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

Yuichi Michikawa^a, Tadashi Baba^{a,*}, Yuji Arai^a, Teruyo Sakakura^b, Moriaki Kusakabe^{b,*}

^aInstitute of Applied Biochemistry, University of Tsukuba, Tsukuba Science City, Ibaraki 305, Japan ^bLaboratory 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_1 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

*Corresponding authors. aFax: (81) (298) 536 632. bFax: (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.

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 stressinduced 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, [α-32P]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.

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 $5 \times SSPE$ ($1 \times SSPE = 10$ 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 32 P-labeled probe. The filters were washed in $2 \times SSC$ ($1 \times SSC = 15$ mM sodium citrate, pH 7.0, and 0.15 M NaCl) at room temperature for 10 min, $2 \times SSC$ containing 0.1% SDS at 60°C for 10 min, and in $2 \times SSC$ 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'-GAAGACCGCAGGAAGAAGGAACGTGTT-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 MgCl₂, 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, CSP1 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 chaintermination 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 9×10^5 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 λ Mmt70G101 clone. Since λ 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, CSP1 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 192nucleotide AluI-AluI fragment in the 5'-end intron was used to screen 9×10^5 plaques from a BALB/c mouse genomic library. Two positive clones, \(\lambda \text{Mmt70G207} \) 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 6×10^5 plaques from the BALB/c mouse genomic library, using an RsaI-HindIII DNA fragment in the first intron, yielded four positive clones, \(\lambda \text{Mmt70G301}, \(\lambda \text{MmtG302}, \) λMmt70G303, and λMmt70G304. Thus, three geno- λ Mmt70G207, λ Mmt70G209, clones, λMmt70G303, 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-

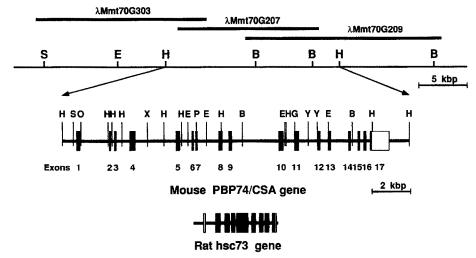


Fig. 1. Organization of the mouse PBP74/CSA gene. The gene contains 17 exons (exons 1-17) interrupted by 16 introns. The protein-coding region, and the 5'- and 3'-untranslated regions of the PBP74/CSA mRNA are shown by closed and open boxes, respectively. For comparison, the organization of the rat hsc73 gene [12] is also indicated. The sites of restriction enzymes are shown as follows: B, BamHI; E, EcoRI; H, HindIII; G, Bg/II; O, XhoI; P, PvuII; S, SmaI; X, XhaI; Y, StyI.

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) ^a	Sequence	Donor	Term	Size (nt) ^a	Phase class	Acceptor	Sequence	Term	
1	154	CCGTCCCCAG	gtgagaag	A	~1,600	0	cccccag	GATGGCTGGA	2	
		aArgProGln						AspGlyTrpA		
2	59	GAGATTATGC	gtaagtac	В	191	2	tcttacag	ATCAGAAGCA	3	
		rgAspTyrA1						aSerGluAla		
3	88	ACAAGCAAAG	gtgagcat	C	~ 800	0	aattttag	GTCCTGGAGA	4	
		sGlnAlaLys						ValLeuGluA		
4	182	AGAAAGACAC	gtgagtaa	D	~2,200	2	tctttaag	TAAGAATGTT	5	
		lnLysAspTh						rLysAsnVal		
5	125	GAGACTGCAG	gtaagtgg	E	~ 700	1	tttcccag	AAAATTACTT	6	
		GluThrAlaG						luAsnTyrLe		
6	74	ACAGCGACAG	gtaaaatt	F	88	0	ttctctag	GCCACTAAGG	7	
		rGlnArgGln						AlaThrLysA		
7	107	AAGATAAAGT	gtaagttg	G	~1,200	2	ttttttag	CATTGCTGTG	8	
		luAspLysVa						lIleAlaVal		
8	163	CAAGAGAGAG	gttagtta	H	~ 400	0	atttatag	ACAGGGGTTG	9	
		eLysArgGlu						ThrGlyValA		
9	93	ATCTGTGCAG	gtgaggga	I	~2,500	0	ctgtgcag	ACTGACATCA	10	
		rSerValGln						ThrAspIleA		
10	210	GATGCCCAAG	gtatggac	J	~ 800	0	atgcttag	GTTCAGCAGA	11	
		gMetProLys						ValGlnGlnT		
11	228	AAAGAGCCAG	gtaagagc	K	~ 900	0	XXXcctag	GTGTTTTCTA	12	
		sLysSerGln					J	ValPheSerT		
12	105	GTTCACTTTG	gtaagtgt	L	528	0	tattttag	ATTGGAATTC	13	
		nPheThrLeu	5 5 5					IleGlyIleP		
13	118	GAGCAACAGA	gtaagtaa	M	~1,000	1	aactacag	TTGTAATCCA	14	
		GluGlnGlnI	3		•			leValIleGl		
14	95	CAGGAAGAAG	gtgattac	N	317	0	tttctaag	GAACGTGTTG	15	
		gArgLysLys	3 3				•	GluArgValG		
15	93	TGCTGATGAG	gtaccatt	0	207	0	gtttccag	TGCAACAAGC	16	
_		oAlaAspGlu	3	_	_ •	-	5	CysAsnLysL		
16	141	GTACAAAAAG	gtacaagg	P	189	0	ctcaacag	ATGGCATCTG	17	
		aTyrLysLys	3 3 3	-		-		MetAlaSerG		
17	963	CTTAATAAAA								

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

^aNucleotide.

		Exon 7	
aagettegtgaateegatgetgeeetagagettaggaacetaaceegacegaageaatge Mt1		gaaagatagatttatttetactttetetag GCC ACT AAG GAT GCT GGC CAG A	631
gctgcccaaatttgtgtgagaccgcgcacattcaggtgtgcacatttcaagcggcgaaca	-694 -634	Ala Thr Lys Asp Ala Gly Gln I TA TCT GGG CTA AAT GTG CTT CGA GTG ATC AAT GAG CCT ACA GCT G	676
aacagcgatggaqaacaaaggtccccqaaacaacqcctcaggacaagatggcqaccgcag caagtcgcggcctccggggacaacgcccacgccctgcctg		le Ser Gly Leu Asn Val Leu Arg Val Ile Asn Glu Pro Thr Ala A	180
CCAAT ccccggcactcgcccgcaaattcgcgtgcactggagcat <u>ccaataaa</u> cgctacagttaaa	-514	CT GCT CTA GCT TAC GGT CTG GAC AAA TCT GAA GAT AAA GT gtaagt la Ala Leu Ala Tyr Gly Leu Asp Lys Ser Glu Asp Lys Va	716 193
cccgaacccgttttctccacgcccctcacacatgcgcttcacgacctctgtccggagcgg NRF-1 Myo-D	-454	tggtcagatgacgtagcattacctgcatttacagggggttgtgtgtg	
agaacqcaggcgcagaacgcacctccttctgcgcaactgacgcaaggagactgtaatctg	-394	gcatggatatggatgtcaaagatt 800 bpaaaccagtagttggtgttac taattttacttgtttctgtatgacatgcttatatgatacgttaacaggcatgccacttag	
ttcgtaatttatcccgtgtgaccttgagtctttcccgtcctaaggaccgggactgtccga TCF-1	-334	gttgataactacattcagttattaagtatttcaggaaatatcatagtt <u>aataattaaact</u> tttggtcttttaattcgttcctttgttttgttccttggaacttaacttactt	
agggatgcctgtccacctctctctctctctctctctctct	-274 -214	EXON 8 tttag C ATT GCT GTG TAT GAT TTA GGT GGT GGA ACC TTT GAC ATT	756
gegeegateeegggageageggacaegtagtetetagteaggeageaegtegggegeete		1 Ile Ala Val Tyr Asp Leu Gly Gly Gly Thr Phe Asp Ile	206
Mt.2 Spl Myc agcagaagaggggttcattgctgcaggggcaagccccaccccaccccaccccacgtggttg	-94	TCT ATC CTG GAA ATT CAG AAA GGA GTG TTT GAG GTG AAA TCT ACC Ser Ile Leu Glu Ile Gin Lys Gly Val Phe Glu Val Lys Ser Thr	801 221
gaggtttccagaagcgtagcAccAccGcTGCACGCACGCTCCGGGCCCCGTGGGGTGTTGGT	-34	AAT GGG GAC ACT TTC TTA GGA GGG GAA GAC TTT GAC CAA GCT TTG Asn Gly Asp Thr Phe Leu Gly Gly Glu Asp Phe Asp Gln Ala Leu	846 236
TCTTGCCCTCGTAACCCCCTCTGTCCAGCCACC ATG ATA AGC GCC AGC AGA GC	20	TTG CGG CAC ATT GTC AAG GAG TTC AAG AGA GAG gttagttaccactgct	879
Met Ile Ser Ala Ser Arg Al	-40 65	Leu Arg Bis Ile Val Lys Glu Phe Lys Arg Glu tagtcaccactggttaaggtgtaggcgttggtgttgagaatttttgtttg	247
a Ala Ala Ala Arg Leu Val Gly Thr Ala Ala Ser Arg Ser Pro Al	-25	gcttttagctttgttaatagcttttttatactaaggtaactaac	
A GCC GCC CGT CCC CAG gtgagaagetgccatgccttccggtgggggctccaggc a Ala Ala Arg Pro Gln	81 -20	gtgttatagtagggagaag 140 bpactactctctcgcagtctatatatgte Exon 9	
ccggactcgagtgaggcaggccttgccttcgggtcagactctaggaaaaatccggagcga agggatgtaacggaccttctgtgggcattgttggccttcttgcagggctttagcttcgaa		tcctaatcctatttatag ACA GGG GTT GAT TTG ACC AAA GAC AAC ATG G Thr Gly Val Asp Leu Thr Lys Asp Asn Met A	910 258
ctgtgctgagtcaccatccttggcgttcctaagtctttaccccgctaattgagacgtctg tcccccctctaacctgtgcgctttgaatgtgcctggacttaggcagtggacgtagtttac tggaaa 1280 bpttttacatactaaggactaaagcttgcatgtgttcctca		CG CTT CAG AGG GTT CGG GAA GCT GCT GAG AAG GCT AAA TGT GAA C la Leu Gln Arg Val Arg Glu Ala Ala Glu Lys Ala Lys Cys Glu L	955 273
gaatatgaaaagcccaagttcttgtgtagttgatggtgatgcaacttttttccccccag Exon 2		TT TCC TCA TCT GTG CAG gtgagggatggaaaaatcccagtactgagcatatttg	972
GAT GGC TGG AAT GGC CTT AGC CAT GAG GCT TTT AGA TTT GTT TCA	126	eu ser ser val Gin	278
Asp Gly Trp Asn Gly Leu Ser His Glu Ala Phe Arg Phe Val Ser AGA AGA GAT TAT GC gtaagtacaacctcagtttctctgagaaaaaaaaaaaaa] -5 140	aatagtgtattetaatttacetaatgteagtgtagetetttacagttttetgttggetga aaacttggggcatgagcaaaggaacaacttgatgatgateagttetttteatttgaatgaatg aagtagatttatggatgtgtgtatettttgeetgeatgtgtgtetgtaetacaatttgtge	
Arg Arg Asp Tyr Al	1	ttggtttctgtggcggccagaagagggtatcagaactgacatgtcagtgttgggga 2100 bpcacactgagtctaaattctgatcatgtctttcagtcgtgttatgttact	
ttattgaacctcaaagcttggatgggttgggtgcgttatacatttgtacttgtagtttat tcaatatgccactggtaacaccaacataaaaacacagttcttcgtattggagaccactgtt Exon 3	_	ttgagtgagtatcaaagatcacgtctcccatctgacgtgtggtcctgtgcag	978
cagatgaccatggaatttcatttcttacag A TCA GAA GCA ATC AAG GGT GCA a Ser Glu Ala Ile Lys Gly Ala	162 8	ATC AAC TTG CCA TAC CTT ACC ATG GAT GCT TCT GGA CCA AAG CAT Ile Asn Leu Pro Tyr Leu Thr Met Asp Ala Ser Gly Pro Lys His	1023 295
GTG GTT GGT ATT GAT TTG GGT ACT ACT AAC TCC TGT GTG GCT GTT Val Val Gly Ile Asp Leu Gly Thr Thr Asn Ser Cys Val Ala Val	207 23	TTG AAT ATG AAG CTG ACT CGA GCT CAG TTT GAA GGC ATT GTC ACA	1068
ATG GAG GGC AAA CAA GCA AAG Met Glu Gly Lys Gln Ala Lys	228 30	Leu Asn Met Lys Leu Thr Arg Ala Gln Phe Glu Gly Ile Val Thr GAT CTA ATC AAG AGA ACT ATT GCT CCG TGT CAG AAA GCT ATG CAG Asp Leu Ile Lys Arg Thr Ile Ala Pro Cys Gln Lys Ala Met Gln	310 1113 325
atacccagtctggcattaagtacataggaatgctgagtcgggccaggttagtggtggg cactttaatcctagtaaaggcagagggatctctgagttcaggaccagcctagagtacaaa Exon 4		GAT GCA GAA GTC AGC AAG AGT GAC ATA GGA GAA GTG ATT CTG GTT Asp Ala Glu Val Ser Lys Ser Asp Ile Gly Glu Val Ile Leu Val	1158 340
gtgag 600 bpactttaatcagtgtctttgggaattttag GTC CTG GAG Val Leu Glu	237 33	GGT GGC ATG ACA AGG ATG CCC AAG gtatggactcatggtatttctcctagag Gly Gly Met Thr Arg Met Pro Lys	1182 348
AAT GCT GAA GGT GCC AGA ACT ACC CCT TCT GTG GTT GCC TTT ACA Asn Ala Glu Gly Ala Arg Thr Thr Pro Ser Val Val Ala Phe Thr	282 48	gaaaaaataacaatgcattcttgaggcaaatggcttgtgttgtgtgtg	
GCA GAT GGA GAA CGA CTT GTT GGT ATG CCA GCA AAA CGG CAA GCT	327	<pre>tagtcgcttytgatggtcatytaggaggatggcttyggttctgggtaagatggcat cagaattctaccttgcacacacacacttaaacccagtctggtaaagagaagttgttaagctt 180 bpcagatgtaaaattgataatacttcttcctaagattgggcgattgtg</pre>	
Ala Asp Gly Glu Arg Leu Val Gly Met Pro Ala Lys Arg Gln Ala GTC ACC AAT CCA AAC AAT ACC TTC TAT GCT ACT AAG CGT CTT ATT	63 372	tagaatatgagaactgtattttataaccottgtcatgtgcottotatgtgtgatatgt cototgggaagctagataattaaattcttototottatgtggacagtggacagtggatgagatgc	
Val Thr Asn Pro Asn Asn Thr Phe Tyr Ala Thr Lys Arg Leu Ile	78	tetgaagttgteaaatacaaacaagtetgeagtettggatätgaatetetetgaettget gtetggeaggegtattetgtttggeteeateagtegeeetgggtgttggetaacaggtte	
GGA CGA CGA TAT GAT GAC CCT GAA GTA CAG AAA GAC AC gtgagtaat Gly Arg Arg Tyr Asp Asp Pro Glu Val Gln Lys Asp Th	410 91	Exon 11 tttctccctgatgcttag GTT CAG CAG ACT GTA CAA GAT CTT TTT GGC A	1213
<pre>aggaaaatcagtccagaagactggtgctttgatcaaagttctgtggataccttgagttct gtggatcaccttggatcactttttcattatttctgcttgggaagaaatcaccaccatc</pre>		Val Gln Gln Thr Val Gln Asp Leu Phe Gly A	359
<pre>agaggcatataggtttttttttttttttttttttttttt</pre>		GA GCC CCG AGT ANA GCT GTT ANT CCT GAT GAG GCT GTA GCC ATC G rg Ala Pro Ser Lys Ala Val Asn Pro Asp Glu Ala Val Ala Ile G	1258 374
Exon 5	1	GA GCT GCC ATC CAG GGA GGT GTG TTG GCT GGT GAC GTT ACA GAC G ly Ala Ala Ile Gln Gly Gly Val Leu Ala Gly Asp Val Thr Asp V	1303 389
agggatagtttacttacataatctttctgtctttaag T AAG AAT CTT CCT TTT r Lys Asn Val Pro Phe	426 96	TG CTG CTC CTG GAT GTC ACT CCC CTC TCT CTG GGT ATT GAG ACT C al Leu Leu Leu Asp Val Thr Pro Leu Ser Leu Gly 11e Glu Thr L	1348 404
AAA ATT GTC CGT GCC TCC AAT GGT GAT GCT TGG GTT GAG GCT CAT Lys Ile Val Arg Ala Ser Asn Gly Asp Ala Trp Val Glu Ala His	471 111	TG GGA GGC GTC TTT ACC AAA CTT ATT AAT AGG AAC ACC ACT ATT C	1393
GGA AAA CTC TAT TCT CCA AGT CAG ATT GGA GCA TTT GTG TTG ATG Gly Lys Leu Tyr Ser Pro Ser Gln Ile Gly Ala Phe Val Leu Met	516 126	eu Gly Gly Val Phe Thr Lys Leu Ile Asn Arg Asn Thr Thr Ile P CA ACC AAA AAG AGC CAG gtaagagccattctttttcctgcctattaacagtcc	1410
AAG ATG AAA GAG ACT GCA G gtaagtggatttatttcacatttaggaaaattgg lys Met Lys Glu Thr Ala G	535 133	ro Thr Lys Lys Ser Gln caagttgtacaagtgctgtttacaatcactttatgaactctttaaaactttgtttctaag	424
aatgtgctgtttatttctctgcattaatactgattaacttcatattctgtagataatgga		Exon 12	1404
gtetgaagett 400 bpgaatteatgeteeteetgetgeeteetgggtgeta agttgeaaggtgatageteaggtaeteaetetagtgtettteettggggtgetettteea aggeetetetaeatattaageeacaaaggagtetgttgeceeteaagaggatgagatga		Val Phe Ser Thr Al	1424 429
gaatattaggetacagtttigttgeetittittittättiteetaacaiggtaceacaaitg Exon 6	1	T GCT GAT GGA CAA ACT CAA GTA GAG ATT AAA GTG TGT CAG GGG GA a Ala Asp Gly Gln Thr Gln Val Glu Ile Lys Val Cys Gln Gly Gl	1469 444
aattttattetttgttteecag AA AAT TAC TTG GGC CAC ACA GCA AAA AA lu Asn Tyr Lau Gly His Thr Ala Lys As	563 142	A CGA GAG ATG GCT GGA GAC AAA CTT CTA GGA CAG TTC ACT TT u Arg Glu Met Ala Gly Asp Asn Lys Leu Leu Gly Gln Phe Thr Le	1514 459
T GCT GTG ATC ACA GTC CCT GCT TAT TTC AAT GAT TCA CAG CGA CA n Ala Val lle Thr Val Pro Ala Tyr Phe Asn Asp Ser Gln Arg Gl	608 157	G gtaagtgtttttgagtctgagtatcatgcttttggggtctatagcttgcaaagctcca u	1515 459
G gtaaaattagatctcttgtttgctgggagtggagtggggtacctgagttaaaggatg	609 157	<pre>aactgetgacattacaggcatattgtgttattttttaaaaagaacgttatgtacatgagt tatgaaacccatgattttagttttttacctaaagtgctttgtgtttttcagaatttgaaat</pre>	

sus GT/AG sequence at the donor and acceptor sites of RNA splicing [25]. The sizes of introns range from 88 to approximately 2,500 nucleotides. The largest exon (exon 17) encodes the carboxyl-terminal region of PBP74/CSA and the 3'-untranslated region.

```
Exon 13
                                                                    ATT GGA A
Ile Gly I
 TT CCC CCA GCC CCT CGT GGA GTG CCC CAG ATT GAA GTT ACA TTT G le Pro Pro Ala Pro Arg Gly Val Pro Gln Ile Glu Val Thr Phe A
                                                                                   1567
477
AC ATT GAT GCC AAT GGG ATT GTG CAC GTT TCT GCC AAA GAT AAA sp Ile Asp Ala Asn Gly Ile Val His Val Ser Ala Lys Asp Lys
GC ACT GGT CGT GAG CAA CAG A gtaagtaatcagactgtaaatgatctgtcca ly Thr Gly Arg Glu Gln Gln I
gtt... 780 bp ...dtgtagdtatggdgdddgadtdtgatdtgddtgdtttdtaa
tgdtgtaatcaaaggtgtgtacdttdatacdtagtgaagtaatttddcagtgcagtaag
                                                    Exon 14
 agaactaaagagctgcctttttttttttttaactacag
T GGT GGA TTA AGC AAA GAT GAT ATT GAA AAT ATG GTT AAA AAT GC r Gly Gly Leu Ser Lys Asp Asp Ile Glu Asn Met Val Lys Asn Al
A GAG AAG TAC GCT GAG GAA GAC CGC AGG AAG AAG gtgattacttttge a Glu Lys Tyr Ala Glu Glu Asp Arg Lys Lys
tgacagtagagtttggttacagatacccagtctctctcacctctatctctatatgctt
tgtttaatgtgagattacactatgtagtcagggtgtttgagacaacctccctagtacctg
ggatccagggcacatactgggtttgggatgatctctttcgtcttccactatcaaacttc
catgtcacaaaagccggctcaactctcagacatgccaagtactcagctaagctaactgtg
tggcaagtaacttctcaattgttactgaagttaatgaaactttggttgattaacttttct
      GAA CGT GTT GAA GCA GTT AAT ATG GCT GAA GGA ATT ATT CAT Glu Arg Val Glu Ala Val Asn Met Ala Glu Gly Ile Ile Bis
GAC ACA GAA ACC AAG ATG GAA GAA TIT AAG GAC CAG TTG CCT GCT
Asp Thr Glu Thr Lys Met Glu Glu Phe Lys Asp Gln Leu Pro Ala
GAT GAG gtaccattcactggttctttgaagctcatttttcttgggttggtagtgattcAap Glu
agtttettaaacatatgaataagggttggaggtggagaggcaaaaaccatttagaaatet
aaatagtagatgtaatagtaaatetaaatgtaatgagacaagatacetagattaaaaaga
                                                Exon 16
                                                TGC AAC AAG CTA AAG GAA
Cys Asn Lys Leu Lys Glu
tgtgtctgtgtatctgtgtgttttgccgtttccag
ACA GGA GAG AAC ATC AGG CAG GCA GCA TCT TCC CTA CAG CAG GCG
Thr Gly Glu Asn Ile Arg Gln Ala Ala Ser Ser Leu Gln Gln Ala
TCA TTG AAA CTC TTC GAA ATG GCG TAC AAA AAG gtacaagggctggagg
Ser Leu Lys Leu Phe Glu Met Ala Tyr Lys Lys
                                                                                  1962
608
  {\tt gaacttgtggatacctgtgaaatcttagtgatcctcaagtattttctcaacag} \begin{array}{l} {\tt ATG} \ \ {\tt GC} \\ {\tt Met} \ \ {\tt Al} \end{array}
A TCT GAA CGG GAA GGT TCT GGA AGT TCT GGC ACT GGG GAA CAG AA a Ser Glu Arg Glu Gly Ser Gly Ser Ser Gly Thr Gly Glu Gln Ly
                                                                                  2012
G GAA GAT CAG AAG GAA GAG AAA CAG TAATCGTGGCAGTGCATTGTGGAGCC s Glu Asp Gln Lys Glu Glu Lys Gln ***
                                                                                  2063
633
AGAAGGACATACTATGAAGCTTGGGACTAAAGGGACTTCCTGAGCAGAAAAGGGGCAGAC
                                                                                  2123
2183
TAGTGACAATTGCTAACTCATTTAATGGGTAATAAAGTCAGCAATAGCAGGTTCATACTG
                                                                                  2243
TTCTGTCACTAGCCTGTTATTTTCAGCTGCATGTAAAGGGGTGGGATGGGGCTGTGAACC
                                                                                 2303
AATCATTAAGGTAGATTTGGTTTGTGCTGAAATGGCTGTGATTTCAAGGTGGGAAGCCCA
TTTCACATGCAGTGGAGGTAGTCTGTCATTGACCTTGAATTGAGATCATATGCAGATGCT
                                                                                 2423
TGTTGGCCAAGAGCACTACTATAAAGAATGACCTCTGTATATTTGCTCCTACAACTAATG
                                                                                 2483
CCTTTAAGACTGAGCTACCTGTACCATGGTCTGTAGGTGCAGAAGCTAGGTCAGTGGATA
                                                                                 2543
      TTGTGTTAGCCATAGCTTAAAGTATGATATGAGAATGATATAAGCCTCTCATGGGC
                                                                                 2603
{\tt CTGAGGCATACTTCTCTAGCCACCCTCTTGGTTGGCCAATGTCTGGCATCTGTATTCTTG}
2723
TCTCCT TGTGAATGGTGATAGCTCAAGGATTATGACTGCTATCAGTTTTGTAGGGAGAA
                                                                                 2783
AAATCACTGGCTAAAAGGTTGAACAAATGAAACATGGGGAGTGACTAATAAAATGCTGGC
                                                                                 2843
GTTAGACTCCGTCTTAATAAAAgtactggcataaacaaatccagtcgttgtaatcttcta 2925
```

totagaaaaacttcagtgctttt

PBP74/CSA possesses a 46-residue leader peptide at the amino-terminus that probably functions as a signal necessary for targeting and transport of the PBP74/CSA precursor into mitochondria [2,3]. As shown in Fig. 3, the exon-intron organization within the amino-terminal leader sequence in PBP74/CSA is highly similar to that of the corresponding region in human cytochrome c_1 , an inner-membrane protein protruding into the intermembrane space of the mitochondria [15]. The phase class of first intron, which is located at the central part of the amino-terminal leader sequence, is in phase 0 for PBP74/CSA and cytochrome c_1 . The carboxyl-terminal sequence (Arg-Pro-Gln and Arg-Thr-Pro-Gln in PBP74/CSA and cytochrome c_1 , respectively), which is encoded by the first exon, is highly conserved in these two proteins. Moreover, the locations of positively charged residues are similar among PBP74/CSA, cytochrome c_1 , and Ssc1p [26], located in the mitochondria of yeast. However, the leader sequence encoded by the second exon in PBP74/CSA is clearly distinguishable from that in cytochrome c_1 by the absence of a highly hydrophobic stretch (data not shown). These results suggest that PBP74/CSA is not located in the inner-membrane of the mitochondria.

4. DISCUSSION

In this study, we have characterized a mouse genomic region containing the PBP74/CSA gene, including approximately 700 bp of the 5'-flanking region. Sequence analysis shows that this gene is separated into 17 exons by 16 introns (Figs. 1 and 2). The transcription initiation site of the PBP74/CSA gene is not clear at the present time. However, an additional exon(s) is unlikely to be present in the 5'-flanking region since the combined size of the 17 exons is close to the length of the gene transcript (3.1 kb) determined by Northern blot analysis [2]. Two dimeric repeats of the consensus sequence of HSE [15,16] are located in the putative promoter region of the gene (Fig. 2). Recently, the interaction of recombinant mouse heat-shock transcription factors 1 and 2 with the HSE has been characterized; the binding of these two factors to dimeric HSEs is weak, and stable binding occurs when at least three adjacent sites of the HSE are present [27]. Thus, identification of

Fig. 2. DNA sequence of the mouse PBP74/CSA gene. The sequences encoded by exons are given by capital letters. The amino acid sequence is shown below the nucleotide sequence numbered in the 5'- to 3'-direction. Only the nucleotides in exons are numbered from a translation initiation codon, ATG, and the nucleotides in the 5'-end region from the ATG codon are indicated by negative numbers. The polyadenylation signal sequence, AATAAA, is underlined with wavy lines. The consensus sequences for potential regulatory elements, Mt1 [19,20], CCAAT [41], NRF-1 [21,22], MyoD [23,24], TCF-1 [42,43], HSE [15,16], F-ACT1 [44], Mt2 [19], Spl [45], and Myc [46,47], are also underlined.

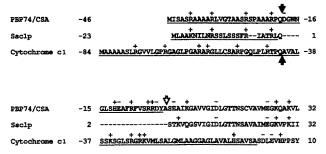


Fig. 3. Comparison of intron insertion in the amino-terminal leader sequences of PBP74/CSA and cytochrome c_1 . The amino-terminal leader sequences of PBP74/CSA [2,3], human cytochrome c_1 [15], and yeast Ssc1p [26] are underlined, and the positively and negatively charged residues are also indicated, 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 nuclearencoded 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

- 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., Kosic-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.