Volume 230, number 1,2 25–30 FEB 05661 March 1988

A zebrafish homologue of the murine *Hox*-2.1 gene

Pål R. Njølstad, Anders Molven and Anders Fjose

Laboratory for Biotechnology, University of Bergen, PO Box 3152, Årstad, N-5029 Bergen, Norway

Received 15 December 1987; revised version received 23 January 1988

Homeobox-containing sequences were isolated from a genomic library of zebrafish *(Brachydanio rerio)*. A λ clone containing two homeobox cross-hybridizing regions was characterized. DNA sequencing of one of these regions (ZF-21) revealed that it contains a homeobox closely related to the Antennapedia class of *Drosophila* homeobox sequences. Moreover, the deduced amino acid sequence of the C-terminal end (81 residues including the homeobox) is identical to the corresponding part of the murine Hox-2.1 protein. Similar to Hox-2.1, a ZF-21 derived transcript of 2.3 kb is present in embryos at the somite forming stages.

Homeobox; Protein homology; Embryogenesis; *(Brachydanio rerio)*

1. INTRODUCTION

The homeobox is a conserved DNA sequence present in multiple copies in higher organisms including the vertebrates [1–4]. This DNA sequence encodes a protein domain thought to be involved in DNA binding [5,6]. In the fruit fly *Drosophila melanogaster* these proteins are known to be involved in regulating the generation and morphological differentiation of body segments [7–9]. Therefore, speculations have been proposed that also the pattern forming processes of vertebrate embryos are regulated by homeobox-containing genes. However, it has proved difficult to obtain conclusive results concerning the developmental role of mammalian homeobox genes due to the inaccessibility of early embryonic stages, the scarcity of mutants and the complex body pattern of mammals.

If the basic regulatory mechanisms of vertebrate development are very similar, studies of easily accessible embryos of more primitive vertebrate species may help to elucidate the functional role of mammalian homeobox genes. The quickly developing zebrafish (*Brachydanio rerio*) embryos which are easy to observe through highly transparent chorion membranes, are particularly well suited for such investigations [10–12].

The present study reports the molecular cloning of two closely linked zebrafish homeobox genes. One of these seems to be the zebrafish homologue of the murine *Hox*-2.1 gene [13]. The time of embryonic expression of this zebrafish *Hox*-2.1 homologue coincides with the period of somite formation.

Correspondence address: A. Fjose, Laboratory for Biotechnology, University of Bergen, PO Box 3152, Årstad, 5029 Bergen, Norway

2. EXPERIMENTAL

Total DNA from adult zebrafish was partially digested with the restriction enzyme *Sau*3A, and DNA fragments of 15–20 kb were used to construct a genomic library in the λEMBL3 vector as described by Eiken et al. [14]. 5×10^5 clones were screened using ^{32}P-labelled homeobox-containing DNA fragments from the *Drosophila* genes 'Antennapedia' and 'Sex combs reduced' [15,16]. DNA fragments were subcloned into the plasmid vector pGem-4 (Promega Biotec, USA) and sequenced as described [14].

Total RNA was extracted from staged zebrafish embryos as described for embryos of Atlantic salmon [17]. Aliquots of 15 μg RNA from different embryonic stages were electrophoresed on a 1.0% agarose/formaldehyde gel and transferred to a nylon filter by blotting [18]. The filters were prehybridized at 58°C for 4 h in a high stringency solution according to the protocol of McGinnis et al. [15]. DNA fragments cloned in the pGem-4 vector (Promega Biotec, USA) were used

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50 © 1988 Federation of European Biochemical Societies

to make single stranded RNA probes which were added directly to the prehybridization solution. Following hybridization at 58°C for 48 h, the filters were washed in 0.1 × SSC/0.1% SDS at 70°C for a total of 90 min with three changes of washing solution.

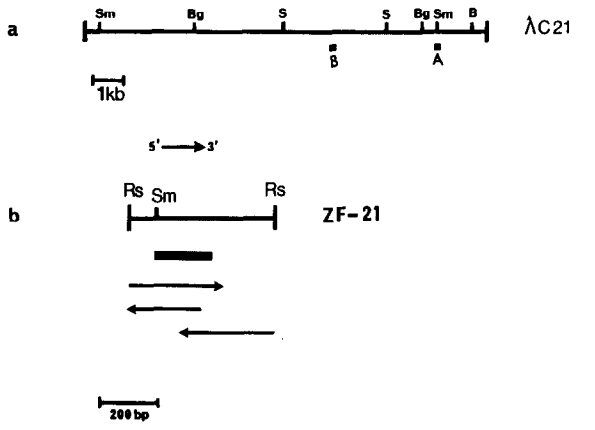

Fig.1. Restriction maps and DNA sequencing strategies. (a) Distribution of restriction enzyme sites in the homeobox-containing λ clone C21. The approximate positions of homeobox sequences are indicated by black boxes labelled A and B. (b) Localization (black bar) and 5′-3′ orientation (thick arrow) of the A homeobox sequences within a small subregion of λC21. Sequencing by the dideoxy method was done as illustrated by the thin arrows. Restriction enzymes: B, *Bam*HI; Bg, *Bgl*II; R, *Rsa*I; S, *Sal*I; Sm, *Sma*I.

```
(a)      -60                                                                -1
ZF-21    TGC TAA TGT TTG TTC CCC TTT GCG TTT TGT CCA CAT TAT TCA GAT ATG ACT GGA CCA GAC
Hox-2.1  CCA C-G ATA --C CC- TGG A-G AG- AAG CT- -AC ATC AGC CAC --- --- --- --G --- ---
Hox-1.3  CAG ATC -AC CCC -GC ATG CGC AA- C-G CAC ATT AG- C-C GAC A-- --A GG- --C --- --A
Antp     -CG GGG ATG CC- -CT --A C-G TAT CCC --G ATG -GA AG- CAG TT- GGT -AG T-T -A- --A

         1                                                                  60
ZF-21    GGC AAA AGG GCC CGA ACT GCA TAT ACC CGC TAT CAG ACG CTA GAG CTG GAA AAA GAG TTC
Hox-2.1  --A --- --- --- --G --C --C --- --T --- --C --- --C --G --- --- --- --G --A ---
Hox-1.3  --- --- --- --- --G --G --C --C --T --- --C --- --C --G --- --- --G --- --A ---
Antp     C-- --- C-C -GA A-G CAG A-- --C --- --G --C --- --T --- --- --A --G --G --- --T

         61                                                                 120
ZF-21    CAT TTC AAT AGA TAC CTC ACC CGA AGA AGG AGG ATA GAG ATA GCC CAC GCT CTA TGC CTC
Hox-2.1  --C --- --- C-C --- --G --- --G C-G C-A C-T --C --- --C --- --- --G --T --- --G
Hox-1.3  --C --- --C C-C --- --G --- --C C-- --A --- --C --A --- --T --T --C --T --- ---
Antp     --C --- --- C-C --- T-G --- --T C-G C-A --- --C --- --C --- --- --C --G --- ---

         121                                                                180
ZF-21    TCA GAG CGT CAA ATT AAG ATA TGG TTC CAA AAC CGG CGG ATG AAA TGG AAA AAG GAC AAT
Hox-2.1  --C --- --- --G --C --A --C --- --- --G --- --T --C --- --G --- --G --A --- --C
Hox-1.3  --C --- A-A --- --- --A --C --- --- --- --- A-- A-- --- --G --- --- --A --T ---
Antp     A-G --- --C --G --A --- --T --- --- --G --T --- --C --- --G --- --G --- --G --C

         181                                                        231
ZF-21    AAA CTG AAG AGC ATG AGT CTG GCT ACC GCA GGT AGC GCT TTC CAA CCA TAG
Hox-2.1  --- --- --A --T --- --- --- --- --A --A --C --- --C --- --- --T -GA
Hox-1.3  --G --- --A --- --- --- A-- --C G-G --- --G G-G --- --- -GC --C -GA
Antp     --G AC- --- G-- GA- CCG GGA T-C GGA -GC -AA G-- -AC GAG AT- A-- CCA
```

Fig.2. Comparison of DNA (a) and deduced protein (b) sequences from the ZF-21 region with cDNA sequences from Antennapedia and two murine *Hox* genes. The homeobox extends from position 1 to 180. Homologies are denoted with hyphens and sequence

3. RESULTS

3.1. *Isolation and characterization of a zebrafish homeobox sequence*

Homeobox-containing genes have recently been described for Atlantic salmon [17] and zebrafish [14]. We have cloned a total of eight different zebrafish genes [14] in cross-hybridization experiments with *Drosophila* homeobox sequences of the Antennapedia class [19]. One of the λ clones isolated, λC21, carries two regions of cross-hybridizing sequences (fig.1a) and overlapping clones contain additional homeoboxes (Njølstad and Fjose, unpublished). This implies that λC21 is derived from a gene cluster similar to the *Hox* loci of mouse [20]. One of the homeobox-containing regions of λC21 has been sequenced according to the strategy illustrated in fig.1b. A nucleotide sequence which includes a 180 bp homeobox, 60 bp of 5′-flanking and 51 bp 3′-flanking sequences is in fig.2a compared to the Antennapedia gene (*Antp*) of *Drosophila* [21] and two mouse homeobox genes, *Hox*-1.3 and *Hox*-2.1 [22,23].

Comparison of ZF-21 with *Antp* reveals DNA and protein homologies as high as 74.4 and 88.8%, respectively. Therefore, the zebrafish sequence ap-

```
(b)                                                                  -5                -1
ZF-21                                                                Met Thr Gly Pro Asp
Hox-2.1                                                              --- --- --- --- ---
Hox-1.3                                                              Ile Gly --- --- Glu
Antp                                                                 Gly Lys Cys Gln Glu

          1                                                                            20
ZF-21     Gly Lys Arg Ala Arg Thr Ala Tyr Thr Arg Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe
Hox-2.1   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Hox-1.3   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Antp      Arg Gly --- Gly --- Gln Thr --- --- --- --- --- --- --- --- --- --- --- --- ---

          21                                                                           40
ZF-21     His Phe Asn Arg Tyr Leu Thr Arg Arg Arg Arg Ile Glu Ile Ala His Ala Leu Cys Leu
Hox-2.1   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Hox-1.3   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Antp      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

          41                                                                           60
ZF-21     Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys Asp Asn
Hox-2.1   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Hox-1.3   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Antp      Thr --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- Glu ---

          61                                                        76
ZF-21     Lys Leu Lys Ser Met Ser Leu Ala Thr Ala Gly Ser Ala Phe Gln Pro ***
Hox-2.1   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ***
Hox-1.3   --- --- --- --- --- --- Met --- Ala --- --- Gly --- --- Arg Pro ***
Antp      --- Thr --- Gly Glu Pro Gly Ser Gly Gly Glu Gly Asp Glu Ile Thr Pro
```

deviations are indicated by the substituting nucleotide or amino acid. Arrowheads indicate the positions for splicing in the three vertebrate genes. A potential stop codon in the genomic ZF-21 sequence is underlined.

pears to be a typical member of the Antennapedia class [19] of homeoboxes. A much more outstanding feature of this homeobox gene is its remarkable similarity to *Hox*-1.3 [23] and *Hox*-2.1 [13,22]. Interestingly, these two murine genes seem to be closely related through a relatively recent duplication event [23].

In fig.2a, the genomic ZF-21 sequence is compared to cDNA derived from the two *Hox* genes. The sequence homology starts at a consensus splice acceptor site 24 bp upstream of the homeobox of ZF-21 and extends to a position 51 bp beyond the 5'-end where all the three vertebrate genes have signals for termination of protein synthesis (fig.2a,b). Interestingly, the splicing junctions in the two *Hox* cDNA sequences are located in almost identical positions as the potential splice site of ZF-21. Upstream of this splicing signal the ZF-21 genomic DNA contains a translation stop codon. Therefore, the C-terminal encoding regions of these three genes seem to be organized in an almost identical manner.

Throughout the 255 bp region shared between ZF-21 and *Hox*-2.1 only silent nucleotide changes are found. As a result, the corresponding protein sequences (fig.2b) are completely identical in spite of DNA homologies of only 80%. Similarly, the deduced amino acid sequences of the ZF-21 and *Hox*-1.3 homeoboxes are identical. However, beyond the borders of the homeobox these two genes diverge somewhat.

3.2. *Expression during embryogenesis*

To obtain further information on the evolutionary relationship between the three vertebrate genes, the embryonic transcription of ZF-21 was compared to the expression patterns of *Hox*-2.1 and *Hox*-1.3. This was achieved by Northern analysis of embryonic zebrafish RNA. An RNA probe complementary to the homeobox region of ZF-21 (fig.1b) was hybridized to a Northern blot of total RNA from different embryonic stages (fig.3).

Transcripts of the ZF-21 gene are only present at very low levels in RNA isolated from fertilized eggs that have developed for 2 and 7 h, respectively (fig.3, lanes 1 and 2). An increased accumulation of ZF-21 transcripts is first seen at the 16 h stage (lane 3) when the embryos undergo an early phase of somite formation [24]. Persistent expression of

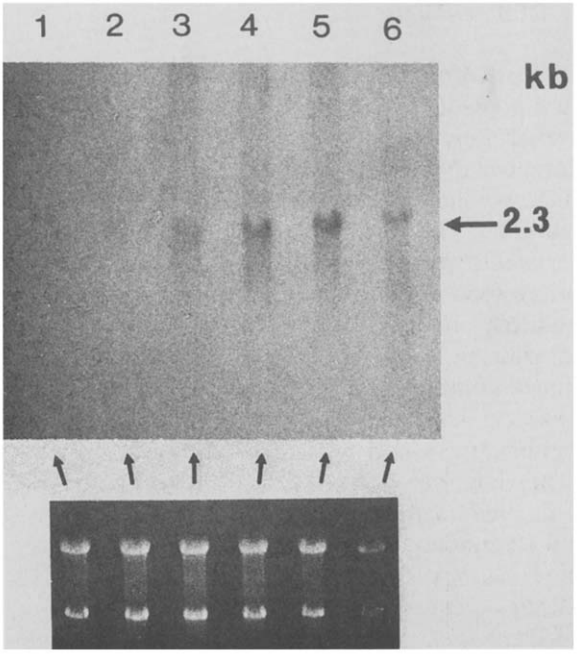

Fig.3. Temporal pattern of ZF-21 expression during development. Lanes 1–5 contain total RNA from embryos of stages 2, 7, 16, 29 and 48 h, respectively. Lane 6 contains an equal amount of total RNA isolated from adult zebrafish. The corresponding RNA lanes stained with ethidium bromide are shown below the blot. A single transcript of 2.3 kb is detected.

the ZF-21 gene seems to occur in older developmental stages and a peak level of the 2.3 kb transcript is obtained about 29 h after fertilization. At this stage the embryo has formed 30 somites and the morphological differentiation is considerable as judged from the presence of optic lenses and a functioning heart [24]. A similar amount of transcripts is present in 48 h embryos (lane 5) when the process of differentiation of internal and external organs is almost completed. Using poly(A)$^+$ RNA for Northern analysis of ZF-21, both the transcript size and temporal expression pattern were confirmed (not shown).

Studies on the developmental expression of *Hox*-1.3 and *Hox*-2.1 have demonstrated that both genes are transcribed in mouse embryos during the somite forming stages [13,22,23,25]. Similar to ZF-21, the *Hox*-2.1 gene generates a major mRNA species of about 2.3 kb [13,22]. In contrast, the major transcript derived from *Hox*-1.3 is 1.85 kb [23]. Thus, ZF-21 and *Hox*-2.1 are also more closely related with respect to transcript size.

4. DISCUSSION

This report describes a zebrafish homeobox sequence (ZF-21) of the Antennapedia class which is located in a gene cluster. Surprisingly, ZF-21 encodes the C-terminal part of a putative protein sequence which is identical to the corresponding domain of *Hox*-2.1 in mouse [13,22]. ZF-21 shares the same sequence identity with the human homologue of *Hox*-2.1 [4,13]. Some sequence homology has previously been observed between homeobox-containing genes of distantly related vertebrate species (frog and human). In that case, however, the amino acid sequence outside the borders of the homeodomain were not found to be completely identical [26].

The identical protein domains found in ZF-21 and its mammalian homologues are likely to interact with other strongly conserved DNA sequence elements and/or regulatory proteins. Consequently, many of the components in the regulatory circuit in which the ZF-21 gene is an integral part, must have closely related equivalents in mammals. This implies that information obtained from studies on zebrafish homeobox genes probably can be applied to gain insight in the developmental regulatory mechanisms of vertebrates in general.

A possible relatedness between the promoter elements of ZF-21 and its mouse equivalent might be revealed by comparing temporal and spatial expression patterns in both species. We have made a first attempt along these lines by analysing the embryonic expression of ZF-21 by Northern blotting. The highest concentration of the 2.3 kb ZF-21 transcript was found in embryos which undergo somite formation and this corresponds to the developmental stage in mouse embryos at which a *Hox*-2.1-derived mRNA of the same size is present at high levels [13,22].

We also detected the ZF-21 transcript in embryos that have developed for only 2 and 7 h, respectively. In contrast, the expression of *Hox*-2.1 has not been monitored in the earliest phase of embryonic development due to the difficulty of obtaining experimental material [13].

Although the *Hox*-2.1 transcription data on early presomite stages are incomplete, its resemblance to ZF-21 with respect to DNA sequence, transcript size and somite stage expression suggests that the functional roles of these genes are essentially the same. It remains, however, to be seen whether the spatial expression of the zebrafish gene corresponds to the regionally restricted distribution of *Hox*-2.1 transcripts observed within the central nervous system of mouse embryos [22,27]. These and other questions concerning the functions of ZF-21 can now be further investigated by performing in situ hybridization and/or immunolocalization experiments on zebrafish embryos of all developmental stages.

Acknowledgements: We thank K. Kleppe and J. Apold for support and encouragement. The work was made possible by grants from the Norwegian Research Councils NAVF and NTNF.

REFERENCES

[1] McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A. and Gehring, W.J. (1984) Cell 37, 403–408.
[2] Carrasco, A.E., McGinnis, W., Gehring, W.J. and De Robertis, E.M. (1984) Cell 37, 409–414.
[3] McGinnis, W., Hart, C.P., Gehring, W.J. and Ruddle, F.H. (1984) Cell 38, 675–680.
[4] Levine, M., Rubin, G.M. and Tjian, R. (1984) Cell 38, 667–673.
[5] Laughon, A. and Scott, M.P. (1984) Nature 310, 25–31.
[6] Desplan, C., Theis, J. and O'Farrell, P.H. (1985) Nature 318, 630–635.
[7] Garcia-Bellido, A. (1977) Am. Zool. 17, 613–629.
[8] Lewis, E.B. (1978) Nature 276, 565–570.
[9] Nusslein-Volhard, C. and Wieschaus, E. (1980) Nature 287, 795–801.
[10] Streisinger, G., Walker, C., Dower, N. and Singer, F. (1981) Nature 291, 293–296.
[11] Kimmel, C.B. and Warga, R.M. (1986) Science 231, 365–368.
[12] Marcey, D. and Nusslein-Volhard, C. (1986) Nature 321, 380–381.
[13] Hauser, C.A., Joyner, A.L., Klein, R.D., Learned, T.K., Martin, G. and Tjian, R. (1985) Cell 43, 19–28.
[14] Eiken, H.G., Njølstad, P.R., Molven, A. and Fjose, A. (1987) Biochem. Biophys. Res. Commun., in press.
[15] McGinnis, W., Levine, M., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) Nature 308, 428–433.
[16] Kuroiwa, A., Kloter, U., Baumgartner, P. and Gehring, W.J. (1985) EMBO J. 4, 3757–3764.
[17] Fjose, A., Molven, A. and Eiken, H.G. (1987) Gene, in press.
[18] Thomas, P. (1980) Proc. Natl. Acad. Sci. USA 77, 5201–5205.
[19] McGinnis, W. (1985) Cold Spring Harb. Symp. Quant. Biol. 50, 263–270.

[20] Martin, G.R., Boncinelli, E., Duboule, D., Gruss, P., Jackson, I., Krumlauf, R., Lonai, P., McGinnis, W., Ruddle, F.H. and Wolgemuth, D. (1986) Nature 325, 21–22.
[21] Schneuwly, S., Kuroiwa, A., Baumgartner, P. and Gehring, W.J. (1986) EMBO J. 5, 733–739.
[22] Krumlauf, B., Holland, P.W., McVey, J.H. and Hogan, B.L.M. (1987) Development 99, 603–617.
[23] Odenwald, W.F., Taylor, C.F., Palmer-Hill, F.J., Friedrich, V., Tani, M. and Lazzarini, R.A. (1987) Gene Dev. 1, 482–496.
[24] Hisaoka, K.K. and Battle, H.I. (1958) J. Morphol. 102, 311–321.
[25] Dony, C. and Gruss, P. (1987) EMBO J. 6, 2965–2975.
[26] Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987) EMBO J. 6, 767–777.
[27] Utset, M.F., Awgulewitsch, A., Ruddle, F.H. and McGinnis, W. (1987) Science 235, 1379–1382.