The EMBO Journal vol.11 no.10 pp.3663–3671, 1992

Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene

Hiroaki Imataka, Kazuhiro Sogawa, Ken-ichi Yasumoto, Yasuo Kikuchi, Kiminobu Sasano[1], Akira Kobayashi, Masanori Hayami[2] and Yoshiaki Fujii-Kuriyama[3]

Department of Chemistry, Faculty of Science, [1]Department of Pathology, Medical School, Tohoku University, Sendai 980 and [2]Institute for Virus Research, Kyoto University, Kyoto, Japan
[3]Corresponding author

Communicated by A.Tsugita

The cDNAs for two DNA binding proteins of BTE, a GC box sequence in the promoter region of the P-450IA1(CYP1A1) gene, have been isolated from a rat liver cDNA library by using the BTE sequence as a binding probe. While one is for the rat equivalent to human Sp1, the other encodes a primary structure of 244 amino acids, a novel DNA binding protein designated BTEB. Both proteins contain a zinc finger domain of Cys-Cys/His-His motif that is repeated three times with sequence similarity of 72% to each other, otherwise they share little or no similarity. The function of BTEB was analysed by transfection of plasmids expressing BTEB and/or Sp1 with appropriate reporter plasmids into a monkey cell line CV-1 and compared with Sp1. BTEB and Sp1 activated the expression of genes with repeated GC box sequences in promoters such as the simian virus 40 early promoter and the human immunodeficiency virus-1 long terminal repeat promoter. In contrast, BTEB repressed the activity of a promoter containing BTE, a single GC box of the CYP1A1 gene that is stimulated by Sp1. When the BTE sequence was repeated five times, however, BTEB turned out to be an activator of the promoter. RNA blot analysis showed that mRNAs for BTEB and Sp1 were expressed in all tissues tested, but their concentrations varied independently in tissues. The former mRNA was rich in the brain, kidney, lung and testis, while the latter was relatively abundant in the thymus and spleen. These results suggest that BTEB and Sp1, which recognize the same DNA sequence, exert different effects on the transcription of the genes with different number and arrangement of GC box sequences in the promoter region.

Key words: cytochrome P-450/GC box/Sp1/transcription factor

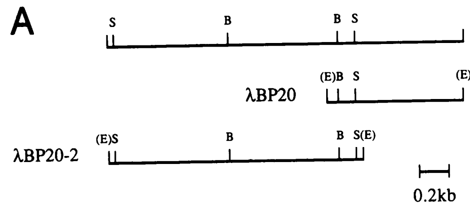
Introduction

Transcription is a major regulatory point in gene expression and its frequency depends largely on interaction of the regulatory proteins with the cognate DNA elements in gene promoters (for review, see Mitchell and Tjian, 1989). Analysis of promoters in a variety of genes has identified a number of distinct DNA elements required for gene expression. A GC box sequence, one of the most common regulatory DNA elements of eukaryotic genes, is recognized by the Sp1 transcription factor; its consensus sequence is represented as 5'-G/T G/A GGCG G/T G/A G/A C/T-3' (Briggs *et al.*, 1986). Arrangement of the GC box sequence in the promoter region varies in number and spatial array among genes. Several genes, including adenovirus E1b (Schmidt *et al.*, 1989) and the transcription factor NFI (Ammendola *et al.*, 1990), contain a single GC box just upstream of the TATA box, while simian virus 40 (SV40) early promoter (Briggs *et al.*, 1986) and human immunodeficiency virus-1 long terminal repeat (HIV-1 LTR) (Jones *et al.*, 1986) contain six and three tandem repeats of the GC box sequence, respectively, just upstream of the TATA box. Multiple GC box sequences are often found within a few hundred base pairs upstream of the transcription start site(s) in promoters of some cell growth-related genes such as genes for H-*ras* (Ishii *et al.*, 1986), c-*myb* (Dvorçk *et al.*, 1989) and insulin receptor (Araki *et al.*, 1991), and genes encoding housekeeping enzymes such as 3-hydroxy-3-methylglutaryl CoA reductase (Reynolds *et al.*, 1984), hypoxanthine phosphoribosyl transferase (Kim *et al.* 1986) and dihydrofolate reductase (Swick *et al.*, 1989). This type of promoter often lacks the TATA box sequence. Although all the above-mentioned gene promoters are considered to be activated by SpI, some GC box sequences could be targets for factors other than Sp1 (Kageyama *et al.*, 1988; Chavrier *et al.*, 1990; Huang *et al.*, 1990).

We have been studying the mechanisms of gene regulation of cytochrome P-450IA1 (CYP1A1), a drug-metabolizing P-450 whose expression is induced by polycyclic aromatic hydrocarbons such as 3-methylcholanthrene and TCDD (Kawajiri *et al.*, 1984). Using chimeric genes consisting of the 5'-flanking sequence of the CYP1A1 gene and CAT structural gene, we found an inducible enhancer element, xenobiotic responsive element (XRE), which is essential for the inducible expression of the gene (Fujisawa-Sehara *et al.*, 1987) and another element, basic transcriptional element (BTE), which is necessary for the constitutive expression of the gene (Yanagida *et al.*, 1990). The BTE sequence of ~20 bp contains a GC box. Here we cloned two cDNAs for proteins that bind to BTE. One is for Sp1 and the other encodes a novel DNA binding factor (designated BTEB or basic transcription element binding protein). Transient transfection analysis indicates that while BTEB, as well as Sp1, is a transcription factor that stimulates promoters with repeated GC boxes, the promoter of CYP1A1 gene, which has a single GC box sequence, is activated by Sp1 but repressed by BTEB.

© Oxford University Press

A



B

```
         .         .         .         .         .         .         .         .         .
GGACAGGACCCCCTTGAGCTCGCCCCTCAGCTGCCACCATGAGCGACCATGAGCGACCAAGATCACTCAATGGATGAAGTGACAGCTGTG   90
                                                M  S  D  Q  D  H  S  M  D  E  V  T  A  V     14

AAGATTGAAAAAGGTGTTGGTGGCAATAACGGGGGTAGCGGGAACGGTGGTGGTGCCGCCTTTTCTCAGACTCGAAGCAGCAGCACAGGC   180
K  I  E  K  G  V  G  G  N  N  G  G  S  G  N  G  G  G  A  A  F  S  Q  T  R  S  S  S  T  G     44

AGTAGCAGCAGCAGTGGTGGCGGAGGAGGGCAGGAATCCCAGCCATCTCCTTTGGCTCTGCTGGCAGCAACCTGCAGCAGAATTGAGTCA   270
S  S  S  S  S  G  G  G  G  G  Q  E  S  Q  P  S  P  L  A  L  L  A  A  T  C  S  R  I  E  S     74

CCCAATGAGAATGAGAACAGCAACAACTCCCAGGGTCCGAGTCAGTCAGGGGGCACAGGTGAACTTGACCTCACAGCCACACAACTTTCA   360
P  N  E  N  E  N  S  N  N  S  Q  G  P  S  Q  S  G  G  T  G  E  L  D  L  T  A  T  Q  L  S     104
                                                      -------->
CAGGGTGCCAATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAACAGTACCAATGGCAGC   450
Q  G  A  N  G  W  Q  I  I  S  S  S  S  G  A  T  P  T  L  K  E  Q  S  G  N  S  T  N  G  S     134
                                                      S                 S
AATGGCAGTGAGTCTTCCAAGAACCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTG   540
N  G  S  E  S  S  K  N  R  T  V  S  G  G  Q  Y  V  V  A  A  T  P  N  L  Q  N  Q  Q  V  L     164

ACAGGTCTACCTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGATGGGCAACAGCTGCAGTTTGCTGCC   630
T  G  L  P  G  V  M  P  N  I  Q  Y  Q  V  I  P  Q  F  Q  T  V  D  G  Q  Q  L  Q  F  A  A     194

ACTGGGGCCCAAGTGCAGCAGGATGGTTCTGGTCAAATACAGATCATACCAGGTGCAAATCAACAGATCATCACAAATAGAGGAAGTGGG   720
T  G  A  Q  V  Q  Q  D  G  S  G  Q  I  Q  I  I  P  G  A  N  Q  Q  I  I  T  N  R  G  S  G     224

GGCAACATTATTGCTGCTATGCCAAATCTACTCCAGCAAGCTGTCCCCCTCCAAGGCCTTGCTAATAATGTGCTCTCAGGACAGACTCAG   810
G  N  I  I  A  A  M  P  N  L  L  Q  Q  A  V  P  L  Q  G  L  A  N  N  V  L  S  G  Q  T  Q     254

TATGTGACCAATGTACCAGTGGCCCTGAATGGGAACATCACCTTGCTACCTGTCAACAGCGTTTCTGCAGCTACCCTGACTCCCAGCTCT   900
Y  V  T  N  V  P  V  A  L  N  G  N  I  T  L  L  P  V  N  S  V  S  A  A  T  L  T  P  S  S     284

CAGGCAGGCACGATCAGCAGCTCTGGATCCCAGGAGAGTGGCTCACAGCCTGTCACCTCAGGGACTGCCATCAGTTCTGCCAGCTTGGTG   990
Q  A  G  T  I  S  S  S  G  S  Q  E  S  G  S  Q  P  V  T  S  G  T  A  I  S  S  A  A  L  V     314
      V                                                           T              S
TCATCACAAGCCAGTTCCAGCTCCTTTTTCACCAATGCCAACAGCTATTCAACAACTACCACCACCAGCAACATGGGAATTATGAACTTT   1080
S  S  Q  A  S  S  S  S  F  F  T  N  A  N  S  Y  S  T  T  T  T  T  S  N  M  G  I  M  N  F     344

ACCAGCAGCGGATCATCAGGGACTAGTTCTCAAGGCCAGACATCCCAGAGAGTTGGTGGGCTGCAAGGGTCTGATTCTCTAAACATCCAG   1170
T  S  S  G  S  S  G  T  S  S  Q  G  Q  T  S  Q  R  V  G  G  L  Q  G  S  D  S  L  N  I  Q     374
   T                    N                 P           S                    A
CAGAACCAGACATCAGGAGGATCACTGCAAGGAAGTCAGCAAAAAGAGGGAGAGCAAAGTCAGCAGACACAGCAACAACAAATTCTTATT   1260
Q  N  Q  T  S  G  G  S  L  Q  G  S  Q  Q  K  E  G  E  Q  S  Q  Q  T  Q  Q  Q  Q  I  L  I     404
                              A  G                       N
CAGCCTCAGCTAGTTCAAGGTGGACAAGCTCTTCAGGCCCTCCAAGCAGCACCATTGTCAGGACAGACCTTCACAACTCAAGCTATTTCC   1350
Q  P  Q  L  V  Q  G  G  Q  A  L  Q  A  L  Q  A  A  P  L  S  G  Q  T  F  T  T  Q  A  I  S     434

CAGGAAACCCTCCAGAACCTCCAGCTTCAGGCTGTTCAAAACTCTGGCCCCATCATCATTCGGACACCAACCGTGGGGCCCAATGGACAG   1440
Q  E  T  L  Q  N  L  Q  L  Q  A  V  Q  N  S  G  P  I  I  I  R  T  P  T  V  G  P  N  G  Q     464
                                    P
GTCAGTTGGCAGACCCTTCAGCTGCAGAACCTTCAAGTTCAGAACCCACAAGCCCAGACAATCACCTTGGCCCCTATGCAGGGTGTTTCC   1530
V  S  W  Q  T  L  Q  L  Q  N  L  Q  V  Q  N  P  Q  A  Q  T  I  T  L  A  P  M  Q  G  V  S     494

TTGGGGCAGACTAGCAGCAGCAATACCACCCTTACACCCATTGCCTCAGCTGCCTCCATTCCTGCTGGCACAGTCACTGTGAATGCTGCT   1620
L  G  Q  T  S  S  S  N  T  T  L  T  P  I  A  S  A  A  S  I  P  A  G  T  V  T  V  N  A  A     524

CAACTGTCCTCCATGCCAGGCCTCCAGACCATTAACCTCAGTGCATTGGGTACTTCAGGGATCCAGGTGCACCAGCTTCCAGGCCTGCCT   1710
Q  L  S  S  M  P  G  L  Q  T  I  N  L  S  A  L  G  T  S  G  I  Q  V  H  Q  L  P  G  L  P     554
                                                                        P  I  Q
TTGGCTATAGCAAACACCCCAGGTGATCATGGAGCTCAACTTGGCCTTCATGGACCTGGCGGTGATGGAATACATGATGAGACAGCAGGT   1800
L  A  I  A  N  T  P  G  D  H  G  A  Q  L  G  L  H  G  P  G  G  D  G  I  H  D  E  T  A  G     584
               A                                      A                       D
GGAGAAGAAGGAGAGAACAGCCCTGATCCCCAACCCCAAGCTGGTCGCAGGACTCGGCGGGAAGCATGTACATGCCCCTATTGTAAAGAC   1890
G  E  E  G  E  N  S  P  D  P  Q  P  Q  A  G  R  R  T  R  R  E  A  C  T  C  P  Y  C  K  D     614
                           A
AGTGAGGGGAGGGGCTCTGGAGATCCTGGCAAAAAGAAACAGCACATTTGTCATATCCAAGGATGCGGCAAAGTATATGGCAAAACCTCA   1980
S  E  G  R  G  S  G  D  P  G  K  K  K  Q  H  I  C  H  I  Q  G  C  G  K  V  Y  G  K  T  S     644

CATCTCCGTGCACACTTGCGCTGGCATACAGGGGAGAGGCCATTCATGTGTAATTGGTCATACTGTGGGAAGCGCTTTACTCGTTCGGAT   2070
H  L  R  A  H  L  R  W  H  T  G  E  R  P  F  M  C  N  W  S  Y  C  G  K  R  F  T  R  S  D     674
                                                   T
GAGCTTCAGAGACATAAACGTACACATACAGGAGAGAAGAAATTTGCCTGCCCTGAGTGCCCTAAGCGTTTCATGAGGAGCGATCATCTG   2160
E  L  Q  R  H  K  R  T  H  T  G  E  K  K  F  A  C  P  E  C  P  K  R  F  M  R  S  D  H  L     704

TCAAAGCATATCAAGACCCACCAGAACAAAAAGGGAGGCCCAGGTGTAGCCTTGAGCGTGGGCACATTGCCCCTGGACAGTGGGGCAGGT   2250
S  K  H  I  K  T  H  Q  N  K  K  G  G  P  G  V  A  L  S  V  G  T  L  P  L  D  S  G  A  G     734

TCAGAAGGCAGTGGCACTGCCACTCCTTCAGCCCTTATTACCACCAATATGGTAGCCATGGAGGCCATCTGTCCAGAGGGTATTGCCCGT   2340
S  E  G  S  G  T  A  T  P  S  A  L  I  T  T  N  M  V  A  M  E  A  I  C  P  E  G  I  A  R     764

CTTGCCAACAGTGGCATCAACGTCATGCAGGTGACAGAGCTGCAGTCCATTAATATCAGTGGCAATGGTTTCTGAGATTAGACGCCAGGG   2430
L  A  N  S  G  I  N  V  M  Q  V  T  E  L  Q  S  I  N  I  S  G  N  G  F  ***                  788
                                 A  D
CCAGAGAGACATATGGGCCAATACCCAAGCCTGGGATGCAAGTAG                                               2475
```

Results

cDNA cloning of BTE binding factors

A rat liver cDNA library was constructed in the λgt11 expression vector. About 1.2×10^6 plaques were screened with ^{32}P-labelled BTE as a probe to isolate two positive clones, λBP20 and λBP26. Sequence analysis of the cloned cDNAs suggested that λBP20 is a cDNA clone equivalent to human Sp1, and λBP26 is a novel protein binding to BTE. To obtain cDNAs that encode the remaining two-thirds of the N-terminal portion of rat Sp1, we constructed another cDNA library by the primer extension method using a primer that was complementary to a 5′ part of λBP20 and screened it. An extended cDNA clone, λBP20-2, which covered the remaining sequence, was obtained (Figure 1).

To examine whether a protein encoded by λBP26 specifically binds to the BTE sequence, cDNA of λBP26 was inserted into an expression vector with the T7 promoter to express the binding protein in *Escherichia coli* (Studier and Moffatt, 1986). The extracts were prepared from the bacteria and subjected to gel mobility shift assay and DNase I footprint analysis using a DNA fragment containing the BTE sequence. When extracts prepared from bacteria expressing a protein from λBP26 cDNA were incubated with the ^{32}P-labelled BTE and resolved by gel electrophoresis, one shifted band was observed, while extracts from bacteria harbouring the vector without cDNA gave no shifted bands (Figure 2A). The shifted band must be due to specific interaction of the DNA binding protein with the BTE sequence, since a 60-fold molar excess of the non-labelled BTE competed efficiently with the labelled BTE for the formation of the shifted band, whereas the mutated BTE had essentially no effect (Figure 2A). When a 600- or 2000-fold molar excess of competitors was included, a very fast shifted band occasionally appeared (Figure 2A). We do not know the nature of this band, which was not reproducibly observed. DNase I footprint analysis using a promoter region sequence (−166 to +42 bp relative to the transcription start site) of the CYP1A1 gene showed that bacterial extracts of the DNA binding protein protected a sequence from −43 to −66 bp, containing the BTE sequence that was identical to that given by nuclear extracts from Hepa-1 cells (Figure 2B). These DNA binding analyses indicated that cDNA of λBP26 encodes a protein factor that specifically binds the BTE sequence and we tentatively name this factor BTEB or BTE binding factor. Bacterially expressed Sp1 also showed the same protected pattern as the nuclear extracts of Hepa-1 and the BTEB protein in the DNase I footprint analysis (data not shown). In a previous paper (Yanagida *et al.*, 1990), we reported that an oligonucleotide of the GC box (10 bp) did not compete with the BTE sequence for binding to BTE-binding proteins in the nuclear extract of the cultured cells. This non-competition was presumably due

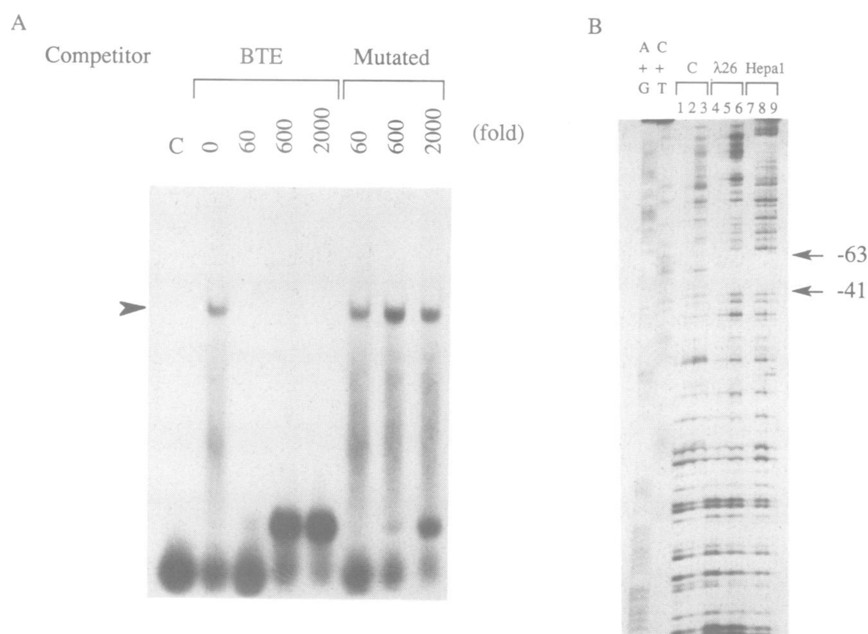


Fig. 2. Binding specificity of bacterially synthesized BTEB protein. (**A**) Gel shift assay. Extracts from bacteria not harbouring (lane C), or harbouring (other lanes) plasmids of λBP26 cDNA were incubated with [^{32}P]BTE in the presence of a 0- to 2000-fold molar excess of cold BTE or mutated BTE (Yanagida *et al.*, 1990). An arrow head indicates the shifted band. (**B**) DNase I footprint analysis. The bacterial extracts containing (lanes 4, 5 and 6) or not containing (lanes 1, 2 and 3) λBp26 cDNA product, or nuclear extracts from Hepa-I cells (lanes 7, 8 and 9) were incubated with a DNA fragment of the P-450IA1 gene promoter and were treated with 0.6 (lanes 1, 4 and 7), 0.2 (lanes 2, 5 and 8) or 0.06 U (lanes 3, 6 and 9) of DNase I. The protected region is indicated by arrows and nucleotide positions of the P-450IA1 gene relative to the RNA start site (Sogawa *et al.*, 1984). G + A and C + T, sequence ladders of chemical cleavages of the DNA fragments by the method of Maxam and Gilbert (Maxam and Gilbert, 1977).

Fig. 1. Nucleotide and amino acid sequences of rat Sp1 cDNA. (**A**) Restriction cleavage map of Sp1 cDNA. E, *Eco*RI; S, *Sac*I; B, *Bam*HI. (**B**) Amino acids are represented by single letter symbols. Rat Sp1 amino acid sequence is compared with the 696 C-terminal amino acids of human Sp1 (Kadonaga *et al.*, 1987). Only the amino acids of the human Sp1 that do not match the sequence of rat Sp1 are shown below the rat sequence. The N-terminus of the published human Sp1 is indicated by an arrow. An oligonucleotide sequence whose complementary sequence was used for a primer-extended cDNA library is underlined.

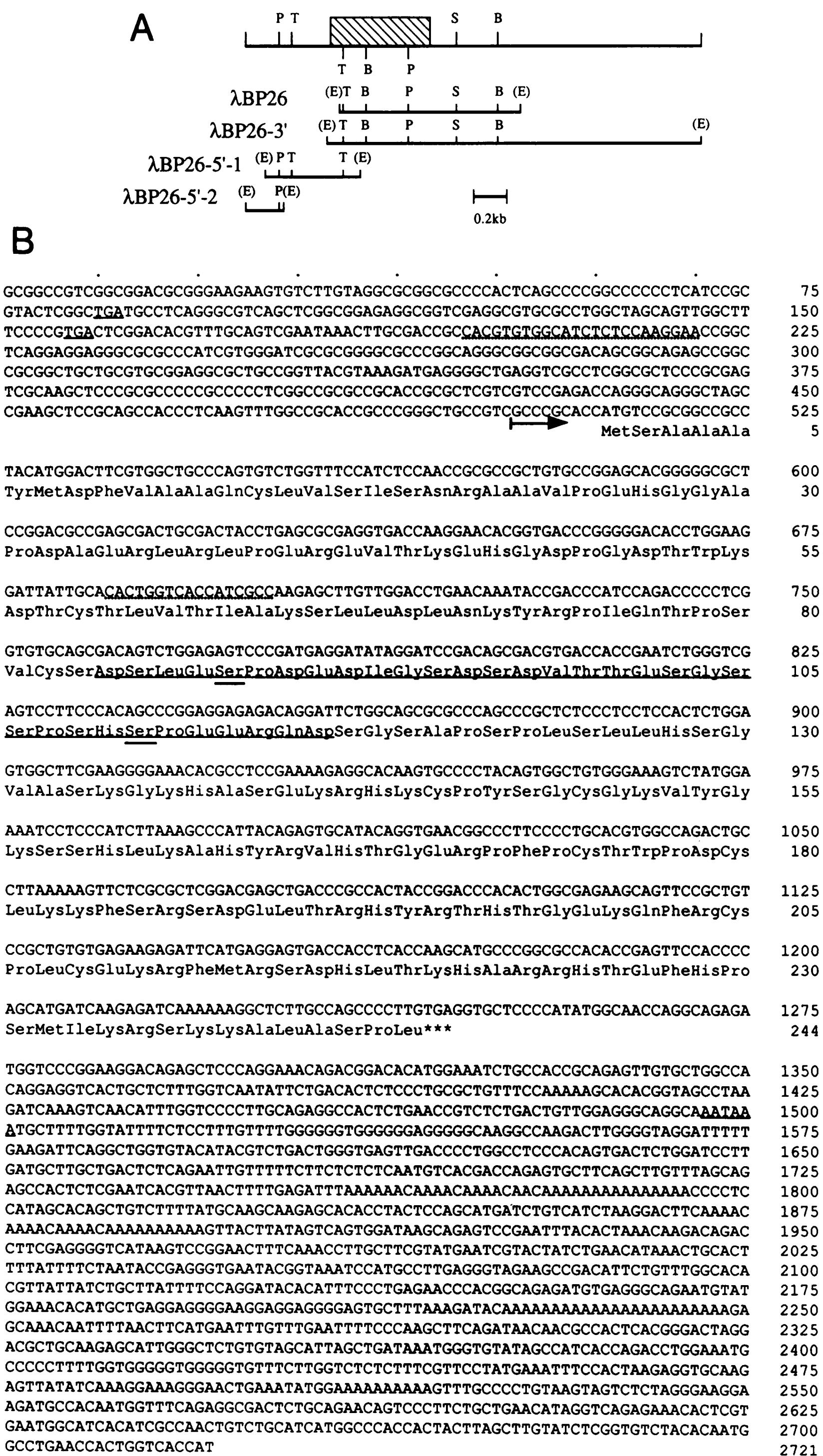


Fig. 3. Nucleotide sequence and deduced primary structure of rat BTEB cDNA. (**A**) Restiction cleavage map of BTEB cDNA. E, *Eco*RI; S, *Sac*I; B, *Bam*HI; P, *Pma*Cl; T, *Eco*T14I. The box represents the ORF. (**B**) The acidic region mentioned in the text is underlined. Serine residues that are possible targets of casein kinase II (Edelman *et al.*, 1987) are doubly underlined. In-frame stop codons in the 5'-UTR as to the open reading frame and AATAAA, a possible polyadenylation signal in the 3'-UTR are underlined. An arrow indicates the 5' end of a truncated BTEB cDNA used for expression in CV-1 and SL2. Nucleotides whose complementary sequences were used for the primer extension of cDNA are indicated by a dotted line.

to the fact that the GC box used as the competitor was too short for competition.

Since the BTEB cDNA (1.2 kb) was shorter than the BTEB mRNA (~5 kb) as judged by RNA blot analysis (Figure 4), we extended the BTEB cDNA in the 5' and 3' directions. For extension to the 5' end, we performed several cycles of screening of primer-extended cDNA libraries with appropriate DNA fragments as a probe to isolate several cDNA clones (Figure 3). Altogether, these cloned cDNA covered a 2.6 kb sequence, which is still not long enough for the entire length of the mRNA.

Structure of rat BTEB and Sp1

Two cDNA clones of λBP20 and 20-2 in combination encoded a primary structure of 788 amino acids that is highly similar (98%) to that of human Sp1 previously reported

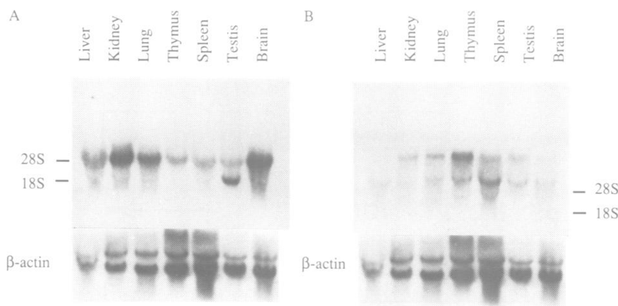


Fig. 4. Northern blot analysis of BTEB and Sp1. Total RNA (20 μg) obtained from indicated organs of rats were fractionated and transferred to a nitrocellulose filter. The filter was hybridized with a BTEB- (**A**) or Sp1-specific (**B**) probe. The same filter was rehybridized with a β-actin probe. The positions of 18S and 28S RNAs are indicated.

(Figure 1) (Kadonaga *et al.*, 1987). Different amino acids (22/788) between the two sequences appear to be clustered in several parts.

The 2.6 kb cDNA of BTEB has an open reading frame (ORF) that encodes a polypeptide of 244 amino acids flanked by a long (>500 bp) stretch of GC-rich (73%) 5'-untranslated region (UTR) and (>1.4 kb) 3'-UTR. The nucleotide sequence surrounding the putative initiator methionine codon (GCACCATGT) matches fairly well the consensus sequence for a eukaryotic translation start site (CCA/GCCATGG) (Kozak, 1984). From the deduced primary structure, BTEB was found to possess a three-times repeated zinc finger domain with a Cys-Cys/His-His motif. This feature is shared by Sp1; sequence similarity between the zinc finger domains of the two transcription factors is 72% (58/81 amino acids) at the amino acid level. A survey of amino acid sequences in the EMBO protein database shows that no equivalent protein with sufficient similarity to BTEB has been reported except for the zinc finger domains. Therefore, BTEB is considered to be a novel DNA binding protein.

Expression of BTEB and Sp1 mRNA

Expression of rat BTEB and Sp1 in various tissues was examined by RNA blot analysis. Both BTEB and Sp1 mRNAs were detectable in all the organs examined, suggesting that these two transcription factors have fundamental roles in cells (Figure 4). BTEB mRNA was most abundant in the kidney, lung, brain and testis. The main transcript for BTEB was a 5 kb mRNA in all organs examined except for the testis where a shorter transcript (2 kb) was dominant. This short mRNA was probably due to alternative usage of the polyadenylation signal at nucleotide position 1400 (Figure 3), since probes 3' downstream of this signal did not hybridize with the testicular shorter mRNA (data not shown). Sp1 mRNA was ~8.3 and 5.2 kb in all the tissues examined and was most abundant in the thymus and spleen. The concentrations of the two mRNAs were apparently different from each other in some tissues.

Functional analysis of BTEB and Sp1

To explore the function of BTEB and compare it with that of Sp1, we constructed the expression vectors (pRSVBTEB and pRSVSp1) of BTEB and Sp1. This was achieved by inserting each of the cDNAs under the RSV promoter in the plasmid pRSV(Xho) and transfecting them with a reporter plasmid, pSV/MC53, in which a promoter region of the CYP1A1 gene had been inserted between the SV40 enhancer and the CAT gene (Yanagida *et al.*, 1990; Figure 5). This reporter gene contains a single GC box sequence derived from BTE just upstream of the TATA box. While cotransfection with pRSVSp1 enhanced the expression of CAT activity from the reporter gene, the plasmid expressing BTEB repressed the expression of CAT activity, which was probably driven by endogenous Sp1 acting on the BTE. When the two plasmids, pRSVSp1 and pRSVBTEB, were simultaneously transfected into CV-1 cells with the reporter gene, the Sp1-stimulated CAT activity was also repressed to some extent by cotransfection with pRSVBTEB (Figure 5). The binding affinity of BTEB protein for the BTE sequence was found to be almost equal to that of Sp1 (Sogawa,K., Imataka,H. and Fujii-Kuriyama,Y., unpublished observation), suggesting that BTEB competed with Sp1 existing endogenously or expressed from the transfected pRSVSp1 for the BTE sequence, resulting in the apparent repression of the CAT expression. It has been reported that a yeast transcription factor, GAL4, required reiteration of GAL4 binding sites for prominent transcription enhancer activity (Carey *et al.*, 1990). This observation led us to construct pSV/MC53 + 4GC, which contains five repeats of the BTE-derived GC box (5'-GAGGCGTGGC-3') in the promoter (Figure 5A), and test it for responsiveness to BTEB. As shown in Figure 5C, BTEB elevated CAT activity from this promoter. The effects of BTEB on natural promoters containing repeated GC boxes such as SV40 early promoter (pSV2CAT) and HIV-I-LTR(pHIV-CAT) were then examined. Figure 6 shows that BTEB, as well as Sp1, was able to stimulate the expression of the CAT activity driven by the multiple GC box sequences. RNase protection assay using an antisense RNA probe to a 5'-untranslated region of HIV-1-LTR confirmed that the mRNA content was changed in parallel with the CAT activity (data not shown). Taken together with the results of a gel mobility shift assay, in which the SV40 early promoter and HIV-1-LTR promoter specifically bound the bacterially expressed BTEB (data not shown), these results suggest that BTEB serves as a transcriptional activator through the reiterated GC box sequences. It was quite difficult to find an appropriate control expression vector for normalization of transfection efficiency and we did not use an internal control. Many promoters that are employed to express indicator proteins have GC boxes and could be thereby influenced by expressed Sp1 and BTEB. Even the Rous sarcoma virus promoter, which has no GC box, has been shown to be activated by Sp1 (Swick *et al.*, 1989) and by BTEB (Imataka,H., Sogawa,K. and Fujii-Kuriyama,Y., unpublished data). Instead we repeated transfection experiments with different batches of plasmid DNAs to confirm the reproducibility of the experiments.

Discussion

Two regulatory proteins binding to a GC-box sequence, BTE, and their structural features

We have cloned two cDNAs encoding regulatory proteins binding to the BTE sequence in the promoter region of the rat CYP1A1 gene that contains a GC box consensus sequence. Sequence analysis of the cloned cDNAs revealed that one of these cDNAs encodes a primary structure of 788 amino acids, which from its close sequence similarity (98%)

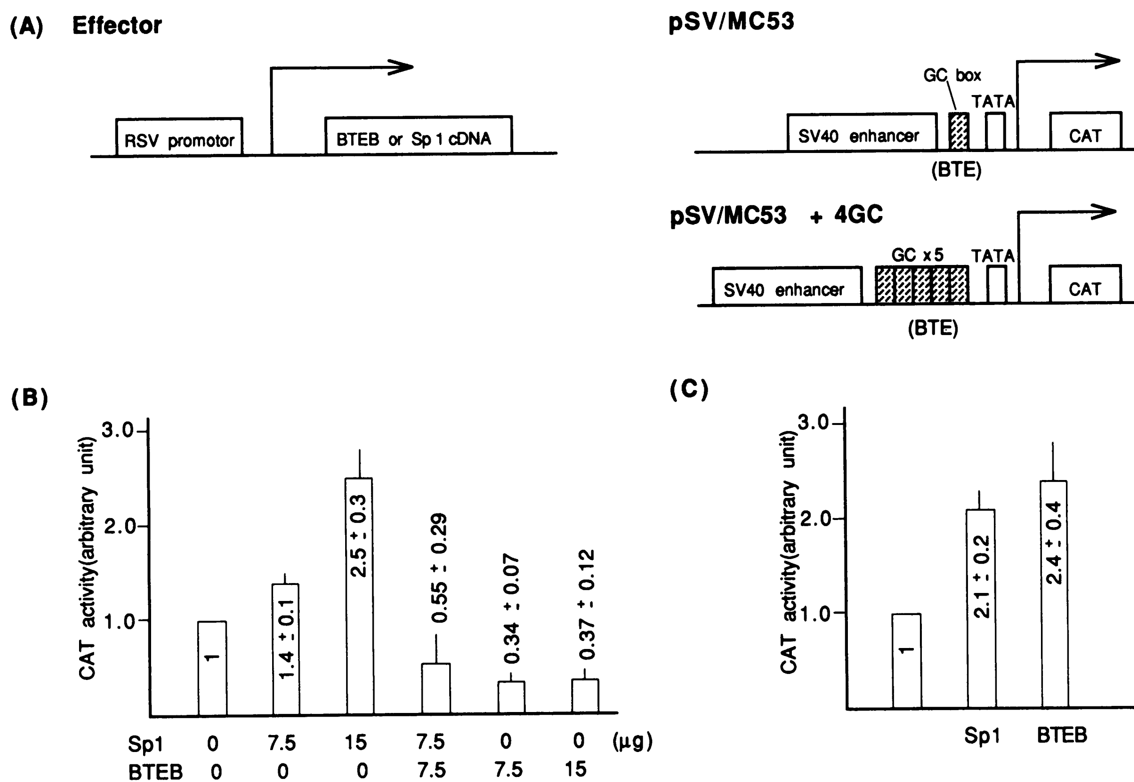


Fig. 5. Effects of BTEB and Sp1 on the expression of CAT activity driven by promoters with a single or repeated GC box sequence. (**A**) Schematic representations of the effectors, pRSVBTEB and pRSVSp1 and of reporters, pSV/MC53 and pSV/MC53 + 4GC. 8 µg of pSV/MC53 (**B**) and pSV/MC53 + 4GC (**C**) were cotransfected into CV-1 with indicated amounts of pRSVSp1 and/or pRSVBTEB in a total amount of DNA (18 µg) adjusted with pRSV(Xho). The CAT activity of extracts from cells transfected with the reporter and pRSV(Xho) was arbitrarily assigned a value of 1.0. Other CAT activities were normalized accordingly. Each column and bar represents the mean and standard deviation of three transfection experiments.

to human Sp1, is considered to be a rat transcription factor; the other encodes a protein of 244 amino acids, a novel DNA binding protein. Recently, cDNA for human BTEB has been isolated in our laboratory and its deduced amino acid sequence showed remarkable homology with the rat counterpart protein (98% amino acid identity). Therefore, BTEB may be functionally important. These two proteins have three consecutive zinc finger domains with Cys-Cys/His-His motifs. Figure 7 compares the amino acid sequences of the zinc finger domains of rat BTEB, rat Sp1, mouse Krox20 (Chavrier *et al.*, 1988) and mouse Krox24 (Lemaire *et al.*, 1988). Krox20 and Krox24 are serum-inducible transcriptional activators that have three consecutive zinc finger motifs and are known to recognize GC-rich regulatory sequences that are similar to, but distinct from, the target sequence of Sp1 (Chavrier *et al.*, 1990; Lemaire *et al.*, 1990).

Sequence similarity (72%) in the zinc finger domains betweer BTEB and Sp1 is much higher than that in any of the compared pairs, except for the pair of Krox20 and Krox24. From site-directed mutagenesis studies, Nardelli *et al.* (1990) suggested that the amino acids in the zinc finger domains are responsible for the recognition of the varied nucleotides between target sequences of Sp1 and Krox20. Interestingly, these amino acids are conserved between BTEB and Sp1, in agreement with recent findings that no notable difference has been found between the apparent affinities of BTEB and Sp1 for various GC box sequences (Sogawa,K., Imataka,H. and Fujii-Kuriyama,Y., unpublished observation).

As described, Sp1 is highly conserved for such a large molecule, suggesting that Sp1 plays an important role in interacting with other transcription factors. Two glutamine-rich domains and serine/threonine domains are important to the structure of Sp1 and it has been proved that the former function as an activator domain (Courey and Tjian, 1988). In contrast, BTEB is a small molecule with no glutamine-rich region. Instead, it has an acidic region from position 84 to 116 N-terminal of the zinc finger domains where aspartic acid and glutamic acid account for 28% (11/39) of the total amino acids, giving a net charge of -10 in this region. Several transcriptional activators, including the yeast factor GAL4, are equipped with acidic domains as an effector (Mitchell and Tjian, 1989). Serine residues at the 88 and 110 positions are possible targets for casein kinase II (Edelman *et al.*, 1987).

BTEB is functionally distinct from Sp1

BTEB seems to require tandem repeats of GC boxes for promoter activation, while Sp1 can actvate transcription from a promoter with a single GC box. Constitutive transcription from the CYP1A1 gene with a single GC box sequence in the promoter is shown to be activated by Sp1, but repressed by BTEB. These functional differences between the two factors may stem partly from differences between the activation domains as discussed above. It should be noted that GAL4, which has acidic domains, also requires reiteration of GAL4 target elements for its transcriptional enhancer activity (Carey *et al.*, 1990), apparently similar to the case with BTEB. It has been suggested that different

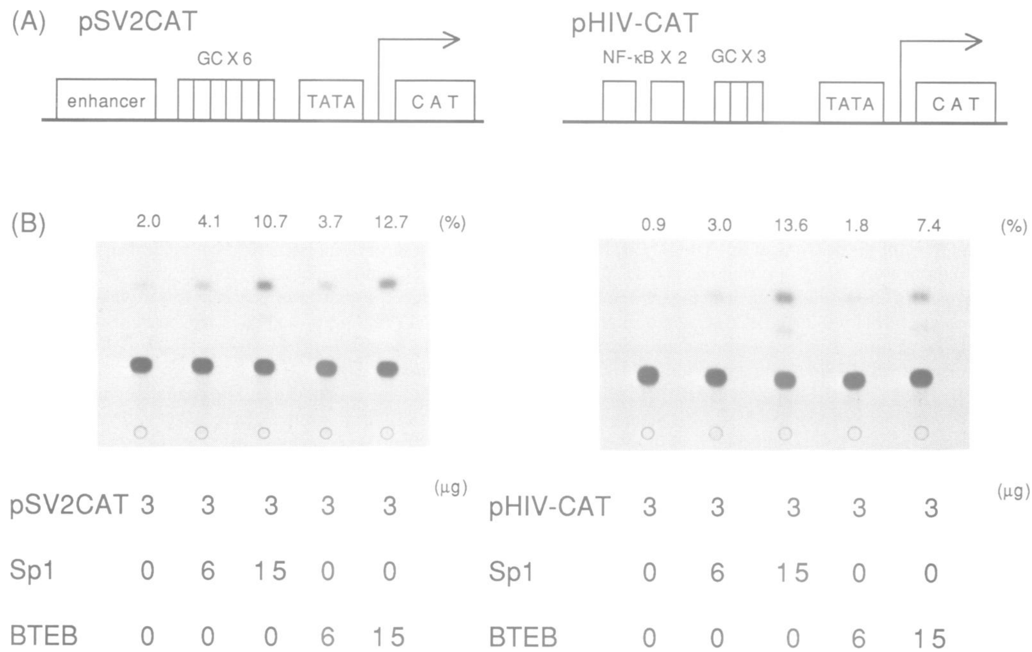


Fig. 6. Effects of BTEB and Sp1 on the expression of CAT activity driven by SV40 early and HIV-I-LTR promoters. (**A**) Schematic representations of pSV2CAT and pHIV-CAT. (**B**) Analysis of expressed CAT activities. 3 μg of pSV2CAT or pHIV-CAT were cotransfected into CV-1 with the indicated amounts of pRSVSp1 or pRSVBTEB in a total of 18 μg of DNA adjusted with pRSV(Xho).

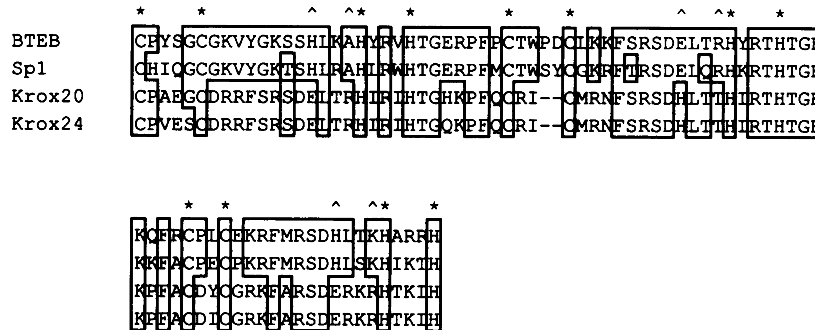


Fig. 7. Comparison of amino acid sequences of zinc finger domains of BTEB, Sp1, Krox20 and Krox24. Asterisks indicate cysteine and histidine residues that possibly coordinate zinc. ^ indicates amino acids that are considered to be important for DNA recognition (Nardelli *et al.*, 1991). Amino acids of BTEB that are shared by any of the other protein sequences are blocked.

activation domains of transcription factors interact with different adaptors or coactivators that mediate between the transcriptional regulator and general transcriptional machinary including TFIID, RNA polymerase II and others (Martin *et al.*, 1990; Pugh and Tjian, 1990; Lin and Green, 1991). *In vitro* transcription using purified BTEB will clarify the mechanism by which BTEB functions.

Expression of BTEB and its significance

BTEB expression could activate expression of genes with repeated GC boxes. Insulin receptor (Araki *et al.*, 1991) and H-*ras* genes (Ishii *et al.*, 1986) have multiple GC boxes, some of which are tandemly arranged. It is intriguing to test whether these genes, and other housekeeping genes that contain multiple GC boxes, respond to BTEB. On the other hand, BTEB expression could repress Sp1-supported transcription from promoters with a single GC box sequence such as the CYP1A1 gene promoter. BTEB and Sp1 in combination would create a variety of cellular states by selectively activating or repressing the expression of genes with various numbers of the GC box sequence in the promoter. BTEB and Sp1 mRNAs were detectable in all organs examined. However, the abundance of BTEB and Sp1 mRNAs seems to vary independently among various organs. It remains to be seen whether and how these variations in the expression of the two mRNAs are related to the cellular states. It could be speculated that BTEB expression is under post-transcriptional control, since BTEB mRNA has a long stretch of GC-rich 5′-UTR. A number of genes that are involved in cell growth control have a long GC-rich 5′-UTR (Kozak, 1989). In some genes, including the ornithine decarboxylase (Manzella and Blackshear, 1990) and PDGF genes (Rao *et al.*, 1988), secondary structure formed in the 5′-UTR of the mRNAs seems to inhibit translation. It will also be interesting to investigate whether the mRNA of BTEB is always translated into protein or whether it is subject to translational control.

Materials and methods

Cloning of cDNAs

Total RNA was isolated from rat liver by the guanidium thiocyanate treatment followed by centrifugation in caesium chloride solution, and poly(A)$^+$ RNA

was obtained by passing the total RNA through an oligo(dT) cellulose column (Maniatis *et al.*, 1982). cDNA synthesis and subsequent insertion of the cDNA into the λgt11 vector were performed using commercial kits (Amersham). The procedure for screening bacteriophages was essentially the same as that of Vinson *et al.* (1988).

BTE, 5′gatc^{-59}GAGAAGGAGGCGTGGCCAAC^{-40}
CTCTTCCTCCGCACCGGTTGctag

was tandemly (10–20 times) ligated and ^{32}P-labelled by the nick translation method (specific activity, 5×10^9 d.p.m./μg). To obtain cDNAs missing in the λBP20 and BP26 cDNA for the mRNAs of Sp1 and BTEB, rat liver cDNA libraries were constructed using an oligonucleotide (17mer) complementary to the 5′ part of the λBP20 sequence (Figure 1B) and an oligonucleotide (17mer) complementary to the λBP26 sequence (Figure 3B) as primers, respectively. The resulting libraries were screened with a 0.14 kb *Bam*HI–*Eco*RI fragment of λBP20 (nucleotide positions 1541–1680, Figure 1) or with a 0.45 kb *Pma*CI–*Eco*RI fragment of λBP26 (584–1036, Figure 3). For further 5′ extension from λBP26–5′-1, a cDNA library was constructed using a primer (26mer) complementary to a 5′ part of λBP26–5′-1 (Figure 3) and poly(A)$^+$ RNA, which had been enriched in the mRNA species hybridized with λBP26 by centrifugation in sucrose density gradient (Maniatis *et al.*, 1982). This library (5×10^5 plaques) was screened with a 0.08 kb *Eco*RI–*Eco*T14I of λBP26–5′-1 (144–233, Figure 3). Nucleotide sequence was determined by the dideoxy chain termination method (Sanger *et al.*, 1977).

Expression of cDNA in E.coli

An expression vector, pAR2106 (Studier and Moffatt, 1986), was digested with *Bam*HI, filled in with Klenow enzyme and ligated with *Eco*RI linkers (10mer). λBP26 cDNA was inserted into the *Eco*RI site of the plasmid. A bacterial strain, BL-21 (DE8) (Studier and Moffatt, 1986), harbouring the recombinant was grown until A_{600} reached 0.1. Then IPTG was added to 0.4 mM and incubation was continued for another 3 h. Bacteria were harvested and lysed by freezing and thawing three times in a buffer consisting of 25 mM HEPES (pH 7.8), 12.5 mM $MgCl_2$, 1 mM DTT, 0.1% Triton X-100 and 0.5 mM PMSF. The cell lysates were centrifuged at 30 000 r.p.m. for 1 h after addition of glycerol to 20%. The supernatants were referred to as bacterial extracts. Protein concentrations were determined according to Bradford (1986) using bovine serum albumin as standard.

Gel mobility shift assay and DNase I footprint analysis

For gel mobility shift assays, bacterial extracts (15 μg protein) were mixed with unlabelled BTE (0–1 μg) for competition experiments and double-stranded poly(dI–dC) (4 μg) in a buffer containing 10 mM HEPES (pH 7.6), 1 mM $MgCl_2$ and 1 mM DTT and kept on ice for 10 min. A ^{32}P-labelled BTE probe (2×10^4 d.p.m.) was added to the mixture and was incubated at 24°C for 20 min. The reaction mixture was resolved by electrophoresis in 4.6% native polyacrylamide gel. For DNase I footprint analysis, a DNA fragment spanning from −166 to +42 bp of the P-450IA1 gene was used as a probe (Yanagida *et al.*, 1990). The 5′ end of the non-coding strand of the fragment was labelled with [γ-^{32}P]ATP. Bacterial extracts (60 μg protein) or Hepa-1 cell nuclear extracts (20 μg) (Yanagida *et al.*, 1990) were pre-treated with 1 μg double-stranded poly(dI–dC) on ice for 10 min and then incubated with the labelled probe at 20°C for 10 min. The reaction mixture was treated with various concentrations of DNase I at 25°C for 90 s and was resolved on a 7 M urea–6% polyacrylamide gel.

Expression of cDNAs in eukaryotic cells and CAT assay

BTEB cDNA was truncated as follows. Most of the 5′-UTR of BTEB cDNA was removed by digestion with Bal31 and the region downstream of the *Sac*I site was eliminated. The resulting BTEB cDNA begins at nucleotide position 496 and ends at 1192 (Figure 3). The truncated BTEB and the Sp1 cDNA were ligated with *Xho*I linkers. For expression of cDNAs in CV-1 cells, pRSVCAT (Gorman *et al.*, 1982a) was modified and used as a vector. From pRSVCAT, a 0.6 kb *Hin*dIII–*Sca*I fragment containing most of the CAT coding region was eliminated and the *Hin*dIII and *Sca*I sites of the truncated vector were converted to a *Xho*I site. The resulting plasmid was named pRSV(Xho). The *Xho*I-linked cDNAs were inserted into the *Xho*I site of pRSV(Xho) in the sense orientation, generating pRSVBTEB and pRSVSp1. As reporter plasmids, pSV/MC53 (Yanagida *et al.*, 1990), pSV/MC53 + 4GC (Figure 5A) pSV2CAT (Gorman *et al.* 1982b) and pHIVCAT (Shibata *et al.* 1990) were employed.

To generate pSV/MC53 + 4GC,
5′CTGAGGCGTGGCTGAGGCGTGG
GACTCCGCACCGACTCCGCACC

was inserted twice into the Bal31 site in the GC box of pSV/MC53 (Yanagida *et al.*, 1990). Insertion of the sequences was confirmed by sequencing.

CV-I cells were maintained in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and were distributed at a density of 2.5×10^5/6 cm dish at 12–24 h before transfection. The calcium phosphate precipitation method (Gorman *et al.*, 1982b) was employed for transfection. Fixed amounts of a reporter plasmid were mixed with various amounts of pRSVBTEB or pRSVSp1. The total amount of DNA was kept constant in a series of transfection experiments by adding pRSV(Xho) as a carrier. At 4 h after transfection, each dish was subjected to glycerol shock and then incubated for 40 h. The CAT assay was performed as described (Gorman *et al.*, 1982b). Briefly, cells were lysed by freeze–thaw and were centrifuged to obtain supernatant fractions. The suprenatant (30 μg protein) was incubated with [^{14}C]chloramphenicol for 15 min at 37°C, when pSV/MC53, pSV/MC53 + 4GC and pSV2CAT were used as reporter plasmids. For pHIVCAT, 80 μg protein was incubated for 1 h.

RNA blot analysis

Total RNA was obtained by the guanidine thiocyanate method (Chirgwin *et al.*, 1979) from rat liver, lung, kidney, thymus, spleen, testis and brain. RNA (20 μg) was denatured, electrophoresed and transferred to nitrocellulose filters. The fragment spanning from nucleotide 1 to 790 (*Bam*HI) of BTEB cDNA and the *Sac*I–*Bam*HI (21–925) fragment of Sp1 cDNA were ^{32}P-labelled by the random priming method using a commercial labelling kit (Boehringer Mannheim). The filter was hybridized with the BTEB probe (0.9×10^9 d.p.m./μg) or Sp1 probe (1.0×10^9 d.p.m./μg). The hybridized filter was washed at 50°C for 30 min in 0.1×SSC and 0.2% SDS, and exposed to an X-ray film at −80°C for 20 h. The filters were rehybridized with a ^{32}P-labelled β-actin probe (Tokunaga *et al.*, 1986).

Acknowledgements

We thank Dr M.Iwabuchi (Kyoto University) for pAR2106 and useful advice on expression of cDNA in *E.coli*. We are also grateful to Drs I.Nishida and Mi-AeYou (Aichi Cancer Center) for pGEMActex-2, to Dr M.Obinata (Tohoku Univeristy) for pRSVCAT, to Dr T.Miyake (Mitsubishi Kasei Institute for Life Science) for SL2 cells and to Dr A.Adachi (Kyoto University) for pHIVCAT. H.I. was partly supported by a fellowship of the Japan Society for Promotion of Science for Japanese junior scientists. This work was partly supported by a Grant-in-Aid for Scientific Research on Priority Areas from Ministry of Education, Science and Culture in Japan and by a grand from The Mitsubishi Foundation.

References

Ammendola,R., Gounari,F., Piagio,G., De Imone,V. and Cortese,R. (1990) *Mol. Cell. Biol.*, **10**, 387–390.

Araki,E., Murakami,T., Shirotani,T., Kanai,F., Shinohara,Y., Shimada,F., Mori,M., Shichiri,M. and Ebina,Y. (1991) *J. Biol. Chem.*, **266**, 3944–3948.

Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.

Briggs,M.R., Kadonaga,T., Bell,S.P. and Tjian,R. (1986) *Science*, **234**, 47–52.

Carey,M., Lin,Y.-S., Green,M.R. and Ptashne,M. (1990) *Nature*, **345**, 361–364.

Chavrier,P., Zerial,M., Lemaire,P., Almendral,J., Bravo,R. and Charnay,P. (1988) *EMBO J.*, **7**, 29–35.

Chavrier,P., Vesque,C., Galliot,B., Vigneron,M., Dolle,P., Duboule,D. and Charnay,P. (1990) *EMBO J.*, **9**, 1209–1218.

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

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

Dvorçk,M., Urbçnek,P., Bartunek,P., Paces,V., Vlach,J., Pecenka,V., Arnold,L., Trçvnicek,M. and Riman,J. (1989) *Nucleic Acids Res.*, **17**, 5651–5664.

Edelman,A.M., Blumenthal,D.K. and Krebs,E.G. (1987) *Annu. Rev. Biochem.*, **56**, 567–613.

Fujisawa-Sehara,A., Sogawa,K., Yamane,M. and Fujii-Kuriyama,Y. (1987) *Nucleic Acids Res.*, **15**, 4179–4191.

Gorman,C.M., Merlino,G.T., Willingham,M.C., Pastan,I. and Howard,B.H. (1982a) *Proc. Natl. Acad. Sci. USA*, **79**, 6777–6781.

Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982b) *Mol. Cell. Biol.*, **2**, 1044–1051.

Huang,H.-C., Sundseth,R. and Hansen,U. (1990) *Genes Dev.*, **4**, 287–298.

Ishii,S., Kadonaga,J.T., Tijian,R., Brady,J.H., Merlino,G.T. and Pastan,I. (1986) *Science*, **232**, 1410–1413.

Jones,K.A., Kadonaga,J.T., Luciw,P.A. and Tjian,R. (1986) *Science*, **232**, 755–759.

Kadonaga,J.T., Carner,K.R., Masiarz,F.R. and Tjian,R. (1987) *Cell*, **51**, 1079–1090.
Kageyama,R., Merlino,G.T. and Pastan,I. (1988) *J. Biol. Chem.*, **263**, 6329–6336.
Kawajiri,K., Gotoh,O., Tagashira,Y., Sogawa,K. and Fujii-Kuriyama,Y. (1984) J. *Biol. Chem.*, **259**, 10145–10149.
Kim,S.H., Moores,J.C., David,D., Respess,J.G., Jolly,D.J. and Friedmann,T. (1986) *Nucleic Acids Res.*, **14**, 3103–3118.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Kozak,M. (1989) *J. Cell Biol.*, **108**, 229–241.
Lemaire,P., Revelant,O., Bravo,R. and Charnay,P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4691–4695.
Lemaire,P., Vesqne,C., Schmitt,J., Stunnenberg,H., Frank,R. and Charnay,P. (1990) *Mol. Cell. Biol.*, **10**, 3456–3467.
Lin,Y.-S. and Grenn,M.R. (1991) *Cell*, **64**, 971–981.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Manzella,J.M. and Blackshear,P.J. (1990) *J. Biol. Chem.*, **265**, 11817–11822.
Martin,K.J., Lillie,J.W. and Green,M.R. (1990) *Nature*, **346**, 147–152.
Maxam,A.M. and Gibert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560–564.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Pugh,B.F. and Tjian,R. (1990) *Cell*, **61**, 1187–1197.
Nardelli,J., Gibson,T.J., Vesque,C. and Charnay,P. (1991) *Nature*, **349**, 175–178.
Rao,C.D., Pech,M., Robbins,K.C. and Aaronson,S.A. (1988) *Mol. Cell. Biol.*, **8**, 284–292.
Reynolds,G.A., Basu, S.K., Osborne,T.F., Chin,D.J., Gil,G., Brown,M.S., Godstein,J.L. and Luskey,K.L. (1984) *Cell*, **38**, 275–285.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schmidt,M.C., Zhou,Q. and Berk,A.J. (1989) *Mol. Cell. Biol.*, **9**, 3299–3307.
Shibata,R., Miura,T., Hayami,M., Ogawa,K., Sakai,H., Kiyomasu,T., Ishimoto,A. and Adachi,A. (1990) *J. Virol.*, **64**, 742–747.
Sogawa,K., Gotoh,O., Kawajiri,K. and Fujii-Kuriyama,Y. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5066–5070.
Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
Swick,A.G., Blake,M.C., Kahn,J.W. Azizkhan,J.C. (1989) *Nucleic Acids Res.*, **17**, 9291–9304.
Tokunaga,K., Taniguchi,H., Yoda,K., Shimizu,M. and Sakiyama,S. (1986) *Nucleic Acids Res.*, **14**, 2829.
Vinson C.R.,LaMarco, K.L., Johnson,P.F., Landschulz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.
Yanagida,A., Sogawa,K., Yasumoto,K. and Fujii-Kuriyama,Y. (1990) *Mol. Cell. Biol.*, **10**, 1470–1475.

Received on July 10, 1991; revised on June 23, 1992

Note added in proof

The rat Spl and BTEB sequences are available from the EMBL/GenBank/DDBJ databases under accession numbers D12768 and D12769 respectively.