ELSEVIER

Biochimica et Biophysica Acta 1494 (2000) 144–148

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.elsevier.com/locate/bba

Promoter paper

Genomic organization and the promoter region of the round-spotted pufferfish (*Tetraodon fluviatilis*) *CDC*37 gene[1]

Shu-Chuan Tsai [a], Jiann-Horng Leu [b], Chih-Ming Chou [b], Maw-Sheng Yeh [b], Fore-Lien Huang [a], Chang-Jen Huang [b,*]

[a] *Department of Zoology, National Taiwan University, Taipei, Taiwan*
[b] *Institute of Biological Chemistry, Academia Sinica, Nankang Borough, Taipei 115, Taiwan*

Received 3 April 2000; accepted 15 May 2000

Abstract

The *CDC*37 gene was isolated from a round-spotted pufferfish genomic library and characterized. This gene is composed of nine exons spanning 3.5 kb. Exon 1 contains the 5′-untranslated region and exon 2 contains the putative translation initiation site. By 5′-RACE (rapid amplication of cDNA ends) and sequence analysis, we deduced the promoter region for the *CDC*37 gene and found that it does not contain typical TATA or CCAAT box. The 1.8 kb DNA fragment upstream of the putative transcription initiation site contains numerous potential binding sites for transcription factors including CREB, E2A, Ets-1, GATA, NF-IL6 and PEA3. When this DNA fragment was placed upstream of the chloramphenicol acetyltransferase (CAT) reporter gene and transfected into a carp CF cell line, it could drive the synthesis of CAT enzyme four times more efficiently than the promoterless pCAT-Basic did. In addition, the *CDC*37 gene is linked to the *TYK*2 gene in a tail-to-head manner with a small intergenic region of 292 bp.

Keywords: *CDC*37; Genomic structure; Promoter; Round-spotted pufferfish (*Tetraodon fluviatilis*)

*CDC*37 was first identified as an important gene required for the cell division cycle in budding yeast [1]. Using the yeast two-hybrid system, Cdc37 has been shown to specifically interact with Cdk4 and Cdk6 [2]. Subsequently, Cdc37 and the associated heat-shock protein Hsp90 have been found to act as protein kinase chaperones to maintain the activity of diverse protein kinases, such as Cdk4, v-Src and Raf-1 [3–5]. The formation of a ternary Raf-1–p50^{cdc37}–Hsp90 complex is crucial for Raf-1 function [6,7]. Moreover, it is interesting to notice that the chick Cdc37 contains binding motifs for the polysaccharide, hyaluronan and other related glycosaminoglycans [8].

We have previously used the round-spotted pufferfish as a model organism, which has a compact genome size (380 Mb) and small introns [9,10], for the complete genomic sequence analysis of all four *JAK* kinase genes [11,12]. In the course of subcloning and partial sequencing of regions beyond the *TYK*2 gene, we found some DNA sequences highly homologous to those of human *CDC*37 gene [13]. We therefore proceeded to determine the sequences of the 5 kb DNA fragment upstream of the *TYK*2 gene and the results confirmed the presence of the pufferfish *CDC*37 gene. We reported here the genome organization and characterization of the promoter region of the pufferfish *CDC*37 gene, which is arranged in a tail-to-head fashion with the *TYK*2 gene with a small intergenic region of 292 bp.

Using lambda FIX II as a cloning vector (Stratagene, La Jolla, CA, USA), a round-spotted pufferfish (*Tetraodon fluviatilis)* liver genomic DNA library has been constructed and amplified as previously reported [14]. DNA fragments containing the pufferfish *TYK*2 sequences including one intron [12,15], were used as probes to screen this library. The probe was labeled using a DIG DNA Labeling Kit (Boehringer Mannheim, Mannheim, Germany). Hybridization and washing were carried out as previously described [14,16]. As shown in Fig. 1, the restriction map of one representative clone was constructed by digesting the phage DNA with a panel of restriction enzymes separately or in various combinations. A total of 15 kb of this clone was completely sequenced by conventional subcloning strategy combined with automated sequencing. The resulting sequences indicated that there

* Corresponding author. Fax: +886-2-2788-9759;
E-mail: cjibc@gate.sinica.edu.tw

[1] The sequence data in this paper has been submitted to the EMBL/GenBank Data Libraries under the accession number AF091237.

0167-4781/00/$ – see front matter
PII: S0167-4781(00)00138-X

Table 1
Exon–intron organization of the puffer fish *CDC37* gene

exon number	exon size (bp)	3'end of the exon	5'end of the intron	intron size (bp)	3'end of the intron	5'end of the exon	amino acid interrupted
1	164	AAAATCACAAAAG	gtttgggtca	73	cttcctgcag	GCATC ATG TCC	
2	113	TGG AGA CAT CAG	gtaatacacg	243	tgtatgtcag	GCA CGG GTC	Gln36 (3)
3	282	GGT TTC AGT AAG	gtgaaagaaa	273	tcatccttag	AGC ATT GTT	Lys130(3)
4	106	AAA CAT TTT G	gtggggcaca	94	ttttgtgtag	GC ATG TTG CAT	Gly166(1)
5	116	GAA GTC GAG GAG	gtaaaatcta	122	ttgtttttag	AAA CAT GCC	Glu204(3)
6	123	GAA AAG ATT AAG	gtaacttcct	191	tgtttaccag	ACA GCG GAT	Lys245(3)
7	183	TCC CTG CCA CCT	gtaagttgac	87	tgaattgcag	GAG ATG CAG	Pro306(3)
8	72	ATG GAT CCA ACT	gtaagttcat	407	cgtttgttag	GAG GCA AAG	Thr330(3)
9	525	GTTCAAATAAAAATGTGCTACATTCAAAAAAAAAAAAAAAA (the end of puffer fish *CDC37* gene)					

were two genes contained in this clone. Two-third sequences of this clone have been confirmed to be homologous to those of human *TYK*2 gene [17]. The remaining sequences upstream of the *TYK*2 gene were found to be homologous to those of human *CDC*37 gene [13]. These two genes are arranged in a tail-to-head fashion. The end of the 3′-untranslated region (3′-UTR) of *CDC*37 and the start of the 5′-UTR of the *TYK*2 gene are separated by 292 bp as reported previously [12].

To characterize the 5′- and 3′- end of the round-spotted pufferfish *CDC*37 mRNA, 5′- and 3′-RACE assays were performed (data not shown). The coding region of the *CDC*37 gene spans 3.5 kb and consists of nine exons and eight introns. The sequences around the exon/intron boundaries were determined and are shown in Table 1. All exon/intron boundaries identified conformed to the GT/AG splice donor/acceptor rule [18]. All exons of the coding region are relatively small ranging from 72 to 282 bp. The sizes of introns are also small, ranging from 73 bp (intron 1) to 407 bp (intron 8), with an average size of 186 bp.

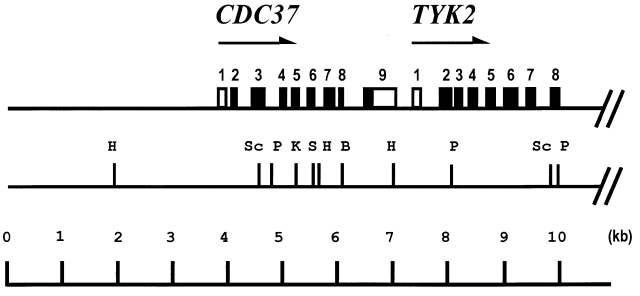

Fig. 1. Organization and physical map of the round-spotted pufferfish *CDC*37 gene. This gene is orientated in a tail-to-head manner with *TYK*2 kinase gene which is partially shown. The structure of the pufferfish *CDC*37 gene is represented with regard to the organization of the exons and introns. Exons are indicated to scale by boxes numbered 1 to 9. Solid boxes indicate the pufferfish *CDC*37 coding region whereas open boxes represent the 5′- and 3′-untranslated region. Introns and the 5′- and 3′-flanking regions are indicated by the solid lines. A restriction map is indicated to contain several cleavage sites. Restriction endonuclease sites are B, *Bam*HI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sal*I; Sc, *Sac*I. The representative phage clone, PFJ1, was isolated from a Lambda FIX II round-spotted pufferfish genomic library.

The open reading frame encoded by the eight exons of the pufferfish *CDC*37 gene produces a protein of 375 amino acids with a molecular mass of 42 375 Da. As shown in Fig. 2, the overall amino acid sequence identity between the pufferfish and mouse, human and chicken Cdc37 protein is 73, 74 and 68%, respectively. The comparison also reveals the presence of two hyaluronan-binding motifs [19] (amino acid residues 158–168 and 263–273 in Fig. 2) and the pRB-binding motif (LVCEE) [20] that are conserved among vertebrate species. Other motifs, such as one putative tyrosine phosphorylation site (WDDSQK**Y**) and a potential serine phosphorylation site (IEV**S**DD) by casein kinase II are also found in pufferfish Cdc37 protein. The change of this residue from serine to alanine reduces Cdc37 activity as reported previously [21].

In order to identify sequence elements involved in the control of the pufferfish *CDC*37 gene expression, a DNA fragment corresponding to the 5′-upstream region of the *CDC*37 gene was sequenced. Analysis of these sequences revealed numerous putative binding sites for transcription factors. As shown in Fig. 3, no TATA box, which is generally located at a position about 30 bp upstream of the RNA start site [22], was identified. Four putative binding sites for a ubiquitously expressed transcription factor, NF-IL6 [23] was observed at −36, −127, −1082 and −1450. Two proto-oncogene product E2A [24] and Ets-1 [25] binding sites were found at −1774 and −1418, respectively. In addition, the flanking region also contained DNA motifs for CREB [26] at −28; GATA [27,28] at −306, −753 and −1110; and PEA3 [29] at −19, −748, −1079, −1203 and −1448.

To verify whether the 5′-flanking region of the round-spotted pufferfish *CDC*37 gene exhibits functional promoter activity, the genomic DNA fragment containing the 5′-upstream region (−1836 to +164) was fused to the CAT reporter gene in pCAT-Basic (Promega) to create pCDC37-CAT. Following transfection into carp CF cells [30], the CAT activity of pCDC37-CAT was about 4 and

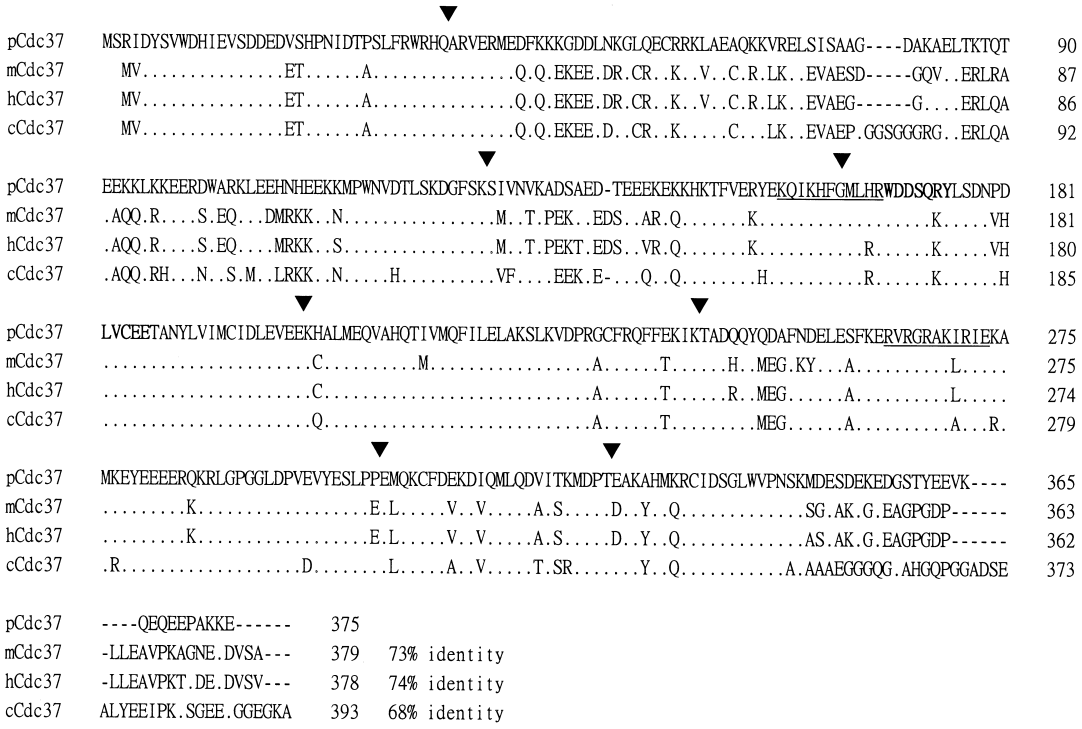

Fig. 2. Amino acid sequence comparison of the pufferfish, human, mouse and chick Cdc37 proteins. The dashes indicate amino acids identical to the pufferfish Cdc37 whereas dots indicate gaps introduced to generate an optimal alignment. The sequences of Cdc37 proteins from human [13], mouse [13] and chick [8,19] are retrieved from the databases. The identity of other Cdc37 proteins to the pufferfish Cdc37 is indicated at the end of each sequence. The putative tyrosine phosphorylation site and pRB-binding motif are in bold whereas the two hyaluronan-binding motifs are underlined. Spliced sites are indicated by down-pointing arrowheads. It is obvious that all spliced sites in the coding region are identical.

0.05 times the promoter activity of pCAT-Basic and pRSV-CAT [31], respectively (Fig. 4). Therefore, the promoter region of the round-spotted pufferfish *CDC*37 gene is functional.

In this study, we have cloned and characterized the *CDC*37 gene from the round-spotted pufferfish *T. fluviatilis*, whose genome size is 3.2 and 8 times smaller than that of chick and human [32,10], respectively. A different species of the pufferfish *Fugu rubripes* (*Fugu*) has been used as a model for vertebrate genome analysis [9]. *Fugu* belongs to the order Tetraodontiformes and is widely consumed in Japan. Its genome is estimated to be approximately 404 Mb, 7.5 times smaller than that of human. We [11,12,33,34] and others [35] have provided several lines of evidence that the round-spotted pufferfish *Tetraodon fluviatilis*, can also be used as another model organism for comparative genomic structure analysis, in addition to *Fugu*.

Table 2
Comparison of intron size (bp) of pufferfish and chick *CDC*37 genes

Intron	Pufferfish	Chick
1	73	0
2	243	1500
3	273	400
4	94	2000
5	122	1200
6	191	110
7	87	140
8	407	2000

The exon–intron organization of the pufferfish and chick *CDC*37 genes are compared (Fig. 2). The overall amino acid sequence identity between the pufferfish and chick Cdc37 is 75%. In addition, all of the splice sites of both genes are identical except for intron 1, which is only present in the pufferfish *CDC*37 gene. Comparison of the intron sizes of both genes is shown in Table 2. There are four introns larger than 1 kb for chick *CDC*37 gene, while all introns of the pufferfish *CDC*37 gene are less than 500 bp. Therefore, both pufferfish and chick *CDC*37 genes are highly conserved not only at the amino acid sequence level but also in their genomic structure.

Although pufferfish and chick *CDC*37 genes do not have TATA or CCAAT boxes at appropriate positions, the regulatory elements in the promoter region of both genes are different. In chick, there are several potential transcription factor binding sites, such as Sp1, E2A and MZF1 [19]. On the contrary, the putative promoter region of the pufferfish *CDC*37 gene not only has E2A, but also has other potential binding sites for transcription factors including NF-IL6, Ets-1, CREB, GATA and PEA3. In this study, the promoter activity of the 5′-flanking 1.8 kb DNA fragment of the pufferfish *CDC*37 gene assayed by CAT activity has been tested to be low but significant for transfection into a carp cell line (Fig. 4). The relevance of CAT activity to the presence of potential regulatory ele-

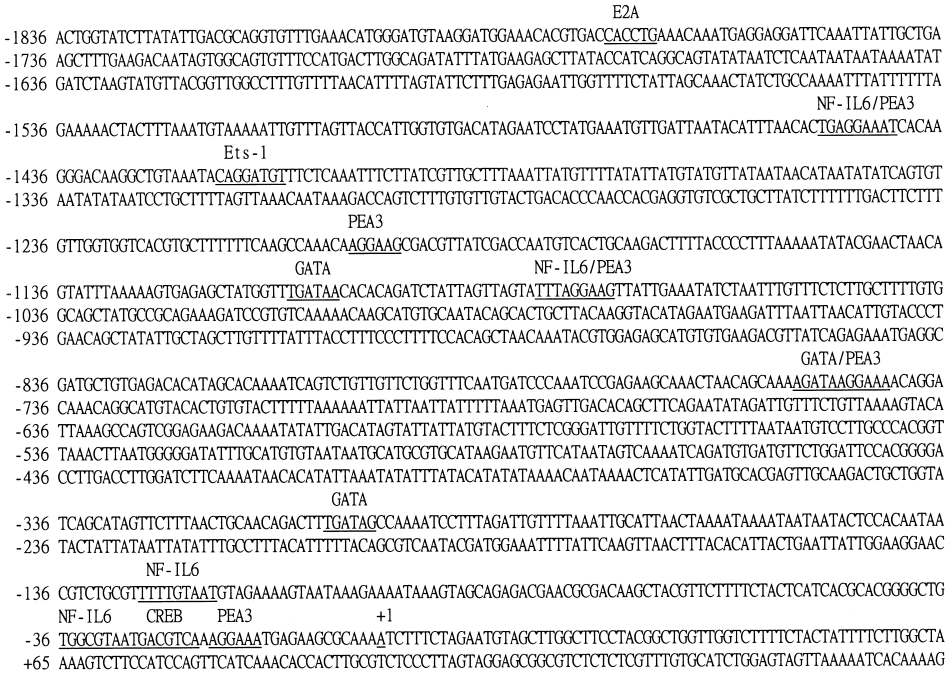

Fig. 3. Nucleotide sequence of the 5′-flanking region of the round-spotted pufferfish *CDC*37 gene. Bases are numbered with respect to the presumptive transcription start site (designated as +1), which was determined by 5′-RACE. Potential binding sites for a variety of transcription factors are also marked and underlined. Nucleotide +238 in this figure is the same as the last nucleotide of the first exon which contains only the 5′-UTR.

ments such as enhancer or suppressor in the 5′-flanking region of the pufferfish *CDC*37 gene needs further deletion analysis. So far, this is the first report to demonstrate the promoter activity of the *CDC*37 gene. In chick, although its genome structure has been described first, there is no data about the 2 kb promoter region of the chick *CDC*37 gene [19].

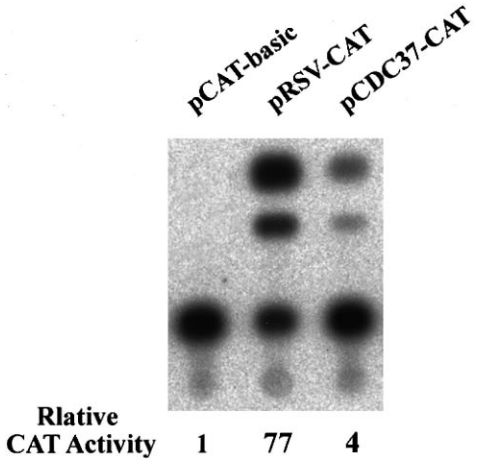

Fig. 4. Analysis of the promoter activity of the 5′-flanking region of the round-spotted pufferfish *CDC*37 gene fused to the CAT reporter gene. Each chimeric gene was cotransfected with pSV-β-galactosidase DNA into CF cells and assayed for CAT and β-galactosidase activities as previously described [14]. CAT activity in an individual experiment was corrected for variation in transfection efficiency by normalizing the value to the β-galactosidase activity in the same extract. The data represented the mean of triplicate transfection experiments for each plasmid. The acetylated products of the CAT assay were separated by thin layer chromatography developed with chloroform:methanol (95:5, v/v), visualized by autoradiography and quantitated by using the PhosphoImager (Bio-Imaging Analyzer BAS 2000, Fuji, Japan).

The interesting result in this study and the previous report [12] is that the pufferfish *CDC*37 gene is linked on the same chromosome to the *TYK*2 gene, separated only by a 292 bp intergenic sequence. These two genes are transcribed in the same direction by their own promoter. Similar tail-to-head linkage has been described for genes encoding human interleukin-4 and interleukin-13 with an intergenic region of 12 kb [36]. Their gene expression has been suggested to be coregulated. As for Cdc37 protein and TYK2 kinase, their function may be related because Cdc37 has been shown to specifically interact with Cdk4, Cdk6 [2] and Raf-1 [7]. Thus, Cdc37 may be able to interact with TYK2 or other JAK kinases and this issue remains to be studied.

Acknowledgements

This research was supported by grants from the National Science Council (NSC-87-2311-B-001-092 and NSC-89-2311-B-001-073), Taiwan, Republic of China. We are grateful to Dr. Sheng-Ping Hwang and Dr. Wen-Chang Lin for critically reading the manuscript and helpful discussions.

References

[1] S.I. Reed, Genetics 95 (1980) 579–588.

[2] L. Lamphere, F. Fiore, X. Xu, L. Brizuela, S. Keezer, C. Sardet, G.F. Draetta, J. Gyuris, Oncogene 14 (1997) 1999–2004.
[3] T. Ozaki, S. Sakiyama, FEBS Lett. 375 (1995) 155–158.
[4] Y. Kimura, S.L. Rutherford, Y. Miyata, I. Yahara, B.C. Freeman, L. Yue, R.I. Morimoto, S. Lindquist, Genes. Dev. 11 (1997) 1775–1785.
[5] G.H. Perdew, H. Wiegand, J.P. Van den Heuvel, C. Mitchell, S.S. Singh, Biochemistry 36 (1997) 3600–3607.
[6] A.M. Silverstein, N. Grammatikakis, B.H. Cochran, M. Chinkers, W.B. Pratt, J. Biol. Chem. 273 (1998) 20090–20095.
[7] N. Grammatikakis, J.H. Lin, A. Grammatikakis, P.N. Tsichlis, B.H. Cochran, Mol. Cell. Biol. 19 (1999) 1661–1672.
[8] N. Grammatikakis, A. Grammatikakis, M. Yoneda, Q. Yu, S.D. Banerjee, B.P. Toole, J. Biol. Chem. 270 (1995) 16198–16205.
[9] S. Brenner, G. Elgar, R. Sandford, A. Macrae, B. Venkatesh, S. Aparicio, Nature 366 (1993) 567–571.
[10] R. Hinegardner, D.E. Rosen, Am. Nat. 106 (1972) 621–644.
[11] J.H. Leu, M.S. Chang, J.W. Yao, C.K. Chou, C.J. Huang, Biochim. Biophys. Acta 1395 (1998) 50–56.
[12] J.H. Leu, S.J. Yan, T.F. Lee, C.M. Chou, S.T. Chen, P.P. Hwang, C.K. Chou, C.J. Huang, DNA Cell Biol. 19 (2000) 431–446.
[13] L. Stepanova, X. Leng, S.B. Parker, J.W. Harper, Genes Dev. 10 (1996) 1491–1502.
[14] M.S. Chang, G.D. Chang, J.H. Leu, F.L. Huang, C.K. Chou, C.J. Huang, T.B. Lo, DNA Cell Biol. 15 (1996) 827–844.
[15] C.M. Chou, W.C. Lin, J.H. Leu, T.L. Su, C.K. Chou, C.J. Huang, J. Biomed. Sci. 5 (1998) 127–134.
[16] J. Sambrook, E.F. Fritsch, T. Maniatis, in: N. Irwin, N. Ford, C. Nolan (Eds.), Molecular Cloning: a Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[17] I. Firmbach-Kraft, M. Byers, T. Shows, R. Dalla-Favera, J.J. Krolewski, Oncogene 5 (1990) 1329–1336.
[18] R. Breathnach, C. Benoist, K. O'hare, F. Gannon, P. Chambon, Proc. Natl. Acad. Sci. USA 75 (1978) 4853–4857.
[19] L. Huang, N. Grammatikakis, B.P. Toole, J. Biol. Chem. 273 (1998) 3598–3603.
[20] T. Ozaki, S. Sakiyama, DNA Cell Biol. 15 (1996) 975–979.
[21] A.E. Fliss, Y. Fang, F. Boschelli, A.J. Caplan, Mol. Biol. Cell 8 (1997) 2501–2509.
[22] R. Breathnach, P. Chambon, Annu. Rev. Biochem. 50 (1981) 349–383.
[23] S. Akira, H. Isshiki, T. Sugita, O. Tanabe, S. Kinoshita, Y. Nishio, T. Nakajima, T. Hirano, T. Kishimoto, EMBO J. 9 (1990) 97–1906.
[24] C. Murre, P.S. McCaw, D. Baltimore, Cell 56 (1989) 777–783.
[25] B. Wasylyk, S.L. Hahn, A. Giovane, Eur. J. Biochem. 211 (1993) 7–18.
[26] G.A. Gonzalez, M.R. Montminy, Cell 59 (1989) 675–680.
[27] T. Evans, M. Reitman, G. Felsenfeld, Proc. Natl. Acad. Sci. USA 85 (1988) 5976–5980.
[28] S.H. Orkin, Blood 80 (1992) 575–581.
[29] R.C. O'Hagan, J.A. Hassell, Oncogene 16 (1998) 301–310.
[30] S.N. Chen, G.H. Kou, in: E. Kurstak, Y. Kuroda (Eds.), Invertebrate and Fish Tissue Culture, Japan Societies Press, Springer, New York, 1986, pp. 218–227.
[31] C. Gorman, C. Padmanabhan, B.H. Howard, Science 221 (1983) 551–553.
[32] K. Ladjali-Mohammedi, J.J. Bitgood, M. Tixier-Boichard, F.A. Ponce De Leon, Anim. Genet. 29 (1998) 85–90.
[33] M.S. Chang, J.H. Leu, C.W. Yao, F.L. Huang, C.J. Huang, Zool. Stud. 36 (1997) 159–169.
[34] J.W. Yao, J.H. Leu, C. Chin, C.K. Chou, C.J. Huang, DNA Cell Biol. 17 (1998) 69–81.
[35] T. Crnogorac-Jurcevic, J.R. Brown, H. Lehrach, L.C. Schalkwyk, Genomics 41 (1997) 177–181.
[36] D.V. Smirnov, M.G. Smirnova, V.G. Korobko, E.I. Frolova, Gene 155 (1995) 277–281.