Nucleic Acids Research, 1996, Vol. 24, No. 19 **3821–3828**

The mouse poly (C)-binding protein exists in multiple isoforms and interacts with several RNA-binding proteins

Birgit Funke[1], Brigitte Zuleger[1], Ricardo Benavente[2], Tillman Schuster[3], Martin Goller[1], James Stévenin[4] and Ivan Horak[1,5,*]

[1]Institut für Virologie und Immunbiologie der Universität Würzburg, Germany, [2]Lehrstuhl für Zell- und Entwicklungsbiologie, Theodor Boveri Institut für Biowissenschaften, Würzburg, Germany, [3]Institut für Medizinische Strahlenkunde und Zellforschung der Universität Würzburg, Würzburg, Germany, [4]Institut de Génétique et de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg Cedex, France and [5]Forschungsinstitut für Molekulare Pharmakologie, Krahmerstraße 6, 12207 Berlin, Germany

Received May 20, 1996; Revised and Accepted August 15, 1996

DDBJ/EMBL/GenBank accession no. X75947

ABSTRACT

The murine poly(C)-binding protein (mCBP) was previously shown to belong to the group of K-homology (KH) proteins by virtue of its homology to hnRNP-K. We have isolated cDNA-splice variants of mCBP which differ by two variable regions of 93 bp and/or 39 ± 3 bp respectively. Both variable regions are located between the second and third KH-domain of mCBP. The characterization of a partial genomic clone enabled us to propose a model for the generation of the second variable region by the use of a putative alternative splice signal. The mCBP mRNA is expressed ubiquitously and the protein is found predominantly in the nucleus with the exception of the nucleoli. We have identified five proteins which interact with mCBP in the yeast two hybrid system: mouse y-box protein 1 (msy-1), y-box-binding protein, hnRNP-L, filamin and splicing factor 9G8. The interaction between mCBP and splicing factor 9G8 was confirmed *in vivo*. These results suggest a function of mCBP in RNA metabolism.

INTRODUCTION

We have previously described the purification and characterization of the single-stranded (ss) DNA binding protein mCBP (murine poly(C)-binding protein) which was initially identified by virtue of its binding to a recombination hotspot in murine retrotransposons (1). mCBP was shown to bind preferentially to poly(dC) and, to a lesser extent, to poly(rC) *in vitro*. Cloning and characterization of one mCBP cDNA revealed a similarity to the hnRNP-K protein (2). Both mCBP and hnRNP-K contain three KH (K-Homology) domains which were first described for the hnRNP-K protein (3). Furthermore, the KH domains are similarly distributed in both proteins. However, the homology is restricted to the KH motifs, there is only an overall homology of 36% at amino acid level between mCBP and hnRNP-K (2).

Several mCBP homologues have been isolated by other investigators. A murine cDNA variant differing by 31 amino acids was described as hnRNP-X (4). Human homologues have been isolated by several groups (5,6). One of them, αCP-2 (α complex protein 2), was identified as a component of a multiprotein complex stabilizing the human α-globin-mRNA (6). All of these proteins show a strong poly(C)-binding activity.

Interestingly, hnRNP-K also binds to poly(dC) and poly(rC) *in vitro*. It was described as the major C-binding activity in HeLa cells (7). hnRNP-K was identified as a component of the hnRNP complexes which assemble on nascent hnRNA concurrent with transcription and which are thought to be involved in RNA processing and RNA transport (8). Immunohistochemical studies revealed that hnRNP proteins are predominantly localized in the nucleus (9) although some of them were also detected in the cytoplasm where they are involved in mRNA stabilization (10). All members of the hnRNP family which have been characterized so far bind specifically to RNA as well as to ssDNA *in vitro* (11). Binding was shown to be mediated by a RNA recognition motif (RRM) found in most hnRNP proteins (12).

Although hnRNP-K is found in hnRNP complexes, it is not a typical member of the hnRNP family as it lacks the classical RNA recognition motif. Binding of the hnRNP-K protein to single-stranded nucleic acids is mediated by its three KH domains (13).

KH domains appear to be highly conserved in evolution. They are found in many species from *E.coli* to mammals (14–16) including the human FMR1 (fragile X mental retardation) gene product which is associated with the Fragile-X-Syndrome (13). A reported mutation in one of the KH domains of the FMR1-protein impairs its RNA binding activity and results in mental retardation (13,14). Other examples of KH proteins include Nova, a neuronal protein which is recognized by an autoantibody of breast and lung cancer patients who develop a disorder of motor control (17) and gld-1, a tumor suppressor gene identified

*To whom correspondence should be addressed at: Forschungsinstitut für Molekulare Pharmakologie, Krahmerstraße 6, 12207 Berlin, Germany

in *Caenorhabditis elegans* (18). A common feature of many KH proteins is their ability to bind RNA and ssDNA *in vitro* (2,7,13,19). Altogether, there is increasing evidence that proteins containing KH domains are involved in important regulatory mechanisms which in some cases are associated with human diseases.

hnRNP-K exists as a group of highly related isoforms varying by small insertions of five and/or 20 amino acids, respectively (20). These variants are likely to be generated by alternative splicing of a single primary transcript—a phenomenon which seems to be a common theme among hnRNP proteins (21,22). Surprisingly, only one of the hnRNP-K variants was shown to act as a transcriptional activator of the human c-*myc* gene by binding in a sequence specific manner to a regulatory element upstream of the c-*myc* promotor P1 (23,24). Furthermore, hnRNP-K has been shown to interact directly with the SH3 domains of s-src, lyn, fyn and the vav protooncogene product and thus may be involved in translating cell surface receptor signals into regulatory events on the level of gene expression and RNA processing (25–27).

At present, not much is known about the function of mCBP *in vivo*. In the present study, we describe the isolation of mCBP binding proteins by the yeast two hybrid system as a step to elucidate the role of mCBP *in vivo*. Furthermore we report the isolation and characterization of several novel mCBP splice variants.

MATERIALS AND METHODS

Isolation of mCBP splice variants

A partial mCBP cDNA was isolated by RACE (rapid amplification of cDNA ends) with degenerated primers based on the mCBP N-terminal amino acid sequence as described (2). The amplified product was subcloned into pBluescript II KS +/– and was used to screen a cDNA library (mouse liver/λ ZAP vector, Stratagene) to obtain full length clones.

DNA sequencing

Nucleotide sequences were determined by double-stranded sequencing using the T7 Sequencing kit (Pharmacia) according to the manufacturers instructions.

RNA isolation and Northern blotting

Poly A^+ RNA was isolated out of different mouse tissues (129 Ola) with the QuickPrep Micro mRNA purification kit (Pharmacia) according to the manufacturers instructions. Total RNA was extracted as described (28).

Poly A^+ RNA (1.5 µg) or 15 µg total RNA were size fractionated on 1% agarose/formaldehyde gels, transferred to nylon membranes (Biodyne B, 0.45 µm, PALL) and baked at 80°C for 15 min. Prehybridization and hybridization was performed in hybridization buffer (50% formamide, 5× SSPE, 5× Denhardt's solution, 0.3% SDS, 0.3 mg/ml herring sperm DNA) at 42°C. Membranes were hybridized overnight with [α-^{32}P]dCTP labelled mCBP-KL cDNA (Rediprime, Amersham) and then washed with 0.2× SSC/0.1% SDS at 62°C until excess probe was removed. After autoradiography, membranes were stripped by washing with boiling 0.1% aqueous SDS solution and rehybridized with a β-actin cDNA probe as described above.

RT–PCR

Poly A^+ RNA was digested twice with 10 U RNase free DNase (Promega). An aliquot of 5 µg was used for first strand cDNA synthesis with 1 µg of a dT_{18} primer using Superscript II reverse transcriptase (Gibco, BRL). PCR reactions (25 µM of each primer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 2 U *Taq* polymerase, Promega) were performed as follows: 5 min 94°C; 1 min 94°C, 1 min 60°C, 2 min 72°C, 35 cycles; 10 min 72°C. Primers used in PCR: **A**, 5′-CAACTCAACTGAGCGGCAA-3′; **Ai**, 5′-CTATTGCTGGCATTCCGCAG-3′; **B2**, 5′-TCCATGTGCCTCAACCCTGA-3′; **B1**, 5′-TCAGGG TTGAGGCACATGGA-3′; **C1**, 5′-GGAGAGCTGGATTCAATGCC-3′; **D**, 5′-CCAGACATCTGACGGATCTC-3′; **Di**, 5′-TGATTTTGGCGCCTTGACGC-3′.

Isolation and characterization of a partial mCBP genomic clone

A mouse genomic cosmid library (Supercos1/129J) was screened with the entire mCBP-KL cDNA. Colonies were transferred to nitrocellulose filters (Protran BA 85, 0.45 µm, Schleicher & Schuell). The DNA was denatured (5 min, 0.5 M NaOH/1.5 M NaCl), neutralized (5 min, 0.5 M Tris–HCl pH 7.5/1.5 M NaCl) and crosslinked to the membranes by baking at 80°C for 2 h under vacuum. Prehybridization and hybridization were performed in aqueous solution at 68°C (3× SSC, 50 mM NaH_2PO_4 pH 6.5, 0.5% SDS, 7 mg/ml non-fat milk powder and 3× SSC, 20 mM NaH_2PO_4, 0.5% SDS, 7 mg/ml non-fat milk powder, respectively). Positive colonies were analyzed by Southern blotting with the same probe. Hybridizing fragments were subcloned into pBluescript II KS +/–. The exon/intron structure as well as the pseudogene sequence were determined by double-stranded sequencing using primers derived from the cDNA sequence.

Cell lysis and immunoprecipitation

Jurkat cells were washed twice in PBS and resuspended in lysis buffer containing 150 mM NaCl, 50 mM HEPES pH 7.4, 1% NP-40, 1 mM PMSF, 1 mM NaF and 0.1 TIU/ml aprotonin (Sigma). Cells were disrupted by short vortexing and incubation on ice for 15 min followed by a centrifugation at 11 000 *g* for 5 min at 4°C. The supernatant was used for immunoprecipitation.

Cellular lysates containing 500 µg of proteins were precleared with 20 µl Protein A/G PLUS agarose (Santa Cruz) followed by incubation for 2 h at 4°C with 20 µg purified polyclonal mCBP antibody. The agarose beads were washed three times with lysis buffer and resuspended in SDS sample buffer. After denaturation at 100°C for 5 min, proteins were size fractionated by PAGE and analysed by western blotting.

Antibodies and western blotting

The mCBP polyclonal antiserum was generated from guinea pigs, which were immunized with the mCBP-KL protein fused to six N-terminal histidines. Immunoprecipitations were performed with 20 µg of purified mCBP antibody or 5 µl of purified anti maltose binding protein serum (Biolabs), 3 µg/ml of the monoclonal 9G8 antibody (29) were used for the western blots.

Proteins were transferred on nitrocellulose membranes (Schleicher & Schuell) by semidry electroblotting and the membrane was blocked for 30 min in 2% non-fat milk powder in PBS. The first antibody was added in the same solution and incubated for 16 h at 4°C, followed by three washes with 0.1% Tween-20 in PBS.

The membrane was then incubated with a 1:10 000 dilution of the POD-coupled second antibody (0.8 mg/ml, Dianova) and finally washed four times. The presence of protein was detected by enhanced chemoluminescense reaction (Amersham).

Cultivation and immunostaining of RV-SMC cells

RV-SMC cells (smooth muscle cells of rat vein) were cultivated on cover slips as described (30). After 10 min fixation at –20°C in methanol and 1 min at –20°C in acetone, cells were air dried for 10 min at room temperature. Immunostaining was performed with a 1:100 dilution of anti mCBP serum in PBS for 15 min at room temperature followed by a 10 min washing in PBS and incubation of a goat anti guinea pig antibody coupled to Texas red (Dianova) diluted 1:200 for 15 min at room temperature. After a final washing for 10 min, cells were desiccated in ethanol and embedded in Mowiol. In parallel, cells were stained with Hoechst DNA fluorochrome no. 33258 according to standard procedures.

Plasmids and two hybrid library

pGBT9-KL contains the full length cDNA of mCBP-KL cloned into the yeast expression vector pGBT9 (31). The yeast strain HF7c which contains two GAL4 inducible reporters (His3 and LacZ) was transformed with the bait plasmid pGBT9-KL and then retransformed with a 9.5–19.5 day mouse embryo cDNA library. This library contains cDNA inserts ranging from 350–700 bp generated by random primed cDNA synthesis (32). cDNA fragments were inserted into a pVP16 vector that directs the synthesis of fusion proteins containing the mouse embryo peptides and the VP16 activation domain. The plasmid pLEXA-Lamin (32), coding for a fusion protein between LEXA and human LaminC (amino acids 66–230), was used for yeast control transformations.

Library screen

HF7c cells (33) were transformed to tryptophan prototrophy with pGBT9-KL and afterwards retransformed with the mouse cDNA library by standard procedures (31,34). Histidine prototroph clones were assayed for β-galactosidase (lacZ) activity by a filter assay (35). The library plasmids of the lacZ positive clones were recovered by isolation of total DNA from the colonies followed by electroporation into *E.coli* and subsequent restriction analysis of both plasmids. The cDNA inserts were sequenced by using the primer: 5′-GGTACCGAGCTCAATTGCGG-3′ positioned in the polylinker region upstream the 5′-end of the cDNA insert.

RESULTS

Isolation and characterization of mCBP splice variants

We previously reported the isolation of an mCBP cDNA (referred to as mCBP-KL in this paper) by RACE followed by screening of a mouse cDNA library with the RACE product (2). In the present study we analyzed additional mCBP cDNAs and detected two small variable regions by comparing the sequences of two full length clones (mCBP-KL and mCBP-E, Fig. 1A). Unlike mCBP-KL, mCBP-E contains a 93 bp insertion, whereas further downstream, 39 bp are missing (Fig. 1B). Both regions are located between the second and third KH-domain of mCBP. A characterization of additional mCBP cDNAs as well as the RACE product pRACE 1.2 at these precise regions gave a more complex

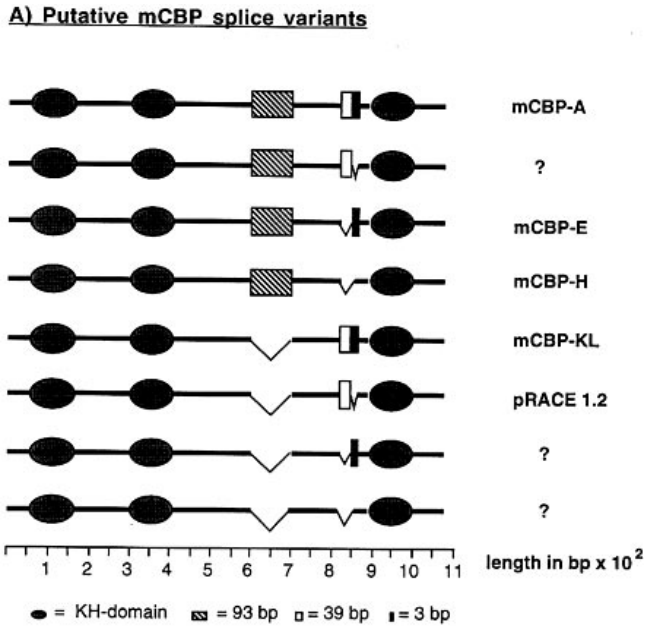

Figure 1. Characterization of mCBP splice variants. (**A**) The coding regions of the different cDNAs are represented by solid lines. KH domains are symbolized by ovals, variable regions are symbolized by striped (±93 bp), white (±39 bp) and black (±3 bp) boxes, respectively. The entire cDNA sequence was determined for mCBP-E, -H and -KL. Only the variable regions were sequenced for mCBP-A and pRACE1.2; ?, putative splice variants which have not been isolated so far. (**B**) DNA sequences of the first and second variable region.

pattern than initially expected. Concerning the first variable region of ±93 bp, all clones were identical either to mCBP-KL or mCBP-E whereas the second variable region showed a more diverse phenotype: the maximum size of this region turned out to be 42 bp (compare mCBP-H versus mCBP-KL, Fig. 1A) which have to be further divided into 39 + 3 bp as there are variants lacking either the first 39 bp or the last 3 bp (mCBP-E and pRACE1.2, respectively). Provided that the ±39 ± 3 bp phenotype occurs independently from the first variable region and that there are no further undetected variable regions, a total of eight different mCBP variants (five of which have been characterized) would exist, which are most likely generated by alternative splicing of a single primary transcript.

In order to elucidate the generation of the mCBP variants mentioned above, we isolated a part of the mCBP genomic locus from a mouse genomic library (see Materials and Methods). A detailed characterization revealed that the second and third exon from the 3′-end of the mCBP gene (denoted n–2 and n–3) contribute to the second variable region of ±39 ± 3 bp (Fig. 2A). Interestingly, the variable three base pairs (CAG) do not seem to

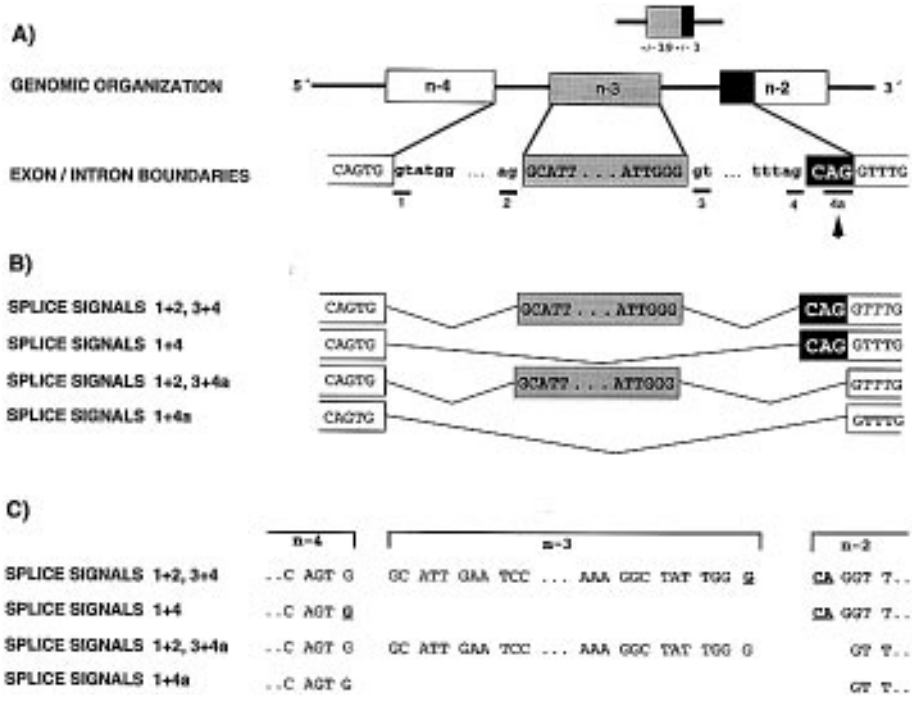

Figure 2. Model for the generation of mCBP splice variants. (**A**) The variable region of ±39 ± 3 bp is schematically shown on top. The organization of those parts of the mCBP gene which contribute to this sequence is schematically drawn. The second, third and fourth exon from the 3′-end of the mCBP gene (exons n–2, n–3 and n–4) are symbolized by boxes, intron sequences by solid lines. The sequence of the exon/intron boundaries is written below. Capital letters are used for exon sequences. Splice donor and acceptor sites 1–4 are underlined. The putative alternative splice signal 4a is indicated by an arrow. The alternatively spliced three bases of exon n–2 are highlighted. (**B**) The sequences which result from different usage of the splice donor and acceptor sites are shown. (**C**) The sequence of the variable region of the four possible phenotypes is written in frame. The exons however are not spliced in frame. Therefore, the alternative splicing of the first three bases of exon n–2 leads to the generation of the codon GCA (Alanin).

be encoded by an individual exon: usage of the regular splice donor and acceptor sites 1–4 explains the generation of the +39 + 3 bp phenotype (splice sites 1 + 2, 3 + 4, Fig. 2B) as well as the –39 + 3 bp phenotype (splice sites 1 + 4, Fig. 2B). The generation of those variants lacking the last 3 bp, however, cannot be explained only by the use of splice sites 1–4. We detected a putative alternative splice acceptor site 4a (Fig. 2A) within exon n-2 which closely matches the consensus for splice acceptor sequences (36), resulting in the occurrence of two contiguous acceptor sites in the TAG/CAG sequence. This alternative splice acceptor signal would explain the generation of the +39 – 3 bp and the –39 – 3 bp phenotype (Fig. 2B). The first three bases of exon n–2 however, do not represent a codon because exons n–4, n–3 and n–2 are not spliced in frame (Fig. 2C). The alternative splicing of the first three bases of exon n–2 therefore leads to the presence or absence of the amino acid alanine (GCA) on protein level.

Isolation and characterization of an mCBP processed pseudogene

Apart from the genomic clone described above, we obtained two cosmids whose inserts hybridized to probes from the 5′ and the 3′ coding region of mCBP-KL (data not shown). The hybridizing fragments were subcloned and analyzed for mCBP coding sequences. They contained two different intronless mCBP sequences, which are highly degenerated compared with the cDNA sequence (data not shown). This indicates that they represent processed pseudogenes which resemble reverse transcribed copies of mRNAs integrated into the genome (37). One of them resembles a mCBP cDNA with the –93 bp and –39 + 3 bp phenotype concerning the first and second variable region, respectively, which is consistent with the model presented in Figure 1.

Expression of mCBP splice variants in different mouse tissues

mCBP expression was investigated by Northern blotting and showed the appearance of a 1.9 kb transcript in mouse testes, thymus, brain and liver (Fig. 3). Expression of mCBP was also confirmed in mouse lymph nodes and colon as well as in EL4 cells (data not shown). Highest expression levels upon comparison with actin expression were found in testes and EL4 cells whereas thymus, lymph nodes, brain and colon showed lower mCBP expression. Only weak hybridization was detected in liver.

As the different mCBP splice variants could not be resolved by Northern blotting, we performed RT–PCR with RNA from different mouse tissues. Different primers positioned within and next to the first and second variable region of the mCBP cDNA were used in PCR. All primer combinations as well as the predicted sizes of the amplified fragments are shown in Figure 4A.

RT–PCR performed with RNA isolated from eight different mouse tissues (testes, thymus, spleen, kidney, liver, colon, heart and brain) revealed the same pattern of amplification in all tissues examined which is shown for thymus in Figure 4B. RNA, which was not reverse transcribed, served as a negative control to verify that the RNA was not contaminated by cDNAs or pseudogenes.

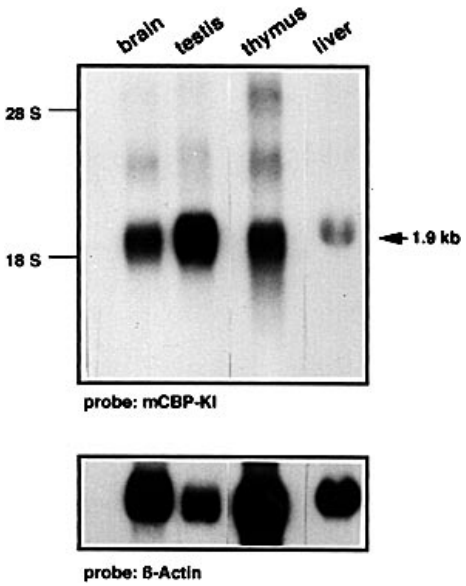

Figure 3. Expression of mCBP in different mouse tissues. 1.5 µg Poly A^+ RNA from brain, testes and thymus as well as 15 µg total RNA from liver (18S and 28S rRNA serve as standard) were separated and blotted as described in Materials and Methods. Membranes were probed with the mCBP-KL cDNA (on top) and reprobed with β-actin (below).

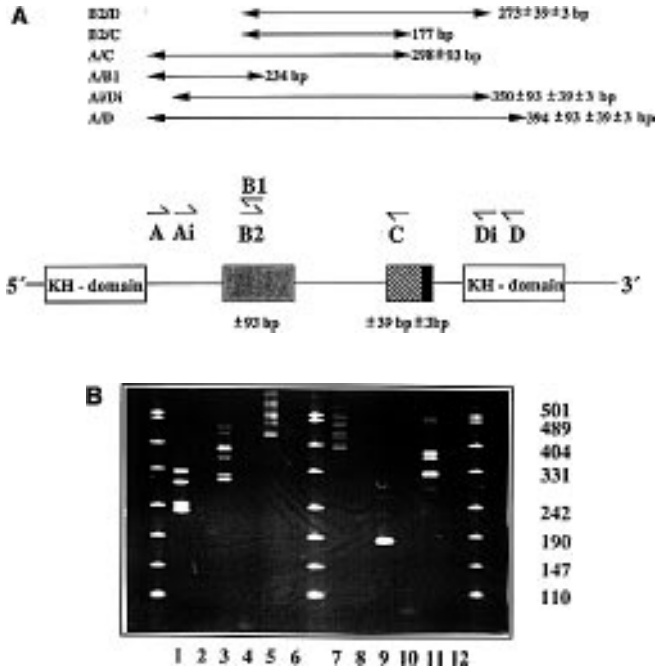

Figure 4. RT–PCR of mouse thymus RNA. (**A**) Selection of mCBP specific primers. The figure shows the position of different primers used in RT–PCR according to the cDNA sequence of mCBP including second and third KH-domain as well as the two insertions of the +93 bp and +39 + 3 bp phenotype. The predicted PCR products as well as their lengths are shown above. (**B**) Expression of different mCBP splice variants in mouse thymus. Reverse transcribed RNA isolated from mouse thymus was amplified by PCR according to experimental procedures using the primer combinations: AB1 (1), A/C (3), A/D (5), Ai/Di (7), B2/C (9) and B2/D (11). Lanes 2, 4, 6, 8, 10 and 12, negative controls. PCR products were size-fractionated on a 4% non-denaturating polyacrylamide gel. The length of the marker fragments is shown in bp.

The primer combination B2/C gives rise to the predicted PCR product of 177 bp length (Fig. 4B, lane 9), which corresponds to mCBP splice variants with both variable regions. In contrast, all the other primer combinations amplified several DNA fragments in addition to those expected (Fig. 4B, lanes 1, 3, 5, 7 and 11).

It seems unlikely that the additional DNA fragments were generated by non-specific amplification, because the negative controls show no amplified DNA and the nested primer combinations (A/D and Ai/Di) generated the same banding pattern. This suggests the existence of other splice variants in addition to those already described in this paper.

Subcellular localization of mCBP

A polyclonal anti mCBP serum was used to detect the subcellular localization of mCBP in different cell lines. Mouse 3T3 cells as well as rat RV-SMC cells were immunostained with the mCBP and preimmune serum combined with a concurrent staining of nuclei with a DNA fluorochrome. In RV-SMC cells and 3T3 cells the mCBP protein is localized predominantly to the nucleus with the exception of the nucleoli (Fig. 5 and data not shown). During mitosis, mCBP is found dispersed in the cytoplasm, where it migrates back to the nucleus during telophase.

Isolation of mCBP-binding proteins

Apart from the detailed characterization of the nucleic acid binding properties of mCBP (2) and its identification as a component of a multiprotein complex stabilizing the human α-globin mRNA (6), nothing is known about the function of mCBP *in vivo*. As nucleic acid binding proteins often exert their function by the interaction with other proteins, we employed the yeast two hybrid system, originally described by Fields and Song (38), to identify proteins which interact with mCBP.

A full length cDNA of mCBP-KL fused to the GAL4 DNA binding domain was used as a 'bait' named pGBT9-KL. Yeast strain HF7c was retransformed with a 9.5–10.5 day mouse embryo cDNA library (see Materials and Methods). The expression of mCBP in day 2.5 mouse embryonic stem cells as well as in day 14 mouse embryonic fibroblasts was confirmed by EMSA (data not shown).

A total of 4×10^6 transformants were placed on selective medium and then screened as described in the Materials and Methods. False positives were excluded by transfecting HF7c cells either with the library plasmid alone or together with pGBT9, pGBT9-KL or pLexALamin (negative control). Fifteen out of 400 clones showed β-galactosidase expression only when transformed with pGBT9-KL and the library plasmid.

mCBP interacts with a series of RNA-binding proteins in the yeast two hybrid system

Sequencing of the 5′-ends of the 15 cDNAs which were identified in the yeast two hybrid system revealed eight different sequences, three of them being unknown. The remaining five cDNA fragments correspond to the following already described sequences (Table 1): msy-1 (bp 251–600), murine y-box-binding protein (bp 170–520), the murine homologues of human hnRNP-L (bp 220–580), human splicing factor 9G8 (bp 120–500) and human actin binding protein 280 (also known as filamin, bp 2400–2800).

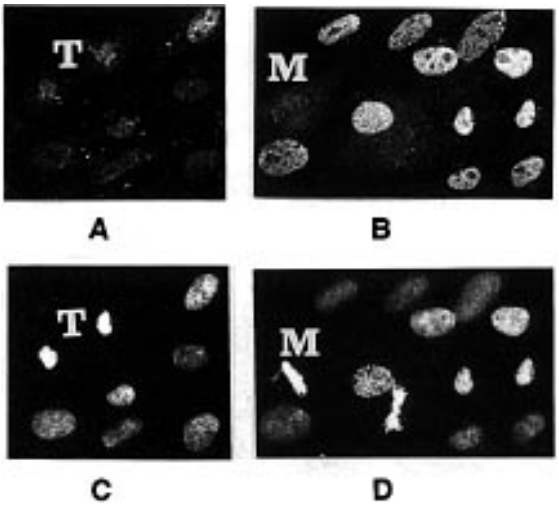

Figure 5. Subcellular localization of mCBP in rat RV-SMC cells. RV-SMC cells were immunostained as described in experimental procedures with preimmune serum (**A**) and anti mCBP serum (**B**). The nuclei of the same cells were stained simultaneously with a DNA fluorochrome. (**C**) Stained nuclei of A; (**D**) stained nuclei of B; M, metaphase chromosomes; T, late telophase chromosomes.

Interestingly, four of the five cDNA inserts code for peptides involved in the binding of nucleic acids. Msy-1 and y-box-binding protein both contain an 80 amino acid cold shock domain, which is responsible for the binding of DNA and RNA. The isolated cDNA fragment of murine hnRNP-L spans the region of its human counterpart containing one of the two RNP-1 motifs responsible for the binding of RNA. An RNP-1 as well as an RNP-2 motif is also present in the murine homologue of the human splicing factor 9G8.

Table 1. Identification of mCBP-binding proteins

mCBP binding proteins	Characteristic protein domain	EMBL accession no.
msy-1	CSD	M62867
y-box-binding protein	CSD	L35599
hnRNP-L	RNP-CS	X16135
splicing factor 9G8	RNP-CS	L22253
filamin	–	X53416

The table shows mCBP-binding proteins identified by the yeast two hybrid system as well as protein motifs encoded by the cDNA inserts of the corresponding library plasmids. CSD, cold shock domain; RNP-CS, ribonucleoprotein consensus sequence.

mCBP interacts *in vivo* with the splicing factor 9G8

The interaction between mCBP and splicing factor 9G8 was confirmed *in vivo* by co-immunoprecipitation. The specificity of the mCBP antibody was demonstrated by Western analysis (data not shown). This antibody was then used to precipitate mCBP from total cellular lysates as described in Materials and Methods. Western blot was performed with the 9G8 antibody (Fig 6).

The 35 kDa 9G8 protein co-purified with mCBP when the anti mCBP antibody was used (lane 2), whereas it was not co-immunoprecipitated with a mock antibody (lane 3). Jurkat cell extract served as a positive control (lane 1). These results strongly suggest that the interaction between mCBP and the splicing factor 9G8 also occurs *in vivo*.

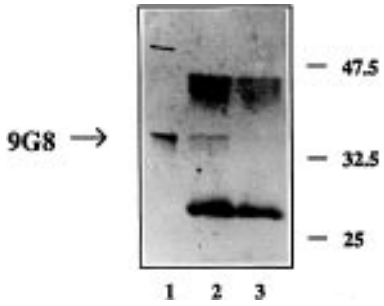

Figure 6. Co-immunoprecipitation of mCBP and splicing factor 9G8. Co-immunoprecipitation of mCBP and splicing factor 9G8. mCBP was precipitated from 500 µg Jurkat total cellular lysate with 20 µg purified mCBP antibody (2) as described in experimental procedures. The same precipitation was performed with a mock antibody (5 µl of anti maltose binding protein serum, Biolabs) in lane 3; lane 1, 10 µg of Jurkat total cellular lysate. Western blotting was performed as described in experimental procedures, protein marker is shown in kDa.

DISCUSSION

mCBP belongs to the poly(C)-binding proteins in vertebrates and, by the existence of three KH domains, shows similarity to hnRNP-K, a member of the group of hnRNP proteins. In this paper, we have described mCBP splice variants which differ by two variable regions of 93 bp and/or 39 ± 3 bp respectively, both of them being located between the second and third KH domain.

Characterization of a partial genomic mCBP clone revealed that the alternatively spliced 3 bp do not represent an individual mini-exon. According to the model we present in Figure 2, the usage of an alternative splice site within exon n–2 leads to the absence of its first three bases. However, they do not represent a codon. The exons are not spliced in frame so that the usage of the alternative splice site 4a leads to the absence of the codon GCA which codes for alanine (Fig. 2C). Alternative splicing of ±1 amino acid has also been described for hnRNP-K (20) and FBP, another member of the KH domain protein family (39); however, in none of these cases the genomic situation has been characterized.

The existence of multiple variants generated by alternative splicing is a common feature of many hnRNP proteins: four different variants have been described for hnRNP-K differing by two miniexons of 15 and/or 60 bp, respectively (20). Other examples are hnRNP-B1 and hnRNP-A2, which are identical except for 12 amino acids at their N-terminus (22), hnRNP-C1 and C2 which differ by 13 amino acids (22) and PTB, also described as hnRNP-I (40), which contains two variable regions of 19 and/or 7 amino acids respectively (21). The initial purification of mCBP (2) by ssDNA affinity chromatography resulted in two separate bands, which are likely to represent two isoforms of the protein. Interestingly, both isoforms lack the 93 bp variable region, as determined by protein sequencing. This could indicate that mCBP isoforms, which contain this region, do not bind ssDNA as strongly as the mCBP isoforms without that region. This would be reminiscent of the hnRNP-K protein whose different splice variants seem to have different ss nucleic acid

binding properties as only one of the four isoforms was found to interact with a regulatory element of the c-*myc* promotor (23). The ability of a protein to interact with both RNA and DNA is not unusual: in *Xenopus laevis*, the transcription factor TFIIIA, which activates the 5S rRNA transcription, also binds the 5S rRNA transcript to form a 7S particle (41). Further experiments will have to be performed to determine if the different mCBP variants have different nucleic acid binding properties.

Why do so many variants exist which differ only by a few amino acids? One possibility is an alteration of their nucleic acid binding properties. The observation that only one of the four splice variants of hnRNP-K has been identified as a transactivator of human c-*myc* gene (23) supports this hypothesis. Another possibility would be the modification of protein–protein interactions, as hnRNP proteins are components of large multiprotein complexes (8).

The expression of different mCBP variants was examined by Northern blotting and RT–PCR. The same banding pattern of amplified DNA products was observed in all tissues tested, which is consistent with the Northern blot results. Interestingly, with the exception of one primer combination (Fig. 4B, lane 9), we obtained more amplified PCR products as expected. The possibility that these bands correspond to processed pseudogenes can be excluded because the negative controls showed no amplified DNA. For that reason, it seems possible that mCBP variants, in addition to those already described in this paper, do exist. A human mCBP homologue has been described as αCP2 (6) and hnRNP-E2 (42). This protein contains an additional variation of four amino acids compared with the mCBP clone pRACE 1,2 (Fig. 1A) but is otherwise identical. These four amino acids are located within the region amplified by RT–PCR. A comparable variant could also exist in mouse. This will have to be confirmed by the isolation and characterization of further mCBP cDNAs. Another human variant (αCP1/hnRNP-E1) has been characterized, which is only 89% identical to mCBP-E (6,43). Highest diversity is found dispersed in the region between the second and third KH domain, which again is the region amplified by RT–PCR (Fig. 4A). This variant was probably not detected by us because of the stringent screening conditions we employed for the isolation of mCBP cDNAs. However, it is possible that this and other variants could have been amplified in the RT–PCR assay.

In order to get insight into a possible function of mCBP *in vivo*, we employed the yeast two hybrid system to identify mCBP-binding proteins. Interestingly, five of the isolated sequences encode parts of proteins which have been shown to participate in RNA-binding or metabolism. Both the msy-1 protein as well as the y-box-binding protein belong to the family of y-box-proteins originally identified by virtue of their binding to promoters containing reverse CCAAT boxes. They have been further demonstrated to participate in the packaging of RNA (44). msy-1 was shown to bind and stabilize paternal mRNA in mouse spermatocytes as a component of the 60–80S mRNP particles (45). hnRNP-L is found to be associated with nascent transcripts of amphibian lampbrush chromosomes (46) and binds to a *cis*-acting element of a HSV-TK transcript (47). Filamin (48) belongs to the group of actin crosslinking proteins. Because *in situ* hybridization revealed a co-localization of poly A^+ RNA, the actin crosslinking proteins are hypothesized to be involved in the transport and positioning of mRNP particles (49). The splicing factor 9G8 belongs to the family of SR splicing factors (29) which are associated with spliceosomes that assemble on hnRNA after transcription (50).

The interaction of mCBP with the splicing factor 9G8 was verified *in vivo* by co-immunoprecipitation (Fig. 6B). Further experiments will reveal if the interaction of mCBP with msy-1, y-box-binding protein, hnRNP-L and filamin also occur *in vivo*. Interestingly, all of them are components of large multiprotein complexes involved in RNA metabolism. In this context, it is interesting to note that the human mCBP homologues αCP1 and αCP2 have also been characterized as parts of an mRNA stability complex which binds to the 3′-UTR of α-globin-mRNA in erythrocytes (6).

The above results favour a participation of mCBP in RNA metabolism whereas mCBP was described as a ssDNA binding protein based on its preferential interaction with poly(C) sequences *in vitro* by Goller *et al.* (2). This is not necessarily a contradiction as there are examples for bifunctional proteins, including hnRNP-K, which bind to RNA as well as to DNA (7,23,41).

ACKNOWLEDGEMENTS

We thank Dr Michael Zimmer for providing the cosmid library and for helpful discussion. We also thank Dr Andreas Kalmes and Carsten Hagemann for their support with the Yeast Two Hybrid System. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 165).

REFERENCES

1 Edelmann,W., Kröger,B., Goller,M. and Horak,I. (1989) *Cell*, **57**, 937–946.
2 Goller,M., Funke,B., Gehe-Becker,C., Kröger,B., Lottspeich,F. and Horak,I. (1994) *Nucleic Acids Res.*, **22**, 1885–1889.
3 Siomi,H., Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1993) *Nucleic Acids Res.*, **21**, 1193–1198.
4 Hahm,K., Kim,G., Turck,C.W. and Smale,S.T. (1993) *Nucleic Acids Res.*, **21**, 3894.
5 Leffers,H., Dejgaard,K. and Celis,J.E. (1995) *Eur. J. Biochem.*, **230**, 447–453.
6 Kiledjian,M., Wang,X. and Liebhaber,S.A. (1995) *EMBO J.*, **14**, 4357–4364.
7 Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1992) *Mol. Cell. Biol.*, **12**, 164–171.
8 Dreyfuss,G., Matunis,M.J., Piñol-Roma,S. and Burd,C.G. (1993) *Annu. Rev. Biochem.*, **62**, 289–321.
9 Kamma,H., Portman,D.S. and Dreyfuss,G. (1995) *Exp. Cell. Res.*, **221**, 187–196.
10 Zaidi,S.H.E. and Malter,J. (1995) *J. Biol. Chem.*, **270**, 17292–17298.
11 Piñol-Roma,S., Choi,Y.D., Matunis,M.J. and Dreyfuss,G. (1988) *Genes Dev.*, **2**, 215–227.
12 Burd,C.G. and Dreyfuss,G. (1994) *Science*, **265**, 615–621.
13 Siomi,H., Choi,M., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1994) *Cell*, **77**, 33–39.
14 Siomi,H., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1993) *Cell*, **74**, 291–298.
15 Brauer,D. and Roemig,R. (1979) *FEBS Lett.*, **106**, 352–357.
16 Mahone,M., Saffman,E.E. and Lasko,P.F. (1995) *EMBO J.*, **14**, 2043–2055.
17 Buckanovich,R.J., Posner,J.B. and Darnell,R.B. (1993) *Neuron*, **11**, 657–672.
18 Jones,A.R. and Schedl,T. (1995) *Genes Dev.*, **9**, 1491–1504.
19 Wong,G., Müller,O., Clark,R., Conroy,L., Moran,M.F., Polakis,P. and McCormick,F. (1992) *Cell*, **69**, 551–558.
20 Dejgaard,K., Leffers,H., Rasmussen,H.H., Madsen,P., Kruse,T.A., Gesser, B., Nielsen,H. and Celis,J.E. (1994) *J. Mol. Biol.*, **236**, 33–48.

21 Gil,A., Sharp,P.A., Jamison,S.F. and Garcia-Blanco,M.A. (1991) *Genes Dev.*, **5**, 1224–1236.
22 Burd,C.G., Swanson,M.S., Görlach,M. and Dreyfuss,G. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9788–9792.
23 Takimoto,M., Tomonaga,T., Matunis,M. J., Avigan,M., Krutzsch,H., Dreyfuss,G. and Levens,D. (1993) *J. Biol. Chem.*, **268**, 18249–18258.
24 Tomonaga,T. and Levens,D. (1995) *J. Biol. Chem.*, **270**, 4875–4881.
25 Taylor,S.J. and Shalloway,D. (1994) *Nature*, **368**, 867–871.
26 Bustelo,X.R., Suen,K.L., Michael,W.M., Dreyfuss,G. and Barbacid,M. (1995) *Mol. Cell. Biol.*, **15**, 1324–1332.
27 Weng,Z., Thomas,S.M., Rickles,R. J., Taylor,J.A., Brauer,A.W., Seidel-Dugan,C., Michael,W.M., Dreyfuss,G. and Brugge,J.S. (1994) *Mol. Cell. Biol.*, **14**, 4509–4521.
28 Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156–159.
29 Cavaloc,Y., Popielarz,M., Fuchs,J.P., Gattoni,R. and Stévenin,J. (1994) *EMBO J.*, **13**, 2639–2649.
30 Franke,W.W., Schmid,E., Vandekerkhove,J. and Weber,K. (1980) *J. Cell. Biol.*, **87**, 594–600.
31 Bartel,P.L., Chien,C.-T., Sternglanz,R. and Fields,S. (1993) Hartley,D.A. (ed.) *Cellular Interactions in Development—A practical approach*. Oxford University Press, pp. 153–179.
32 Vojtek,A.B., Hollenberg,I.M. and Cooper,J.A. (1993) *Cell*, **74**, 205–214.
33 Feilotter,H.E., Hannas,G.J., Rudell,C.J. and Beach,D. (1994) *Nucleic Acids Res.*, **22**, 1502–1503.
34 Gietz,D., St Jean,A., Woods,R.A. and Schiestl,R.H. (1992) *Nucleic Acids Res.*, **20**, 1425.
35 Breeden,L. and Nasmyth,K. (1985) *Cold Spring Harbour Symp. Quant. Biol.*, **50**, 643–650.
36 Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
37 Weiner,A.M., Deininger,P.L. and Efstratiadis,A. (1986) *Annu. Rev. Biochem.*, **55**, 631–661.
38 Fields,S. and Song,O.-K. (1989) *Nature*, **340**, 245–246.
39 Duncan,R., Bazar,L., Michelotti,G., Tomonaga,T., Krutzsch,H., Avigan,M. and Levens,D. (1994) *Genes Dev.*, **8**, 465–480.
40 Ghetti,A., Piñol-Roma,S., Michael,W.M., Morandi,C. and Dreyfuss,G. (1992) *Nucleic Acids Res.*, **20**, 3671–3678.
41 Honda,B.M. and Roeder,R.G. (1980) *Cell*, **22**, 119–126.
42 EMBL accession no. X78136
43 EMBL accession no. X78137
44 Wolffe,A.P. (1994) *BioEssays*, **16**, 245–251.
45 Tafuri,S.R., Familari,M. and Wolffe,A.P. (1993) *J. Biol. Chem.*, **268**, 12213–12220.
46 Piñol-Roma,S., Swanson,M.S., Gall,J.G. and Dreyfuss,G. (1989) *J. Cell. Biol.*, **109**, 2575–2587.
47 Liu,X. and Mertz,J.E. (1995) *Genes Dev.*, **9**, 1766–1780.
48 Hock,R.S., Davis,G. and Speicher,D.W. (1990) *Biochemistry*, **29**, 9441–9451.
49 Bassell,G.J., Powers,C.M., Taneja,K.L. and Singer,R. (1994) *J. Cell. Biol.*, **126**, 863–876.
50 Kiseleva,E., Wurtz,T., Visa,N. and Daneholt,B. (1994) *EMBO J.*, **13**, 6052–6061.