FEBS Letters 354 (1994) 62–66

FEBS 14704

Identification of two *ets* related genes in a marine worm, the polychaete annelid *Nereis diversicolor*

Anne Lelièvre-Chotteau[a,*], Vincent Laudet[b], Anne Flourens[b], Agnès Bègue[b], Dominique Leprince[b], François Fontaine[a]

[a]*CNRS ER 87, Bât. SN3, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France*
[b]*CNRS UA 1160, Institut Pasteur, 1 rue Calmette, 59019 Lille Cedex, France*

Received 4 July 1994; revised version received 15 September 1994

Abstract The ETS family includes a growing number of transcription factors with a highly conserved DNA-binding domain, the ETS domain. We have used PCR amplification with degenerated oligonucleotides to isolate two putative ETS DNA-binding coding domains in a primitive form of cœlomate, the polychaete annelid *Nereis diversicolor*. These sequences are highly related to the ETS and ERG groups of the *ets* gene family. For the *erg* sequence an adjacent region encoding for 91 amino acids has been characterized after library screening, and we show an expression in cells isolated from the cœlomic cavity of the animal. A phylogenic analysis confirms that the *ets-1/ets-2* and the *erg/fli* dichotomy arose specifically in the vertebrate lineage.

Key words: ets-related gene; ETS phylogeny; Polychaete annelid (*Nereis diversicolor*)

1. Introduction

The founder of the *ets* gene family is the c-*ets-1* proto-oncogene, which was transduced by the E26 avian erythroblastosis virus. E26 carries two distinct oncogenes, v-*myb* and v-*ets* [1,2] and can cause erythroblastosis and myeloblastosis to chickens [3,4]. The ETS family includes genes isolated from species phylogenetically divergent as man, sea urchin and *Drosophila* (reviewed in [5,6]). Recently, partial ETS domains have been detected in lower invertebrate metazoans [7].

All the members of this growing family of transcription factors have a highly conserved region of approximatively 85 amino acids, named the ETS domain, localized in most cases in the carboxy-terminus of the protein. This domain defines a new sequence-specific DNA-binding motif [8,9]. Divergence rate analysis between the different ETS domains revealed that the *ets* gene family can be divided into nine groups, namely the *ETS, ERG, ELG, PEA3, ELK, ELF, D ETS-4, POK* and *SPI* [10,11].

The roles of the ETs related proteins are still not clear. The ETs proteins have been shown to be transcriptional activators, cooperating with other nuclear oncogenes and forming nucleoprotein complexes through protein-protein interactions. These transcription factors bind in a specific manner to different but related purine-rich sequences, including the GGA core [12].

In vertebrate organisms, *ets* genes are differently regulated and expressed in a variety of tissues: some of these play a role in proliferation, differentiation and maturation of hematopoietic lineage cells [13], whereas others are expressed during embryological development and are associated with morphogenesis and tissues modelage [14–16].

The finding that most oncogenes captured by retroviruses are highly conserved at the amino acid level among many organisms has led us to search for oncogene-related sequences in lower organisms. Evolutionary studies provide a means to pinpoint important structural or functional domains conserved throughout many diverse species and to elucidate their biological function. In this study, we present predicted amino acid sequences of two ETS putative DNA-binding domains, related to the ERG and ETS groups, and of an Erg-specific domain (designed 'R' by Lautenberger et al. [10]). So, these genes have been named, respectively, Nd *erg* and Nd *ets*.

Northern blot analysis reveals an expression of Nd *erg* gene in the figurated elements of the cœlomic cavity of the animal.

The phylogenic position of the Erg sequence, comprising the ETS domain and the 'R' region, is analyzed and discussed.

2. Materials and methods

2.1. Polymerase chain reaction and cloning

100 ng of genomic DNA were amplified. The amplification conditions consisted in a 1-min denaturation at 94°C, a 1-min annealing at 52°C and a 2-min extension at 72°C. Each reaction contained 50 pmol of degenerated oligonucleotides previously published [17] and derived from two conserved regions of the ETS domain. Each oligonucleotide contained a *Xba*I site to facilitate cloning of amplified fragments. Thirty cycles were performed using an automated thermal cycler. The products of this reaction were electrophoresed on a 4% NuSieve LMP agarose (FMC). The amplified DNA was purified from the gel and cloned into the plasmid vector pUC 18 for sequencing.

2.2. Library screening and subcloning of the erg gene fragment

After partial digestion by *Sau*3A, *Nereis* genomic DNA was packaged into the λDASHII vector (Gigapack II Plus Packaging Extract, Stratagene). The resulting library ($5 \cdot 10^5$ phages) was screened using the 180 bp *Xba*I–*Xba*I fragment obtained by PCR amplification.

Hybridization was performed at 42°C in a solution of 3 × SSC, phosphate buffer 25 mM, 1 × Denhardt's solution, 50% formamide and 100 $mg \cdot ml^{-1}$ denatured salmon sperm DNA. Filters were washed at 50°C twice with 2 × SSC, 0.1% SDS and twice with 0.1 × SSC, 0.1% SDS.

From 750,000 phages, seven independant overlapping clones were selected by their hybridization to the probe. From one of them, a 0.8 kb *Hae*III–*Hae*III fragment was isolated, subcloned into the pUC 18 vector and then sequenced with universal primers or with the specific primers GGTGATGATGTGCGAGTT and CTGGTCGACCCGGATGAA, by the dideoxyribonucleotide chain termination method [18] and the T7 sequencing kit (USBC). The entire nucleotide sequence was determined on both strands with overlapping regions.

2.3. RNA isolation and RT PCR analysis

The polychaete annelid *Nereis diversicolor* has a good ability to regenerate its caudal extremity. From regenerating tissues, poly(A^+)

*Corresponding author. Fax: (33) 20 43 68 49.

SSDI 0014-5793(94)01091-9

RNA were purified with the QuickPrep mRNA purification kit (Pharmacia). Single-strand cDNA was prepared using an oligo(dT) primer and M-MLV reverse transcriptase (Clontech). An aliquot of the mixture was amplified as described above.

2.4. Northern blot analysis

Polyadenylated RNAs were also isolated from cœlomic cells collected from the cœlomic cavity of about 200 worms. Ten μg of mRNA were fractionated by electrophoresis through a denaturating 1.2% agarose gel containing 0.66 M formaldehyde and transferred to a nylon Hybond N^+ (Amersham) membrane. The filter was hybridized with the 180 bp *XbaI–XbaI*. Nd *erg* fragment at 42°C in a solution containing 50% formamide, 10% dextran sulfate, 5 × SSC, 50 mM Na_2PO_4, 5 × Denhardt's solution and 250 $\mu g \cdot ml^{-1}$ of denaturated salmon sperm DNA. The filter was washed sequentially in 2 × SSC, 0.1% SDS at 55°C for 1 h, in 0.5 × SSC, 0.1% SDS at 55°C for 30 min and in 0.1 × SSC, 0.1% SDS at 65°C for 30 min.

2.5. Phylogenic analysis of Nd Erg

Deduced amino acid sequences were treated with the Neighbor-Joining method [19] and the Parsimony analysis [20]. The solidity of each branch was calculated by the bootstrap procedure [21].

3. Results and discussion

ETS family members are defined by the presence of a highly conserved region termed the ETS domain. In the present study by PCR amplification with degenerate primers (and library screening for the Nd Erg sequence), we have identified two sequences encoding ETS domains in the marine worm *Nereis diversicolor* (Figs. 1A and 2).

Alignment of the deduced amino acid sequence named Nd Erg concerns two regions, respectively designed as the ETS

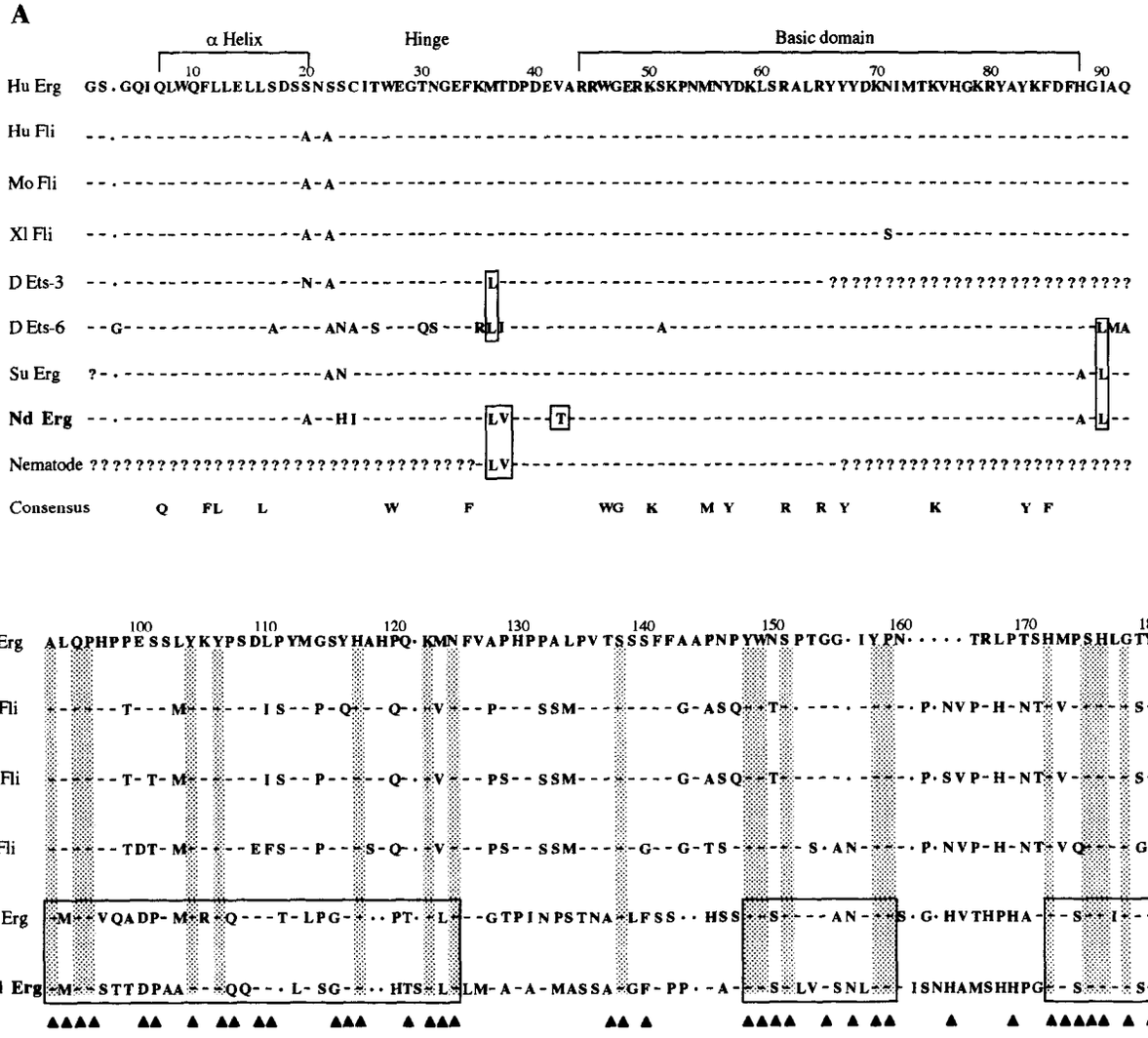

Fig. 1. Alignment of ETS domains (A) and of 'R' region (B) of the ERG group proteins actually known: Hu Erg [22], Hu Fli [23,24], Mo Fli [25], Xl Fli [26], D Ets-3 and D Ets-6 [17], Su Erg [10] and in a nematode [7], and comparison with the Nd Erg deduced amino acid sequence. The species abbreviations are: Hu, human; Mo, mouse; Xl, *Xenopus laevis*; D, *Drosophila*; Su, sea urchin; Nd, *Nereis diversicolor*. Dashes indicate identity with the human Erg sequence. For the ETS domain (A), boxes indicate positions where Nd Erg sequence is identical with the sequences isolated from invertebrate organisms. A consensus sequence of invariant and conserved amino acids for all the ETS domains is listed at the bottom. For the 'R' region (B), the residues conserved between Su Erg and Nd Erg are marked with a black triangle and the regions showing the more important homology are boxed. Amino acids that match Erg proteins are shaded.

domain and the 'R' region, whereas the deduced amino acid sequence of Nd Ets represents only a part of the ETS domain.

3.1. The ETS domain of Nd Erg and Nd Ets

Comparison of ETS domain sequences of the ERG group (Fig. 1A) shows that the Nd Erg sequence most closely resembles the Hu Erg [22], Hu Fli [23,24], Mo Fli [25], Xl Fli [26], D Ets-3 and D Ets-6 [17] or Su Erg [10] proteins. So, the ETS domain of Nd Erg displays the typical features of the ETS domain of proteins related to the ERG group, according to Laudet et al. [11]. It contains conserved amino acid residues found in all the ETS family members and particularly conserved amino acid residues which define the ERG group. The high degree of conservation suggests that this region could bind DNA, as the ETS domains of other Erg proteins, but the nucleotidic sequence specifically recognized by Nd Erg has to be precised. The homology between Nd Erg and the ERG group members ranges from 89% to 91%. The most important homology is found with the sea urchin Su Erg. In other Invertebrate organisms, sequences belonging to the ERG group have been described only in *Drosophila* (D Ets-3 and D Ets-6) and in a nematode worm. For the nematode, Degnan et al. [7] have published a 33 amino acid sequence similar to Nd Erg and with the same two specific amino acids (leucine and valine in positions 36 and 37). We can also notice that the leucine residue in position 36 is conserved in the protostome lineage (in the nematode, in *Nereis* and *Drosophila*) and that the leucine (position 90) is conserved in all Erg proteins from invertebrate organisms known so far. The valine (position 46) found among all Erg proteins and in nearly all the Ets proteins is here substituted by a threonine: the biological significance of this specific mutation in *Nereis* remains to be studied.

A fragment named Nd Ets (Fig. 2) obtained by RT PCR from regenerates of the caudal extremity shows a deduced amino acid sequence highly related to a part of the ETS domain (amino acids 12–52) of the ETS group proteins (the homology is around 85%). This sequence exhibits the isoleucine to lysine substitution (position 48) found in D Ets-2. This isoleucine was also observed in Ets sequences of a mollusc and a cnidarian previously described by Degnan et al. [7]. Instead of a serine, we observed at position 23 an histidine in Nd Ets as in Su Ets-2.

3.2. The 'R' region of Nd Erg

In the ERG group, the region following the ETS domain of the products of the human being, mouse, *Xenopus* and sea urchin *erg* genes (called 'R' by Lautenberger et al. [10]) is much conserved. In the deduced amino acid sequence of the product of the *erg* gene of *N. diversicolor* we observe the same conserved 'R' region (Fig. 1B). This region seems to be responsible for

```
                    20        30        40        50
Hu Ets-1    LLELLTDKSCQSFISWTGDGWEFKLSDPDEVARRWGKRK
Hu Ets-2    -----S----------------A----------
D Ets-2/Pointed -----L--T----------T----------I--
Su Ets-2    --------T--HI----------
Nd Ets      --------N--H----S--------A----------I--
```

Fig. 2. Alignment of ETS domains of the ETS group proteins. Abbreviations are as in Fig. 1. Dashes indicate identity with Hu Ets-1 sequence.

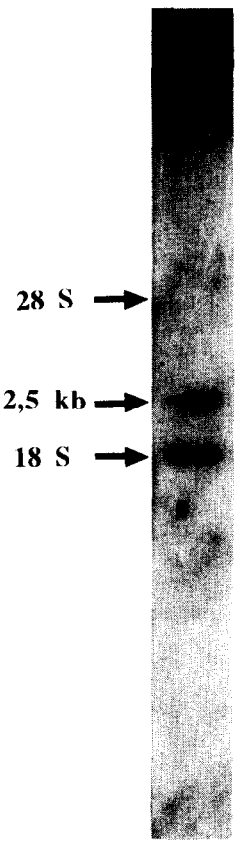

Fig. 3. Ten μg of polyadenylated RNA from cœlomic cells were analysed through denaturating gel electrophoresis, transferred to a nylon membrane and hybridized to a ^{32}P-labeled 180 bp *XbaI–XbaI* Nd *erg* probe. The size of the Nd *erg* mRNA and positions of RNA 28S and 18S are indicated on the left.

specific activity of the Erg proteins. In *N. diversicolor* the ETS domain and the 'R' region are coded by the same exon as in human Erg proteins [27].

We observed an even more important homology between 'R' regions of Erg proteins from *Nereis* and sea urchin (Su Erg), which are the only ones known so far in invertebrate organisms, than with vertebrate 'R' regions. Between the Nd Erg and Su Erg 'R' region the percentage of homology observed is 47%, whereas the homologies between the invertebrate 'R' region compared to vertebrate 'R' region range from 37.5% to 39.5%. We observed between the Nd Erg and Su Erg 'R' region two domains of high homology located at the N-terminal and at the C-terminal extremity, respectively. This carboxy-terminal domain is referred to as the CTA (carboxy-terminal transcriptional activation) domain [28,29]: it is well conserved between mammalian Erg proteins and corresponds to a transcriptional regulatory domain. Amino acid composition analysis of this CTA domain in vertebrates shows that it is particularly rich in proline and serine residues. In invertebrates we observe the same proline-serine rich region along the 'R' domain (17 to 19% of serine residues and about 10% of proline residues). Secondary structure analysis of the amino acid sequence of the CTA domain of the Erg and Fli proteins predicts a turn–loop–turn structure [28,29] which could be involved in transcriptional activation by molecular interactions with other proteins. In the amino-terminal region, the alignment of deduced amino acid

sequences shows residues that are more particularly matched across Su Erg and Nd Erg: it can be supposed that this ATR (amino-terminal 'R') region refers to a new so far undescribed functional domain. Thus the weaker homology between the vertebrate and invertebrate 'R' region could be in relation with a specific function of this region in these two zoological groups.

3.3. Expression of Nd erg

After hybridization with a fragment overlapping a part of the ETS domain (corresponding to amino acids 12 to 52), we detected a 2.5 kb mRNA in cells isolated from cœlomic cavity (Fig. 3). These cells include exclusively gametes and nurse cells, and granulocytes. Ultrastructural and immunohistochemical methods have led to the recognition of different types of granulocytes constituting a primitive form of immune system, some of them being implicated in a cooperative cellular phenomenon (i.e. cicatrization, encapsulation and non-self recognition) [30,31].

In vertebrates, some ETS family genes have been shown to be expressed during oogenesis or in hematopoietic cells [13,32]. The putative role of the product(s) of Nd *erg* in oogenesis or cellular immunity mediated through cœlomic cells is to be explored.

3.4. Phylogenic analysis of the Nd Erg sequence

The *ets* gene family is very diversified, constituted by numerous members identified in metazoan diploblast and triploblast organisms (Fig. 4). Degnan et al. [7] have shown deduced amino acid sequences of the ETS group in diploblast organisms but the more ancient member of the ERG group actually known has been found in a nematode worm, an acœlomat triploblast.

Due to its small size, the Nd Ets sequence was not studied by molecular phylogeny. But an observation of the sequence reveals that Nd Ets is very close to the *Drosophila ets* gene. This observation is fully consistent with the origin of Nd Ets since annelids are phylogenically situated close to arthropods [33].

The phylogenic position of the Nd Erg sequence (including ETS and 'R' domains) was established using distance matrix or parsimony analysis. Both types of analyse give the same relationships inside the ERG group with high bootstrap values which suggests that the tree is congruent. Due to the fact that the 'R' domain (and for D Ets-3 the C-terminal part of the ETS domain) are not known, we excluded D Ets-3 and D Ets-6 from our analysis. In fact a tree constructed with only the ETS domains and containing these sequences gives the same result but with low bootstrap values (data not shown). In the tree shown in Fig. 5, Nd Erg appears to be, as the Su Erg, closely related to the ancestral non duplicated *erg/fli* gene. Indeed,

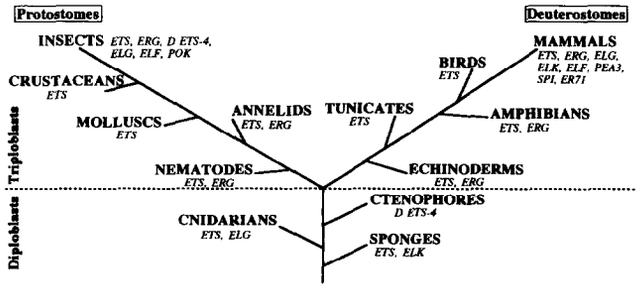

Fig. 4. Schematic genealogical tree showing the different ETS family groups actually known in metazoan organisms.

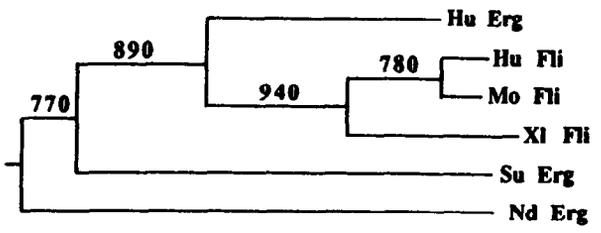

Fig. 5. Phylogenic tree connecting all members of the ERG group based on sequences of the ETS domain and 'R' region. The solidity of each branch of the tree was tested by 1000 bootstrap replicates.

phylogenical tree reconstruction [10,11] together with chromosomal mapping suggest that early in evolution an *ets/erg* gene was duplicated to give one ancestral *ets* and one ancestral *erg* gene, these two genes being linked on the same chromosomal locus. Later on, and specifically on the vertebrate lineage, this gene complex was duplicated to give *ets-1* and *ets-2* and *erg* and *fli* [11]. The phylogenic position of Nd Erg reinforces this model since Nd Erg appears as an 'uncommitted' *erg/fli* gene. It is highly probable that a longer Nd Ets sequence will also behave as an 'uncommitted' *ets-1/ets-2* gene as the *Drosophila* and sea urchin *ets* genes. In this model the moment of the first duplication of the *ets/erg* ancestor is not known but it is likely that this duplication is very old since partial *ets* sequences have been identified in a variety of early metazoans such as sponges, cnidarians and ctenophores, confirming the existence of the *ets* gene family prior to the dichotomy of diploblasts and triploblasts. Whether these *ets* sequences represent real *ets* genes or *ets/erg* ancestor related sequences is difficult to say in view of the small sequences available. Cloning and sequencing of *erg* genes in these organisms as well as other ETS family members in *N. diversicolor* will in the future shed light on the evolution and diversification of this family.

3.5. Conclusion

The function of *ets* genes in development is still the object of numerous investigations. The existence of *ets* genes has so far not been detected in protozoa, fungi and plants [7], but they are diversified in diploblasts. This multigene family may be essential for the establishment of the cellular environment and to allow cells to coordinate the correct patterns of gene expression during ontogenesis. Among products of proto-oncogenes a large number of conserved domains have been shown to play essential roles in intracellular signalling or regulation of transcription during embryogenesis and cell differentiation. Identification of these functional domains in more simple organisms than higher metazoans by phylogenic analysis similar to the one described in this paper will provide an insight towards the understanding of the functional properties of the Erg and Ets proteins.

Acknowledgements: We are very grateful to Prof. André Dhainaut, Prof. Maurice Porchet and Prof. Dominique Stéhelin for their interest in this project. We also thank Françoise Bonet for patient typing and Christian Wattez for critical reading of the manuscript.

References

[1] Leprince, D., Gégonne A., Coll, J., de Taisne, C., Schneeberger, A., Lagrou, C. and Stéhelin, D. (1983) Nature 306, 395–397.

[2] Nunn, M.F., Seeburg, P.H., Moscovici C. and Duesberg, P.H. (1983) Nature 306, 391–395.
[3] Golay, J., Introna, M. and Graf, T. (1988) Cell 55, 1147–1158.
[4] Graf, T., McNagny, K., Brady, G. and Frampton, J. (1992) Cell 70, 201–213.
[5] Mac Leod, K., Leprince, D. and Stéhelin, D. (1992) Trends Biol. Sci. 17, 251–256.
[6] Wasylyk, B., Hahn, S.L. and Giovane, A. (1993) Eur. J. Biochem. 211, 7–18.
[7] Degnan, B.M., Degnan, S.M., Naganuma, T. and Morse, D.E. (1993) Nucleic Acids Res. 21, 3479–3484.
[8] Karim, F.D., Urness, L.D., Thummel, C.S., Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C., Maki, R.A., Gunther, C.V., Nye, J.A. and Graves, B.J. (1990) Genes Dev. 4, 1451–1453.
[9] Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D. and Graves, B.J. (1992) Genes Dev. 6, 975–990.
[10] Lautenberger, J.A., Burdett, L.A., Gunnell, M.A., Qi, S., Watson, D.K., O'Brien, S.J. and Papas, T.S. (1992) Oncogene 7, 1713–1719.
[11] Laudet, V., Niel, C., Duterque-Coquillaud, M., Leprince, D. and Stéhelin, D. (1993) Biochem. Biophys. Res. Commun. 190, 8–14.
[12] Janknecht, R. and Nordheim, A. (1993) Biochim. Biophys. Acta 1155, 346–356.
[13] Seth, A., Ascione, R., Fisher, R.J., Mavrothalassitis, G.J., Bhat, N.K. and Papas, T.S. (1992) Cell Growth Differ. 3, 327–334.
[14] Vandenbunder, B., Pardanaud, L., Jaffredo, T., Mirabel, M.A. and Stehelin, D. (1989) Development 106, 265–274.
[15] Desbiens, X., Quéva, C., Jaffredo, T., Stéhelin, D. and Vandenbunder, B. (1991) Development 111, 699–713.
[16] Quéva, C., Leprince, D., Stéhelin, D. and Vandenbender, B. (1993) Oncogene 8, 2511– 2520.
[17] Chen, T., Bunting, M., Karim, F.D. and Thummel, C.S. (1992) Dev. Biol. 151, 176–191.
[18] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5464.
[19] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[20] Swofford, D.L., (1992) Illinois Natural History Survey, Champaign.
[21] Felsenstein, J. (1988) Annu. Rev. Genet. 22, 521–565.
[22] Rao, V.N., Papas, T.S. and Reddy, E.S.P. (1987) Science 237, 635–639.
[23] Prasad, D.D.K., Rao, V.N. and Reddy, E.S.P. (1992) Cancer Res. 52, 5833–5837.
[24] Watson, D.K., Smyth, F.E., Thompson D.M., Cheng, J.Q., Testa, J.R., Papas, T.S. and Seth, A. (1992) Cell Growth Diff. 3, 705–713.
[25] Ben-David, Y., Giddens, E.B., Letwin K. and Bernstein, A. (1991) Genes Dev. 5, 908–918.
[26] Meyer, D., Wolff, C.M., Stiegler, P., Senan, F., Befort, N., Befort, J.J. and Remy, P. (1993) Biol. Dev. 44, 109–121.
[27] Sorensen, P.H.B., Lessnick, S.L., Lopez-Terrada, D.L., Liu, X.F., Triche, T.J. and Denny, C.T. (1994) Nature Genetics 6, 146–151.
[28] Rao, V.N., Ohno, T., Prasad, D.D.K., Bhattacharya, G. and Reddy, E.S.P. (1993) Oncogene 8, 2167–2173.
[29] Siddique, H.R., Rao, V.N., Lee, L. and Reddy, E.S.P. (1993) Oncogene 8, 1751–1755.
[30] Porchet-Henneré, E., M'Beri, M., Dhainaut, A. and Porchet, M. (1987) Cell Tissue Res. 248, 463–471.
[31] Porchet-Henneré, E. (1990) J. Invert. Pathol. 56, 353–361.
[32] Murakami, K., Mavrathalassitis, G., Bhat, N.K., Fisher, R.J. and Papas, T.S. (1993) Oncogene 8, 1559–1566.
[33] Adoutte, A. and Philippe, H. (1993) Comp. Mol. Neurobiol. 2, 75–104.