

# Structure and organization of the gene encoding a mouse mitochondrial stress-70 protein

Yuichi Michikawa<sup>a</sup>, Tadashi Baba<sup>a,\*</sup>, Yuji Arai<sup>a</sup>, Teruyo Sakakura<sup>b</sup>, Moriaki Kusakabe<sup>b,\*</sup>

<sup>a</sup>*Institute of Applied Biochemistry, University of Tsukuba, Tsukuba Science City, Ibaraki 305, Japan*

<sup>b</sup>*Laboratory of Cell Biology, RIKEN, Tsukuba Science City, Ibaraki 305, Japan*

Received 25 October 1993

We have previously found that an antigenic protein specific for C3H strain mouse (C3H strain-specific antigen, CSA) is identical to peptide-binding protein 74 (PBP74). PBP74/CSA is a novel member of the stress-70 protein family in mitochondria. In this study, mouse genomic clones encoding PBP74/CSA, including the 5'- and 3'-flanking regions of the gene, have been isolated and sequenced. The PBP74/CSA gene contained 17 exons interrupted by 16 introns. Two dimeric repeats of the consensus sequence of the heat-shock element are present in the 5'-flanking region of the PBP74/CSA gene. Moreover, the first intron is interrupted within the amino-terminal leader sequence, the pattern of which is similar to that of cytochrome *c*<sub>1</sub> located in the mitochondria.

Stress-70 protein; Peptide-binding protein74; C3H mouse strain-specific antigen (CSA); Genomic DNA; Gene organization

## 1. INTRODUCTION

An antigenic protein specific for C3H strain mouse (C3H strain-specific antigen, termed CSA) is a genetic marker in mouse for the study of cell lineage, homeostasis in tissue architecture, and cell-cell interactions in chimeric animals [1]. Antibody against CSA was prepared by immunization of partially purified proteins from muscle and liver extracts of C3H/HeN strain mice into (BALB/c×SJL/J) F1 mice [1]. The monoclonal antibody, as well as the polyclonal antibody, specifically recognizes cells derived only from the C3H strain of mouse. Cells from other strains of mice, including BALB/c and C57BL/6, show no immunoreactivity. Immunohistochemical studies using anti-CSA monoclonal antibody reveals the localization of this protein in mitochondria [2]. Moreover, the deduced amino acid sequence demonstrates that CSA is essentially identical to peptide-binding protein 74 (PBP74), a novel member of the stress-70 protein family [2,3]. However, the substitution of two amino acids is present in the PBP74/CSA sequences between C3H/HeN and BALB/c strain mice [2]. Western blot analysis indicates that arginine at residue 578 in the PBP74/CSA sequence of the C3H strain contributes to the immunogenicity of CSA [2]. It is thus

interesting that CSA, which was originally isolated as a genetic marker in mice, is a mitochondrial stress-70 protein.

Stress-70 proteins in mammalian species are divided into two groups: constitutively expressed and stress-induced proteins (for reviews, see [4,5]). Most of the genes encoding the stress-induced stress-70 proteins do not contain, intron sequences [6–10]. This fact may be due to the requirement of rapid gene expression for exogenous stress [9]. The protein-coding regions of the constitutively expressed proteins are interrupted by introns, and the exon-intron organization is highly homologous to each other [11–14]. At any rate, the consensus sequence of the heat-shock element, HSE (5'-NGAAN-3' or 5'-NTTCN-3'), is located in the 5'-flanking region of the genes coding for both types of stress-70 proteins [15,16].

In order to facilitate further studies of PBP74/CSA, the genomic structure and organization need to be established. We now describe the exon-intron organization of the mouse PBP74/CSA gene. The sequence of the putative promoter region is also reported.

## 2. EXPERIMENTAL

### 2.1. Materials

A mouse genomic library prepared from the liver of the C57BL/6 strain of mouse in Charon28 was the kind gift of Dr. Akiyoshi Fukamizu at the Institute of Applied Biochemistry, University of Tsukuba. Another genomic library from BALB/c liver was purchased from Clontech (Palo Alto, CA). Radioisotope, [ $\alpha$ -<sup>32</sup>P]dCTP (3,000 Ci/mmol), was purchased from Bresatec (Adelaide, Australia). Restriction endonucleases and modifying enzymes were purchased from Nippon Gene (Toyama, Japan) or Takara Shuzo (Kyoto). All other reagents were of the highest purity available.

\*Corresponding authors. \*Fax: (81) (298) 536 632. \*Fax: (81) (298) 369 010.

The nucleotide sequences reported in this paper have been submitted to the DDBJ, EMBL, and GenBank Nucleotide Sequence Databases with the accession numbers D17655–D17666.

## 2.2. Screening of mouse genomic libraries

Mouse genomic libraries were screened by the plaque hybridization method [17]. Plaque lifts were prehybridized at 42°C in 5 × SSPE (1 × SSPE = 10 mM sodium phosphate, pH 7.7, 0.18 M NaCl, and 1 mM EDTA), 0.02% Ficoll 400, 0.02% polyvinylpyrrolidone, 0.02% BSA, and 0.1% SDS, followed by hybridization at 60°C overnight in the prehybridization buffer containing denatured salmon testis DNA (0.1 mg/ml) and <sup>32</sup>P-labeled probe. The filters were washed in 2 × SSC (1 × SSC = 15 mM sodium citrate, pH 7.0, and 0.15 M NaCl) at room temperature for 10 min, 2 × SSC containing 0.1% SDS at 60°C for 10 min, and in 2 × SSC at room temperature for 10 min, prior to autoradiography at -80°C. Positive clones were plaque-purified, and the DNA fragments were subcloned into the appropriate sites of pUC19 for further characterization.

## 2.3. Polymerase-chain reaction (PCR)

Oligonucleotides, CSPI (5'-GAAGACCGCAGGAAGAAGGAA-CGTGTT-3') and CSP2 (5'-GTAGGAGCAAATATACAGAGGT-CATTCTTT-3'), were synthesized using a Cyclone-Plus DNA synthesizer (Millipore, Bedford, MA). PCR amplification was carried out in a mixture (50 µl) containing 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.01% gelatin, 0.1% Triton X-100, 0.2 mM each of deoxynucleotides (dATP, dCTP, dGTP, and dTTP), 1 µg of genomic DNA, 0.01 mM each of oligonucleotide primers, CSPI and CSP2, and 2.5 U of *Taq* DNA polymerase (Nippon Gene). The reaction program consisted of 35 cycles of 93°C for 1 min, 55°C for 2 min, and 72°C for 3 min in a DNA Thermal Cycler reactor (Perkin-Elmer Model PJ2000). The PCR products were purified by PAGE, and subcloned into a pCR II vector (Invitrogen, San Diego, CA).

## 2.4. Analytical procedures

Nucleotide sequence analysis was carried out by the dideoxy chain-termination method [18], using a commercial kit of BcaBest (Takara Shuzo, Kyoto). Computer-aided analysis of nucleotide and protein sequences was carried out using a GENETYX program (Software Development Co., Tokyo).

## 3. RESULTS

A single positive clone, termed  $\lambda$ Mmt70G101, was initially isolated by screening  $9 \times 10^5$  plaques from a genomic library of C57BL/6 mouse, using an entire re-

gion of the TM7 cDNA sequence encoding PBP74/CSA [2] as a probe. The insert DNA was sequenced and compared with the cDNA sequence of PBP74/CSA. The genomic region encoded by  $\lambda$ Mmt70G101 shared a high degree of sequence identity with PBP74/CSA (data not shown). However, some insertions, deletions, and substitutions were found between the cDNA sequence of PBP74/CSA and the nucleotide sequence of the  $\lambda$ Mmt70G101 clone. Since  $\lambda$ Mmt70G101 did not encode either enough open-reading frame or a possible intron-like sequence, we conclude that this clone codes for a pseudogene of the PBP74/CSA gene (the details will be reported elsewhere). On the basis of the deleted sequence in the pseudogene, two oligonucleotide primers, CSPI and CSP2, were synthesized, and the corresponding genomic region in BALB/c mouse was amplified by PCR, and sequenced. The DNA fragment contained three exons interrupted by two introns. A 192-nucleotide *AhaI*-*AhaI* fragment in the 5'-end intron was used to screen  $9 \times 10^5$  plaques from a BALB/c mouse genomic library. Two positive clones,  $\lambda$ Mmt70G207 and  $\lambda$ Mmt70G209, were isolated. These two clones overlapped each other but lacked a 5'-end region of the PBP74/CSA gene. Further screening of  $6 \times 10^5$  plaques from the BALB/c mouse genomic library, using an *RsaI*-*HindIII* DNA fragment in the first intron, yielded four positive clones,  $\lambda$ Mmt70G301,  $\lambda$ Mmt70G302,  $\lambda$ Mmt70G303, and  $\lambda$ Mmt70G304. Thus, three genomic clones,  $\lambda$ Mmt70G207,  $\lambda$ Mmt70G209, and  $\lambda$ Mmt70G303, were selected for sequence analysis (Fig. 1).

Organization of the mouse PBP74/CSA gene and the genomic DNA sequence including 5'- and 3'-flanking sequences are shown in Figs. 1 and 2, respectively. This gene is approximately 17 kbp in length and consists of 17 exons separated by 16 introns. No significant similar-

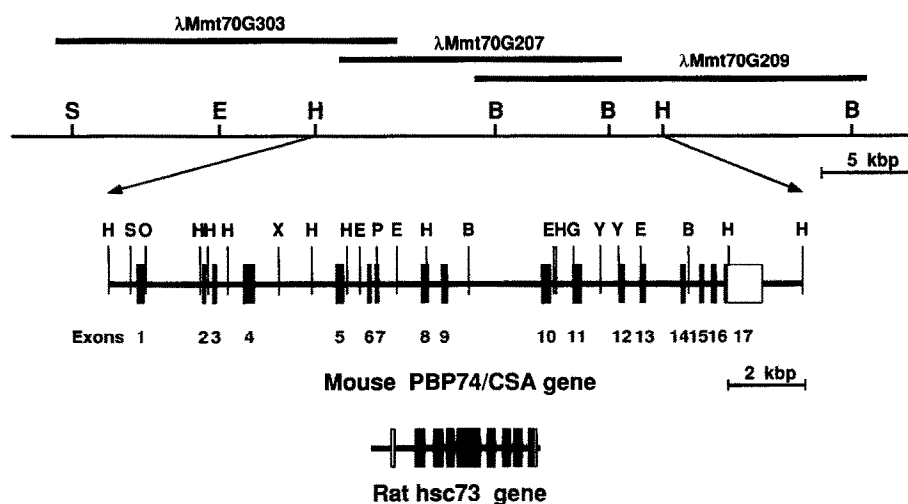


Fig. 1. Organization of the mouse PBP74/CSA gene. The gene contains 17 exons (exons 1–17) interrupted by 16 introns. The protein-coding region, and the 5'- and 3'-untranslated regions of the PBP74/CSA mRNA are shown by closed and open boxes, respectively. For comparison, the organization of the rat hsc73 gene [12] is also indicated. The sites of restriction enzymes are shown as follows: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; G, *Bgl*II; O, *Xho*I; P, *Pvu*II; S, *Sma*I; X, *Xba*I; Y, *Sty*I.

ity of the gene organization is found between PBP74/CSA and rat hsc73 [12] (Fig. 1). Also, there is no nucleotide substitution between the genomic and cDNA sequences in BALB/c strain mouse. As described previously [2], the substitution of two nucleotides at positions 1,925 and 1,943 in the cDNA sequence encoding PBP74/CSA is present between the C3H strain and other strains of mice lacking immunoreactivity with anti-CSA antibody, including BALB/c and C57BL/6 mice. To verify whether these substitutions occur in the genomic sequence of C3H mouse strain, PCR was carried out using the genomic DNAs from C3H and C57BL/6 strain mice, and CSP1 and CSP2 as primers. The nucleotide sequence of the amplified fragment of C57BL/6 mouse is identical to that of BALB/c, whereas the same nucleotide substitutions were confirmed in the genomic sequence of C3H strain mouse (data not shown).

Analysis of primer extension and S1 mapping failed

to clarify the transcription initiation site of the PBP74/CSA gene. This may be due to the fact that the 5'-untranslated region of the PBP74/CSA mRNA is highly rich in guanine and cytosine, and forms a stable secondary structure according to the computer-aided structure analysis (not shown). Two sequence stretches containing dimeric HSEs [15,16] are found in the putative promoter region of the gene, 6 and 174 nucleotides upstream from the 5'-end of the known cDNA sequence at position -73 (Fig. 2). There is no typical TATA sequence in the promoter region, whereas a consensus CCAAT sequence is located at nucleotides -530 to -534. Some *cis*-acting sequences, including the consensus sequences for Mt1 [19,20], Mt2 [19], NRF-1 [21,22], and MyoD [23,24], are also present in the promoter region.

The sizes of exons and introns, and intron phase classes are summarized in Table I. All sequences at the exon-intron boundaries are consistent with the consen-

Table I  
Nucleotide sequences of splice junctions in the mouse PBP74/CSA gene

Exon			Intron				Exon		
Term	Size (nt) <sup>a</sup>	Sequence	Donor	Term	Size (nt) <sup>a</sup>	Phase class	Acceptor	Sequence	Term
1	154	CCGTCCCCAG aArgProGln	gtgagaag	A	~1,600	0	ccccccag	GATGGCTGGA AspGlyTrpA	2
2	59	GAGATTATGC rgAspTyrAl	gtaagtac	B	191	2	tcttacag	ATCAGAAGCA aSerGluAla	3
3	88	ACAAGCAAAG sGlnAlaLys	gtgagcat	C	~800	0	aatttttag	GTCCTGGAGA ValLeuGluA	4
4	182	AGAAAGACAC lnLysAspTh	gtgagtaa	D	~2,200	2	tctttaag	TAAGAATGTT rLysAsnVal	5
5	125	GAGACTGCAG GluThrAlaG	gtaagtgg	E	~700	1	tttcccag	AAAATTACTT luAsnTyrLe	6
6	74	ACAGCGACAG rGlnArgGln	gtaaaatt	F	88	0	ttctctag	GCCACTAAGG AlaThrLysA	7
7	107	AAGATAAAGT luAspLysVa	gtaagttg	G	~1,200	2	tttttttag	CATTGCTGTG lileAlaVal	8
8	163	CAAGAGAGAG eLysArgGlu	gttagtta	H	~400	0	atttatag	ACAGGGGTTG ThrGlyValA	9
9	93	ATCTGTGCAG rSerValGln	gtgaggga	I	~2,500	0	ctgtgcag	ACTGACATCA ThrAspIleA	10
10	210	GATGCCCAAG gMetProLys	gtatggac	J	~800	0	atgccttag	GTTCAGCAGA ValGlnGlnT	11
11	228	AAAGAGCCAG sLysSerGln	gtaagagc	K	~900	0	XXXcctag	GTGTTTCTTA ValPheSerT	12
12	105	GTTCACTTTG nPheThrLeu	gtaagtgt	L	528	0	tatttttag	ATTGGAATTC IleGlyIleP	13
13	118	GAGCAACAGA GluGlnGlnI	gtaagtaa	M	~1,000	1	aactacag	TTGTAATCCA leValIleGl	14
14	95	CAGGAAGAAG gArgLysLys	gtgattac	N	317	0	tttctaag	GAACGTGTTG GluArgValG	15
15	93	TGCTGATGAG oAlaAspGlu	gtaccatt	O	207	0	gtttccag	TGCAACAAGC CysAsnLysL	16
16	141	GTACAAAAAG aTyrLysLys	gtacaagg	P	189	0	ctcaacag	ATGGCATCTG MetAlaSerG	17
17	963	CTTAATAAAA							

The sequences of exons and introns are indicated by capital and small letters, respectively.

<sup>a</sup>Nucleotide.

aactgctgacattacaggcatattgtgtattttttaaaaagaacgttatgtacatgagt  
tatgaaaccatgatttttagttttttacctaagtgctttgtgttttcagaatttgaat

31

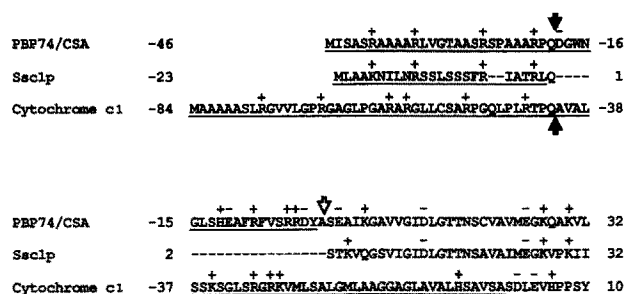


Fig. 3. Comparison of intron insertion in the amino-terminal leader sequences of PBP74/CSA and cytochrome  $c_1$ . The amino-terminal leader sequences of PBP74/CSA [2,3], human cytochrome  $c_1$  [15], and yeast Ssc1p [26] are underlined, and the positively and negatively charged residues are also indicated, by (+) and (-), respectively. The insertion sites of first (phase 0) and second (phase 2) introns are shown by closed and open arrows, respectively.

the two repeats of dimeric HSEs, located in the promoter region of the PBP74/CSA gene, as a transcriptional element(s) is tentative, until the direct evidence is obtained.

Comparison of the promoter regions between the PBP74/CSA gene and other genes encoding nuclear-encoded mitochondrial proteins reveals significant similarities (Fig. 2); the absence of a TATA box, and the presence of consensus sequences for regulatory elements, including Mt1 [19,20], Mt2 [19], and NRF-1 [21,22], as in the case of the gene encoding human cytochrome  $c_1$  [19,20,28]. The PBP74/CSA gene is ubiquitously expressed in all tissues of mouse, but there is a variation in the levels of gene expression among the tissues [2]. The expression pattern of the PBP74/CSA gene is similar to that of the gene encoding rat cytochrome  $c$  oxidase subunit IV in mitochondria [29]. The genes coding for nuclear-encoded, mitochondrial proteins are simply classified by the presence of NRF-1 or NRF-2 in the promoter region [21,22,30]. Although the cytochrome  $c$  oxidase subunit IV gene possesses the consensus NRF-2 sequence [30], the NRF-1 sequence, instead of NRF-2, is present in the PBP74/CSA gene (Fig. 2). Therefore, other elements may contribute to the expression of the PBP74/CSA gene. Moreover, the nucleotide sequence highly rich in guanine and cytosine is located at the 5'-untranslated region of the PBP74/CSA gene (Fig. 2). Computer-aided analysis of this sequence demonstrates a stable secondary structure, as described above. Other genes for TGF- $\beta$ 1 [31], TGF- $\beta$ 3 [32], ferritin [33,34], and c-sis/platelet-derived growth factor 2 [35,36] have been reported to contain the similar sequence with the stable secondary structure in the 5'-untranslated region, which acts as a translation regulator [37,38]. Thus, further experiments are required to elucidate the regulatory mechanism of expression of the PBP74/CSA gene.

Our immunohistochemical observation reveals that PBP74/CSA is a mitochondrial stress-70 protein [2].

However, the localization of this protein in the mitochondria is not certain at present. The gene organization within the amino-terminal region of PBP74/CSA may suggest that the amino-terminal leader sequence contains a bipartite targeting signal, as found in cytochrome  $c_1$  [39] (Fig. 3). The amino acid sequence encoded by the first exon of the PBP74/CSA gene appears to be necessary for targeting to the mitochondria. It is also interesting to suppose that the sequence encoded by second exon functions as a signal for re-localization of this protein within the mitochondria. The phase classes of first and second introns in both genes encoding PBP74/CSA and cytochrome  $c_1$  are in phase 0 and phase 2, respectively (Table I and Fig. 3). According to the exon shuffling model [40], it is unlikely that the second exon has been inserted into an intron sequence between the first and third exons during evolution. At any rate, cloning of the PBP74/CSA gene allows us to examine the gene regulation, and the role of the stress-70 protein in the mitochondria.

**Acknowledgements:** We acknowledge the technical assistance of Ms. Noriko Shioda. This study was partly supported by a grant from the Ministry of Education, Science, and Culture of Japan to M.K. and T.B.

## REFERENCES

- [1] Kusakabe, M., Yokoyama, M., Sakakura, T., Nomura, T., Hosick, H.L. and Nishizuka, Y. (1988) *J. Cell Biol.* 107, 257-265.
- [2] Michikawa, Y., Baba, T., Arai, Y., Sakakura, T., Tanaka, M. and Kusakabe, M. (1993) *Biochem. Biophys. Res. Commun.* (in press).
- [3] Domanico, S.Z., DeNagel, D.C., Dahlseid, J.N., Green, J.M. and Pierce, S.K. (1993) *Mol. Cell. Biol.* 13, 3598-3610.
- [4] Lindquist, S. and Craig, E.A. (1988) *Annu. Rev. Genet.* 22, 631-677.
- [5] Gething, M.-J. and Sambrook, J. (1992) *Nature* 355, 33-45.
- [6] Holmgren, R., Livak, K., Morimoto, R., Freund, R. and Meselson, M. (1979) *Cell* 18, 1359-1370.
- [7] Ingolia, T.D., Craig, E.A. and McCarthy, B.J. (1980) *Cell* 21, 669-679.
- [8] Ingolia, T.D. and Craig, E.A. (1982) *Proc. Natl. Acad. Sci. USA* 79, 525-529.
- [9] Hunt, C. and Morimoto, R.I. (1985) *Proc. Natl. Acad. Sci. USA* 82, 6455-6459.
- [10] Hunt, C. and Calderwood, S. (1990) *Gene* 87, 199-204.
- [11] Craig, E.A., Ingolia, T.D. and Manseau, L.J. (1983) *Dev. Biol.* 99, 418-426.
- [12] Sorger, P.K. and Pelham, H.R.B. (1987) *EMBO J.* 6, 993-998.
- [13] Dworniczak, B. and Mirault, M.-E. (1987) *Nucleic Acid Res.* 15, 5181-5197.
- [14] La Rosa, M., Sconzo, G., Giudice, G., Roccheri, M.C. and Di Carlo, M. (1990) *Gene* 96, 295-300.
- [15] Amin, J., Ananthan, J. and Voellmy, R. (1988) *Mol. Cell. Biol.* 8, 3761-3769.
- [16] Xiao, H. and Lis, J.T. (1988) *Science* 239, 1139-1142.
- [17] Benton, W.D. and Davis, R.W. (1977) *Science* 196, 180-182.
- [18] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- [19] Suzuki, H., Hosokawa, Y., Nishikimi, M. and Ozawa, T. (1989) *J. Biol. Chem.* 264, 1368-1374.
- [20] Suzuki, H., Hosokawa, Y., Toda, H., Nishikimi, M. and Ozawa, T. (1990) *J. Biol. Chem.* 265, 8159-8163.

- [21] Evans, M.J. and Scarpulla, R.C. (1990) *Genes Dev.* 4, 1023–1034.
- [22] Chau, C.A., Evans, M.J. and Scarpulla, R.C. (1992) *J. Biol. Chem.* 267, 6999–7006.
- [23] Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.-F., Weintraub, H. and Lassar, A.B. (1988) *Science* 242, 405–411.
- [24] Davis, R.L., Cheng, P.-F., Lassar, A.B. and Weintraub, H. (1990) *Cell* 60, 733–746.
- [25] Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem.* 50, 349–383.
- [26] Craig, E.A., Kramer, J., Shilling, J., Werner-Washburne, M., Holmes, S., Kosc-Smithers, J. and Nicolet, C.M. (1989) *Mol. Cell. Biol.* 9, 3000–3008.
- [27] Kroeger, P.E., Sarge, K.D. and Morimoto, R.I. (1993) *Mol. Cell. Biol.* 13, 3370–3383.
- [28] Evans, M.J. and Scarpulla, R.C. (1989) *J. Biol. Chem.* 264, 14361–14368.
- [29] Virbasius, J.V. and Scarpulla, R.C. (1990) *Nucleic Acids Res.* 18, 6581–6586.
- [30] Virbasius, J.V., Virbasius, C.A. and Scarpulla, R.C. (1993) *Genes Dev.* 7, 380–392.
- [31] Kim, S.-J., Park, K., Koeller, D., Kim, K.Y., Wakefield, L.M., Sporn, M.B. and Robers, A.B. (1992) *J. Biol. Chem.* 267, 13702–13707.
- [32] Arrick, B.A., Lee, A.L., Grendell, R.L. and Derynck, R. (1991) *Mol. Cell. Biol.* 11, 4306–4313.
- [33] Hentze, M.W., Rouault, T.A., Caughman, S.W., Dancis, A., Harford, J.B. and Klausner, R.D. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6730–6734.
- [34] Caughman, S.W., Hentze, M.W., Rouault, T.A., Harford, J.B. and Klausner, R.D. (1988) *J. Biol. Chem.* 263, 19048–19052.
- [35] Ratner, L., Thielan, B. and Collins, T. (1987) *Nucleic Acids Res.* 15, 6017–6036.
- [36] Rao, C.D., Pech, M., Robbins, K.C. and Aaronson, S.A. (1988) *Mol. Cell. Biol.* 284–292.
- [37] Klausner, R.D. and Harford, J.B. (1989) *Science* 246, 870–872.
- [38] Kozak, M. (1989) *Mol. Cell. Biol.* 9, 5134–5142.
- [39] Glick, B.S., Beasley, E.M. and Schatz, G. (1992) *Trends Biochem. Sci.* 17, 453–459.
- [40] Patthy, L. (1987) *FEBS Lett.* 214, 1–7.
- [41] Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) *Cell* 53, 11–24.
- [42] van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) *EMBO J.* 10, 123–132.
- [43] Waterman, M.L., Fischer, W.H. and Jones, K.A. (1991) *Genes Dev.* 5, 656–669.
- [44] Lee, C.Q., Yun, Y., Hoeffler, J.P. and Habener, J.F. (1990) *EMBO J.* 9, 4455–4465.
- [45] Dynan, W.S. and Tjian, R. (1983) *Cell* 35, 79–87.
- [46] Ariga, H., Imamura, Y. and Iguchi-Ariga, S.M.M. (1989) *EMBO J.* 8, 4273–4279.
- [47] Blackwood, E.M. and Eisenman, R.N. (1991) *Science* 251, 1211–1217.