FEBS Letters 351 (1994) 380–384

FEBS 14509

Molecular cloning and evolutional analysis of a mammalian homologue of the *Distal-less 3* (*Dlx-3*) homeobox gene

Takuji Shirasawa[a,*], Kyoichi Sakamoto[a], Hiroshi Tkahashi[b]

Departments of [a]Molecular Pathology and [b]Immuno-pthology, Tokyo Metropolitan Institute of Gerontology, Itabashi-ku, Tokyo-173, Japan

Received 13 June 1994

Abstract We have cloned and sequenced a cDNA encoding the mammalian homologue of *Distal-less 3* (*Dlx-3*) from a rat embryonic brain cDNA library. The primary structure of rat *Dlx-3* showed a cDNA with an open reading frame of 290 amino acids with a molecular weight of 31 kDa harboring a homeodomain sequence characteristic of the *Distal-less* homoeobox gene. In addition to the highly conserved homeodomain sequence, we newly found the consensus motifs in the N-terminal and C-terminal region, which were specifically conserved among *Distal-less* members of different species. Phylogenetic analysis of the *Distal-less* homeodomain sequence also showed a possible association between the sequence similarities in the homeodomain and the spatial specifications of homeobox genes expressed in the developing central nervous system.

Key words: Distal-less 3; Homeodomain protein; cDNA cloning; Consensus motif; Molecular evolution

1. Introduction

In many different species homeobox genes have been shown to function as important regulators of developmental processes by giving cells positional and functional identities. Such specific action by homeodomain proteins (homeoproteins) is believed to control the transcription of target genes. Accordingly, homeoproteins can be considered as components of a regulatory network that simultaneously coordinates ongoing processes during the development of an organism. On the basis of their similarities and functions, homeodomain proteins have been grouped into various classes such as *Hox*, *Pax*, *POU*, *En*, *Otx*, *Emx*, and *Dll* proteins [1].

Distal-less (*Dll*) proteins form a unique class of homeoprotein initially identified and characterized by the *Drosophila* mutant termed *Distal-less* (*Dll*), in which limb development is completely impaired [2]. Further genetic analysis showed that the *Dll* gene plays a pivotal role in the development of thoracic legs and the peripheral embryonic sensory organs in *Drosophila* [3,4]. Recently, the homologues of *Dll* were widely identified in rodents, *Xenopus*, newt and zebrafish, showing the ubiquitous role in the development. For instance, rodent *Dll* (*Dlx*) show the restricted pattern of expression in developing forebrain, developing retina, and developing tooth [5–7]. *Xenopus Dll* (*XDll*) is dominantly expressed in the anterior part of the embryo and adult ovary [8]. Newt *Box* genes, the newt homologues of *Dll*, were specifically expressed in the skin of forelimbs, hindlimbs, and tail as well as in brain [9]. In zebrafish, the *Dll* gene is regionally expressed in the developing inner ear in association with other homeobox genes, *msh-C* and *msh-D* [10]. Cohen and Jürgen argued that the *Dll* gene product may give the precursor cells the positional information along the proximal-distal axis [11]. As the homeodomain sequence of *Distal-less* proteins is highly conserved among different species, a similar or relevant function may be present in developing forebrain, developing tooth, developing sensory organs of rodents, developing inner ear of zebrafish, and regenerating tail of newt, in which some specific populations of cells may actually need the positional information along the proximal-distal axis during the development of the organs.

We present the isolation of the mammalian homologue of *XDll-3* and show for the first time that there are consensus motifs in N-terminal and C-terminal portion of *Dll* homeoprotein, which are highly specific for the *Distal-less* class of homeobox. Although the function of the N-terminal or C-terminal region in homeoproteins is still poorly understood, two motifs in the N-terminal region and two motifs in the C-terminal region are found to be conserved among members of the *Dll* homeoprotein and then may confer the specificity of *Dll* homeoprotein in addition to the highly conserved homeodomain sequence. We also present a phylogenetic association between the sequence similarity in the homeodomain and the regional specification of various homeobox genes expressed in the developing central nervous system.

2. Materials and methods

2.1. Polymerase chain reaction

Double strand cDNA was synthesized from poly(A)$^+$ RNA of day 15 p.c. rat embryonic brain using Amersham's cDNA synthesis system (Amersham, UK) as described previously [13]. Twenty ng of ds cDNA was used as template and following oligonucleotide primers were present at a final concentration of 0.25 μM in a standard PCR mixture. Primers used in this study were dlx2–1, 5′-ATG ACT GGA GTC TTT GAC AGT-3′; dlx2–2, 5′-AAC AAT GTC TCC TAC TCC GCC AAA AGC AGC-3′; and dlx2–4, 5′-GAA CTT GGA TCG GCG GTT CTG GAA CCA GAT-3′; corresponding to nt 1–21, 301–329 and 629–601 of mouse *Dlx-2*/*Tes-1* [6], respectively. PCR was carried out using a 0.75 min denaturation at 96°C, 1 min annealing at 55°C and 2 min extension at 72°C, for a total of 40 cycles. The PCR product was electrophoresed in 2% low melting point agarose and the amplified fragments were excised, isolated and subcloned into pBluescript (Stratagene, La Jolla, CA) for sequencing using the dideoxy chain termination method [12].

2.2. Isolation and computer analysis of rat Dlx-3 cDNA clone

To isolate the full length cDNA, rat embryonic brain cDNA library [13] was screened with PCR fragment encoding rat *Dlx-3* as a probe. The nucleotide sequence of rat full-length *Dlx-3* was determined by dideoxy chain termination method [12]. Multiple alignment was performed using the Clastal method contained in Lasergene software package (DNASTAR, London, UK) and a phylogenetic tree was made from the multiple alignment.

*Corresponding author. Fax: (81) (3) 3579 4776.

SSDI 0014-5793(94)00896-5

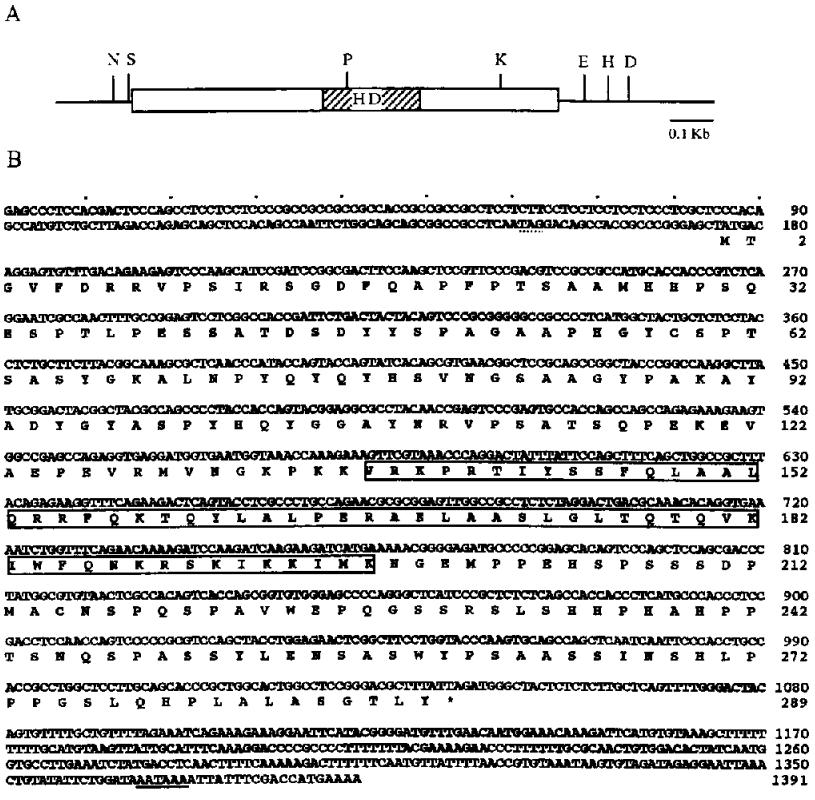

Fig. 1. Primary structure of rat *Dlx-3*. (A) Scheme of the rat *Dlx-3* clone. The box indicates the open reading frame. The shaded box indicates the homeodomain. The relevant restriction enzyme sites are shown in the scheme. N, *Not*I; S, *Sma*I; P, *Pst*I; K, *Kpn*I; E, *Eco*RI; H, *Hin*dIII; D, *Dra*II. (B) Nucleotide and deduced amino acid sequence of rat *Dlx-3* cDNA. Amino acid sequence is shown below the nucleotide sequence. Numbering of the nucleotide begins with the first nucleotide of the cDNA clone. The amino acid sequence encoding the homeodomain is boxed. The polyadenylation signal is underlined. The in-frame stop codon is denoted by a dotted underline. The nucleotide sequence is available from GSDB/DDBJ/EMBL/NCBI databases under the accession number D31743.

3. Results and discussion

3.1. Cloning of rat Distal-less 3 (Dlx-3) gene

To identify the members of the *Distal-less* homeodomain protein (homeoprotein) expressed in rat embryonic brain (*Dlx*), we amplified *Dlx* genes from mRNA of the embryonic brain by PCR reaction with primer dlx2–1 and dlx2–4 or dlx2–2 and dlx2–4 designed based on the published sequence of mouse *Dlx-2/Tes-1* [6]. The nucleotide sequence of 330 bp PCR fragment amplified between primer dlx2–2 and dlx2–4 showed that this PCR fragment encodes the rat homologue of *Dlx-2/Tes-1*. The nucleotide sequence of 530 bp PCR fragment amplified between primer dlx2–1 and dlx2–4, showed 74% identity to mouse *Dlx-2/Tes-1* and 73% identity to mouse *Dlx-1* [5], suggesting that this clone represents a novel member of mammalian *Distal-less*. Then we isolated the full length cDNA encoding this novel member of the *Distal-less* homeobox gene from a rat embryonic brain cDNA library. Fig. 1 shows the 1,391 bp full-length cDNA of a novel *Distal-less* gene. The cDNA encodes a putative open reading frame of 290 amino acids with a calculated molecular weight of 31 kDa harboring a homeodomain sequence (Val^{137}–Lys^{197}). Neighboring sequences around a putative methionine initiation codon at nt 176 are in good agreement with Kozak's consensus sequence [14]. There is an in-frame stop codon at nt 151 in the 5′-untranslated region and a polyadenylation signal at nt 1,367 in the 3′-untranslated region (Fig. 1). The deduced amino acid sequence was searched for the homologous sequences in NBRF database (release 39), revealing that this sequence has 74.1% identity with *Xenopus Dll-3* (L09729, GenBank), 45.8% identity with newt *Box-4* [9], 46.1% identity with zebrafish *Dll-3* [10], 41.8% identity with *Xenopus Dll-4* (L09728, GenBank), 40.7% identity with mouse *Dlx-2/Tes-1* [6,7], 33.6% identity with *Xenopus Dll*[8], and 32.6% identity with newt *Box-5* [9]. To date, the nomenclature among the members of the *Distal-less* homeobox genes in different species has not been unified (Table 1). In order to clarify the interspecies relationship among *Distal-less* homeoproteins of different species we studied the phylogenetic relationship among them (Fig. 2). Newt *Box-4* and zebrafish *Dll-3* may be the homologue of *Xenopus Dll-2* (Fig. 2, Table 1). Mouse *Dlx-2* and the novel member may be the murine homologue of *Xenopus Dll-4* and *Dll-3*, respectively (Fig. 2, Table 1). There-

Table 1
Interspecies relationship among members of *Distal-less* homeobox family

Rodent	*Xenopus*	Newt	Zebrafish
Dlx-1	–	–	–
Dlx-2	*Dll-4*	–	–
*Dlx-3**	*Dll-3*	–	–
–	*Dll*	*Box-4*	*Dlx-3*
–	–	*Box-5*	–

*The novel member reported in this letter is designated as the rat *Dlx-3* gene.

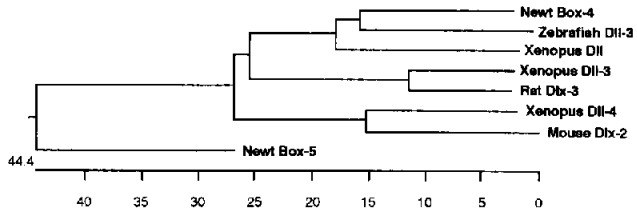

Fig. 2. Phylogenetic tree of *Distal-less* homeobox. The entire sequence of rat *Dlx-3* was compared to those of *Xenopus Dll-3*, newt *Box-4*, zebrafish *Dll-3*, *Xenopus Dll-4*, mouse *Dlx-2*, *Xenopus Dll*, and newt *Box-5* gene. The branch order represents the structural similarity and the branch length represents the sequence divergence. The scale beneath the tree measures the relative distance between sequences.

fore, the novel member was most closely related to *Xenopus Dll-3* and was designated here as rat *Dlx-3* (Fig. 2, Table 1). Since at least 4 distinct members of *Distal-less* are known to be expressed in mouse embryonic brain [7], as yet uncharacterized members of mouse *Distal-less* may correspond to the homologue of *Xenopus Dll-2* or newt *Box-5* (Table 1).

3.2. *Characteristic motifs found in the primary structure of the Distal-less family homeobox genes*

To clarify the characteristic features in the primary structure of *Distal-less* (*Dll*), the published members of *Dll* homeoprotein were aligned with rat *Dlx-3*. As shown in Fig. 3, the homeodomain was highly conserved in all members of *Dll* homeoprotein. In addition, we found for the first time a consensus motif of 18 amino acid residues, SQXSPTLPESXATDSGYY, which was located in the N-terminal region and well conserved among *Dll* members. This consensus motif, designated as consensus motif A, was searched for the homologous sequences in the protein database, revealing that consensus motif A is highly specific for *Distal-less* (*Dll*) homeodomain protein, but not found in homeoproteins of other classes nor in other proteins. Among *Dll* members, newt *Box-4* shows the consensus motif A while *Box-5* fails to show this motif (Fig. 3), although both genes have apparently evolved from a same ancestral gene [9]. It is then interesting to speculate that only newt *Box-4* might retain some specificities shared by other *Dll* members during molecular evolution while *Box-5* might have evolved in a different manner and acquired a different specificity. In the N-terminus, between consensus motif A and the homeodomain, another characteristic feature was found and is shown in Fig. 3 as consensus motif B, in which 8–11 tyrosine residues appear within 50 amino acids. Some of these tyrosine residues locate at the interval of 7–8 amino acid residues or 14–15 amino acid residues, which would fit in 2 or 4 turns of a α-helix, if this domain represents a α-helix structure. Analogous to the leucine zipper structure, the α-helix formation of this domain would display these tyrosine residues at the same surface in the tertial structure, which

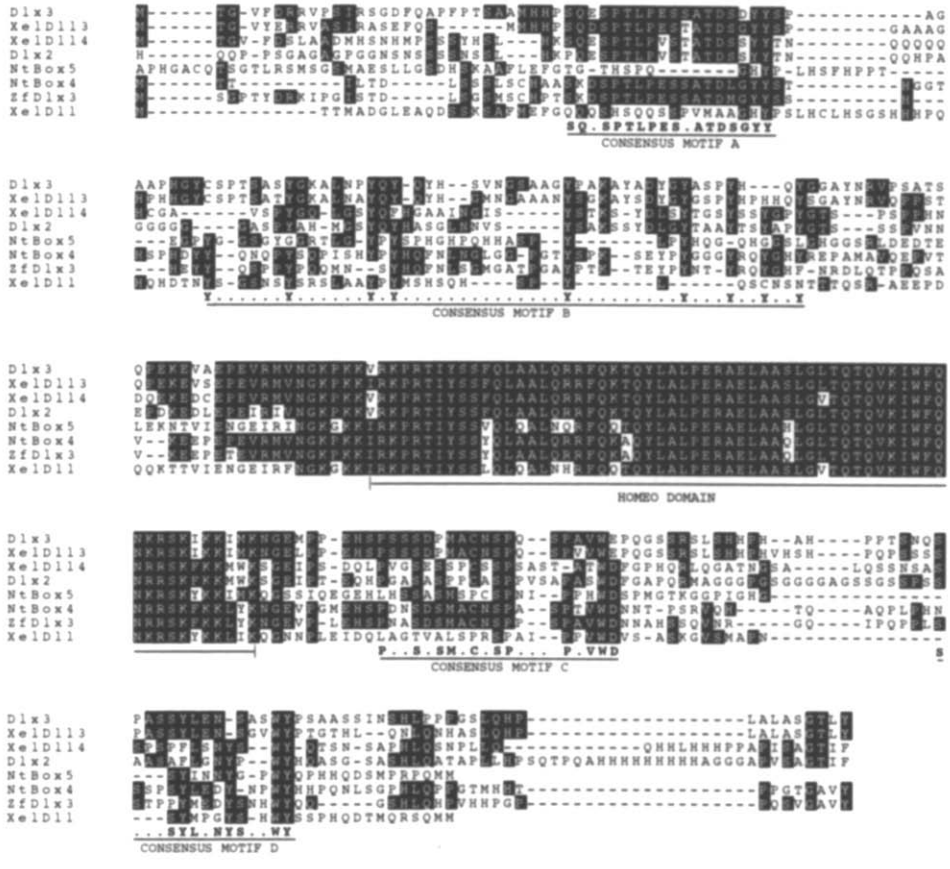

Fig. 3. Amino acid alignment among members of *Distal-less* homeodomain protein. The amino acid sequence (in single-letter code) of rat *Dlx-3* has been aligned with those of *Xenopus Dll-3* (L09729, GenBank), newt *Box-4* [9], zebrafish *Dll-3* [10], *Xenopus Dll-4* (L09728, GenBank), mouse *Dlx-2* [6], *Xenopus Dll* [8], and newt *Box-5* [9] gene using a computer program contained in Lasergene; dashes denote gaps that have been introduced to maximize the alignment. Positions at which at least three of the sequences are identical are shown by black boxes.

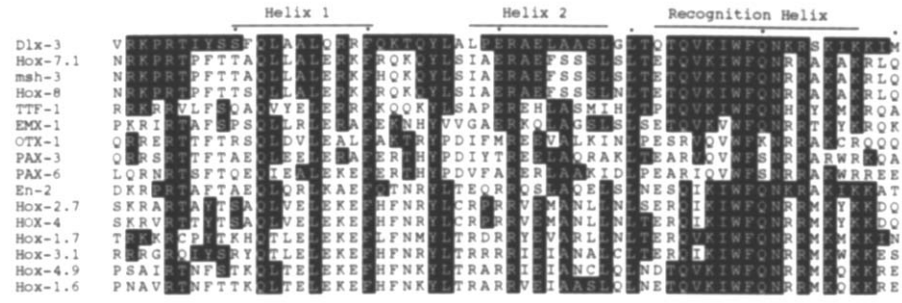

Fig. 4. Amino acid alignment of homeodomain sequence among various homeodomain proteins. The amino acid sequence (in single-letter code) of rat *Dlx-3* has been aligned with those of mouse *Hox-7.1*, *msh-3*, *Hox-8*, *TTF-1*, *Emx-1*, *Otx-1*, *Pax-3*, *Pax-6*, *En-2*, *Hox-2.7*, *Hox-4*, *Hox-1.7*, *Hox-3.1*, *Hox-4.9*, and *Hox-1.6*. Sequences which are identical to that of rat *Dlx-3* are shown by black boxes. The region corresponding helix 1, helix 2, and recognition helix of homeodomain are indicated above the sequence.

may represent the reactive surface for the protein–protein interaction. Although the function of the N-terminal portion of the homeoprotein is still poorly understood, several lines of evidence suggest that the N-terminal domain may modulate DNA–homeodomain interaction [15], interact with effector molecules [16], interact with DNA [17], or interact with other proteins to form the nuclear protein complex [18]. In addition to the consensus motifs in N-terminal region, we also found the consensus motifs in C-terminal region of *Dll* homeoproteins. These consensus motifs, designated as motif C and motif D (Fig. 3), are rich in Ser residues which might be modified by the protein phosphorylation. Taken together, consensus motifs in the N-terminal region of *Distal-less* may be involved in protein–protein interaction for the transcriptional activation and motifs in the C-terminal region may be involved in some modifications of the *Dll* homeoprotein. Further study of the N-terminal and C-terminal domains would clarify the function and role of *Distal-less* homeoprotein as a transcriptional activator.

3.3. Evolutional analysis of Distal-less homeodomain sequence

Homeodomain functions as DNA-binding domain of the transcriptional activator [1]. In general, 60 amino acid homeodomain sequence is highly conserved, irrespective of their classes, and show a helix–turn–helix motif as a common structure which confers the sequence-specific DNA-binding. To examine and compare their DNA-binding specificity in the homeodomain region, the homeodomain sequence of *Dlx-3* was aligned to those of other homeoproteins (Fig. 4). The helix–turn–helix (HTH) motif was deduced from similar sequences present in numerous prokaryotic regulatory proteins as described [19]. As shown in Fig. 4, the recognition helix is the most highly conserved helix structure among various homeoproteins. We found the extensive homology between *Dll* homeoprotein and homeoproteins of other classes. The sequence of Arg[2]-Lys-Pro-Arg-Thr[5] in the homeodomain of *Dlx-3* was shared with other homeoproteins such as *Hox-7.1*, *msh-3*, and *Hox-8*, which are known to be expressed in the developing sensory organs of the head [10,20,28]. Inside the turn between helix 1 and helix 2, on the other hand, *Distal-less* shares Gln[11] with *En-2*, Lys[12] with *Emx-1* and *Otx-1*, and Thr[13] with *Otx-1* and *PAX*, which are known to be expressed in the developing midbrain and forebrain [21,23–25,29]. In helix 2, Leu[35] is highly specific for *Dlx-3*, *TTF-1*, *Emx-1*, *Pax-3*, and *Pax-6*, all of which are expressed in the developing forebrain [21,22,29]. Accordingly, in the aligned sequences a possible association between the spatial specification of gene expression in central nervous system and the sequence similarities in homeodomain would be implied. To further clarify the phylogenetic association between homeodomain sequences and the localization of their gene expression, we made a phylogenetic tree of homeobox genes expressed in developing nervous system (Fig. 5). Homeoproteins are phylogenetically classified into two groups as illustrated in Fig. 5. In group A three homeobox classes, *Hox-7*, *Msh-3*, and *Hox-8*, are expressed in the sensory organs of the head; and the remaining homeobox classes, *TTF-1*, *Dlx*, *Emx*, *Otx*, and *PAX* are expressed in developing forebrain. All these classes are expressed more rostral than homeobox classes in group B; Hox-1, Hox-2, Hox-3, and Hox-4, all of which are mainly expressed in the developing spinal cord with sharp anterior boundary of expression [27]. Interestingly, *Engrailed* homeoprotein, which is the intermediate in the tree, shows the restricted spatial expression in the boundary between developing brain and developing spinal cord, i.e. embryonic midbrain and hindbrain [26]. Taken together, the data suggests that the spatial specification of the homeobox gene may be associated with the sequence similarity in the homeodomain of each homeoprotein.

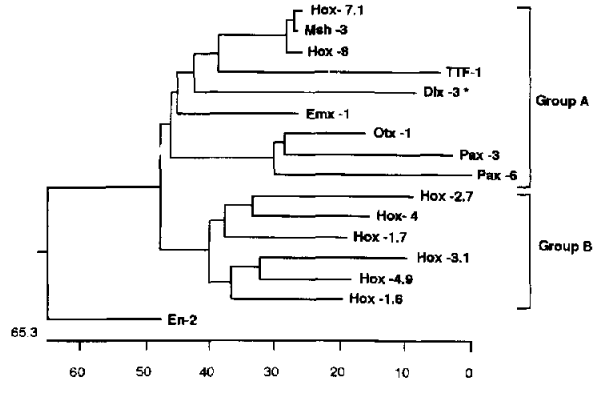

Fig. 5. Phylogenetic tree of homeodomain proteins. The homeodomain sequence of rat *Dlx-3* was compared to those of *Hox-7.1*, *msh-3*, *Hox-8*, *TTF-1*, *Emx-1*, *Otx-1*, *Pax-3*, *Pax-6*, *En-2*, *Hox-2.7*, *Hox-4*, *Hox-1.7*, *Hox-3.1*, *Hox-4.9*, and *Hox-1.6*. The scale beneath the tree measures the relative distance between sequences.

Acknowledgements: We would like to thank Drs. H. Mori, N. Maruyama, T. Takahashi, and A. Graybiel for discussions and encouragement. We would also thank Drs. T. Karasawa, Y. X. Zeng and Ms. K. Miyake for technical assistance.

References

[1] Treisman, J., Harris, E., Wilson, D. and Desplan, C. (1992) BioEssays 14, 145–150.
[2] Cohen, S.M., Bronner, G., Kuttner, F., Jürgens, G. and Jackle, H. (1989) Nature 338, 432–434.
[3] Cohen, S.M. and Jürgens, G. (1989) EMBO J. 8, 2045–2055.
[4] Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M.E., Botas, J. and Cohen, S.M. (1992) Cell 71, 437–450.
[5] Price, M., Lemaistre, M., Pischetola, M., Di Lauro, R. and Duboule, D. (1991) Nature 351, 748–751.
[6] Porteus, M.H., Bulfone, A., Ciaranello, R.D. and Rubenstein, J.L.R. (1991) Neuron 7, 221–229.
[7] Robinson, G.W., Wray, S. and Mahon, K.A. (1991) New Biol. 3, 1183–1194.
[8] Asano, M., Emori, Y., Saigo, K. and Shiokawa, K. (1992) J. Biol. Chem. 267, 5044–7.
[9] Beauchemin, M. and Savard, P. (1992) Dev. Biol. 154, 55–65.
[10] Ekker, M., Akimenko, M.A., Bremiller, R. and Westerfield, M. (1992) Neuron 9, 27–35.
[11] Cohen, S.M. and Jürgens, G. (1990) Nature 346, 482–485.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Shirasawa, T., Akashi, T., Sakamoto, K., Takahashi, H., Maruyama, N. and Hirokawa, K. (1993) Dev. Dynamics 198, 1–13.
[14] Kozak, M. (1986) Cell 44, 283–292.
[15] Brand, A.H. and Perrimon, N. (1993) EMBO J. 12, 4243–4250.
[16] Boulikas, T. (1993) Proc. Natl. Acad. Sci. USA. 90, 4166–4170.
[17] Perez, F., Joliot, A., Bloch, G.E., Zahraoui, A., Triller, A. and Prochiantz, A. (1993) Plant J. 4, 137–150.
[18] Gay, N.J., Poole, S. and Kornberg, T. (1988) EMBO J. 7, 4291–4297.
[19] Laughon, A. and Scott, M.P. (1984) Nature 310, 25–31.
[20] Holland, P.W. (1991) Gene 98, 253–257.
[21] Goulding, M.D., Chalepakis, G., Deutsch, U., Erselius, J.R. and Gruss, P. (1991) EMBO J. 10, 1135–1147.
[22] Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M.G. and Di Lauro, R. (1990) EMBO J. 9, 3631–3639.
[23] Simeone, A., Gulisano, M., Acampora, D., Stornaiuolo, A., Rambaldi, M. and Boncinelli, E. (1992) EMBO J. 11, 2541–2550.
[24] Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A. and Boncinelli, E. (1992) Nature 358, 687–690.
[25] Holland, P.W. and Williams, N.A. (1990) FEBS Lett. 277, 250–252.
[26] Davis, C.A., Holmyard, D.P., Millen, K.J. and Joyner, A.L. (1991) Development 111, 287–198.
[27] Duboule, D. and Dolle, P. (1989) EMBO J. 5, 1497–1505.
[28] Monaghan, A.P., Davidso, D.R., Sime, C., Graham, E., Baldock, R., Bhattacharya, S.S. and Hill, R.E. (1991) Development 112, 1053–1061.
[29] Puschel, A.W., Gruss, P. and Westerfield, M. (1992) Development 114, 643–651.