Volume 231, number 2, 355–360 FEB 05792 April 1988

A zebrafish engrailed-like homeobox sequence expressed during embryogenesis

Anders Fjose, Hans G. Eiken, Pål R. Njølstad, Anders Molven and Ivar Hordvik

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

Received 22 February 1988

The zebrafish genome was found to contain two sequences which cross-hybridize strongly with the engrailed gene of *Drosophila*. Several independent clones containing one of these cross-hybridizing sequences were isolated from a zebrafish genomic library. Characterization of this region (ZF-EN) by DNA sequencing showed that it shares about 70% sequence identity with the engrailed homeobox. More extensive homeobox homology (>90%) was found relative to the murine *En* genes. The closest relationship exists between ZF-EN and *En-2* where the C-terminal domains (104 amino acids) encoded by these genes are almost identical. We also observed that ZF-EN and *En-2* are very similar with respect to their transcript sizes and termporal expression patterns.

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

1. INTRODUCTION

A conserved protein-encoding DNA sequence of about 180 bp, named the homeobox, is present in multiple copies in the genomes of most higher animal species [1–4]. Genetic and molecular analyses of *Drosophila* homeobox-containing genes have demonstrated that they are involved in regulating the generation and morphological differentiation of body segments [5,6]. Recent studies on mice and frogs have provided preliminary evidence that also the development of vertebrate embryos is in part regulated by homeobox genes [7–10].

Five murine *Hox* loci have been isolated which contain clusters of homeobox sequences that are related to the *Drosophila* Antennapedia class [11–13]. Moreover, similar to *Drosophila* [14,15] only two engrailed-like sequences have been detected in the mouse genome [16,17]. Apart from the homeobox, these two mouse genes, *En-1* and *En-2*, also share additional sequences with the *Drosophila* engrailed gene (*en*) which is segmentally expressed in embryos [15,18]. Therefore, efforts have been made to investigate whether the mouse *En* genes are also transcribed in a segmental manner. However, experimental material of early mouse embryos is very difficult to obtain and detailed information on the spatial expression patterns of *En-1* and *En-2* is therefore not yet available [16,17].

Similarly, problems have been encountered when other murine homeobox genes have been analysed for early embryonic expression [7,8,10]. More primitive vertebrates where the embryos easily can be accessed at all developmental stages, should therefore be amenable for the investigation of homeobox-containing genes. In this connection, fish are well suited, especially since they also retain the segmentation of muscle tissue throughout development.

As a model system when studying fish homeobox genes, we have selected the zebrafish (*Brachydanio rerio*), an almost ideal species for the genetic analyses of development [19–22]. Recently, we have described a zebrafish homologue of the

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

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

murine *Hox*-2.1 gene [23]. This work shows that the zebrafish genome also contains at least two engrailed-like homeoboxes. Molecular cloning and characterization of one of these sequences demonstrate its close resemblance to the *En-2* gene of mouse.

2. EXPERIMENTAL

Genomic DNA was prepared from adult zebrafish as described by Frei et al. [24]. Restriction enzyme digestion of genomic DNA and subsequent Southern blotting of DNA fragments separated on agarose gels were performed according to standard procedures [25].

A genomic library from zebrafish was constructed in the EMBL3 vector as described by Eiken et al. [26]. 5×10^5 clones were screened using a ^{32}P-labelled 0.9 kb *Eco*RI fragment from the *en* gene of *Drosophila* as a probe [15]. DNA fragments were subcloned into the plasmid vectors pGem-3/4 (Promega Biotec, USA) and sequenced as described in [26].

Total RNA was extracted from staged zebrafish embryos by the same method as used for embryos of Atlantic salmon [27], and poly(A)$^+$ RNA was selected by oligo(dT)-cellulose chromatography [28]. The oligo(dT)-purified RNA was quantified and tested for efficient removal of ribosomal RNA by electrophoresis together with RNA standards on agarose gels stained with ethidium bromide. Aliquots of 5 μg poly(A)$^+$ RNA were run on a 1% formaldehyde/agarose gel and transferred to nitrocellulose filters which were hybridized with nick-translated probes according to standard procedures [29].

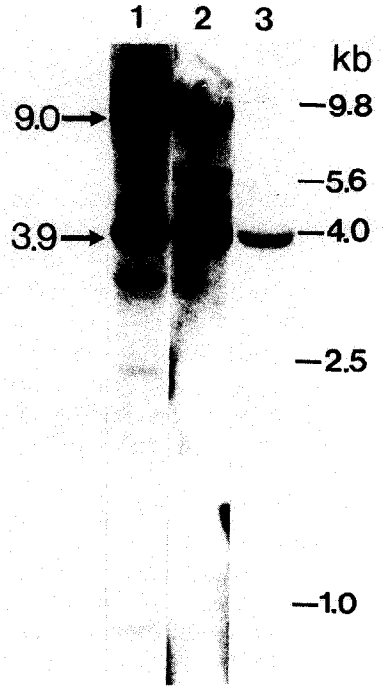

Fig.1. Engrailed-like sequences in the zebrafish genome. Southern blot of *Hin*dIII-digested zebrafish genomic DNA hybridized at conditions of reduced stringency with (lane 1) a homeobox-containing *en* probe from *Drosophila* [15] and (lane 2) the *Alu*I/*Rsa*I fragment of ZF-EN (see fig.2b). Lane 3 shows the hybridization signal obtained with the same ZF-EN probe under stringent conditions. Following removal of hybridized DNA by washing, the same filter was used in all three experiments. Arrows point towards the two strongly cross-hybridizing fragments of 3.9 kb and 9.0 kb, respectively. Size markers are indicated alongside the blots.

3. RESULTS

3.1. *Isolation of cross-hybridizing sequences*

Homeobox-containing sequences of the Antennapedia class have recently been described for both zebrafish [23,26] and Atlantic salmon [27]. In vertebrates engrailed-like genes have so far only been analysed in mammals [16,17]. Therefore, in an initial experiment we analysed total DNA to detect engrailed-homeobox cross-hybridizing sequences. Zebrafish DNA samples digested with a number of restriction enzymes and transferred to nitrocellulose filters were hybridized with a probe containing engrailed homeobox sequences from *Drosophila* under reduced stringency conditions.

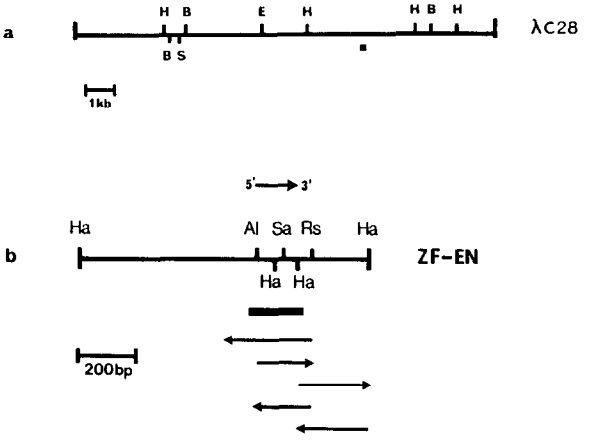

Fig.2. Restriction maps and DNA sequencing strategies. (a) Distribution of restriction enzyme sites in the engrailed cross-hybridizing λ clone C28. The approximate position of the homeobox sequence is indicated by a black box. (b) Localization (black bar) and 5'-3' orientation (thick arrow) of the homeobox sequence within a small sub region of λC28. Sequencing by the dideoxy method was done as illustrated by the thin arrows. Restriction enzymes: Al, *Alu*I; B, *Bam*HI; Bg, *Blg*II; E, *Eco*RI; H, *Hin*dIII; Ha, *Hae*III; Rs, *Rsa*I; S, *Sal*I; Sa, *Sau*3A.

As shown for *Hin*dIII-digested DNA (fig.1; lane 1), two strong hybridization signals and a number of weaker bands were detected. The observation of two strong hybridization bands is consistent with the finding of two engrailed-like homeoboxes in both the *Drosophila* and mouse genomes [14–17].

Using the same *en* probe for screening of a zebrafish genomic library at reduced hybridization stringency, four independent clones were isolated. Further characterization of the *en* positive clones revealed that they contained the same cross-hybridizing *Hae*III restriction fragments (not shown). One of the four clones isolated (λC28) was further characterized by restriction enzyme analysis (fig.2a). The hybridizing region (*Alu*I/*Rsa*I fragment in fig.2b) was used as a probe against a Southern blot of total genomic DNA. When hybridized at stringent conditions only one band which coincides directly in position with the 3.9 kb *en* cross-hybridizing fragment was detected (fig.1; lane 3). Using the same probe under conditions of reduced hybridization stringency, three additional bands of weaker signal intensities were observed (fig.1; lane 2). These results strongly indicated that the isolated clones contain one of the two zebrafish sequences which are closely related to the engrailed homeobox.

3.2. *Sequence cross-homologies*

Sequence comparisons with the corresponding *Drosophila* [14,15] and mouse [16,17] genes reveal a high degree of relatedness with respect to protein homology and gene structure (fig.3a,b). In the region upstream of the homeobox, the genomic DNA sequence of ZF-EN shares considerable homology with the cDNA sequences of *En-1* and *En-2*. However, a sharp divergence of zebrafish and mouse sequences is seen upstream of position −44 where ZF-EN has a consensus splice acceptor site. Interestingly, both the *En* genes of mouse have splice sites in exactly the same position [17]. Other similarities in gene organization are noticed downstream of the homeobox where the translation termination signals of all three genes coincide.

The zebrafish homeodomain encoded by the homeobox is quite related to the *en* and invected (*inv*) genes of *Drosophila*, sharing about 74% homology (fig.3b). An additional stretch of 20 residues following the C-terminal end of the homeodomain has been found to be highly conserved between the mouse and *Drosophila* proteins [16,17] and this is also the case with ZF-EN.

Within the ZF-EN region limited by a consensus splice site and a termination signal, the few differences relative to the corresponding mouse *En* sequences are predominantly third base changes. As a result, part of the zebrafish protein consists of a stretch of 104 amino acid residues that is co-linear and almost identical to the two mouse *En* proteins. The closest resemblance is seen between ZF-EN and *En-2* which share 89.4% homology.

3.3. *Expression during embryogenesis*

RNA can be isolated with equal efficiency from all stages of zebrafish embryos [23,26] and this allowed us to analyse the transcription of ZF-EN also during the earliest period of development. Embryonic expression of ZF-EN gene was investigated by Northern analysis of poly(A)$^+$ RNA from five different developmental stages. As shown in fig.4, a transcript of 3.8 kb was detected only in postgastrula zebrafish embryos (lanes 3–5). ZF-EN transcripts are first detected at the 16 h stage (lane 3) when embryos undergo an early phase of somite formation [19]. Interestingly, the concentration of the transcript appears to remain at a rather constant level until 48 h after fertilization. This coincides with the stages of somite formation when also the mouse *En-2* gene is expressed [17]. Moreover, the ZF-EN mRNA is of almost identical size as the major transcript (3.7 kb) of *En-2* [17]. This gives more support to the assumption that these genes are homologous.

4. DISCUSSION

In a previous report we have demonstrated the presence of a *Hox*-2.1 homologue in the zebrafish genome [23]. Here we provide evidence for the existence of a zebrafish gene (ZF-EN) equivalent to the mouse *En-2* gene [17]. The putative C-terminal parts (104 amino acids) of these two proteins are about 90% identical, and this is suggestive of strong functional conservation. Thus, during the 400 million years since mammals and fish diverged, the time required for a 1% divergence (unit evolutionary period) of the *En-2*/ZF-EN C-terminal parts has been on average 40 million years. This is even considerably higher than the 20

million years calculated for the highly conserved cytochrome *c* protein [30].

In relation to questions concerning evolution it is also pertinent to ask whether the zebrafish genome contains two equally conserved engrailed-like sequences as have been found in mouse and *Drosophila*. Our Southern blot data (fig.1; lane 2) suggest the situation to be more complex since several fainter bands in addition to two strong hybridization signals were obtained. The 3.9 kb hybridization band of high intensity clearly corresponds to the homeobox-containing *Hin*dIII restriction fragment of λC28 (fig.2a). A 9 kb band of similar signal strength is likely to contain a highly conserved engrailed-like sequence which could correspond to the murine *En-1* gene. Perhaps most interesting, however, is the finding of weak hybridization signals over seven additional restriction fragments. Two of these DNA sequences also appear to hybridize with a ZF-EN probe (fig.1; lane 2). This could be due to weak cross-hybridization with some homeoboxes of the

a

```
              -54        ▼                                                   -1
ZF-EN         GTG CAC GCA GGT CCC AGG TCT CGT AAA CCA AAG AAG AAA GCC GCA AGT AAA GAA
En-1          CC- TC- T-T -▼- --A C-C A-C A-G --G -T- --- --- --- AAG AAC GAG --G ---
En-2          CCC TC- T-- -▼- --- --- --C --A --- --- --- --- --G AA- C-T -AC --- --G
en            T-C --G -AC CCC -G- TAC CGC --C CCC CC- C-- CCA --G -A- AAG -CC --C --C
inv           -G- GGT -GG --C GT- GA- AAG G-G G-G G-C GCC G-T GGG -G- -GG GTG CCG --G

        1                                                                         60
ZF-EN   GAC AAA CGA CCA CGC ACG GCC TTC ACG GCG GAG CAG CTT CAG AGA CTC AAG GCC GAG TTC
En-1    --- --G --G --G --G --- --G --- --- --C --- --- --G --- --- --- --- --G --- ---
En-2    --- --G --G --T --- --A --- --- --T --T --- --- --C --- --G --- --- --T --- --T
en      --G --G --T --- --- --C --G --- T-C AGC --- --- T-G GCC C-C --- --- CGG --- ---
inv     --- --- A-G --G --A --- --- --- -GC -GA AC- --- T-G GCC --- --G --- CA- --- ---

        61                                                                       120
ZF-EN   CAG ACC AAC CGC TAC CTG ACC GAG CAG CGG GCG CAA AGC CTG GCG CAG GAA CTG GGC CTC
En-1    --- G-A --- --- --T A-C --G --- --- --- CGA --G -C- --C --C --- --G --C A-- --G
En-2    --- --- --- A-G --- --- --A --- --- --- --C --G --T --- --A --- --G --C A-- --G
en      A-C GAG --T --- --T --- --- --- -G- A-A CGC --G CAG --- AGC AGC --G T-- --- --G
inv     A-C GAG --T --- --T --- --G --- A-- --A CGC --G CAG --- AGC GG- --- --- --A --G

        121                                                                      180
ZF-EN   AAC GAA TCT CAG ATC AAA ATC TGG TTC CAA AAC AAG CGG GCC AAA ATC AAA AAG GCC AGC
En-1    --T --G --C --- --- --G --- --- --- --- --- --- --T --- --G --- --G --G --- -CA
En-2    --- --G --- --- --- --G --T --- --- --G --- --- --- --- --- --- --G --A --- -CG
en      --- --G G-G --- --- --G --- --- --- --G --- --- --- --- --G --- --G --- T-G -CG
inv     --- --G G-G --- --- --G --- --- --- --G --- --A --- --- --G C-G --- --- T-G ---

        181                                                                      240
ZF-EN   GGC GTC AAG AAC GGT CTG GCA ATA CAC CTG ATG GCA CAG GGA CTG TAC AAC CAC AGC ACC
En-1    --- A-- --- --- --C --- --G C-G --- --C --- --C --- --- --- --- --- --- TCT ---
En-2    --- AA- --- --- AC- T-- --G G-G --- --C --- --- --- --C --- --- --- --T TC- ---
en      --- TC- --A --T CCG --- --- C-G --G --- --- --C --- --- T-- --- --- --- ACC ---
inv     --- AC- --- --T CCG --- --G C-G --G --- --- --G --- --- T-- --- --- --- TCG --G

        241                                     273
ZF-EN   ACG TCA AAG GAG GAC AAA TCA GAC AGT GAC TGA
En-1    --C A-G GTT C-- --- --- GAC --G --C --G -AG
En-2    --- G-C --- --- -G- --G --G --- --C --G -AG
en      GT- C-G CT- ACC A-G G-G GAG --G GAG CT- GAG
inv     -TA C-G CT- ACC CG- G-G GAG --G GAG CTG CAG
```

Fig.3. Comparison of DNA (a) and deduced protein (b) sequences from the ZF-EN region with the corresponding parts from the two murine *En* genes [17] and the *Drosophila* genes *en* and *inv* [14]. The homeobox extends from position 1 to 180. Homologies are denoted

Antennapedia class or alternatively, the fish genome contains a subfamily of engrailed-like sequences which has diverged more from the ancestor gene.

Transcripts of the engrailed gene are present in restricted regions of segment progenitors during the process of gastrulation when the segments of *Drosophila* embryos become visible [15,18]. If *En-1*/ZF-EN play some role in the control of vertebrate embryogenesis analogous to that of the *en* gene, a similar temporal correlation between somitogenesis and expression would be expected. In contrast to *Drosophila*, visible segmentation of fish embryos occurs mainly after the gastrulation period and this coincides exactly with the developmental stages at which ZF-EN transcripts were detected (fig.4). Therefore, the previous finding of *En-2* transcripts in PSA-1 cells, which have been used as an in vitro model for cells in the early stages of mouse development [17], might not reflect the embryonic transcriptional state of this murine gene. However, we cannot exclude the

b

```
                                       -14                                                            -1
ZF-EN                                  Pro Arg Ser Arg Lys Pro Lys Lys Lys Ala Ala Ser Lys Glu
En-1                                   --- --- Thr --- --- Leu --- --- --- Lys Asn Glu --- ---
En-2                                   --- --- --- --- --- --- --- --- --- Asn Pro Asn --- ---
en                                     Arg Tyr Arg --- Pro Lys Gln Pro --- Asp Lys Thr Asn Asp
inv                                    Arg Glu Lys Gly Glu Ala Ala Asp Gly Gly Gly Val Pro ---

        1                                                                             20
ZF-EN   Asp Lys Arg Pro Arg Thr Ala Phe Thr Ala Glu Gln Leu Gln Arg Leu Lys Ala Glu Phe
En-1    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
En-2    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
en      Glu --- --- --- --- --- --- --- Ser Ser --- --- --- Ala Arg --- --- Arg --- ---
inv     --- --- --- --- --- --- --- --- Ser Gly Thr --- --- Ala Arg --- --- His --- ---

        21                                                                            40
ZF-EN   Gln Thr Asn Arg Tyr Leu Thr Glu Gln Arg Ala Gln Ser Leu Ala Gln Glu Leu Gly Leu
En-1    --- Ala --- --- --- Ile --- --- --- --- Arg --- Thr --- --- --- --- --- Ser ---
En-2    --- --- --- --- --- --- --- --- --- --- Arg --- --- --- --- --- --- --- Ser ---
en      Asn Glu --- --- --- --- --- --- Arg --- Arg --- Gln --- Ser Ser --- --- --- ---
inv     Asn Glu --- --- --- --- --- --- Lys --- Arg --- Gln --- Ser Gly --- --- --- ---

        41                                                                            60
ZF-EN   Asn Glu Ser Gln Ile Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Ile Lys Lys Ala Ser
En-1    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- Thr
En-2    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
en      --- --- Ala --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- Ser Thr
inv     --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- Leu --- --- Ser ---

        61                                                                            80
ZF-EN   Gly Val Lys Asn Gly Leu Ala Ile His Leu Met Ala Gln Gly Leu Tyr Asn His Ser Thr
En-1    --- Ile --- --- --- --- --- Leu --- --- --- --- --- --- --- --- --- --- --- ---
En-2    --- Asn --- --- Thr --- --- Val --- --- --- --- --- --- --- --- --- --- --- ---
en      --- Ser --- --- Pro --- --- Val Gln --- --- --- --- --- --- --- --- --- Thr ---
inv     --- Thr --- --- Pro --- --- Val Gln --- --- --- --- --- --- --- --- --- --- ---

        81                                  90
ZF-EN   Thr Ser Lys Glu Asp Lys Ser Asp Ser Asp ***
En-1    --- Thr Val Gln --- --- Asp Glu --- Glu ***
En-2    --- Ala --- --- Gly --- --- --- --- Glu ***
en      Val Pro Leu Thr Lys Glu Glu Glu Glu Leu Glu
inv     Ile Pro Leu Thr Arg Glu Glu Glu GLu Leu Gln
```

with hyphens and sequence deviations are indicated by the substituting nucleotide or amino acid. Arrowheads point towards a potential splicing signal in ZF-EN and the known splice sites in the cDNA sequences of *En-1* and *En-2*.

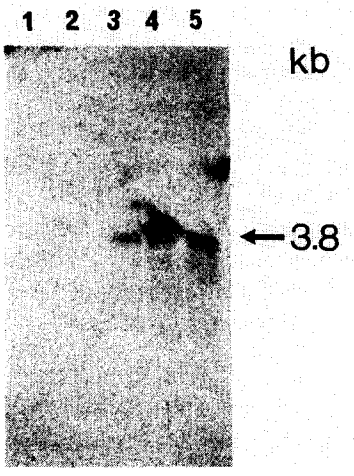

Fig.4. Temporal pattern of ZF-EN expression during development. Lanes 1–5 contain 5 μg poly(A)$^+$ RNA from embryos of stages 2, 7, 16, 29 and 48 h, respectively. The Northern blot was hybridized with a nick-translated probe made from the homeobox-containing *AluI/RsaI* fragment of ZF-EN (fig.2b). A single transcript of 3.8 kb is detected.

possibility that a considerably lower level of ZF-EN expression occurs prior to somite formation in zebrafish embryos.

Zebrafish embryos of all developmental stages can be obtained for further investigations of these aspects by analysing the spatial expression pattern of the ZF-EN gene in tissue sections. Moreover, our finding of a strong sequence conservation between fish and mammalian homeobox-containing genes implies that the results obtained from future zebrafish expression analyses also may provide clues of importance for an understanding of the molecular mechanisms regulating pattern formation in embryos of higher vertebrates.

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] Lewis, E.B. (1978) Nature 276, 565–570.
[6] Nüsslein-Volhard, C. and Wieschaus, E. (1980) Nature 287, 795–801.
[7] Gaunt, S.J., Miller, J.R., Powell, D.J. and Duboule, D. (1986) Nature 324, 662–664.
[8] Utset, M.F., Awgulewitsch, A., Ruddle, F.H. and McGinnis, W. (1987) Science 235, 1379–1382.
[9] Carrasco, A.E. and Malacinski, G.M. (1987) Dev. Biol. 121, 69–81.
[10] Dony, C. and Gruss, P. (1987) EMBO J. 6, 2965–2975.
[11] Hart, C.P., Awgulewitsch, A., Fainsod, A., McGinnis, W. and Ruddle, F.H. (1985) Cell 43, 9–18.
[12] Duboule, D., Baron, A., Mähl, P. and Galliot, B. (1986) EMBO J. 5, 1973–1980.
[13] Martin, G.R., Boncinelli, E., Duboule, D., Gruss, P., Jackson, I., Krumlauf, R., Lonai, P., McGinnis, W., Ruddle, F.H. and Wolgemuth, D. (1987) Nature 325, 21–22.
[14] Poole, S., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) Cell 40, 37–43.
[15] Fjose, A., McGinnis, W. and Gehring, W.J. (1985) Nature 313, 284–289.
[16] Joyner, A.L., Kornberg, T., Coleman, K., Cox, D. and Martin, G.R. (1985) Cell 43, 29–37.
[17] Joyner, A.L. and Martin, G.R. (1987) Genes Dev. 1, 29–38.
[18] Kornberg, T., Siden, I., O'Farrell, P. and Simon, M. (1985) Cell 40, 45–53.
[19] Hisaoka, K.K. and Battle, H.I. (1958) J. Morph. 102, 311–321.
[20] Streisinger, G., Walker, C., Dower, N., Knauber, D. and Singer, F. (1981) 291, 293–296.
[21] Kimmel, C.B. and Warga, R.M. (1986) Science 321, 365–368.
[22] Marcey, D. and Nüsslein-Volhard, C. (1986) Nature 321, 380–381.
[23] Njølstad, P.R., Molven, A. and Fjose, A. (1988) FEBS Lett. 230, 25–30.
[24] Frei, E., Baumgartner, S., Edström, J.E. and Noll, M. (1985) EMBO J. 4, 979–987.
[25] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[26] Eiken, H.G., Njølstad, P.R., Molven, A. and Fjose, A. (1987) Biochem. Biophys. Res. Commun. 149, 1165–1171.
[27] Fjose, A., Molven, A. and Eiken (1988) Gene 62, 141–152.
[28] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[29] Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) in: Basic Methods in Molecular Biology, Elsevier, New York.
[30] Dickerson, R.E. (1971) J. Mol. Evol. 21, 26–45.