The EMBO Journal vol.8 no.7 pp.2023 – 2028, 1989

Leucine zipper structure of the protein CRE-BP1 binding to the cyclic AMP response element in brain

Toshio Maekawa[1,2], Hiroshi Sakura[1,3], Chie Kanei-Ishii[1], Tatsuhiko Sudo[1], Tadashi Yoshimura[4], Jun-ichi Fujisawa[4], Mitsuaki Yoshida[4] and Shunsuke Ishii[1]

[1]Laboratory of Molecular Genetics, [2]Frontier Research Programs, Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), Tsukuba, Ibaraki 305, [3]The Third Department of Internal Medicine, Faculty of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 103 and [4]Department of Viral Oncology, Cancer Institute, Toshima-ku, Tokyo 170, Japan

Communicated by A.Levitzki

By screening a λgt11 library with the multimerized sequence of the cAMP response element (CRE), we isolated human clones encoding the CRE binding protein, CRE-BP1, from a human brain cDNA library. CRE-BP1 expressed in *Escherichia coli* bound not only to the CRE element of the somatostatin and fibronectin genes, but also to the CRE element of the adenovirus E4 gene, suggesting that the protein was not distinguishable from the adenovirus transcription factor, ATF. The human CRE-BP1 clone encoded a 54.5 kd protein similar at its carboxy terminus to the leucine zipper motifs found in other enhancer binding proteins such as C/EBP and c-*jun*/AP-1. CRE-BP1 mRNA was expressed in all of the cells examined and was abundant in brain. The structure of CRE-BP1 and its recognition elements suggest that cellular response to extracellular stimuli is controlled by a family of transcription factors that bind to related *cis*-active elements and that contain several highly conserved domains.

Key words: cAMP response element/enhancer binding protein/leucine zipper/adenovirus transcription factor

Introduction

The second messengers cyclic AMP (cAMP) and diacylglycerol activate a cAMP-dependent protein kinase, protein kinase A (Nairn *et al.*, 1985), and protein kinase C (Nishizuka, 1984) respectively. The activated kinases seem to regulate gene transcription by mechanisms not yet understood. The cAMP response element (CRE; TGACGTCA) is an inducible enhancer of the genes that induce transcription in response to increases in the intracellular cAMP concentration (Comb *et al.*, 1986; Montminy *et al.*, 1986). A nuclear phosphoprotein, CREB, which has a mol. wt of 43 kd and which binds with high affinity to the CRE of the rat somatostatin gene, has been purified from both PC12 pheochromocytoma cells and rat brain and found to stimulate transcription of the cAMP-responsive gene when in dimeric form (Montminy and Bilezikjian, 1987; Yamamoto *et al.*, 1988). CREB is the same as or closely related to the adenovirus transcription factor (ATF) that binds to the CRE in the regulatory regions of adenovirus early genes (Hurst and Jones, 1987; Lee *et al.*, 1987; Hardy and Shenk, 1988; Lin and Green, 1988). To understand the mechanism by which activation of second-messenger pathways results in the transcriptional stimulation of many cellular genes, cloning of the gene that codes for the protein that binds to the CRE would be useful. Here, we report the isolation of human clones encoding the CRE binding protein CRE-BP1 from a human brain cDNA library. CRE-BP1 is a protein of mol. wt 54.5 kd, similar at its carboxy terminus

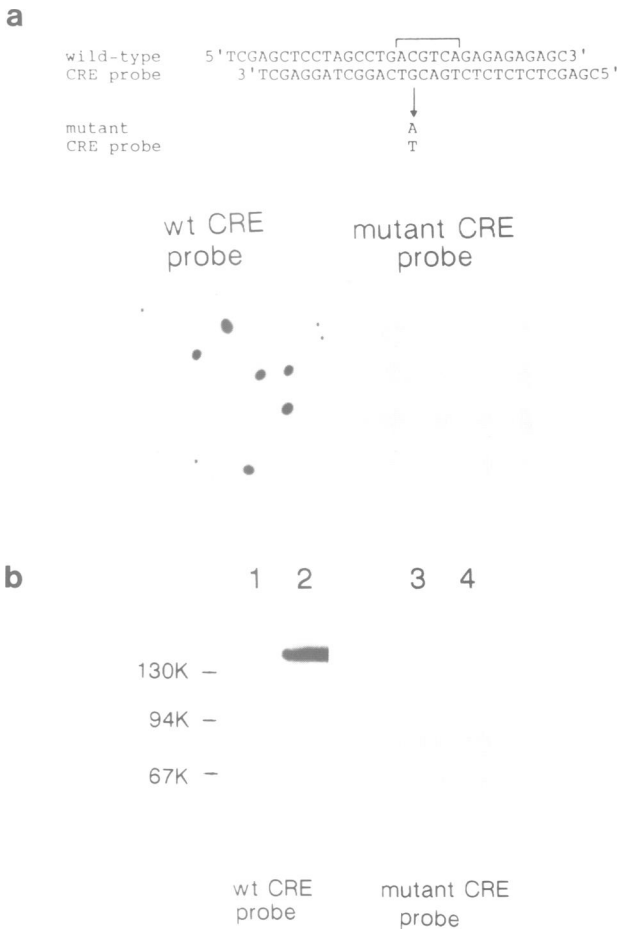

Fig. 1. Southwestern analysis with CRE probes. (**a**) Screening of λgt11 library. A human fetal brain λgt11 cDNA library was screened with a CRE probe containing the sequence shown in Materials and methods corresponding to CRE of the human fibronectin gene. One of the positive clones, λCRE-BP1-1 was used for Southwestern analysis with the synthetic oligonucleotide probe (monomer) containing the wild-type or mutated CRE sequence of the human somatostatin gene shown here. A palindromic CRE sequence is denoted by a box. (**b**) Detection of CRE binding of fusion protein. Proteins from induced cultures of λCRE-BP1-1 lysogens (**lanes 2** and **4**) or bacteria not harboring prophage (**lanes 1** and **3**) were separated by SDS–PAGE and transferred to a nitrocellulose filter. The filter was incubated with the wild-type or mutated CRE probe of the human somatostatin gene shown in (a). Protein size markers are indicated on the left.

to the leucine zipper motifs. CRE-BP1 mRNA was relatively abundant in the brain.

Results

Cloning of the CRE-BP1 cDNA

To isolate cDNA encoding human CRE binding protein, we used the Southwestern method developed by Singh *et al.* (1988) for isolation of a factor binding the H2TF1 element, and modified by Vinson *et al.* (1988) for isolation of C/EBP. In this method, β-galactosidase fusion proteins are screened by their ability to bind with high affinity to specific DNA sequences. A double-stranded oligonucleotide corresponding to the CRE sequence of the human fibronectin gene (see Materials and methods) was multimerized six times and used to screen a λgt11 cDNA library prepared from human brain mRNA. About 1×10^6 recombinants were screened and among the 25 phage clones identified in the first screening, three were found to bind with the DNA probe in the subsequent rounds of screening. One of them, designated λCRE-BP1-1, was analyzed further. To check whether the λCRE-BP1-1 clone bound with the CRE of other genes, the oligonucleotide corresponding to the CRE sequence of the rat or human somatostatin gene was also used in Southwestern analysis. The λCRE-BP1-1 clone bound to both probes containing one of these CRE sequences (Figure 1a). In the control experiment, a mutated CRE probe of the human somatostatin gene, in which the central portion of the CRE was altered by a change of the C:G pair at position −45 into A:T, was used (Figure 1a). This mutation does not form DNA−protein complexes after *in vitro* binding of nuclear extracts (Andrisani *et al.*, 1988). The λCRE-BP1-1 clone did not bind with the mutated CRE probe (Figure 1a).

To characterize the protein product encoded by the clone λCRE-BP1-1, lysogen was prepared and induced to express

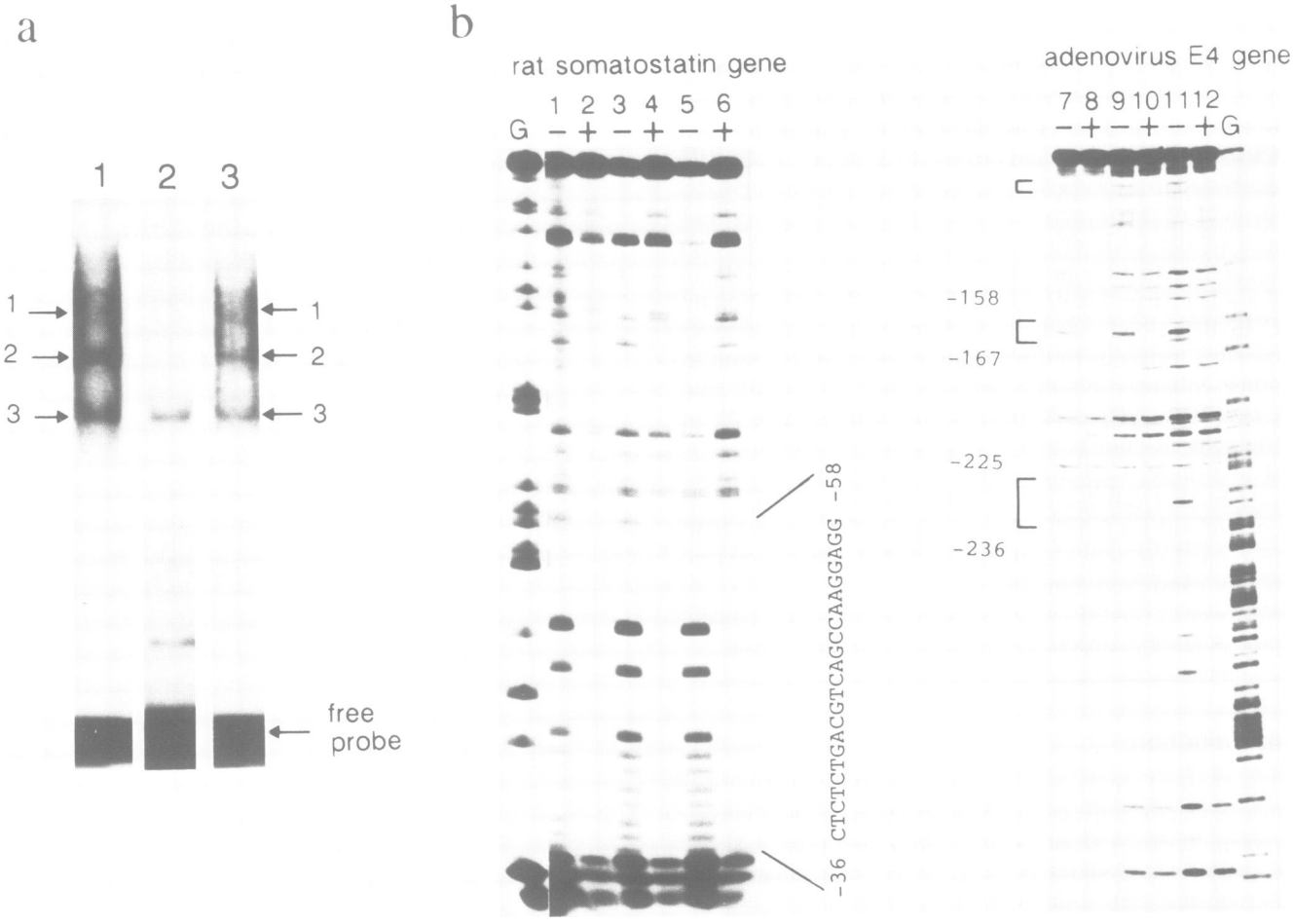

Fig. 2. DNA binding of the protein encoded by λCRE-BP1-1. (**a**) Gel-retardation assay. Approximately 15 ng of fusion protein encoded by λCRE-BP1-1 that had been purified by DNA-affinity chromatography was incubated with the wild-type (**lane 1**) or mutated CRE probe (**lane 2**) of the human somatostatin gene shown in Figure 1a, or with the wild-type CRE probe of the rat somatostatin gene (**lane 3**). The reaction products were separated by electrophoresis in a non-denaturing polyacrylamide gel. The CRE probe of the rat somatostatin gene was made by annealing two complementary oligonucleotides, 5′-GATCCCTGGGGGCGCCTCCTTGGCTGACGTCAGAGAGAGAGG-3′ and 5′-GATCCCTCTCTCTCTGA-CGTCAGCCAAGGAGGCGCCCCCAGG-3′, corresponding to the CRE sequence of the rat somatostatin gene (nt −32 to −67). The X-ray film for lanes 1 and 3 was exposed for 10 h and that for lane 2 was for 50 h. The slow bands (1−3) are shown by arrows. (**b**) DNase I footprinting analysis. The ^{32}P-labeled DNA fragment containing rat somatostatin CRE or the adenovirus E4 CRE sequence was incubated with 60 ng (lanes marked +) of the affinity-purified fusion protein encoded by λCRE-BP1-1. In control experiments, a ^{32}P-labeled DNA fragment was incubated without protein (lanes marked −). The final concentration of DNase I for digestion was 20 (**lanes 1** and **2**), 40 (**lanes 3** and **4**), 80 (**lanes 5** and **6**), 5 (**lanes 7** and **8**), 10 (**lanes 9** and **10**) or 20 μg/ml (**lanes 11** and **12**). The guanine-specified sequence markers prepared by cleavage of the same ^{32}P-labeled DNA fragment are shown here (lanes marked G). The protected regions are depicted on the right or left side of the panel.

cloned cDNA at high levels. Extracts from the induced lysogen were prepared, fractionated by SDS−PAGE and used for Southwestern analysis. The protein, the mol. wt of which was ∼160 kd, bound the ^{32}P-labeled oligonucleotide probe corresponding to the CRE of the human somatostatin gene, but did not bind the mutated probe (Figure 1b, lanes 2 and 4). The 160 kd protein was not detected in proteins from bacteria not harboring prophage (Figure 1b, lane 1), bacteria not treated with isopropyl β-D-thiogalactopyranoside (IPTG) or from lysogens harboring other clones (data not shown).

DNA binding properties of CRE-BP1

To examine the DNA binding properties of the protein encoded by λCRE-BP1-1 cDNA, it was extracted from the induced lysogens and partially purified by ammonium sulfate precipitation and sequence-specific DNA affinity column chromatography (see Materials and methods). The partially purified protein, which accounted for 20−30% of the total protein, was first studied by a gel-retardation assay with use of a ^{32}P-labeled probe containing the CRE sequence of the human or rat somatostatin gene. As shown in Figure 2a (lanes 1 and 3), the fusion proteins gave rise to three shifted

a

```
-26                                             GAATTCTGTGATAAGTTATTCAACTT
                        → λCRE-BP1-1
1     ATG AAA TTC AAG TTA CAT GTG AAT TCT GCC AGG CAA TAC AAG GAC CTG
      Met Lys Phe Lys Leu His Val Asn Ser Ala Arg Gln Tyr Lys Asp Leu    16

49    TGG AAT ATG AGT GAT GAC AAA CCC TTT CTA TGT ACT GCG CCT GGA TGT
      Trp Asn Met Ser Asp Asp Lys Pro Phe Leu Cys Thr Ala Pro Gly Cys    32

97    GGC CAG CGT TTT ACC AAC GAG GAT CAT TTG GCT GTC CAT AAA CAT AAA
      Gly Gln Arg Phe Thr Asn Glu Asp His Leu Ala Val His Lys His Lys    48

145   CAT GAG ATG ACA CTG AAA TTT GGT CCA GCA CGT AAT GAC AGT GTC ATT
      His Glu Met Thr Leu Lys Phe Gly Pro Ala Arg Asn Asp Ser Val Ile    64

193   GTG GCT GAT CAG ACC CCA ACA CCA ACA AGA TTC TTG AAA AAC TGT GAA
      Val Ala Asp Gln Thr Pro Thr Pro Thr Arg Phe Leu Lys Asn Cys Glu    80

241   GAA GTG GGT TTG TTT AAT GAG TTG GCG AGT CCA TTT GAG AAT GAA TTC
      Glu Val Gly Leu Phe Asn Glu Leu Ala Ser Pro Phe Glu Asn Glu Phe    96

289   AAG AAA GCT TCA GAA GAT GAC ATT AAA AAA ATG CCT CTA GAT TTA TCC
      Lys Lys Ala Ser Glu Asp Asp Ile Lys Lys Met Pro Leu Asp Leu Ser   112
    I
337   CCT CTT GCA ACA CCT ATC ATA AGA AGC AAA ATT GAG GAG CCT TCT GTT
      Pro Leu Ala Thr Pro Ile Ile Arg Ser Lys Ile Glu Glu Pro Ser Val   128
                                 II
385   GTA GAA ACA ACT CAC CAG GAT AGT CCT TTA CCT CAC CCA GAG TCT ACT
      Val Glu Thr Thr His Gln Asp Ser Pro Leu Pro His Pro Glu Ser Thr   144

433   ACC AGT GAT GAG AAG GAA GTA CCA TTG GCA CAA ACT GCA CAG CCC ACA
      Thr Ser Asp Glu Lys Glu Val Pro Leu Ala Gln Thr Ala Gln Pro Thr   160

481   TCA GCT ATT GTT CGT CCA GCA TCA TTA CAG GTT CCC AAT GTG CTG CTT
      Ser Ala Ile Val Arg Pro Ala Ser Leu Gln Val Pro Asn Val Leu Leu   176

529   ACA AGT TCT GAC TCA AGT GTA ATT ATT CAG CAG GCA GTA CCT TCA CCA
      Thr Ser Ser Asp Ser Ser Val Ile Ile Gln Gln Ala Val Pro Ser Pro   192

577   ACC TCA AGT ACT GTA ATC ACC CAG GCA CCA TCC TCT AAC AGG CCA ATT
      Thr Ser Ser Thr Val Ile Thr Gln Ala Pro Ser Ser Asn Arg Pro Ile   208

625   GTC CCT GTA CCA GGC CCA TTT CCT CTT CTG TTA CAT CTT CCT AGT GGA
      Val Pro Val Pro Gly Pro Phe Pro Leu Leu Leu His Leu Pro Ser Gly   224

673   CAA ACC ATG CCT GTT GCT ATT CCT GCA TCA ATT ACA AGT TCT AAT GTG
      Gln Thr Met Pro Val Ala Ile Pro Ala Ser Ile Thr Ser Ser Asn Val   240

721   CAT GTT CCA GCT GCA GTC CCA CTC GTT CGA CCA GTC ACC ATG GTG CCT
      His Val Pro Ala Ala Val Pro Leu Val Arg Pro Val Thr Met Val Pro   256

769   AGT GTT CCA GGA ATC CCA GGT CCT TCC TCT CCC CAA CCA GTA CAG TCA
      Ser Val Pro Gly Ile Pro Gly Pro Ser Ser Pro Gln Pro Val Gln Ser   272

817   GAA GCA AAA ATG AGA TTA AAA GCT GCT TTG ACC CAG CAA CAT CCT CCA
      Glu Ala Lys Met Arg Leu Lys Ala Ala Leu Thr Gln Gln His Pro Pro   288

865   GTT ACC AAT GGT GAT ACT GTC AAA GGT CAT GGT AGC GGA TTG GTT AGG
      Val Thr Asn Gly Asp Thr Val Lys Gly His Gly Ser Gly Leu Val Arg   304

913   ACT CAG TCA GAG GAA TCT CGA CCG CAG TCA TTA CAA CAG CCA GCC ACA
      Thr Gln Ser Glu Glu Ser Arg Pro Gln Ser Leu Gln Gln Pro Ala Thr   320

961   TCC ACT ACA GAA ACT CCG GCT TCT CCA GCT CAC ACA ACT CCA CAG ACC
      Ser Thr Thr Glu Thr Pro Ala Ser Pro Ala His Thr Thr Pro Gln Thr   336

1009  CAA AGT ACA AGT GGT CGT CGG AGA AGA GCA GCT AAC GAA GAT CCT GAT
      Gln Ser Thr Ser Gly Arg Arg Arg Arg Ala Ala Asn Glu Asp Pro Asp   352

1057  GAA AAA AGG AGA AAG TTT TTA GAG CGA AAT AGA GCA GCA GCT TCA AGA
      Glu Lys Arg Arg Lys Phe Leu Glu Arg Asn Arg Ala Ala Ala Ser Arg   368

1105  TGC CGA CAA AAA AGG AAA GTC TGG GTT CAG TCT TTA GAG AAG AAA GCT
      Cys Arg Gln Lys Arg Lys Val Trp Val Gln Ser [Leu] Glu Lys Lys Ala   384

1153  GAA GAC TTG AGT TCA TTA AAT GGT CAG CTG CAG AGT GAA GTC ACC CTG
      Glu Asp [Leu] Ser Ser Leu Asn Gly Gln [Leu] Gln Ser Glu Val Thr Leu   400

1201  CTG AGA AAT GAA GTG GCA CAG CTG AAA CAG CTT CTT CTG GCT CAT AAA
      [Leu] Arg Asn Glu Val Ala Gln [Leu] Lys Gln Leu Leu Leu Ala His Lys   416

1249  GAT TGC CCT GTA ACC GCC ATG CAG AAG AAA TCT GGC TAT CAT ACT GCT
      Asp Cys Pro Val Thr Ala Met Gln Lys Lys Ser Gly Tyr His Thr Ala   432

1297  GAT AAA GAT GAT AGT TCA GAA GAC ATT TCA GTG CCG AGT AGT CCA CAT
      Asp Lys Asp Asp Ser Ser Glu Asp Ile Ser Val Pro Ser Ser Pro His   448

1345  ACG GAA GCT ATA CAG CAT AGT TCG GTC AGC ACA TCC AAT GGA GTC AGT
      Thr Glu Ala Ile Gln His Ser Ser Val Ser Thr Ser Asn Gly Val Ser   464

1393  TCA ACC TCC AAG GCA GAA GCT GTA GCC ACT TCA GTC CTC ACC CAG ATG
      Ser Thr Ser Lys Ala Glu Ala Val Ala Thr Ser Val Leu Thr Gln Met   480

1441  GCG GAC CAG AGT ACA GAG CCT GCT CTT TCA CAG ATC GTT ATG GCT CCT
      Ala Asp Gln Ser Thr Glu Pro Ala Leu Ser Gln Ile Val Met Ala Pro   496

1489  TCC TCC CAG TCA CAG CCC TCA GGA AGT TGA TTAAAAACCTGCAGTACAACAGT
      Ser Ser Gln Ser Gln Pro Ser Gly Ser ***                           505

1542  TTAGATACTCATTAGTGACTTCAAAGGGAAATCAAGGAAAGACCAGTTTCCATTTATGCGAAA

1605  TCTGTGGTTGTAAATTT
```

b

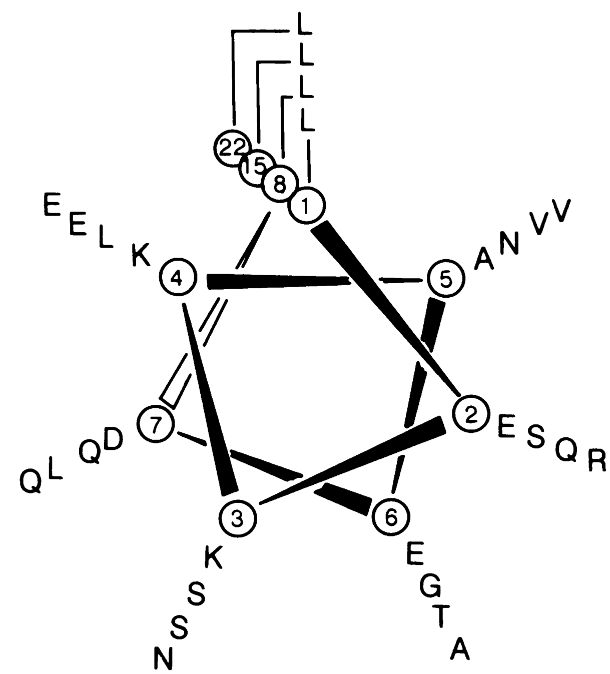

Fig. 3. Structural analysis of cloned CRE-BP1 cDNA. (**a**) Nucleotide sequence of CRE-BP1 cDNA and the primary structure deduced for the protein. The cloned DNAs were sequenced by the dideoxy method (Sanger *et al.*, 1977). The 5′ terminus of cDNA derived from λCRE-BP1-1 is marked by an arrow. Leucine residues in the leucine zipper structure are boxed. The doubly underlined sequences I and II are the putative sites for protein kinase A and protein kinase C respectively. The *Eco*RI site at the junction between the β-galactosidase and cDNA insert is indicated by a line above. The region homologous with the cluster of basic amino acids in the human *fos* protein is indicated by underlining. The sequence indicated by a dotted line is the putative transcriptional activation domain. (**b**) Helical wheel analysis of the leucine zipper portion of CRE-BP1. The amino acid sequence of a portion of CRE-BP1 (amino acids 353−377) is displayed end-to-end down the axis of a schematic α-helix as done for C/EBP by Landschultz *et al.* (1988a). The residue closest to the amino-terminal included here is the leucine residue 353 of CRE-BP1; it is placed at position number one of the idealized helix. The helical wheel consists of seven 'spokes', corresponding to the fit of seven amino acid residues into every two α-helical turns. Amino acid residues are given with one-letter codes.

bands (bands 1−3). When the mutated probe of the human somatostatin gene (see Figure 1a) was used, two of the shifted bands (1 and 2) were not detected and the intensity of band 3 was much less than with the wild-type probe (Figure 2a, lane 2). These results indicated that bands 1 and 2 corresponded to specific DNA−protein complexes.

The purified fusion protein was also analyzed by DNase I footprinting to test its binding to the CRE of the rat somatostatin gene. The purified protein gave 23-nt protection in the region containing this CRE (nt −36 to −58 of the gene; Figure 2b). This protected region in the rat somatostatin gene contained the CRE sequence TGACGTCA in its center, and corresponded well to the region protected by the CREB derived from PC12 cells or rat brain (Montminy and Bilezikjian, 1987; Yamamoto *et al.*, 1988). To find whether the fusion protein encoded by λCRE-BP1-1 also bound to the sequence recognized by ATF of the adenovirus early genes, a ^{32}P-labeled DNA fragment containing the regulatory region of the adenovirus E4 gene was used for DNase I footprinting. The fusion protein protected three regions (nt −236 to −225, −167 to −158 and −147 to −134 in the adenovirus E4 gene; Figure 2b) that corresponded well to the region protected by ATF derived from HeLa cell (Lin and Green, 1988).

Structure of the cDNA encoding CRE-BP1

The λCRE-BP1-1 contained a cDNA insert that was ~3.7 kb long. To obtain clones containing more 5′ sequence, the 41 bp *Eco*RI−*Xba*I cDNA fragment (nt 283−324) of the λCRE-BP1-1 insert was used as a probe to screen the library again, and reactive inserts were detected at a frequency of ~1 in 10 000. The entire coding sequence for CRE-BP1 as predicted from the sequence of the two cDNA clones is shown in Figure 3a.

Analysis of the cDNA sequence showed that a long open reading frame started at nt 1 and a stop codon TAA was localized 15 bp upstream of the ATG sequence at nt 1, which suggests that this ATG sequence is indeed the initiation codon for CRE-BP1 mRNA (Figure 3a). These results meant that the 1515-nt open reading frame between nt 1 and 1515 (Figure 3b) was the coding region for human CRE-BP1. Its length corresponded to an encoded protein of 505 amino acids with a calculated mol. wt of 54.5 kd. This mol. wt is larger than that of CREB purified from PC12 cells (Montminy and Bilezikjian, 1987) and rat brain (Yamamoto *et al.*, 1988), suggesting that CRE-BP1 is different from CREB.

Zinc fingers and helix−turn−helix motifs were not detected in human CRE-BP1. The amimo terminus of CRE-BP1 had no clear homology with other reported proteins, but the carboxy terminus contained a sequence of 29 amino acids similar to the leucine zipper structure (Landschultz *et al.*, 1988a) found in the enhancer binding protein C/EBP (Landschultz *et al.*, 1988b) and nuclear oncogene products such as *jun* and *fos*. Leucines appeared at every seventh position over this region of 29 amino acids (380−408; Figure 3a,b). This distribution was not simply a result of the abundance of this residue, because the region of 29 amino acids contained only two other leucines. When the amino acid sequence of this region was arranged on a schematic α-helix, this segment showed strong amphipathy (Figure 3b) as proposed by Landschultz *et al.* (1988a).

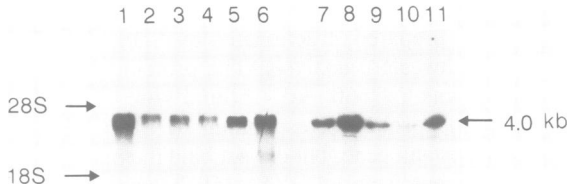

Fig. 4. Northern blot analysis of CRE-BP1 RNAs. Two micrograms (**lanes 1−4**) or 3 μg (**lanes 5** and **6**) of poly(A)$^+$ RNA or else 20 μg of total RNAs (**lanes 7−11**) were studied by Northern blot analysis with a ^{32}P-labeled CRE-BP1 cDNA probe. **Lane 1**, brain; **lane 2**, lung; **lane 3**, liver; **lane 4**, kidney; **lane 5**, HeLa cells; **lane 6**, human adenocarcinoma cell line MKN7; **lane 7**, human epidermoid carcinoma cell line KB; **lane 8**, human B-cell line BALL-1; **lane 9**, rat pheochromocytoma cell line PC12; **lane 10**, mouse fibroblast NIH3T3; **lane 11**, African green monkey kidney cell line CV1. The X-ray film was exposed for 10 h. The position of rRNA and the estimated size of CRE-BP1 mRNA are shown to the left and right respectively.

Expression of CRE-BP1 mRNA

Many genes have a CRE in their regulatory elements (Montminy *et al.*, 1986), and the CRE can act as an inducible or constitutive enhancer depending on the cell type and promoter (Sassone-Corsi, 1988; Kanei-Ishii and Ishii, 1989), so it is likely that CRE-binding protein is involved in the regulation of a set of genes in various cell types. We therefore examined the expression of CRE-BP1 mRNA in human embryo tissues and cell lines (Figure 4). With cDNA as probe, a band corresponding to the length of ~4.0 kb was detected in all of the RNA samples by RNA blotting analysis, suggesting that the sequence of the CRE-BP1 gene was conserved between human and mouse species. The mRNA expression level seemed to be moderate, and the CRE-BP1 mRNA was relatively abundant in brain. This might be because a set of genes specific to neuronal cells such as the genes encoding preproenkephalin (Comb *et al.*, 1986) and somatostatin (Montminy *et al.*, 1986) have CRE in their regulatory regions.

Discussion

In searching for cellular factor involved in transcriptional activation by a *trans*-acting *tax* protein of HTLV-1 (human T cell leukemia virus type 1), we have screened an oligo(dT)-primed HUT102 cDNA library by a Southwestern procedure using an oligonucleotide of *tax*-dependent enhancer (Fujisawa *et al.*, 1986). One of the clones, λHUT-CRE, was found to code for the same protein as CRE-BP1. This coincidence is probably due to the CRE sequence in the *tax*-dependent enhancer sequence. The λHUT-CRE clone has 4.0 kb insert, which is close to the size of mRNA and 285 bp longer than that of λCRE-BP1-1 at the 5′ end, furthermore it has a poly(A) signal together with a poly(A) stretch at the 3′ region. Therefore, the λHUT-CRE clone seems to have a full-sized cDNA sequence of CRE-BP1 mRNA.

Our results indicate that CRE-BP1 is one of the DNA binding proteins containing a leucine zipper structure, as do C/EBP, *jun*/AP-1 and *fos* protein. C/EBP and GCN4 exist in solution as stable dimers (Hope and Struhl, 1987;

```
Human c-fos      138 EKRRIRRERNKMAAAKCRNRR 158
                     ::::   :::  ::  ::  :
CRE-BP1          353 EKRRKFLERNRAAASRCRQKRKVWV 377
                      ::   :  :: ::  :: : :  :
Placental CREB   269 RKREVRLMKNREAARECRRKKKEYV 294
```

Fig. 5. Comparison of the amino acid sequences of the *fos* and human placental CREB proteins. The amino acid numbers are shown at both sides of the sequence, where +1 corresponds to the N-terminal residue of each protein. Dots connect identical amino acid residues.

Landschultz *et al.*, 1988a), so the leucine zipper probably represents the dimerization domain of this class of DNA binding proteins. The amino acid sequence of CRE-BP1 showed that the protein contained a high proportion of basic residues in a region of 33 amino acids (236−268) next to its leucine zipper region. In this basic region, 12 amino acids out of 21 were identical with a region of the human c-*fos* protein (Figure 5). This highly basic region of the two CRE-BP1 polypeptides may be juxtaposed by the leucine zipper in a manner suitable for sequence-specific recognition of DNA, as postulated by Landschultz *et al.* (1988a). Two shifted bands observed in the gel-retardation assay (Figure 2a) could arise from the dimer and monomer forms of CRE-BP1.

Two structural features for the transcriptional activation domain have been reported: an acidic domain (Hope and Struhl, 1986; Ma and Ptashne, 1987) and a glutamine-rich domain (Courey and Tjian, 1988). The N-terminal region (amino acids 124−148) has seven acidic amino acids and a net negative charge of 7. Therefore, this acidic region may be a transcriptional activation domain as reported for the yeast GCN4 and GAL4 transcriptional activator. Findings by Yamamoto *et al.* (1988) indicate that the dimerization and transcriptional efficacy of CREB protein are regulated by phosphorylation. In fact, CREB purified from rat brain is phosphorylated by both protein kinase A and protein kinase C. The CRE-BP1 expressed in *E.coli* by the T7 expression vector was also phosphorylated efficiently by both kinases (data not shown). The serine residue at amino acid 121 of CRE-BP1, which has arginine and lysine at both its amino- and carboxyl-terminal sides, may be the site of phosphorylation by protein kinase C, because the seryl residues in bovine myelin basic protein, which have basic amino acids at both the amino- and carboxyl-terminal sides are phosphorylated efficiently by protein kinase C (Kishimoto *et al.*, 1985). The sequence of Lys-Lys-Ala-Ser at amino acids 97−100 is consistent with the consensus sequence (basic-basic-X-Ser or Thr) of the site phosphorylated by protein kinase A.

During preparation of this manuscript, Hoeffler *et al.*, (1988) reported the isolation of a cDNA clone encoding CREB from a human placental λgt11 library using a Southwestern method. After submission of this manuscript, the isolation of a cDNA clone for rat brain CREB by use of amino acid sequence information from purified CREB was also reported (Gonzalez *et al.*, 1989). The sequences of their cDNAs showed that the CREB also has the leucine zipper structure, but the sequence of the CREB is different from that of CRE-BP1. The human placental CREB and the rat brain CREB have mol. wts of 38 and 43 kd respectively, and both have four leucine residues as the leucine zipper structure. CRE-BP1 had a mol. wt of 54.5 kd and five leucine residues. The only clear homology between CREB and CRE-BP1 was in a region of 24 amino acids (353−377) immediately adjacent to the leucine zipper regions, which also has homology with the human c-*fos* protein, as described above (Figure 5). In this highly basic region, 12 amino acids out of 24 were identical.

Our results suggest the heterogeneity of CRE binding protein and the complexity of the cAMP pathway of signal transduction. AP-1/c-*jun* protein binds to the CRE sequence with affinity one-fifth to one-tenth that of binding to the AP-1 sequence and CRE-BP1 synthesized in *E.coli* binds to the AP-1 sequence in the simian virus 40 (SV40) enhancer with affinity one-fifth or less that of binding to the CRE sequence (data not shown). Therefore, CREB, CRE-BP1 and AP-1/c-*jun*, all of which have the leucine zipper structure, may belong to the same super-gene family. The cDNA clone isolated encoding CRE-BP1 will be a useful probe to study the mechanism by which this protein stimulates transcription.

Materials and methods

Screening of protein replica filters with probe containing CRE oligomer

A λgt11 cDNA library from human fetal brain mRNA (a gift from R.L.Neve; Neve *et al.*, 1986) was screened as described elsewhere (Sakura *et al.*, 1988) with the modifications of Singh *et al.* (1988) and Vinson *et al.* (1988). Two complementary oligonucleotides, 5′-TCGAGACAGTCCCC-CGTGACGTCACCCGGGAGCCCC-3′ and 5′-CGAGGGGCTCCCG-GGTGACGTCACGGGGGACTGTCT-3′, corresponding to the CRE of the human fibronectin gene were annealed, phosphorylated and ligated. Ligated products containing six tandem repeats of the oligonucleotide were isolated by acrylamide gel electrophoresis and cloned into the *Bam*HI site of pUC19. The DNA fragment containing six tandem repeats of the oligonucleotide was isolated from this plasmid, nick-translated and used as a probe for screening.

Analysis of the DNA binding protein produced in E.coli

Y1089 lysogens harboring one of the positive clones, λCRE-BP1-1, were isolated and extracts were prepared from an induced culture as described by Huynh *et al.* (1985). Extracts were separated on 10% SDS−polyacrylamide gel and studied by Southwestern analysis as described previously (Sakura *et al.*, 1988). Extracts prepared from 1 l of an induced culture of λCRE-BP1-1 lysogens were mixed with solid ammonium sulfate to 33% saturation and kept at 4°C for 1 h with gentle stirring. The precipitate formed was recovered by centrifugation, suspended in T buffer [50 mM Tris−HCl, pH 7.9, 1 mM EDTA, 20% glycerol and 1 mM dithiothreitol (DTT)] containing 0.1 M KCl and dialyzed against the same buffer. This protein sample containing ~100 μg of fusion protein was mixed at 4°C for 8 h with 1 ml of the affinity matrix, which was made by use of the rat somatostatin CRE oligonucleotide shown in the legend to Figure 2 as a ligand as described by Kadonaga and Tjian (1986). After the matrix was packed in a column and washed with T buffer containing 0.1 M KCl, bound material was eluted with 1.1 M KCl (volume of each fraction, 1 ml).

A gel-retardation assay was done as described before (Maekawa *et al.*, 1989). In brief, 0.5 μl of the affinity-purified fusion protein (~15 ng) was incubated for 15 min at 25°C with 1 ng of ^{32}P-labeled oligonucleotide and 1 μg of poly(dI-dC) in 15 μl of binding buffer [10 mM Tris−HCl, pH 7.9, 50 mM KCl, 1 mM DTT, 0.04% Nonidet P-40 and 5% (v/v) glycerol]. The reaction mixture was then put onto a 4% polyacrylamide gel in 0.25 × TBE (25 mM Trizma base, 25 mM boric acid and 1 mM EDTA) and electrophoresed.

For DNase I footprinting experiments, either the *Eco*RI−*Hin*dIII fragment of the plasmid pUCRSCRE42 generated by the insertion of the 42 bp of the synthetic oligonucleotide shown in the legend in Figure 2 corresponding to the rat somatostatin CRE into the *Bam*HI site of pUC19 or else the *Eco*RI−*Hae*III fragment (nt −330 to −10 of the E4 gene) of the plasmid pEIV (Hanaka *et al.*, 1987) containing the regulatory region of the adenovirus E4 gene was ^{32}P-labeled at the 5′ end of the lower or upper strand. DNA binding reactions and DNase I digestion were carried out in a volume of 60 μl with 1−4 ng (~10 fmol) of an end-labeled DNA fragment and 60 ng of the purified fusion protein in the same buffer as for the gel-retardation assays. Reaction mixtures were incubated for 15 min at 25°C and then

digested for 1 min at 25°C with different amounts of DNase I (Takara Shuzo Co.) after addition of $MgCl_2$ to the final concentration of 5 mM. The digestion products were separated on a denaturing 7 or 8% polyacrylamide gel and made visible by autoradiography.

RNA blotting analysis

Total RNA was isolated by the guanidium thiocyanate method (Chirgwin *et al.*, 1979). RNA blotting analysis was performed essentially as described by Thomas (1980) with the CRE-BP1 cDNA probe.

Acknowledgements

We thank R.L.Neve for the λgt11 cDNA library prepared from human fetal brain mRNA, Tadashi Yamamoto and Yuji Yamanashi for the RNA samples prepared from human fetal tissues, Hiroshi Handa for the plasmid containing the adenovirus E4 gene, Akira Kishimoto for helpful discussion, and Alfred Johnson and Caroline Latta for critical reading of the manuscript.

References

Andrisani,O.M., Pot,D.A., Zhu,Z. and Dixon,J.E. (1988) *Mol. Cell. Biol.*, **8**, 1947–1956.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.

Comb,M., Birnberg,N.C., Seasholtz,A., Herbert,E. and Goodman,H.M. (1986) *Nature*, **323**, 353–356.

Courey,A.J. and Tjian,R. (1988) *Cell*, **55**, 887–898.

Fujisawa,J., Seiki,M., Sato,M. and Yoshida,M. (1986) *EMBO J.*, **5**, 713–718.

Gonzalez,G.A., Yamamoto,K.K., Fischer,W.H., Darr,D., Menzel,P., Briggs,W., III, Vale,W.W. and Montminy,M.R. (1989) *Nature*, **337**, 749–752.

Hanaka,S., Nishigaki,T., Sharp,P.A. and Handa,H. (1987) *Mol. Cell. Biol.*, **7**, 2578–2587.

Hardy,S. and Shenk,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4171–4175.

Hoeffler,J.P., Meyer,T.E., Yun,Y., Jameson,J.L. and Habener,J.F. (1988) *Science*, **242**, 1430–1433.

Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885–894.

Hope,I.A. and Struhl,K. (1987) *EMBO J.*, **6**, 2781–2784.

Hurst,H.C. and Jones,N.C. (1987) *Genes Dev.*, **1**, 1132–1146.

Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning: A Practical Approach.* IRL Press, Oxford, pp. 49–78.

Kadonaga,J.T. and Tjian,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5889–5893.

Kanei-Ishii,C. and Ishii,S. (1989) *Nucleic Acids Res.*, **17**, 1521–1536.

Kishimoto,A., Nishiyama,K., Nakanishi,H., Uratsuji,Y., Nomura,H., Takeyama,Y. and Nishizuka,Y. (1985) *J. Biol. Chem.*, **257**, 12492–12498.

Landschultz,W.H., Johnson,P.F. and McKnight,S.L. (1988a) *Science*, **240**, 1759–1764.

Landschultz,W.H., Johnson,P.F., Adashi,E.Y., Graves,B.J. and McKnight,S.L. (1988b) *Genes Dev.*, **2**, 786–800.

Lee,K.A.W., Hai,T.-Y., SivaRaman,L., Thimmappaya,B., Hurst,H.C., Jones,N.C. and Green,M.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8355–8359.

Lin,Y.-S. and Green,M.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3396–3400.

Ma,J. and Ptashne,M. (1987) *Cell*, **48**, 847–853.

Maekawa,T., Itoh,F., Okamoto,T., Kurimoto,T., Imamoto,F. and Ishii,S. (1989) *J. Biol. Chem.*, **264**, 2826–2831.

Montminy,M.R. Sevarino,K.A., Wagner,J.A., Mandel,G. and Goodman,R.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6682–6686.

Montminy,M.R. and Bilezikjian,L.M. (1987) *Nature*, **328**, 175–178.

Nairn,A.C., Hemmings,H.C. and Greengard,P. (1985) *Annu. Rev. Biochem.*, **54**, 931–976.

Neve,R.L., Harris,P., Kosik,K.S., Kurnit,D.M. and Donlon,T.A. (1986) *Mol. Brain Res.*, **1**, 271–280.

Nishizuka,Y. (1984) *Nature*, **308**, 693–697.

Sakura,H., Maekawa,T., Imamoto,F., Yasuda,K. and Ishii,S. (1988) *Gene*, **73**, 497–505.

Sanger,F., Nichlen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Sassone-Corsi,P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7192–7196.

Singh,H., LeBowitz,J.H., Baldwin,A.S., Jr and Sharp,P.A. (1988) *Cell*, **52**, 415–429.

Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.

Vinson,C.R., LaMarco,K.L., Johnson,P.F., Landschultz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.

Yamamoto,K.K., Gonzalez,G.A., Biggs,W.H. and Montminy,M.R. (1988) *Nature*, **334**, 494–498.

Received on February 23, 1989; revised on April 20, 1989