Volume 333, number 1,2, 51–55 FEBS 13135 October 1993

Cloning of CRP2, a novel member of the cysteine-rich protein family with two repeats of an unusual LIM/double zinc-finger motif

Ichiro Okano[a], Takeshi Yamamoto[b], Akira Kaji[c], Toru Kimura[a], Kensaku Mizuno[a,*], Toshikazu Nakamura[b]

[a]*Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan*
[b]*Biomedical Research Center, Osaka University Medical School, Suita, Osaka 565, Japan*
[c]*Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6076, USA*

Received 3 September 1993

The cDNA coding for a novel member of the cysteine-rich protein family was isolated from a rat brain cDNA library. It encodes a protein, denoted cysteine-rich protein 2 (CRP2), of 208 amino acid residues containing two tandem repeats of an unusual LIM/double zinc-finger-like motif. The ubiquitous tissue distribution and high level of expression of CRP2 mRNA suggest an important role for CRP2 in cell functions.

Cysteine-rich protein; CRP2; Zinc-finger; LIM motif; ESP1

1. INTRODUCTION

The zinc finger motif, initially found in *Xenopus laevis* transcription factor TFIIIA and then in many gene-regulating and metal-binding proteins, is a small, discrete structural domain with a characteristic placement of cysteine and histidine residues for zinc ion coordination [1–3]. Most of the zinc finger proteins so far characterized function as transcriptional regulators by binding to specific DNA sequences. NMR and X-ray crystallographic studies revealed the three dimensional structure of the zinc finger–DNA complex [3]. Zinc finger proteins are grouped into several classes, such as C_2-H_2, C_2-C_2, and C_x class, according to the number and placement of cysteine and histidine residues [1–3].

Cysteine-rich protein (CRP) was found to have the sequence of two tandem copies of an unusual double zinc finger motif, ($CXXCX_{17}HXXC$)XX ($CXXCX_{17}C$-XXC) [4,5]. Cysteine-rich intestinal protein (CRIP), cloned from weanling rodent intestine, contains a single copy of such a double zinc finger motif, ($CXXCX_{17}H$-XXC)XX($CXXCX_{17}CXHXC$) [6]. Similar motifs were also observed in the LIM consensus motif, named after *lin-11*, *Isl-1*, and *mec-3* homeodomain-containing proteins [7–9], which have two repeats of a double zinc finger domain ($CXXCX_{17-19}$ HXXC)XX($CXXCX_{16-20}$ CXXC/H/D) at the N terminus. These proteins were recently grouped together into a class referred to as the LIM/double finger family, because of the presence of a paired zinc-finger motif characteristically separated by a two-amino-acid linker [5].

Here we report the isolation and sequence of a cDNA encoding a novel member of LIM/double finger protein family, which we term CRP2. It contains two repeats of the sequence, ($CXXCX_{17}HXXC$)XX($CXXCX_{17}CHXX$-C), which is highly homologous to CRP and CRIP, but distinct from these in the spacing of the C and H in the latter finger. Wide tissue distribution of CRP2 mRNA suggests an important role for CRP2 in cell functions.

2. MATERIALS AND METHODS

2.1. *RNA preparation*

Total RNA was isolated by the acid–guanidinium thiocyanate–phenol–chloroform extraction method [10]. Poly(A)$^+$ RNA was prepared using Oligotex-d(T) 30 (Roche, Tokyo) according to the manufacturer's instructions.

2.2. *Polymerase chain reaction (PCR) amplification*

The first strand of cDNA was synthesized with random hexanucleotide primers on poly(A)$^+$ RNAs (2.5 ng) purified from adult male Wistar rat tissues by reverse transcriptase. The single-strand cDNA was initially amplified between the sense primer Q1 (5′-GCCTCTA-GACATGG(A/G)CCCTGGTGCTAC-3′) and a mixture of antisense primers, Q2 (5′-CCGAAGCTTGCCACCATAATCCCCCTC-3′) and Q3 (5′-CCGAAGCTTGCCCCCGTAGTCACCCTC-3′). Amplified DNAs were size-fractionated on 1% agarose gels. Amplified bands of about 700 bp were excised and used as templates for a 2nd PCR reaction, which was performed with the sense primer Q1 and the

*Corresponding author. Fax: (81) (92) 632 2741.

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank Nucleotide Sequence Databases with the accession number D17512.

Abbreviations: CRP, cysteine-rich protein; CRIP, cysteine-rich intestinal protein; ESP1, estradiol-stimulated protein; HGF, hepatocyte growth factor; HLP, HGF-like protein; GAPDH, glyceraldehyde phosphate dehydrogenase; LIM, double zinc finger motif named after *lin-11*, *Isl-1* and *mec-3* homeoproteins.

antisense primer Q5 (5′-AGCTTGAG(C/T)A(A/G)AAC(A/C)A(A/G)(A/C)T-3′). The primers correspond to the amino acid sequences HGPWCY (human hepatocyte growth factor (HGF) (448–453), for Q1), EGDYGG (human HGF (670–675), for Q2 and Q3), and (D/Q)LVL(L/M)KL (human HGF (578–584), for Q5), which are highly conserved sequences shared by human and rat HGF [11–15] and human and mouse HGF-like protein (HLP) [16,17]. PCR was carried out in a DNA thermal cycler (Perkin-Elmer-Cetus) for 40 cycles of denaturation (94°C for 1 min), annealing (53°C for 2 min), and extension (72°C for 2 min) with 1 unit of Taq DNA polymerase. The amplified DNA fragment of about 500 bp (clone B6) was ligated into the TA cloning vector (pCR II) (Invitrogen, San Diego), and sequenced by the dideoxynucleotide chain termination method using a Taq dye primer cycle sequencing kit with 370A DNA sequencer (Applied Biosystems).

2.3. *Construction and screening of a cDNA library*

An oligo(dT)-primed rat brain cDNA library was constructed in λ-ZAP II vector (Stratagene) and screened using ^{32}P-labeled B6 as a probe. Hybridization was performed for 48 h at 42°C in a mixture of 43% formamide, 5 × SSPE (1 × SSPE: 150 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.4), 5 × Denhardt's solution and 50 μg/ml of salmon sperm DNA. The filter was washed twice with 2 × SSPE containing 0.5% SDS at 65°C, and once with 0.2 × SSPE containing 0.1% SDS at 65°C. The filter was autoradiographed at −80°C with an intensifying screen for 24 h. Positive clones were plaque- purified and were allowed in vivo excision and recirculization of pBluescript from λ-ZAP II. The cDNA insert was sequenced on both strands by the dideoxy chain termination method described above.

2.4. *Northern blot hybridization*

Total RNAs (10–20 μg) isolated from rat tissues were denatured with formaldehyde, electrophoresed on 1% agarose/0.7% formaldehyde gel, and transferred to Hybond-N filter (Amersham). The filter was hybridized with ^{32}P-labeled B6 or the control cDNA probe for rat glyceraldehyde phosphate dehydrogenase (GAPDH), and washed under the conditions described above.

3. RESULTS AND DISCUSSION

The initial objective of this study was to isolate clones coding for proteins related to the hepatocyte growth factor (HGF) and HGF-like protein (HLP) [11–17]. cDNAs derived from various rat tissues were subjected to two-step PCR amplification, using primers corresponding to the consensus sequences of HGF and HLP. In addition to the cDNAs coding for rat HGF and HLP, a cDNA of about 500 bp (B6) was amplified from rat brain, pituitary and kidney cDNA templates. Nucleotide sequence analysis of B6 revealed that this cDNA contained a 478-bp fragment, including a 234-bp open reading frame coding for a zinc finger-like motif in the 5′-terminal half (Fig. 1). By screening approximately 3×10^5 phage plaques from a rat brain cDNA library using B6 as a probe, 30 positive plaques were identified. The cDNA clone with a longest insert of about 1.2 kb (named T5) was subjected to DNA sequencing (Fig. 1).

Fig. 2 shows the nucleotide sequence and the predicted amino acid sequence of T5. The 1,170-bp insert contains a 624-bp open reading frame corresponding to an encoded protein of 208 amino acid residues with a calculated molecular weight of 22,695. The initiation codon was assigned to the first in-frame ATG at nucleotides 57–59 which reside in the Kozak consensus sequence, ACCATGG [18]. The coding sequence ends with a TAG stop codon at nucleotides 681–683. The 3′-untranslated region contains a polyadenylation signal, AATAAA, beginning 26 bases upstream of a poly(A) tail. There is an ATTTA sequence at nucleotides 984–988, which is known to be the recognition signal for mRNA instability [19]. This cDNA may have evolved by gene duplication of an ancestral gene as there are two tandem repeats of a highly homologous (82% identical) nucleotide sequence in 66–299 and 429–662 (underlined in Fig. 2).

The deduced amino acid sequence contains four zinc finger-like domains consisting of the characteristic C_2-HC (or C_2-C_2) placement of cysteine and histidine resi-

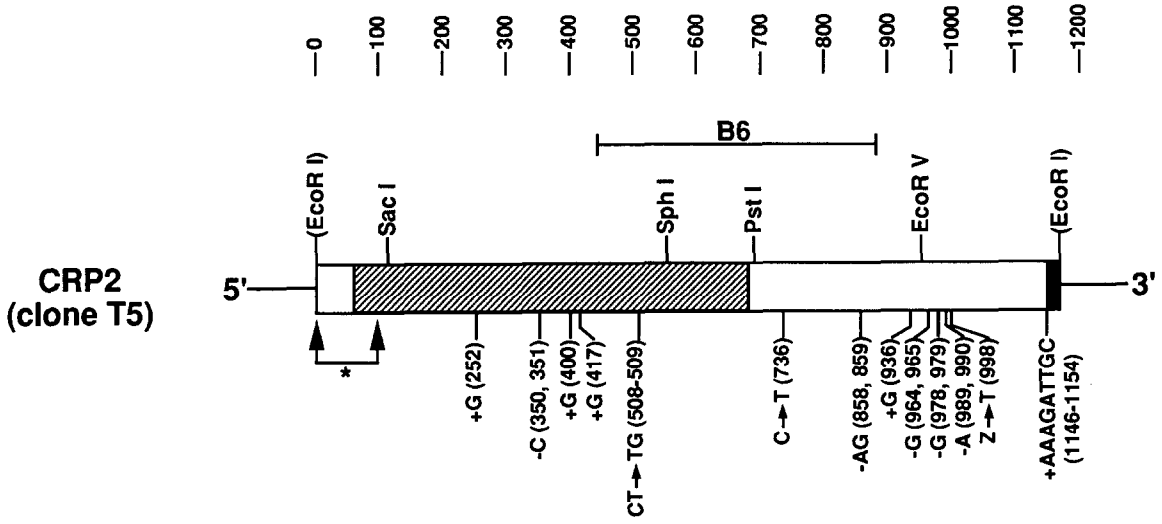

Fig. 1. Structure of the cDNA clone (T5) encoding rat CRP2. The shaded and black boxes indicate the protein-coding region and poly(A) tail, respectively. The open box indicates the 5′- and 3′-noncoding region. The positions of nucleotide insertion, deletion, and substitution in CRP2, compared with ESP1, are indicated. The asterisk denotes the region with no nucleotide sequence homology with ESP1.

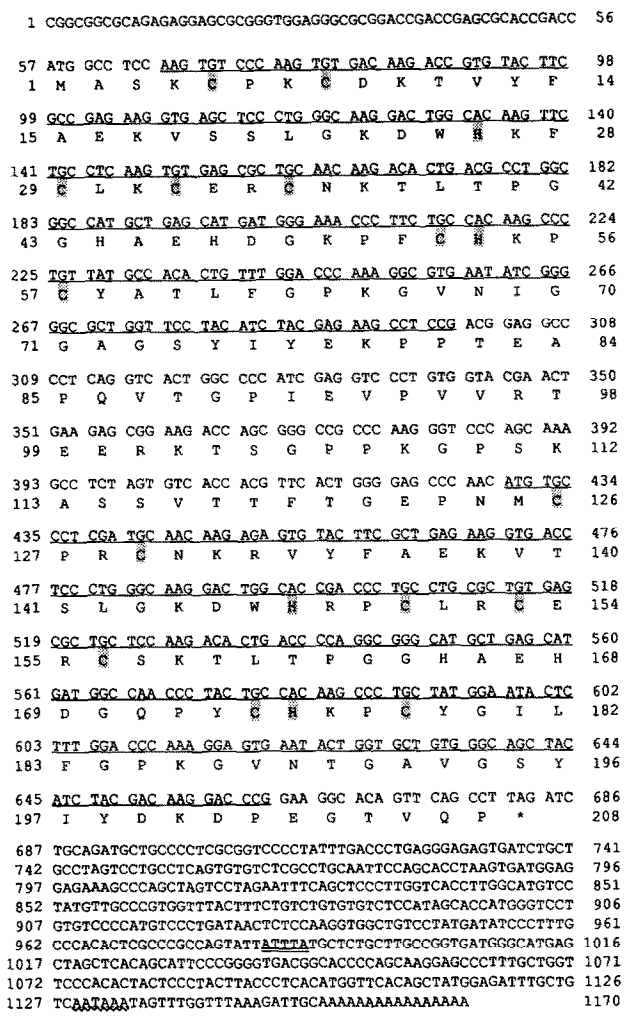

Fig. 2. Nucleotide sequence and deduced amino acid sequence of rat CRP2. Nucleotides and amino acid residues are numbered on the left and right. A single underline represents the regions of repeated sequence. An asterisk denotes the termination codon. The conserved cysteine and histidine residues of the LIM/double finger motif are shaded. A polyadenylation signal is indicated by a wavy line. The putative signal for mRNA instability is doubly underlined.

dues (Fig. 3). The first two zinc fingers and the remaining two constitute a common unusual paired zinc finger motif separated by a two-amino-acid linker, (CXXC-X_{17}HXXC)XX(CXXCX_{17}CHXXC). Reflecting on the gene duplicated structure of this cDNA, the amino acid sequences of (4–81) and (125–202), including a paired finger motif and the C-terminal stretch outside the finger, are significantly homologous and 60 out of 78 amino acid residues are identical (Fig. 4). Among the four finger motifs, the numbers of identical residues between the first and third finger (19/25) and the second and fourth finger (23/26) are significantly higher than that between the paired first and second (9/26) or third and fourth finger (8/26).

Comparison of the nucleotide and deduced amino acid sequence of T5 with GenBank, EMBL, NBRF and SwissProt databases revealed a clear homology with those of rat cysteine-rich intestinal protein (rCRIP) [6] and human cysteine-rich protein (hCRP) [4]. Therefore, we named the T5-encoded protein rat cysteine-rich protein 2 (rCRP2). rCRP2 also has a lesser but distinct homology to the LIM motif-containing proteins, *lin-11*, *Isl-1*, *mec-3*, and rhombotin [7–9,20]. Alignment of two repeated sequences of rCRP2 with the sequence of rCRIP, and the repeated sequences of hCRP and other LIM motif-containing proteins (Fig. 4) clearly shows that rCRP2 is a member of the 'LIM/double finger' motif proteins, defined as unique paired zinc fingers with a two-amino-acid linker [5]. Among these proteins CRP2 is most similar to CRIP (77% identity in the motif) and CRP (43–47% identity in the motif), but the spacing of the Cys and His residues in the 'knuckle' of the second and fourth finger of CRP2, (CXXCX_{17}-CHXXC), differs from that of CRIP (CXXCX_{17}-CXHXC) or CRP (CXXCX_{17}CXXC).

We also found that the overall nucleotide sequence, except for the 5′-terminal 98 nucleotides, is closely related to the reported sequence for rat estradiol-stimulated protein (ESP1), which was cloned from rat C6 glioma cells as the 17β-estradiol responsive cDNA [21]. (This sequence has not been registered in the above databases.) The sequence from nucleotide 99 to the 3′-poly(A) tail of rCRP2 is almost identical to nucleotide −88 to poly(A) tail of ESP1, but there are 13 nucleotide insertions (including a 9 nucleotide insertion adjacent to poly(A) tail), 6 nucleotide deletions, and 4 nucleotide substitutions in CRP2, compared with the sequence of ESP1 (Fig. 1). In particular, three insertions and one deletion in the coding region caused a shift of the reading frame, the result being a total difference in the deduced amino acid sequences between CRP2 (1–121) and ESP1 (1–77) as well as a difference in the translational initiation site (Fig. 1). The 5′-terminal 175 nucleotide sequence of ESP1 (−264 to −89) has no sequence homology with the 5′-terminal 98 nucleotides of CRP2. Thus, it is likely that the cDNAs coding for CRP2 and ESP1 are derived from two highly related genes. The relation between these cDNAs is not clear at present. Wang et al. predicted without any experimental data the presence of two copies of LIM/double finger motif for ESP1 by assuming a one-base deletion at an adequate site of the reported ESP1 sequence [5]. However, the sequence of CRP2 determined in this study differs from their predicted sequence for ESP1.

To determine the size of rCRP2 mRNA and its tissue distribution, we carried out Northern blot analysis using B6 as a probe. As shown in Fig. 5A, a single 1.2-kb mRNA was detected as the major band in various rat tissues. The highest level of expression was observed in heart and lung, a moderate level in liver, stomach, submaxillary gland, pituitary, brain, spleen and kidney, and a lower level in thymus, adrenal gland, and small and large intestine. The wide tissue distribution

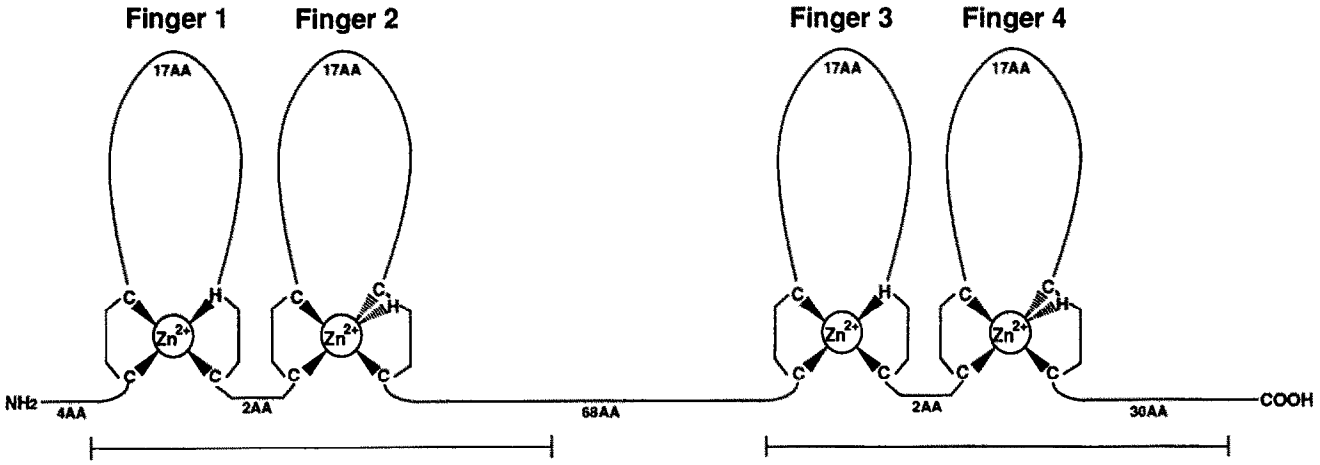

Fig. 3. Schematic representation of the predicted structure of the rCRP2 protein. Four putative zinc finger configurations are illustrated in such a way that the conserved cysteine and histidine residues coordinate zinc ion. Two tandem repeats of the homologous sequence are indicated by a bar.

and high level of expression of CRP2 mRNA indicate that CRP2 has a fundamental role in cellular function. It should be noted that in case of the testis B6 probe hybridized with a band of 2.2-kb mRNA. To characterize the cDNA for testis-specific 2.2-kb mRNA, cloning and sequencing is presently carried out in our laboratory.

The ESP1 cDNA was isolated by differential hybrydization as the 17β-estradiol response gene in rat C6 glioma cells [21]. ESP1 mRNA expression was reported to be sexually dimorphic in most body tissues but not in brain [21]. To compare our results with those of ESP1, we examined the expression of CRP2 mRNA in tissues of adult rats of either sex. As shown in Fig. 5B, CRP2 mRNA is expressed more abundantly in liver and kidney of females, while there is no significant difference between sexes regarding expression in brain, lung and heart. Thus, sexual dimorphism in CRP2 mRNA expression was found only in liver and kidney. Whether or not expression of CRP2 mRNA is regulated by estradiol will need to be determined in studies on the regulation of gene expression of CRP2 using estradiol-responsive cells.

The physiological functions of previously identified cysteine-rich proteins, CRIP and CRP, are not clearly understood. CRIP was isolated from rat intestinal mucosa and was suggested to have a role in intestinal zinc absorption [6,22]. CRP was thought to be a primary response gene induced together with *c-myc* in the cell as it proceeds from G_0 to G_1 or to the S phase of the cell cycle [5]. The sequence similarity suggests that CRP2 may have functions related to those of CRIP and CRP. The presence of the AUUUA sequence in the 3′-noncoding region of CRP2 mRNA indicates that it is the primary response gene regulated in a manner similar to CRP and other short-lived mRNAs [19]. To determine the physiological functions of CRP2, we are examining the regulation of CRP2 gene expression, the metal-

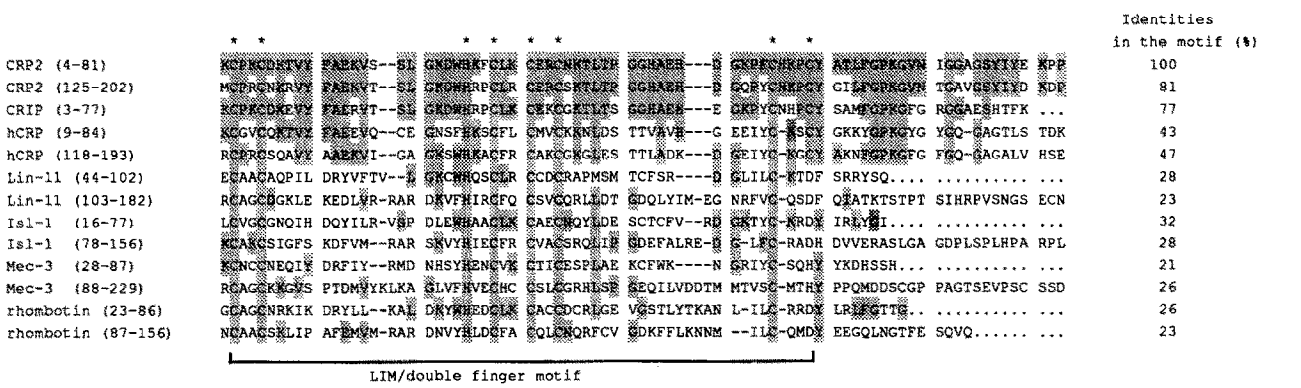

Fig. 4. Comparison of the repeated sequences of rCRP2 with other LIM/double finger proteins. Residues identical with the rCRP2 (4–81) are shaded. The number on the right represents the percentage of identical residues found between the first LIM/double finger motif of rCRP2 and the others. Asterisks are indication of the positions of conserved cysteine and histidine residues.

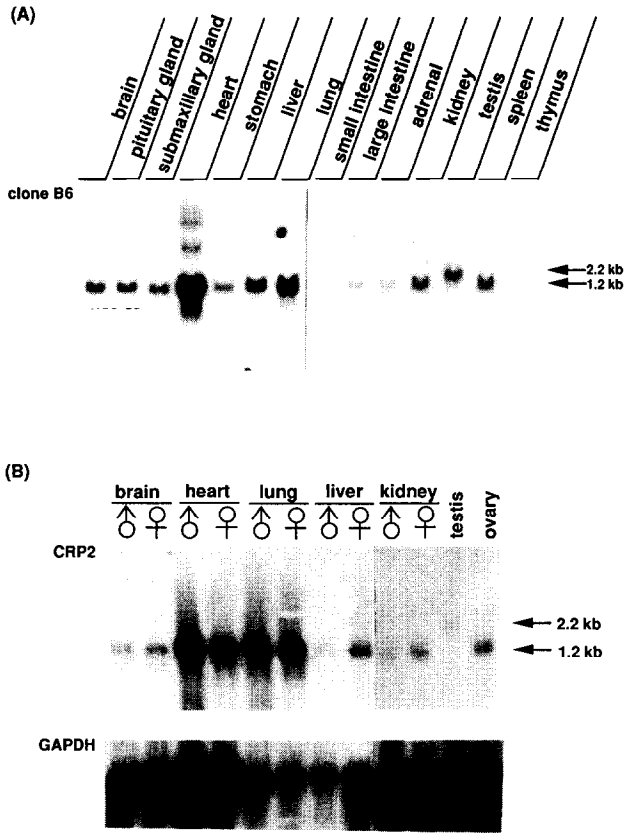

Fig. 5. Northern blot analysis of CRP2 mRNA expression in rat tissues. (A) Total RNA (20 μg) from various tissues of adult male rat was electrophoresed, transferred, and hybridized with ^{32}P-labeled B6 fragment of rCRP2. The blot was exposed at −70°C for 3 days. (B) Total RNA (10 μg) from tissues of adult male and female rats was analyzed as in (A), except that the full-length cDNA of rCRP2 in place of B6 was used as a probe. The same blot was rehybridized with ^{32}P-labeled cDNA for rat GAPDH as a control (lower panel).

chelating and/or DNA-binding ability of the recombinant CRP2 protein, and the subcellular localization of this protein using a specific antibody.

Acknowledgements: We thank M. Ohara for helpful comments. This work was supported by Research Grants for Science and Cancer from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Berg, J.M. (1990) J. Biol. Chem. 265, 6513–6516.
[2] Evans, R.M. and Hollenberg, S.M. (1988) Cell 52, 1–3.
[3] Harrison, S.C. (1991) Nature 353, 715–719.
[4] Liebhaber, S.A., Emery, J.G., Urbanek, M., Wang, X. and Cooke, N.E. (1990) Nucleic Acids Res. 18, 3871–3879.
[5] Wang, X., Lee, G., Liebhaber, S.A. and Cooke, N.E. (1992) J. Biol. Chem. 267, 9176–9184.
[6] Birkenmeier, E.H. and Gordon J.I. (1986) Proc. Natl. Acad. Sci. USA 83, 2516–2520.
[7] Freyd, G., Kim, S.K. and Horvitz, H.R. (1990) Nature 344, 876–879.
[8] Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. and Edlund, T. (1990) Nature 344, 879–882.
[9] Way, J.C. and Chalfie, M. (1988) Cell 54, 5–16.
[10] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[11] Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugiyama, A., Tashiro, K. and Shimizu, S. (1989) Nature 342, 440–443.
[12] Seki, T., Ihara, I., Sugimura, A., Shimonishi, M., Nishizawa, T., Asami, O., Hagiya, M., Nakamura, T. and Shimizu, S. (1990) Biochem. Biophys. Res. Commun. 172, 321–327.
[13] Tashiro, K., Hagiya, M., Nishizawa, T., Seki, T., Shimonishi, M., Shimizu, S. and Nakamura, T. (1990) Proc. Natl. Acad. Sci. USA 87, 3200–3204.
[14] Nakamura, T. (1991) Progr. Growth Factor Res. 3, 67–85.
[15] Mizuno, K. and Nakamura, T. (1993) in: Hepatocyte Growth Factor-Scatter Factor and the c-Met Receptor (Goldberg, I.D. and Rosen, E.M., Eds.) pp. 1–29, Birkhauser, Basel, Boston.
[16] Han, S., Stuart, L.A. and Degen, S.J.F. (1991) Biochemistry 30, 9768–9780.
[17] Degen, S.J.F., Stuart, L.A., Han, S. and Jamison, C.S. (1991) Biochemistry 30, 9781–9791.
[18] Kozak, M. (1989) J. Cell Biol. 108, 229–241.
[19] Shaw, G. and Kamen, R. (1986) Cell, 46, 659–667.
[20] McGuire, E., Hockett, R.D., Pollock, K.M., Bartholdi, M.F., O'Brien, S.J. and Korsmeyer, S.J. (1989) Mol. Cell. Biol. 9, 2124–2132.
[21] Nalik, P., Panayotova-Heiermann, M. and Pong, O. (1989) Mol. Cell. Endocrinol. 62, 235–242.
[22] Hempe, J.M. and Cousins, R.J. (1991) Proc. Natl. Acad. Sci. USA 88, 9671–9674.