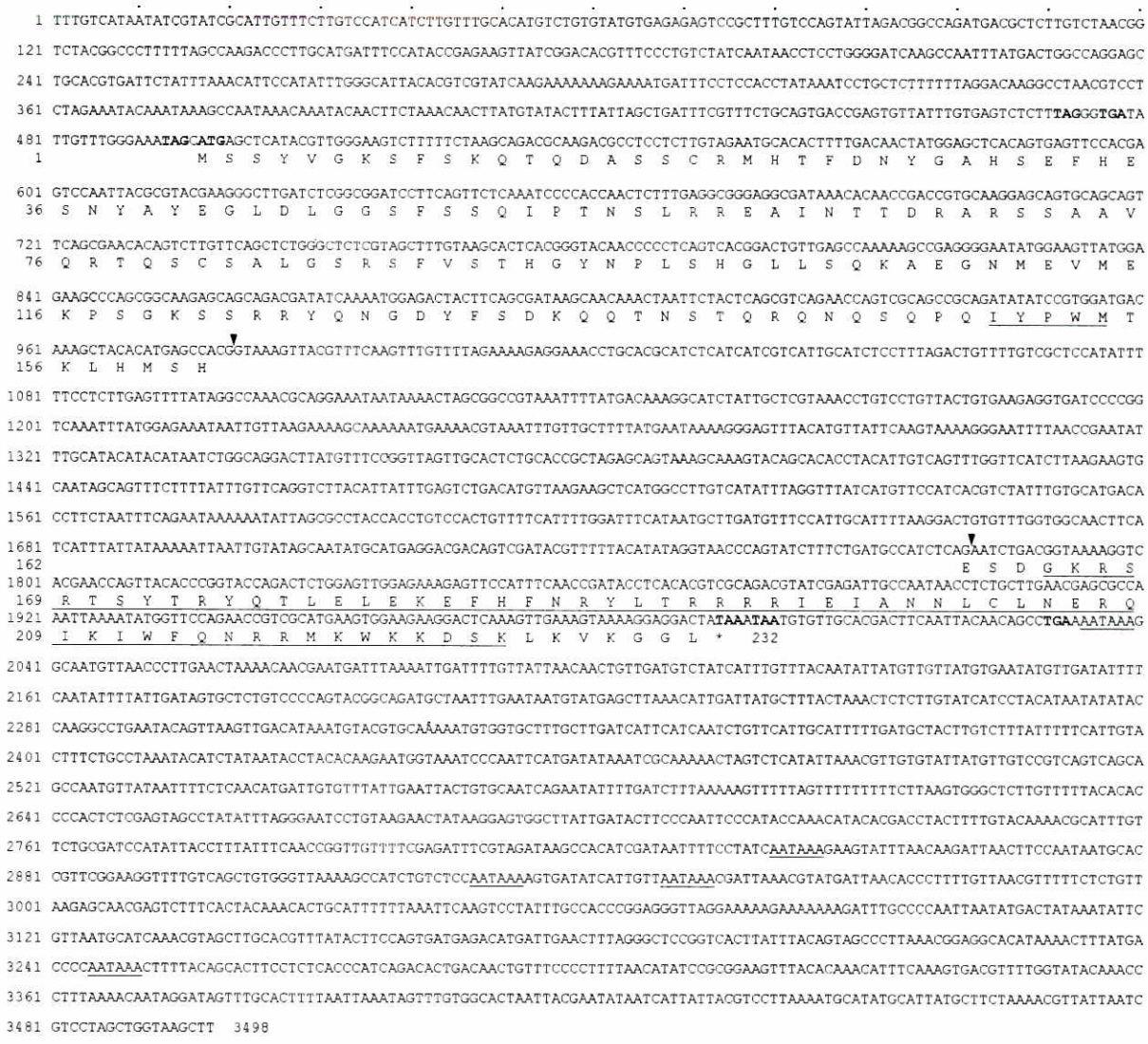


Fig. 2. The genomic sequence including the *hox-3.4* coding region. *Exons 1 and 2 are indicated by translation into the one-letter code below the DNA-sequence. Conserved peptide sequences, the IYPWM-pentapeptide and the homeodomain, are underlined, and two possible splice sites are indicated by arrowheads (▼). In the 5' non-coding region, the promoter proximal ATG, and three stop-codons — in all three reading-frames — preceding the translational start are marked with bold characters, as well as the translational start. In the 3' sequence, five possible poly(A)-addition signals (AATAAA) are also underlined. Furthermore, three translational stop-codons, in all three reading-frames, downstream of exon 2 are marked with bold characters. The EMBL Data Library accession number for the zebrafish hox-3.4 sequence is X68324.*

site, was performed on both strands. Within this DNA sequence two open reading frames (ORFs) were predicted to correspond to the *hox-3.4* coding sequence (Fig. 2; see below). Comparisons to the *HOX3D* coding sequence revealed that the longest ORF (of 161 amino acids) contained exon 1 (Fig. 3). Within this exon, sequences encoding the N-terminal MSSY residues and the IYPWM pentapeptide (preceding the splice site; Mavilio *et al.*, 1986), which are common among homeoproteins, were identified. The second ORF (of 71 aa) encoded a homeodomain identical to that of *HOX3D*, preceded by three and followed by seven residues (Fig. 2). The predicted translational start for exon 1 was found to be similar to the consensus sequence, PuCCATGG (Kozak, 1986 and 1987). Furthermore, translational stop codons were located just upstream of the predicted translational start in all three reading-frames (bold in Fig. 2). An additional ATG was identified in the sequence, more promoter proximal, but this potential translational start site is not related to the consensus sequence and the translated product would only be 17 residues. Other aspects of the *hox-3.4* sequence are consistent with the structural organization of the human *HOX3D* gene. Thus, in both genes, a conserved promoter region is located just upstream of the predicted translational initiation site (see below). Furthermore, the introns of the *HOX3D* and *hox-3.4* genes are of approximately the same length, 699 and 801 nucleotides, respectively. In the zebrafish gene the intron sequence i flanked by a putative donor splice site with a sequence homology of 6 out of 9, and an acceptor splice site with 12 out of 16 nucleotides identical

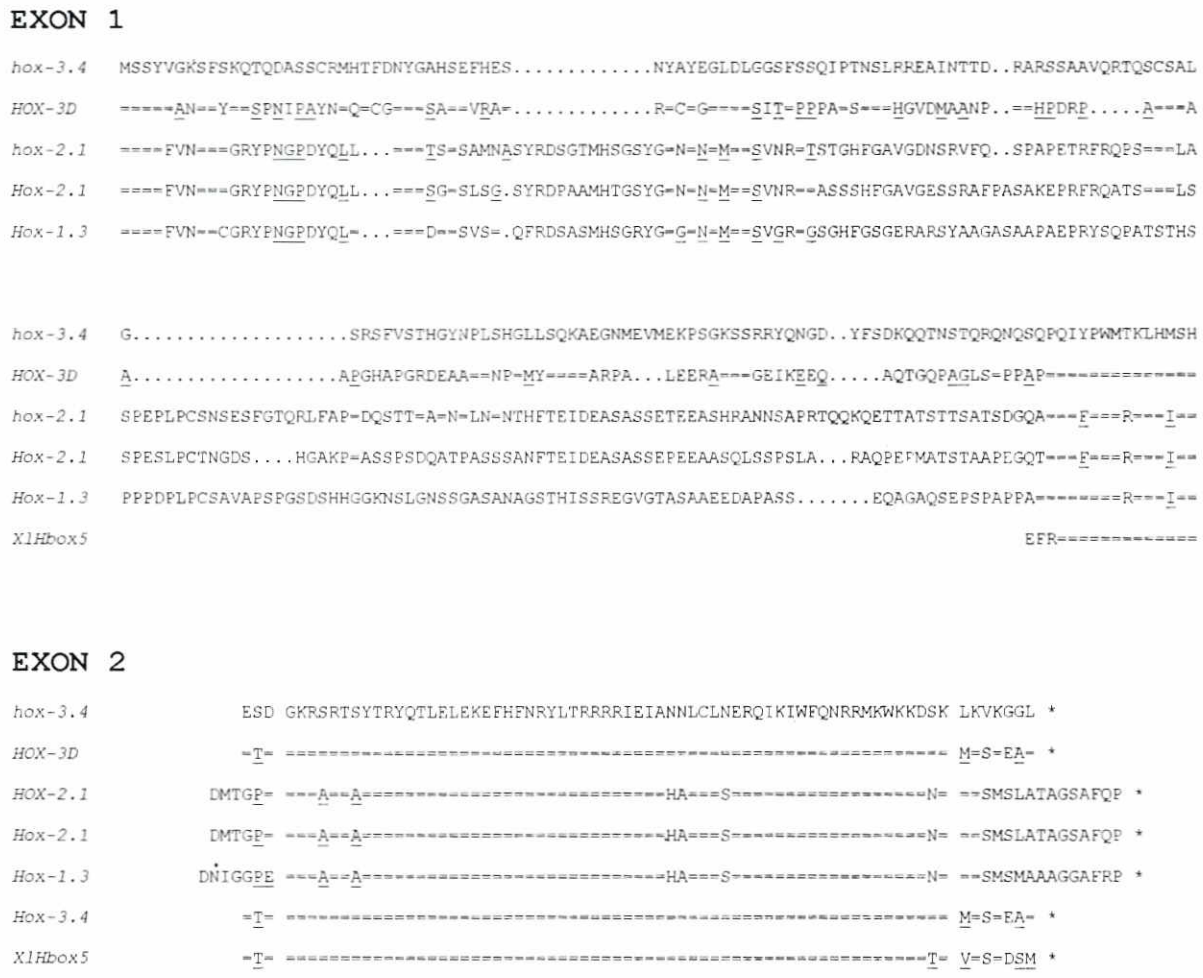


Fig. 3. Comparisons of the putative zebrafish *hox-3.4* protein with homologous and paralogous peptides. *Alignments of the peptide sequences, encoded by exon 1, of hox-3.4, HOX3D (human), hox-2.1 (zebrafish), Hox-2.1 (mouse), and Hox-1.3 (mouse) are shown at the top. Included also are the 16 last residues from exon 1 of XlHbox5. The comparison of the exon 2 encoded peptides from the hox-3.4, HOX3D (Arcioni et al., 1992), XlHbox5 (Fritz and De Robertis, 1988), hox-2.1 (Njølstad et al., 1988b), mouse Hox-3.4 (Gaunt, et al., 1990), Hox-2.1 and Hox-1.3 (Fibi, et al., 1988) is shown below. Amino acids identical to the hox-3.4 sequence are indicated with =, while conserved changes are underlined. The dots have been introduced to achieve maximum alignment of homologous sequences.*

to the respective consensus sequences. Downstream of the predicted coding sequence of *hox-3.4* there are also several possible poly(A)-addition signals (AATAAA, indicated in Fig. 2).

The hox-3.4 gene product

The zebrafish *hox-3.4* homeobox was previously identified to be closely related to the *Scr* homeobox with a high homology to the mouse *Hox-2.1* and *Hox-1.3* homeoboxes (90% on the peptide-level; Eiken *et al.*, 1987). Recently a third member of the same subfamily of mouse *Hox* genes was identified, the *Hox-3.4* gene (Gaunt *et al.*, 1990), which was found to contain a homeodomain identical to that of the zebrafish gene (Fig. 3). Since the human *HOX3D* gene (Simeone *et al.*, 1988) and the *Xenopus XlHbox5* gene (Fritz and De Robertis, 1988) were suggested to be homologues of the mouse *Hox-3.4* gene (Gaunt *et al.*, 1990), we included all three homeodomains in the comparison to *hox-3.4* (Fig. 3). All four homeodomains were found to be identical with one single exception, *XlHbox5*, which has one conserved amino acid substitution in the C-terminal region. The second exon of all four *Hox-3.4* cognates encodes a total of 71 aminoacids, including 3 residues preceding and 7 residues following the homeodomain. The three most N-terminally located amino acids encoded by exon 2 were identical in the human, mouse and frog genes, while the zebrafish contained one conserved amino acid exchange (Fig. 3). The C-terminal, however, is somewhat diverged in the fish and frog cognates as compared to the mammalian derivatives,which are identical. More extensive differences are present in the corresponding region of the *hox-3.4* paralogues *hox-2.1* (zebrafish), *Hox-2.1* (mouse) and *Hox-1.3* (mouse), where the N- and C-terminals are extended by 2 and 8 residues, respectively (Fig. 3). For the *Antp/Scr* proteins the functional specificities, other than DNA-binding, were shown to be located in regions outside the homeodomain (Gibson *et al.*, 1990). Thus, a comparison between functionally related proteins would reveal a higher homology in such regions, while more distant relations would be less homologous.

Comparison of exon 1-encoded sequences from *hox-3.4* with the

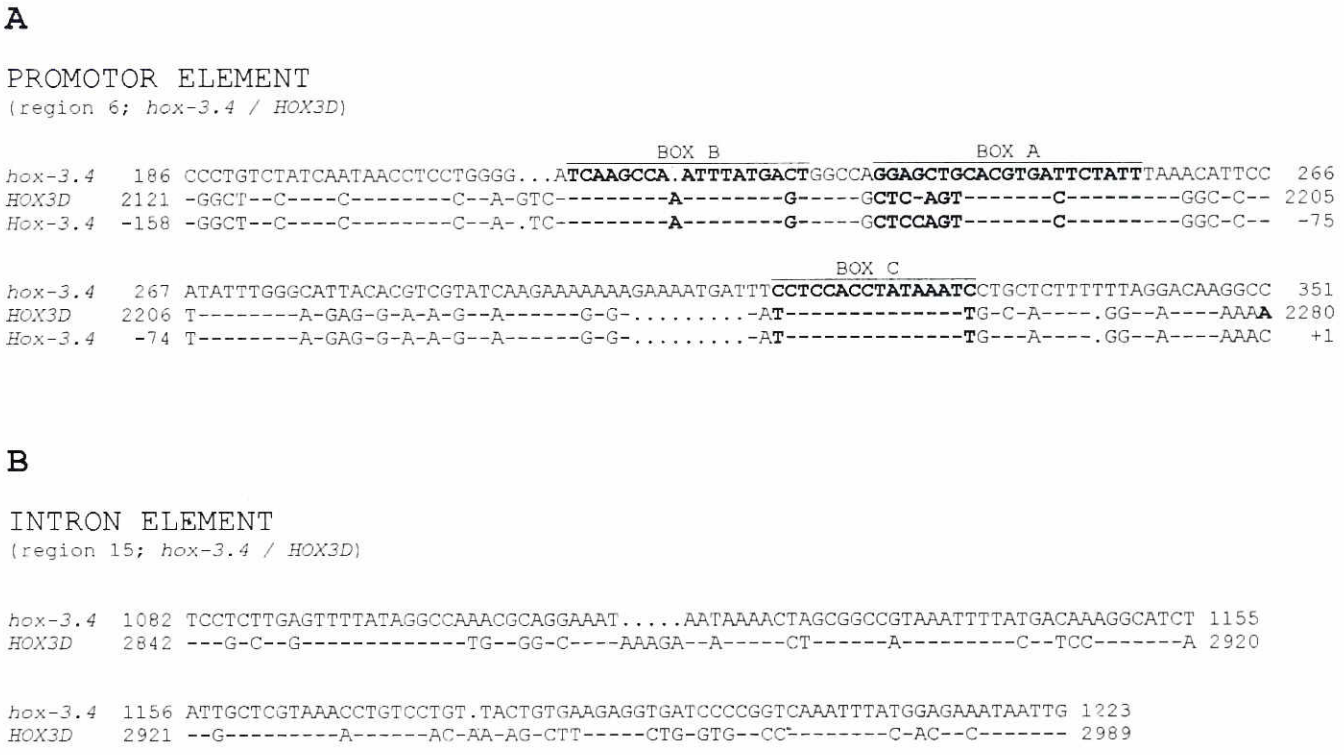


Fig. 4. Conserved DNA sequence-elements in the two *hox-3.4* cognates. *Sequence comparisons of the two conserved DNA elements identified by computer analysis between hox-3.4 and the human HOX3D (Arcioni et al., 1992).* **(A)** *The 180 bp promoter element with the three defined regulatory target sites marked: BOX A, BOX B, and BOX C (see text for details). The mouse Hox-3.4 sequence, homologous to the promoter region, was included according to Arcioni et al. (1992).* **(B)** *The 141 bp intron element homologous in the two cognate genes.*

corresponding parts of *HOX3D* and the paralogues (zebrafish *hox-2.1*, mouse *Hox-2.1* and *Hox-1.3*) revealed extensive differences (Fig. 3). On the basis of the variable exon length, the peptides can be divided into two categories. First the *hox-3.4* cognates, encoding 161 and 151 amino acids-long exon 1-peptides, and secondly the three longer paralogues of 193 residues for the zebrafish, and 187 for the mouse derivatives. The protein sequence alignment (Fig. 3) also showed that the protein sequences encoded by the two ends of exon 1 are the most conserved. In the middle part of this protein region the sequence conservation is quite low, especially when the paralogues are compared. Thus, the overall identity between the *hox-3.4* and *HOX3D* homeoproteins is only 56%, and this homology is mainly concentrated in the N-terminal and the «extended homeodomain» — including the 14 last residues encoded by exon 1 in addition to exon 2. Only 48% of the first 58 N-terminal amino acids are identical between the two cognates. The corresponding comparison to the paralogous peptides gave an identity not exceeding 32%. The «extended homeodomain», on the other hand, showed a 94% identity between the two cognates and 81% to the paralogues.

Identification of conserved sequence elements in non-coding regions of hox-3.4 cognates and paralogues

Computer comparisons of the *hox-3.4* genomic sequence with *HOX3D* and *hox-2.1* sequences revealed many regions of homology. Outside the coding sequence, two conserved DNA elements were identified in the *hox-3.4/HOX3D* cognates. The first corresponds to a 180 bp promoter element of 60% sequence identity (Fig. 4A; see discussion). The second element of 141 bp which is 73% identical is located within the intron sequence in the two cognates (Fig. 4B). The remaining sequence elements, which vary in length between 15 and 40 bp and share a 60-90% sequence identity, are scattered throughout the gene (data not shown), but only a few of these elements are located in similar positions.

In the comparison between the two paralogous zebrafish genes *(hox-2.1* and *hox-3.4),* a similar result was obtained as for the two *hox-3.4* cognates (data not shown). However, the number of conserved elements as well as the average length and the level of sequence conservation are reduced.

Expression of the hox-3.4 gene during early development

To analyze the *hox-3.4* expression pattern, *in situ* hybridization to parasagittal tissue sections from zebrafish embryos was performed. Using a probe that covered the complete *hox-3.4* gene (indicated in Fig. 1), embryos of three different developmental stages were studied (Fig. 5). The *hox-3.4* mRNA level was found to be relatively low at all three stages, and the detection of expression was limited (at least at the 24 and 48 hour stages) to the central nervous system (CNS) with an anterior border located within the caudal hindbrain. This expression extends posteriorly throughout the spinal cord. Reconstruction of the embryos by camera lucida drawings showed that the hybridization signal had its anterior border in the caudal-most (ninth) hindbrain segment, Ca3 (Hanneman *et al.*, 1988), which is located rostral of the first myotome (data not shown). This expression pattern was maintained from the 16 to the

48 h stage. The relative movement of the anterior border during development (cf. Fig. 5B to E) corresponds to the length reduction, or compression, of the hindbrain (Trevarrow *et al.*, 1990). One exception from the CNS-limited expression was observed at the 16 h stage. In the tailbud a restricted area showed an enhanced *hox-3.4* expression (Fig. 5G and H) that was not observed in the older embryos.

Cross-sections of the hatching embryo in the rostral spinal cord (Fig. 6A and B) and at the level of the finbuds (Fig. 6C and D) also revealed the *hox-3.4* expression in the CNS. Hybridization signals were not observed in muscle tissue (Fig. 6B and D). However, the *hox-3.4* expression, possibly present in this tissue, might have been below the detection level. Mesodermal expression of *Hox* genes in vertebrates has previously been identified in regions located posterior relative to the CNS expression (Gaunt *et al.*, 1990; Molven *et al.*, 1990). The mRNA seemed to be more or less uniformly distributed across the neural tube in the rostral parts of the CNS (Fig. 6B), while the expression became more ventrally localized in the sections at the level of the finbuds (Fig. 6D). Further posterior, the hybridization signal was weaker, as observed on the sagittal sections (Fig. 5D and F), and to distinguish between uniform or ventral distribution of *hox-3.4* mRNA became impossible. Finally, no *hox-3.4* mRNA could be detected in the finbuds (data not shown).

Discussion

Conservation of sequence and functional specificity in protein products of the hox-3.4 cognates and paralogues

From comparisons to other recently described vertebrate homeobox-containing genes, we concluded that *hox-3.4* is the zebrafish homologue of the human *HOX3D*, mouse *Hox-3.4* and frog *XlHbox5* genes (Fig. 3), which were previously suggested to be cognates, by Gaunt and co-workers (1990). All four genes contain identical homeodomains, except for one amino acid substitution in the *XlHbox5* protein. Comparisons of the complete gene products was not possible for the mouse and frog derivatives, since only the C-terminal sequences have been published (Fritz and De Robertis, 1988; Gaunt *et al.*, 1990). However, direct alignment of the amino acid sequences derived from *hox-3.4* and the human *HOX3D* genes revealed a total identity of 56%. In the N-terminal region, including the first 58 amino acids, 48% of the amino acid residues are identical. Less conservation is present in the middle part of the protein (position 59-147), where the sequence identity is only 18%. Thus, the level of conservation between these *Hox-3* cognates is considerably lower than for the corresponding cognates in the *Hox-2* complex (Njølstad *et al.*, 1988b).

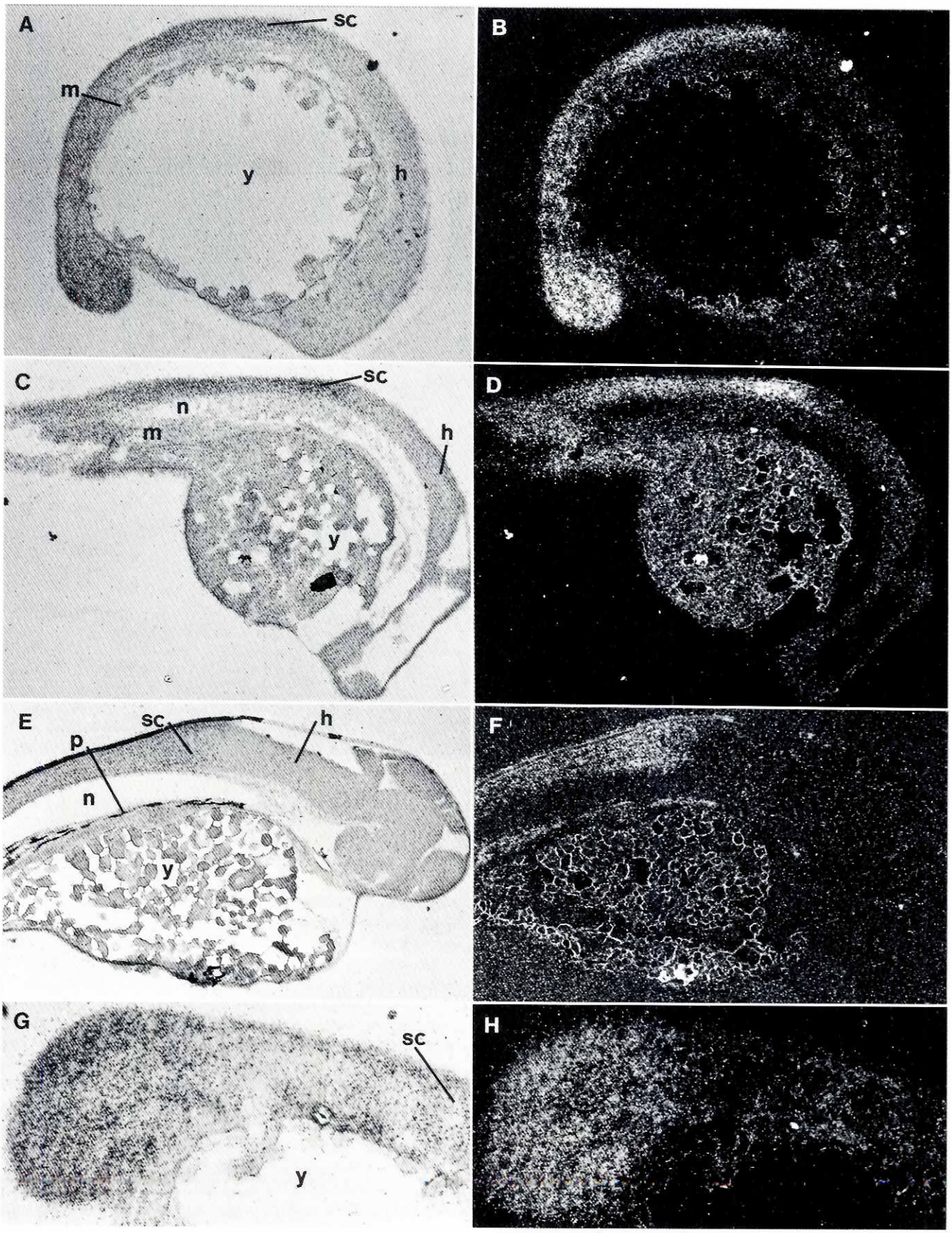
An interesting aspect concerns the relationship between the evolutionary divergence of cognate genes and the conservation of functional specificity of the corresponding protein products. The results obtained from analyses of the functional specificity of homeodomain proteins from *Drosophila* are somewhat contradictory. Studies on hybrid proteins of *Dfd* and *Ubx* indicated that the homeodomain, due to its target-specificity, was responsible for the regulatory specificity (Kuzoira and McGinnis, 1990; Dessain *et al.*, 1992). However, similar studies on hybrid proteins of *Antp* and *Scr* indicated a requirement for additional sequences to completely define the functional specificity (Gibson *et al.*, 1990). On the basis of these results, also a model for the functional structure of the *Antp* protein in *Drosophila* was suggested. According to this model, the Antp-peptide can be divided into two major regions, the potentiating portion, which included the most N-terminal two thirds, and the homeodomain-containing specifying portion, at the C-terminal of the protein (Gibson *et al.*, 1990). It was proposed that the potentiating portion could be further subdivided into an N-terminal general potentiating region, determining the overall levels of Antp activity, and a specific potentiating region, located closer to the middle of the protein, which could be involved in adjusting the level of activity in specific cells. Consistent with this model, the C-terminal specifying portion of the *hox-3.4* and *HOX3D* proteins are almost identical. In the potentiating region, however, only the general portion at the N-terminal is strongly conserved in the two cognates. This may suggest that the mechanisms regulating the overall activity of *hox-3.4* and *HOX3D* have been preserved, while the circuits for specific adjustments in certain cells have been modified through evolution. Similarly, the proteins derived from the zebrafish and mouse *Hox-2.1* cognates share the highest level of sequence identity in the regions corresponding to the location of the specifying portion (C-terminal) and the general potentiating region (Njølstad *et al.*, 1988b). These domains are also partially conserved relative to the paralogues *hox-3.4* and *HOX3D*, while the sequences in the proposed specific potentiating region are almost completely diverged.

Since regions with an elevated level of polar, non-acidic amino acids — serine and threonine or proline — have been shown to be involved in transcriptional activation (Tanaka and Herr, 1990), the distribution of ser/thr and pro residues within the different homeoproteins were calculated. In the predicted *hox-3.4* protein, 24% of the exon 1-encoded amino acids are serine or threonine, while proline residues constitute only 3%. A somewhat different distribution is present in the human cognate, which has equal amounts of the two categories, 13% of each. The paralogous genes encode peptides similar to that of *hox-3.4* with a ser/thr content of 21-28%, while the proline levels vary from 5% (zebrafish *hox-2.1*) to 9-10% for the murine Hox-2.3 and *Hox-1.3* genes. Overall, it seems as if the preference for polar amino acids has been altered towards serine/threonine residues in the zebrafish proteins and more proline residues in mammals.

Multiple non-coding sequence elements are conserved among cognates and paralogues of the hox-3.4 gene

In the DNA-sequence comparison between *hox-3.4* and *HOX3D* additional regions of homology were found outside the coding regions. One of the most interesting conservations was the promoter element of about 180 bp (Fig. 4A) with a 60% homology between the two species. The human promoter was found to be, at least to some extent, cell- or tissue-specific in the expression systems tested for the human derivative, and to contain one retinoic acid-responsive

Fig. 5. Detection of *hox-3.4* transcripts in the CNS of different embryonic stages. *In situ hybridization to parasagittal sections of zebrafish embryos, shown in bright-field* **(A, C, and E)** *and dark-field* **(B, D, and F)**. *Three different developmental stages were used; 16 h (A and B), 24 h (C and D), and 48 h (E and F). The restricted expression identified on parasagittal sections in the tail-bud of 16-h-old embryos.* **(G and H)** *show the tail-bud in bright-field and in dark-field, respectively. Abbreviations used: h, hindbrain; sc, spinal cord; y, yolk; m, mesoderm; n, notochord; and p, pigmentation.*


A
sc
m
y
h
B
C
sc
n
m
h
y
D
E
sc
h
p
n
y
F
G
sc
y
H

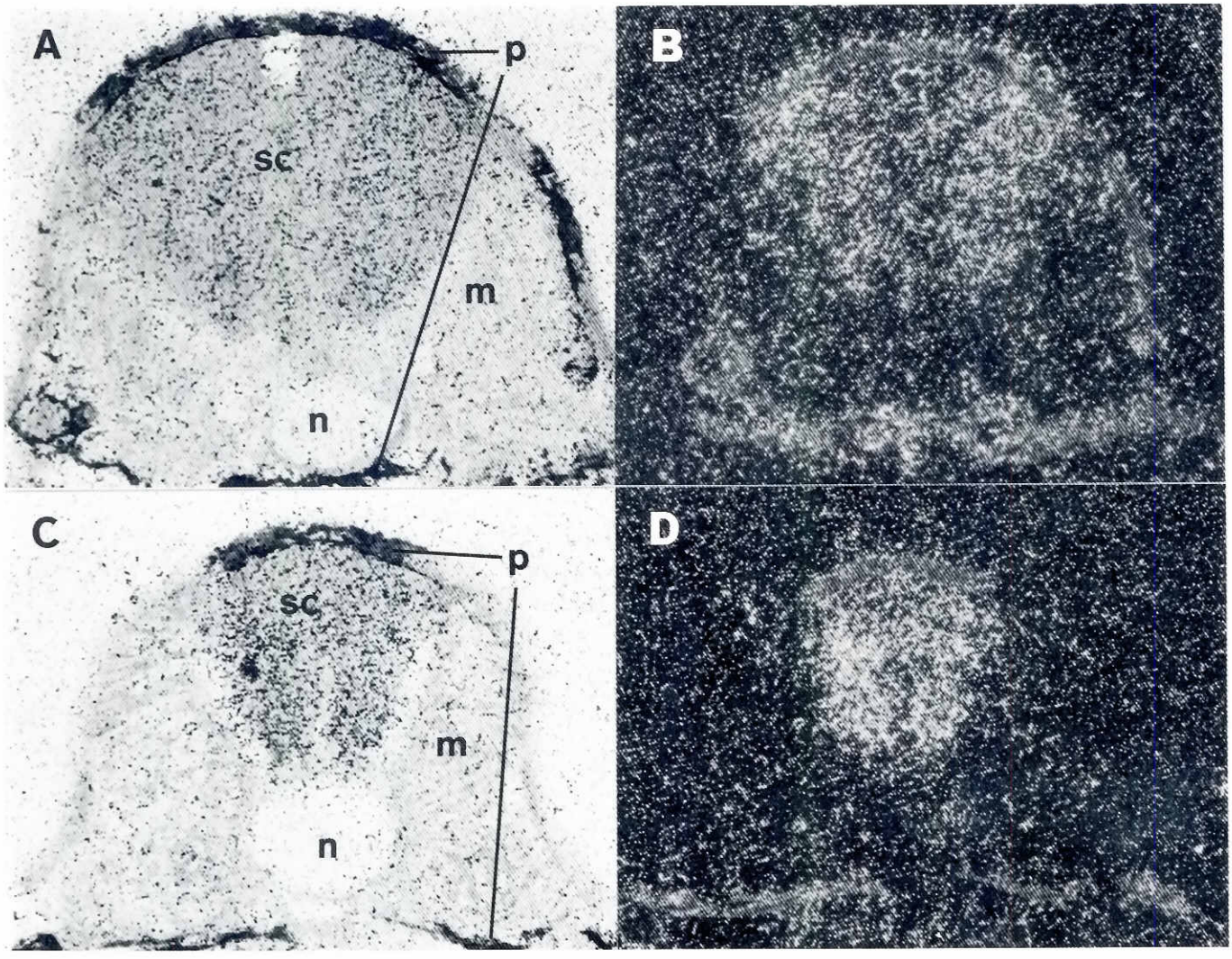


Fig. 6. Analysis of dorsoventral distribution of *hox-3.4* transcripts in the CNS. *In situ hybridization to cross-sections of zebrafish embryos at the 48 h stage, shown in bright-field* **(A and C)** *and dark-field* **(B and D)**. *The section shown in A and B comes from the rostral spinal cord/hindbrain region, while the section in C and D comes from the level of the finbud. Abbreviations used: sc, spinal cord; m, myotome; n, notochord; and p, pigmentation.*

element and two homeoprotein binding sites (Arcioni *et al.*, 1992). When compared to the murine *Hox-3.4/-2.1/-1.3* paralogy group, the *HOX3D* promoter was found to be almost identical to its cognate *(Hox-3.4)*, while the paralogous promoters *(Hox-2.1* and *Hox-1.3)* contained more divergent sequences in all three regulatory elements (Arcioni *et al.*, 1992). Alignment of the *hox-3.4* upstream sequence to the human promoter showed that the two homeodomain-binding sites (box B and C; Fig. 4A) were highly conserved, implying a functional conservation of the regulatory elements. The RA-responsive element, on the other hand, contained less conservation and the 10 base-pairs long palindromic sequence — suggested to be essential for the RA-induced binding (boxA; Fig. 4A; Arcioni *et al.*, 1992) — was found to be reduced to only 6 bp in the *hox-3.4* sequence. However, this may reflect alternative or a more degenerate binding-specificity involved in *Hox* gene regulation in zebrafish. The protein responsible for RA-induced binding in the embryonal carcinoma cells was not identified but previously identified RA receptors were unable to activate *HOX3D* transcription (Arcioni *et al.*, 1992).

Besides the promoter, one surprisingly large region of 142 bp, located within the intron (Fig. 4B), was found to be 73% identical in the two cognates. Additional, shorter stretches of homology (60-90%) are also scattered throughout the genes. The majority of these elements have different locations within the two cognates. Also no common binding sites for known transcription factors were identified in a computer search for such consensus sequences. Therefore, it remains unclear whether it is regulatory or structural demands that have preserved these elements. However, this widely scattered distribution of potential cis-acting elements may relate to the results obtained from transgenic analyses of *Hox-1.1* and *Hox-2.6* regulation (Puschel *et al.*, 1991; Whiting *et al.*, 1991). In both cases, large DNA regions, including the introns and downstream

sequences, were found to be necessary for establishing normal expression patterns. This suggests that the *hox-3.4* cognates may also have several functionally equivalent regulatory elements at different locations within the genes. The identification of a similar but more limited extent of related sequence elements in the zebrafish *hox-2.1* paralogue (data not shown) may reflect a partial conservation of the regulatory network which control the expression of these duplicated genes.

Expression of the hox-3.4 gene in the embryonic CNS

By *in situ* hybridization to parasagittal sections of zebrafish embryos (at 16, 24 and 48 h) we detected *hox-3.4* mRNA in the spinal cord and the caudalmost regions of the hindbrain (Fig. 5). There was no detectable expression in more anterior parts of the CNS. The mouse homologue, *Hox-3.4*, was found to be expressed in the spinal cord with an anterior limit at the spinal cord / hindbrain border, at 12.5 days p.c. (i.e. after somite formation is completed; Gaunt *et al.*, 1990). This was also observed for the *Hox-2.1* expression (Wilkinson *et al.*, 1989), suggesting a similar expression pattern for the different paralogues in the anterior-posterior axis (Gaunt *et al.*, 1990). Northern-blot analysis of poly(A^+) RNA revealed that the zebrafish *hox-2.1* gene was expressed at least as early as from the 12.5 h stage (i.e. at the 6-somite stage; Njølstad *et al.*, 1988b) and by *in situ* hybridization the mRNA was mainly detected in the CNS with an anterior border within the posterior hindbrain (Njølstad *et al.*, 1990). Similar results were obtained for *hox-2.2* (Njølstad *et al.*, 1990). As in the case of *hox-3.4*, mesodermal expression was not detected for the two zebrafish genes. However, immunohistochemical analyses of zebrafish embryos with an antibody raised against the *Xenopus Xlhbox1* protein, which probably corresponds to the *hox-3.3* gene-product (Molven *et al.*, 1990), also revealed expression in mesodermal (myotomal and notochord) tissues of 16 and 22 h embryos. In addition, cells in the proximal finbuds were stained by this antibody. These data demonstrate that zebrafish *Hox* genes are expressed in mesodermal tissues, and judging by the intensity of the antibody stainings, the protein level is approximately the same as in the CNS. However, this might not directly reflect the relative amount of *hox*-mRNA in the two tissues, since post-transcriptional regulation may have increased the protein level in muscle cells. Consistent with this alternative, the three murine Hox paralogues, *Hox-3.4*, *Hox-2.1*, and *Hox-1.3*, all show mesodermal expression, but the hybridization signal was much weaker then in the CNS (Gaunt *et al.*, 1990). These results suggest that the sensitivity of the *in situ* hybridization method used for zebrafish is not sufficient for detection of expression in the mesoderm.

Graham *et al.* (1991) compared the dorsoventral expression pattern for two genes, one from the HOX-2 and one from the HOX-3 cluster, in parallel cross-sections from mouse (12.5 days p.c.). They found the two genes to be expressed in different patterns through the cross-sections, while two genes from the HOX-2 cluster showed an almost identical D-V patterning. The authors suggested that the variety of expression patterns across transverse sections of the neural tube might reflect the *hox* cluster responses to different dorsoventral cues. Thus, the variation in expression pattern might provide the potential to specify different positions in the more complex vertebrate nervous system. The observation that *hox-3.4* is not expressed uniformly within cross-sections at the level of the finbuds, in 48-hour-old embryos (Fig. 6A-D), indicates a similar situation in zebrafish.

The specific *hox-3.4* tailbud expression found in 16 h embryos (Fig. 5G and H) has not been observed previously for vertebrate *Hox* family genes. Tailbud expression was reported for a *Xenopus* homologue to the *Drosophila* pair-rule gene *even skipped* (the *Xhox3*-gene; Ruiz i Altaba and Melton, 1989). However, the tailbud expression of *Xhox3* was observed at the late neurula-early tailbud stage and it disappeared when somitogenesis was completed, while in zebrafish, the *hox-3.4* tailbud expression has disappeared long before somitogenesis is completed at about 30 h after post-fertilization.

Materials and Methods

Subcloning and sequencing of the zebrafish hox-3.4 gene

The *hox-3.4* gene was subcloned from two lambda EMBL3 clones, which were previously shown to contain the *hox-3.4* homeobox region (Eiken *et al.*, 1987) into the pGem7zf(+) vector, obtained from Promega Biotec. Constructs suitable for sequencing of both strands were made and the complete genomic sequence of the gene was determined, using the chain termination procedure (Sanger *et al.*, 1977). The EMBL Data Library accession number to the complete *hox-3.4* sequence is X68324.

Computer comparisons of different Hox sequences

Computer comparisons were performed with the Sequence Analysis Software, package version 7.0, obtained from the Genetics Computer Group (Devereux *et al.*, 1984).

In situ hybridization on tissue sections

Zebrafish embryos, at three different developmental stages, were manually dechorionated and sectioned appropriately for *in situ* hybridization according to Krauss *et al.* (1991).

Acknowledgments

This work was supported by fellowships and grants from the Norwegian Research Council NAVF, Norwegian Cancer Society, Hasselblad Foundation and Nansen Foundation.

References

ACAMPORA, D., D'ESPOSITO, M., FAIELLA, A., PANNESE, A.M., MIGLIACCIO, E., MORELLI, F., STORNAIUOLO, A., NIGRO, V., SIMEONE, A. and BONCINELLI, E. (1989). The human HOX gene family. *Nucleic Acids Res. 17:* 10385-10402.

ARCIONI, L., SIMEONE, A., GUAZZI, S., ZAPPAVIGNA, V., BONCINELLI, E. and MAVILIO, F. (1992). The upstream region of the human homeobox gene *HOX3D* is a target for regulation by retinoic acid and HOX homeoproteins. *EMBO J. 11:* 265-277.

BEEMAN, R.W. (1987). A homeotic gene cluster in the red flour beetle. *Nature 327:* 247-249.

BIGGIN, M.D. and TJIAN, R. (1989). A purified *Drosophila* homeodomain protein represses transcription *in vitro*. *Cell 58:* 433-440.

BONCINELLI, E., SIMEONE, A., ACAMPORA, D. and MAVILIO, F. (1991). *HOX* gene activation by retinoic acid. *Trends Genet. 7:* 329-334.

CARRASCO, A.E., McGINNIS, W., GEHRING, W.J. and DE ROBERTIS, E.M. (1984). Cloning of a *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to *Drosophila* homeotic genes. *Cell 37:* 409-414.

DESSAIN, S., GROSS, C.T., KUZIORA, M.A. and McGINNIS, W. (1992). Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J. 11:* 991-1002.

DEVEREUX, J., HAEBERLI, P. and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res. 12:* 387-395.

DOLLÉ, P., IZPISÚA-BELMONTE, J.-C., FALKENSTEIN, H., RENUCCI, A. and DUBOULE, D. (1989). Coordinate expression of the murine *Hox-5* complex homeobox-containing genes during limb pattern formation. *Nature 342:* 767-772.

EIKEN, H.G., NJØLSTAD, P.R., MOLVEN, A. and FJOSE, A. (1987). A zebrafish homeobox-containing gene with embryonic transcription. *Biochim. Biophys. Res. Commun. 149:* 1165-1171.

FIBI, M., ZINK, B., KESSEL, M., COLBERG-POLEY, A.M., LABEIT, S., LEHRACH, H. and GRUSS, P. (1988). Coding sequence and expression of the homeobox gene *Hox-1.3. Development 102: 349*-359.

FRITZ, M. and DE ROBERTIS, E.M. (1988). *Xenopus* homeobox-containing cDNAs expressed in early development. *Nucleic Acids Res. 16:* 1453-1469.

GAUNT, S.J., COLETTA, P.L., PRAVTCHEVA, D. and SHARPE, P.T. (1990). Mouse *Hox-3.4*: homeobox sequence and embryonic expression patterns compared with other members of the *Hox* gene network. *Development 109:* 329-339.

GIBSON, G., SCHIER, A., LEMOTTE, P. and GEHRING, W. (1990). The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. *Cell 62:* 1087-1103.

GRAHAM, A., MADEN, M. and KRUMLAUF, R. (1991). The murine Hox-2 genes display dynamic dorsoventral patterns of expression during central nervous system development. *Development 112:* 255-264.

GRAHAM, A., PAPALOPULU, N. and KRUMLAUF, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell 57:* 367-378.

HAN, K., LEVINE, M.S. and MANLEY, J.L. (1989). Synergistic activation and repression of transcription by *Drosophila* homeobox-proteins. *Cell 56:* 573-583.

HANNEMAN, E., TREVARROW, B., METCALFE, W.K., KIMMEL, C.B. and WESTERFIELD, M. (1988). Segmental patterns of development of the hindbrain and spinal cord of the zebrafish embryo. *Development 103:* 49-58.

HUNT, P. and KRUMLAUF, R. (1991). Deciphering the Hox-code: clues to patterning branchial regions of the head. *Cell 66:* 1075-1078.

IZPISÚA-BELMONTE, J-C., FALKENSTEIN, H., DOLLÉ,P., RENUCCI, A. and DUBOULE, D. (1991). Murine genes related to the *Drosophila AbdB* homeotic gene are sequentially expressed during development of the posterior part of the body. *EMBO J. 10:* 2279-2289.

JAYNES, J.B. and O'FARRELL, P.H. (1991). Active repression of transcription by the Engrailed homeodomain protein. *EMBO J. 10:* 1427-1433.

KESSEL, M. and GRUSS, P. (1991). Homeotic transformations of murine vertebrae and concomitant alteration of *Hox* codes induced by retinoic acid. *Cell 67:* 89-104.

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell 44:* 283-292.

KOZAK, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res. 15:* 8125-8148.

KRAUSS, S., JOHANSEN, T., KORZH, V., MOENS, U., ERICSON, J.U. and FJOSE, A. (1991). Zebrafish *pax[zf-a]:* a paired box-containing gene expression in the neural tube. *EMBO J. 10:* 3609-3619.

KUZIORA, M.A. and McGINNIS, W. (1990). Altering the regulatory targets of the *Deformed* protein in *Drosophila* embryos by substituting the *Abdominal-B* homeodomain. *Mech. Dev. 33:* 83-94.

LE MOUELLIC, H., LALLEMAND, Y. and BRÛLET, P. (1992). Homeosis in the mouse induced by a null mutation in the *Hox-3.1* gene. *Cell 69:* 251-264.

LEVINE, M., RUBIN, G.M. and TJIAN, R. (1984). Human DNA sequences homologous to a protein coding region conserved between homeotic genes of *Drosophila. Cell 38:* 667-673.

MAVILIO, F., SIMEONE, A., GIAMPAOLO, A., FAIELLA, A., ZAPPAVIGNA, V., ACAMPORA, D., POIANA, G., RUSSO, G., PESCHLE, L. and BONCINELLI, E. (1986). Differential and stage-related expression in embryonic tissues of a new human homeobox gene. *Nature 324:* 664-668.

McGINNIS, W., GARBER, R.L., WIRZ, J., KUROIWA, A. and GEHRING, W.J. (1984a). A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell 37:* 403-408.

McGINNIS, W., HART, C.P., GEHRING, W.J. and RUDDLE, F.H. (1984b). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes in *Drosophila. Cell 38:* 675-680.

McGINNIS, W., LEVINE, M.S., HAFEN, E., KUROIWA, A. and GEHRING, W.J. (1984c). A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and Bithorax complexes. *Nature 308:* 428-433.

MOLVEN, A., HORDVIK, I., NJØLSTAD, P.R., VAN GHELUE, M. and FJOSE, A. (1992). The zebrafish homeobox gene *hox[zf-114]:* primary structure, expression pattern and evolutionary aspects. *Int. J. Dev. Biol. 36:* 229-237.

MOLVEN, A., WRIGHT, C.V.E., BREMILLER, R., DE ROBERTIS, E.M. and KIMMEL, C.B. (1990). Expression of a homeobox gene product in normal and mutant zebrafish embryos: evolution of the tetrapod body plan. *Development 109:* 279-288.

MORRISS-KAY, G.M., MURPHY, P., HILL, R.E. and DAVIDSON, D.R. (1991). Effects of retinoic acid excess on expression of *Hox-2.9* and *Krox-20* and on morphological segmentation in the hindbrain of mouse embryos. *EMBO J. 10:* 2985-2995.

NJØLSTAD, P.R., MOLVEN, A. and FJOSE, A. (1988a). A zebrafish homologue of the murine *Hox-2.1* gene. *FEBS Lett. 230:* 25-30.

NJØLSTAD, P.R., MOLVEN, A., APOLD, J. and FJOSE, A. (1990). The zebrafish homeobox gene *hox-2.2:* transcription unit, potential regulatory regions and *in situ* localization of transcripts. *EMBO J. 9:* 515-524.

NJØLSTAD, P.R., MOLVEN, A., HORDVIK, I., APOLD, J. and FJOSE, A. (1988b). Primary structure, developmentally regulated expression and potential duplication of the zebrafish homeobox gene ZF-21. *Nucleic Acids Res. 16:* 9097-9111.

PUSCHEL, A.W., BALLING, R. and GRUSS, P. (1991). Separate elements cause lineage restriction and specify boundaries of *Hox-1.1* expression. *Development 112:* 279-287.

RANGINI, Z., FRUMKIN, A., SHANI, G., GUTTMAN, M., EYAL-GILADI, H., GRUBENBAUM, Y. and FAINSOD, A. (1989). The chicken homeobox gene *Chox1* and *Chox3:* cloning, sequencing and expression during embryogenesis. *Gene 76:* 61-74.

RUIZ I ALTABA, A., and MELTON, D.A. (1989). Bimodial and graded expression of the *Xenopus* homeobox gene *Xhox3* during embryonic development. *Development 106:* 173-183.

SANGER, F., NICKLEN, S. AND COULSON, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA 74:* 5463-5464.

SIMEONE, A., PANNESE, M., ACAMPORA, D., D'ESPOSITO, M. and BONCINELLI, E. (1988). At least three human homeoboxes on chromosome 12 belong to the same transcription unit. *Nucleic Acids Res. 16:* 5379-5390.

TANAKA, M., and HERR, W. (1990). Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell 60:* 375-386.

TREVARROW, B., MARKS, D.L. and KIMMEL, C.B. (1990). Organization of hindbrain segments in the zebrafish embryo. *Neuron 4:* 669-679.

WHITING, J., MARSHALL, H., COOK, M., KRUMLAUF, R., RIGBY, P.W.J., SCOTT, D. and ALLEMANN, R.K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. *Genes Dev. 5:* 2048-2059.

WILKINSON, D.G., BHATT, S., COOK, M., BONCINELLI, E. and KRUMLAUF, R. (1989). Segmental expression of Hox-2 homeobox-containing genes in the developing mouse hindbrain. *Nature 341:* 405-409.

WINSLOW, G.M., HAYASHI, S., KRASNOW, M., HOGNESS, D.S. and SCOTT, M.P. (1989). Transcriptional activation by the *Antennapedia* and *Fushi tarazu* proteins in cultivated *Drosophila* cells. *Cell 57:* 1017-1030.

WRIGHT, C.V.E., CHO, K.W.Y., HARDWICKE, J., COLLINS, R.H. and DE ROBERTIS, E.M. (1989). Interference with function of a homeobox gene in *Xenopus* embryos produces malformations of the anterior spinal cord. *Cell 59:* 81-93.

Accepted for publication: March 1993