Genomic Cloning of the Gene for an IgE-Binding Lectin Reveals Unusual Utilization of 5′ Untranslated Regions†

Christine A. Gritzmacher, Virginia S. Mehl, and Fu-Tong Liu*

Department of Molecular and Experimental Medicine, SBR-4, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

Received May 15, 1992; Revised Manuscript Received July 30, 1992

ABSTRACT: ϵBP (for ϵ *b*inding *p*rotein) is a M_r 31 000 S-type animal lectin that binds to IgE and has been identified as the homologue of Mac-2, a macrophage cell-surface marker, as well as the lectins RL-29, CBP35, and L-34. The protein is composed of two domains with the amino-terminal portion containing tandem repeats of nine amino acids and the carboxyl-terminal half containing consensus sequences shared by S-type animal lectins. We determined the genomic map in both rat and mouse and isolated overlapping genomic clones that contain the 5′ two-thirds of the murine gene. The remaining portion of the gene was obtained by polymerase chain reaction (PCR) amplification of genomic murine DNA followed by subcloning into plasmid vectors. The ϵBP gene is composed of six exons separated by five introns. The entire amino-terminal repetitive sequence is contained in exon III, and the carboxyl-terminal domain is encoded by the three succeeding exons (IV, V, VI). The latter three exons correspond well in size and share sequence homology with three exons coding for 14-kDa S-type lectins. The sequence in exon I offers an explanation for the generation of two mRNAs differing only in their 5′ untranslated sequences, previously reported in Mac-2 cDNA clones. Using cDNA synthesis and PCR amplification, we determined that two alternative splice sites are used in many different types of cells. This alternative splicing results in different 5′ untranslated regions of the murine ϵBP mRNA.

A large number of lectins have been identified in the animal kingdom and have been shown to be important in multiple cellular functions including cell–cell communication [for reviews, see Drickamer (1988) and Barondes (1988)]. For many years, we have examined one lectin that we described first as a rat protein that binds immunoglobulin E (IgE)[1] and, therefore, called ϵBP for ϵ *b*inding *p*rotein (Liu et al., 1985; Albrandt et al., 1987; Liu, 1990). Molecular cloning and sequencing confirmed that this protein has homologues in other species, many of which were identified on the basis of other characteristics. These include murine Mac-2, a macrophage cell-surface marker (Cherayil et al., 1989); murine L-34, a metastasis-associated lectin (Raz et al., 1989); RL-29, a lactose-binding lectin found in rat lung (Leffler et al., 1989; Leffler & Barondes, 1986); and CBP35, a murine carbohydrate-binding protein (Jia & Wang, 1988). cDNAs for human homologues have also been cloned and sequenced (Robertson et al., 1990; Cherayil et al., 1990; Raz et al., 1991a; Oda et al., 1991). ϵBP described in this report is a galactose-binding lectin of molecular weight approximately 31 000.

This lectin may have multiple biological functions. It is expressed in many different cell types and is found in various locations within cells (Crittenden et al., 1984; Moutsatsos et al., 1986; Gritzmacher et al., 1988; Cherayil et al., 1989). The level of CBP35 increases in rapidly dividing fibroblasts or in cells initiating division following serum starvation (Moutsatsos et al., 1987). Increased expression of L34 and CBP35 occurs in malignant and transformed cells compared to their normal counterparts (Moutsatsos et al., 1986, 1987; Raz et al., 1987, 1991b). Mac-2 protein is detected at higher levels on the surface of thioglycolate-activated macrophages compared to resting macrophages (Ho & Springer, 1982). Recently, Mac-2 was shown to be the major non-integrin laminin-binding protein, suggesting that this protein may facilitate adhesion of macrophages and mast cells to the extracellular matrix during inflammation (Woo et a., 1990).

ϵBP is a member of the S-type lectins (Drickamer, 1988). The protein is composed of two domains: the amino-terminal portion contains tandem repeats of nine amino acids, and the carboxyl-terminal portion shares significant sequence similarities with other S-type lectins, namely, the 14-kDa soluble lectins (Liu, 1990). The latter lectins have also been identified in many species, and the cDNA and derived protein sequences have been reported by many groups [see Abbott and Feizi (1989), Hirabayashi and Kasai (1988), Clerch et al. (1988), and Paroutaud et al. (1987) and references cited therein]. The genomic organization and sequences of three 14-kDa lectin genes have also been determined (Ohyama & Kasai, 1988; Gitt & Barondes, 1991). On the basis of significant sequence similarities, it is reasonable to predict that the 3′ portion of ϵBP and the 14-kDa lectins are derived from the same ancestral gene.

The cloning of Mac-2 cDNA revealed two alternative 5′ untranslated (UT) sequences on Mac-2 mRNA (Cherayil et al., 1989). The molecular basis for generation of these two mRNA species was unkown. The two different 5′ UT regions were detected in cDNA clones derived from macrophages, but only one of them was seen in clones obtained from libraries made from rat basophilic leukemia cells (Albrandt et al., 1987), murine fibroblasts (Jia & Wang, 1988), or murine fibrosarcoma (Raz et al., 1989). These contrasting results raised the possibility that alternative splicing may be unique to expression of this gene in macrophages.

In this paper, we present a restriction map of the gene for ϵBP in both rat and mouse and DNA sequences from

† This work was supported by National Institutes of Health Grants AI19747, AI20958, and AI25086. This is Publication No. 6801-MEM of the Department of Molecular and Experimental Medicine.

[1] Abbreviations: ϵBP, IgE-binding protein; CBP35, carbohydrate-binding protein 35; CHL, chicken hepatic lectin; IgE, immunoglobulin E; PCR, polymerase chain reaction; UT, untranslated (sequence); kb, kilobase(s); bp, base pair(s); SDS, sodium dodecyl sulfate.

0006-2960/92/0431-9533$03.00/0

overlapping genomic clones that contain the murine εBP coding sequence. The genomic organization of the εBP gene was compared to that of the 14-kDa lectin genes. We also identified the sequences that are probably used in alternative splicing of εBP gene transcripts to yield messages that contain different 5′ UT regions but the same coding sequence. On the basis of this information, we analyzed RNA from a number of cell lines and tissues and determined the relative occurrence of the two alternative 5′ UT regions in murine εBP transcripts.

MATERIALS AND METHODS

DNA Blot Hybridization and DNA Probes. High molecular weight DNA was isolated from BALB/c mouse and Lewis X DA (F1) rat liver and subjected to blot hybridization using standard methods (Southern, 1975; Maniatis et al., 1982). The probes used were as follows: a nearly-complete rat εBP cDNA consists of 747 bp from the unique *Xho*I site (at position 94) to the most 3′ *Hae*III site of the rat εBP cDNA (Albrandt et al., 1987); the 5′-specific probe consists of 366 bp from the unique *Xho*I site to the unique *Sph*I site (at position 459); the 3′-specific probe consists of 382 bp from the unique *Sph*I site to the most 3′ *Hae*III site (Albrandt et al., 1987). Probes were labeled with ^{32}P using the nick-translation method (Maniatis et al., 1982).

Genomic Library Construction and Screening. A BALB/c library of *Eco*RI fragments, size-selected from agarose gels in the range of 6–9 kb, was constructed into the vector λZap (from Stratagene, La Jolla, CA) and screened using the nearly-complete rat εBP cDNA probe to yield the 1A1 clone. The insert was released in vivo in the pBluescript vector by growing the clone in the presence of an f1 helper phage according to the supplier's instructions. A BALB/c library of *Mbo*I partially cut DNA in the vector λ EMBL3 (from Clontech Laboratories, Inc., Palo Alto, CA) was screened with the nearly-complete rat εBP cDNA and yielded the MR4b clone. Both libraries were also screened with the 3′-specific rat cDNA probe but yielded only repeat isolates of the 1A1 clone and the MR4b clone or overlapping 5′ clones. Both libraries were probed in 5× SSC (1× SSC is 150 mM NaCl/15 mM sodium citrate, pH 7.0), 2× Denhardt's solution (1× Denhardt's solution is 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone), 0.1% SDS, and 100 μg of salmon sperm DNA/mL at 56 °C, and were washed with 2×–0.1× SSC and 0.1% SDS at 50–65 °C.

Other libraries screened with the nearly-complete εBP cDNA probe and the 3′-specific probe include the following: a rat genomic library made from a partial *Hae*III digest cloned into λ Charon 4A (Clontech Laboratories, Inc.); a BALB/c mouse library made from a partial *Mbo*I digest cloned into λ Charon 28 (37484 from the American Type Culture Collection, Rockville, MD); and a BALB/c library of *Bam*HI fragments (ranging from 15 to 20 kb) cloned into the λZap vector.

Genomic DNA Amplification Using Polymerase Chain Reaction (PCR) and Cloning of PCR Products. Genomic DNA containing exons IV, V, and VI and the intervening introns of the mouse εBP gene were amplified by using PCR with the following primers and conditions. All PCR reactions were done in a final volume of 50 μL containing 250 ng of BALB/c genomic DNA, 25 pmol of each primer, 50 mM KCl, 10 mM Tris-HCl (pH 8.8), 1.5 mM $MgCl_2$, 0.1% Triton X-100, 200 μM of each dNTP, and 2.5 units of *Taq*I DNA polymerase (Promega, Madison, WI). For amplification of the exon IV–intron IV–exon V region, the primers 5′TGCGGAATTCACAGTGAAACCCAACGCAAACAG3′ (containing an *Eco*RI site and the sequence of the 3′ end of exon IV) and 5′CGCGGGATCCACTCTCAAAGGGGAAGGCTGACTG3′ (containing a *Bam*HI site and corresponding to the reverse complement of the 3′ portion of exon V) were used to generate a 2.7-kb fragment in 35 cycles each consisting of 5 s at 92 °C, 1 min at 65 °C, and 4 min at 72 °C, followed by a single cycle of 12 min at 72 °C. When the 2.7-kb fragment was cut with the restriction enzymes *Eco*RI and *Bam*HI, two fragments of 1.9 kb (with *Eco*RI ends) and 0.8 kb (with *Eco*RI and *Bam*HI ends) were generated; these were purified from an agarose gel and subcloned into pUC19 vectors. For amplification of the exon V–intron V–exon VI region, the primers 5′GGCCGAATTCGATTGTTCTAGATTTCAGGAGAGG3′ (containing an *Eco*RI site and a sequence corresponding to the 5′ portion of exon V) and 5′TCTGGGATCCTTTCGGTGCCGCCCCTTCTGGC3′ (containing a *Bam*HI site and corresponding to the reverse complement of the UT sequence 3′ of exon VI) were used to generate a 1.4-kb fragment using 35 cycles each of 5 s at 92 °C, 2 min at 65 °C, and 2 min at 72 °C, followed by one cycle of 12 min at 72 °C. After the 1.4-kb fragment was cut with *Eco*RI and *Bam*HI, it was purified from an agarose gel and subcloned into pUC19.

DNA Sequencing. Subclones were inserted into either pUC19 or pBluescript and sequenced on both strands using the dideoxy-termination method (Sanger et al., 1977). To obtain the entire sequence in overlapping segments, a series of nested deletions were made using exonuclease III and S1 nuclease (Henikoff, 1984) and similarly sequenced. To obtain the sequence of the 2.0-kb *Eco*RI fragment located 3′ to exon I, a number of oligomeric primers were synthesized corresponding to the sequences obtained from the 3′ and 5′ ends of the fragment, and these were used to extend the sequence in a series of overlapping segments. The sequences were analyzed using the GCG computer software program.

Cell Lines and Mouse Tissues. Macrophage cell lines P388D1 and WEHI-3 and fibroblast line 3T3 were obtained from the American Type Culture Collection (Rockville, MD), and SV40-transformed mouse hepatocyte line SVB6 was a gift from Dr. F. Chisari (The Scripps Research Institute); all were maintained in DMEM with 10% bovine calf serum (Hyclone Laboratories, Inc., Logan, UT). Tissues were obtained from 6–8-week-old BALB/c mice and immediately frozen in liquid nitrogen.

Isolation of Poly(A)-Containing RNA and cDNA Synthesis. RNA was isolated using the SDS and proteinase K lysis method followed by binding to and elution from oligo(dT)–cellulose (using the Fast Track kit from Invitrogen Corp., San Diego, CA). First-strand cDNA was synthesized using the cDNA Cycle kit (Invitrogen Corp.) from 0.5 μg of RNA using AMV reverse transcriptase and the primer 5′ACCCGGATATCCTTGAGGGTT3′ which corresponds to the reverse complementary strand of the Mac-2 sequence [amino acid positions 18–24 in Raz et al. (1989)]. The reaction was terminated by adding 480 μL of 10 mM Tris-HCl (pH 7.5)/1 mM EDTA to the reaction and heating at 95 °C for 3 min.

Amplification of cDNA Using PCR. Amplification of the 5′ end of the cDNA was done using a 5′ primer consisting of 5′GGCGGGTGGAGCACTAATCA3′ and a 3′ primer consisting of 5′ACCCGGATATCCTTGAGGGTT3′ (the same 21-mer primer used to produce the cDNA). The 5′ primer corresponds to a sequence in the 5′ UT region that is present on both forms of Mac-2 cDNA (Cherayil et al., 1989). The PCR was done in 50-μL volumes containing 1–20 μL of the first-strand template, 20 pmol of each primer, 50 mM KCl,

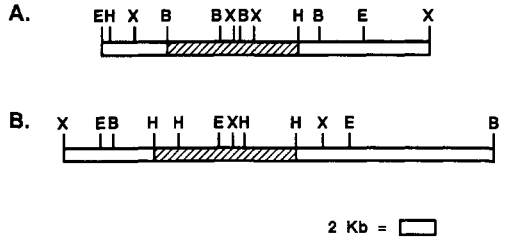

FIGURE 1: Partial restriction maps of murine and rat ϵBP genes. Consensus restriction maps for (A) the rat ϵBP gene and (B) the mouse ϵBP gene. For both, the hatched areas represent the minimum predicted region of coding exons.

10 mM Tris-HCl (pH 8.8), 1.5 mM $MgCl_2$, 0.1% Triton X-100, 200 μM of each dNTP, and 2.5 units of *Taq*I DNA polymerase (Promega, Madison, WI). Amplification using 25 or 40 cycles each of 5 s at 92 °C, 15 s at 55 °C, and 30 s at 72 °C was followed by a final incubation of 5 min at 72 °C. Typically, one-tenth of the reaction was separated on a gel consisting of 1.3% Nusieve and 1% agarose (FMC, Rockland, ME) and transferred to nitrocellulose in 20× SSC. The membranes were probed with either the 5′ amplification primer or an internal probe of 5′GGCACAGAGAGCACTACCCA3′ which is specific for the sequence present only in the longer Mac-2 cDNA (Cherayil et al., 1989). Probes were end-labeled with ^{32}P, and probing was done at 42 °C in 2–10 mL of 6× SSC, 10 mM sodium phosphate (pH 6.8), 1 mM EDTA, 0.5% SDS, 100 μg of salmon sperm DNA/mL, and 0.1% nonfat dry milk (Maniatis et al., 1982).

RESULTS AND DISCUSSION

Partial Restriction Mapping of Mouse and Rat ϵBP Genes. Genomic DNAs from rat and mouse were digested with a variety of restriction enzymes and used to generate restriction maps of the ϵBP genes in both species. Size-separated DNA fragments, immobilized on nitrocellulose, were probed independently with the complete rat cDNA or with the 5′ and 3′ halves of the cDNA (see Materials and Methods). Thus, we were able to orient the many fragments and map the rat and mouse genes as shown in Figure 1. As expected, the maps differ, but *both species have a single gene coding for* ϵBP. The predicted sizes of DNA containing the ϵBP genes determined from these maps are approximately 8.5 kb for the rat and approximately 9 kb for the mouse. These sizes represent the smallest region containing contiguous fragments that hybridized to the rat ϵBP cDNA probe. The actual size of the murine ϵBP gene is approximately 11.5 kb spanning from the 5′ UT region, which our cDNA probes did not detect, to the end of the last coding exon (see below).

When the membranes were washed at low stringency (2× SSC at 42 °C), we detected additional DNA fragments that hybridized weakly to the 3′-half cDNA probe in some of the digests. However, these were not detected when the membranes were washed at higher stringency (1× SSC at 50 °C or higher) (data not shown). These additional fragments probably represent cross-hybridization of the probe with other lectin genes because the 3′ cDNA probe codes for the carboxyl-terminal portion of the ϵBP protein which contains the conserved lectin region.

Isolation and Sequencing of Overlapping Genomic Clones Containing the ϵBP Gene. Two genomic clones representing the 5′ portion of the murine ϵBP gene were isolated from two independent libraries that were probed with rat cDNA. The most 5′ clone (MR4b) was isolated from a partial *Mbo*I library and contains an insert fragment of approximately 15 kb. This

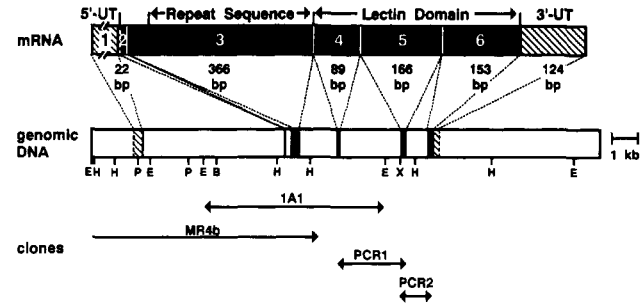

FIGURE 2: Organization of the murine ϵBP gene. (Top frame) Demarcation of exons in mature ϵBP mRNA. The size (bp) of each exon is indicated below. The size of exon 1 is unknown; the sizes for exons 2–6 are delineated from this work, and the size for the 3′ UT region is based on a cDNA clone reported for L-34 (Raz et al., 1989). (Middle frame) Location of the six exons on the genomic map. Restriction sites for *Eco*RI (E), *Hin*dIII (H), *Pst*I (P), *Bam*HI (B), and *Xba*I (X) are shown below. (Bottom frame) Region of the ϵBP gene carried by bacteriophage λ clones (1A1 and MR4b) or plasmid clones generated following PCR amplification of genomic DNA.

clone overlaps another clone (1A1) which contains a 7-kb insert and was isolated from a library containing size-selected *Eco*RI fragments (predicted by the restriction mapping of the mouse ϵBP gene). The two clones overlap by approximately 5 kb as determined by restriction mapping, blot hybridization analysis, and DNA sequencing (Figure 2). The DNA sequences of the entire 1A1 clone and approximately 3.8 kb of the MR4b clone were determined (these sequences have been deposited in GenBank under Accession No. M97896). Analysis of the sequence showed that these two clones contain four exons. Exon I is a 5′ UT region present in poly(A)-containing RNA and some cDNA clones (see below). Exon II contains 4 bp of 5′ UT mRNA sequence and 18 bp coding for the first six amino acids of murine ϵBP including the initial methionine. Exon III (366 bp) encodes 122 amino acids (numbers 7–129), and exon IV (89 bp) codes for 30 amino acids (numbers 130–160) of the cDNA. The sequence of exons II–IV is nearly identical to the corresponding sequence for L-34 (Raz et al., 1989) with the exception that nucleotides 11 and 12 in L-34 cDNA are CG while the corresponding positions in our genomic DNA are GC. The sequence of our genomic DNA differs from that of CBP35 (Jia & Wang, 1988) in two regions, resulting in divergence of the predicted amino acids 91, 92, and 109–111 [the numbering system in Jia and Wang (1988)].

We were unable to isolate any clone containing the DNA encoding the remaining 90 amino acids of the ϵBP protein from a total of 5 independent genomic libraries in λ vectors that were screened using a variety of cDNA probes and the 3′ portion (2.5-kb *Hin*dIII–*Eco*RI fragment) of the 1A1 genomic clone. The inability to obtain a genomic clone of this region in a λ vector probably reflects some artifact in the cloning system (e.g., a functional λ *cos* site may be present in the genomic DNA, leading to cutting of the clone by λ endonuclease). We were able to clone this portion of genomic DNA in plasmid vectors (see below).

The carboxyl-terminal domain of ϵBP shares significant sequence similarities with 14-kDa S-type lectins (Drickamer, 1988; Liu, 1990). The 14-kDa lectin genes consist of four exons of 96, 80–83, 160–172, and 144–150 bp (Gitt & Barondes, 1991; Ohyama & Kasai, 1988). The fourth exon located on our clone 1A1 corresponds to the first part of the carboxyl-terminal domain of ϵBP, and, indeed, it matches in size and shares sequence homology to the second exon of the genes coding for the 14-kDa lectins. On the basis of the genomic organization of the 14-kDa lectin genes, we predicted

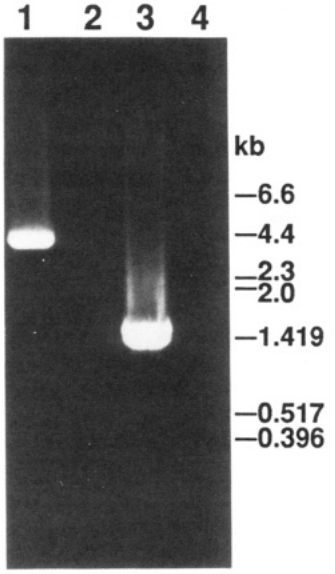

FIGURE 3: Agarose gel electrophoresis analysis of DNA products generated by PCR. Parts of the murine ϵBP gene were amplified by PCR. The products were analyzed on a 1.0% agarose gel and visualized by ethidium bromide staining. Lane 1, PCR product generated by using primers corresponding to the coding sequence from the 3′ end of exon IV and the coding sequence predicted to be in exon V. Lane 3, PCR product generated by using primers corresponding to the 5′ end of exon V and the 5′ end of the 3′ UT region. The positions of size markers (kb) are indicated on the right margin. Lanes 2 and 4 represent negative controls using primers as in lanes 1 and 3, respectively, but without template DNA added.

that there would be two more exons in the ϵBP gene (exons V and VI, corresponding respectively to exons III and IV of the smaller lectins). Because of the technical difficulties we encountered trying to isolate this portion of the gene from genomic libraries, we decided to generate clones containing these exons by using PCR.

Two separate PCRs were performed. In the first, primers corresponding to 23 bp of coding sequence from the 3′ end of exon IV and 24 bp of coding sequence predicted to be in exon V were used. As shown in Figure 3, a 2.7-kb product was formed, which overlaps the 3′ end of clone 1A1 by approximately 1.9 kb and contains an internal *Eco*RI site from the genomic sequence. The DNA was digested with *Eco*RI and *Bam*HI to yield a 1.9-kb *Eco*RI–*Eco*RI fragment and a 0.8-kb *Eco*RI–*Bam*HI fragment. Both fragments were subcloned and partially sequenced. In the second reaction, primers correspond to the first 24 bp of exon V (predicted from the cDNA sequence of the 3′ end of exon IV on the first PCR clone) and 22 bp in the 3′ untranslated region present in cDNA beginning immediately following the TAA stop codon of the cDNA (Jia & Wang, 1988). As shown in Figure 3, a 1.4-kb product was produced, which was subcloned and sequenced. As predicted, this clone contains two exons (166 bp exon V and 153 bp exon VI) of ϵBP. The sequence is identical to the corresponding cDNA sequence of L-34 (Raz et al., 1989) but different from that of CBP35 (Jia & Wang, 1988) by a single base (C→A) at nucleotide 624 [the numbering system in Jia and Wang (1988)]. This single base change does not result in an amino acid difference in the deduced protein sequence. Because our 3′ PCR primer for exon VI is contained within the 3′ UT portion of the exon, we have not cloned or sequenced that portion which occurs 3′ to the primer.

Genomic Organization of the ϵBP Gene and Comparison with the 14-kDa Lectin Gene. From the sequence of MR4b, 1A1, and two PCR clones, the entire organization of the ϵBP gene was established (Figure 2). The gene is composed of six

DONOR	ACCEPTOR
5′ UT (exon I) CTAATCAG/gtgagc	*exon II* cccccttggcttttctctcttag/GAAA**ATG GCA GAC** *Met Ala Asp*
5′ UT (exon I) CTACCCAG/gtgagc	*exon II* cccccttggcttttctctcttag/GAAA**ATG GCA GAC** *Met Ala Asp*
exon II **AGC TTT TCG**/gtaagt *Ser Phe Ser*	*exon III* tttgtctttctttatag/**CTT AAC GAT** *Leu Asn Asp*
exon III **GGA CCA CTG**/gtaaga *Gly Pro Leu*	*exon IV* tgcttgtgtgtcccag/**ACG GTG CCC** *Thr Val Pro*
exon IV **GCA AAC AG**/gtacaa *Ala Asn Ar*	*exon V* tcttctatttctggtttaaaacag/**G ATT GTT CTA** *g Ile Val Leu*
exon V **CCA TTC AAA**/gtaagt *Pro Phe Lys*	*exon VI* tttcccacgtttggttcctag/**ATA CAA GTC** *Ile Gln Val*

FIGURE 4: Sequences of the ϵBP gene at the exon/intron junctions. The boldface characters indicate the protein coding region. The predicted amino acids are indicated below the nucleotide sequences. The slash indicates the splicing junction for each pair of exons.

Table I: Sequence Comparison between Exons of Mouse ϵBP and Mouse L-14[a]

L-14 exon	II	III	IV
ϵBP exon	IV	V	VI
% nucleotide identity	44.4	57.8	43.5
% peptide identity	26.9	47.3	14.3

[a] Nucleotide and peptide sequences for ϵBP are from this work as well as those reported for L-34 (Raz et al., 1989); those for L-14 are from Wilson et al. (1989) (GenBank Assession No. X53067).

exons: the repetitive sequence in the amino-terminal domain is included entirely in exon III, whereas the carboxyl-terminal lectin domain is encoded by three exons: exons IV, V, and VI. For all of the coding exons, the intron–exon splice junctions are readily identifiable, and they generally conform to the consensus sequences AG/GTAAGT at the donor site and $(Y)_{10}$NCAG/G at the acceptor site (Figure 4).

It is remarkable that the first exon coding for the first part of the 5′ UT sequence is nearly 6 kb upstream from the remaining 5′ UT sequence and protein translation initiation site. Exon II, which contains the initiation codon as well as codons for five additional amino acids, is unusually short compared to the average exon length (Smith, 1988). Interestingly, in the genes for 14-kDa lectins that have been cloned, exon I which contains the translation initiation codon is also very short and codes for only the first three amino acids.

The three exons coding for the carboxyl-terminal domain of ϵBP match closely in size to those coding for the 14-kDa lectin. It is also noteworthy that the splice junction of exons IV and V occurs between nucleotides 2 and 3 in a codon, whereas splice junctions of exons III–IV and V–VI do not interrupt a codon (Figure 4). An identical situation is observed for the corresponding intron/exon boundaries for the 14-kDa lectin genes (i.e., intron II does, but introns I and III do not, interrupt a codon) (Gitt & Barondes, 1991; Ohyama & Kasai, 1988). These results indicate an evolutionary common ancestor gene for these two members of S-type lectins. Comparison of DNA and protein sequences of these exons between these two members of the S-type lectins also revealed homology. The highest percentage of sequence identity is found between exon III of the 14-kDa lectin and exon V of ϵBP (Table I). It is possible that these exons code for galactose-binding peptides which are more highly conserved.

Molecular Basis for Generating Two mRNAs with Different 5′UT Sequences. The transcription start site of the ϵBP gene, and thus the exact length of exon I, is not known at this time. The 5′ UT regions of a cDNA clone coding for a lectin designated as L-34 (which is now known to be identical to ϵBP) is 144 bp long. Except for the first five bases, much of

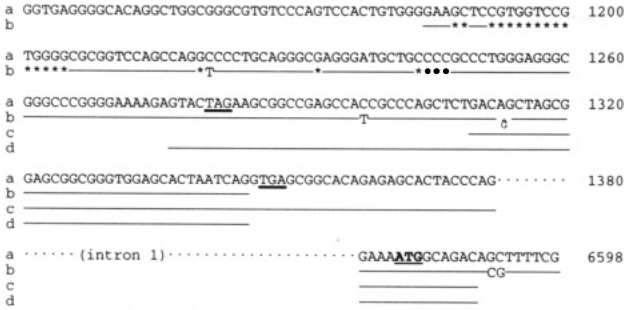

FIGURE 5: Comparison of 5′ UT sequences from various cDNAs of murine ϵBP to the genomic DNA sequence. Line a, murine ϵBP genomic DNA sequence established in this work, corresponding to exons I and II. The numbers on the right margin refer to numbers of base pairs from the 5′ end of the sequenced part of clone MR4b (GenBank Accession No. M97896). Line b, 5′ UT sequence reported for L-34 (Raz et al., 1989). Lines c and d, 5′ UT sequence of cDNA clones for Mac-2 clones, pBC 7 and pBC 12, respectively, obtained from PCR-amplified products (Cherayil et al., 1989). Horizontal lines indicate identical sequences, and asterisks indicate gaps introduced for optimal alignment of sequences. The previously predicted protein translation initiation site (ATG) is boldface, and stop codons in-frame and 5′ to ATG are double-underlined in line a.

the 5′ UT sequence of L-34 matches to a contiguous sequence in the ϵBP genomic DNA (Figure 5). In addition, there are five deletions and two mismatches in the remaining sequence, as compared to the genomic sequence. Two 5′ UT sequences have been predicted from sequencing of two Mac-2 cDNA clones (pBC 7 and pBC 12) (Cherayil et al., 1989). One differs from the other by having 27 bp of additional internal sequence. As shown in Figure 5, lines c and d, both 5′ UT sequences match perfectly with our genomic sequence.

One important reason for scrutinizing the 5′ UT sequence is that a classical signal peptide has not been identified at the amino terminus of protein sequences deduced from cDNA of ϵBP and its homologues. However, these proteins are expressed on the cell surface (Cherayil et al., 1989; Gritzmacher et al., 1988; Frigeri & Liu, 1992; Ho & Springer, 1982) and are secreted (Cherayil et al., 1989). Therefore, it has been questioned whether the first ATG in the cloned cDNA is the translation initiation site. It has been previously noted that a stop codon in-frame to the putative translation initiation ATG codon is found in both the pBC 7 and pBC 12 sequences and there is no other in-frame ATG codon (Cherayil et al., 1989). The sequence of our genomic DNA further strengthens the conclusion that this ATG codon is most likely the protein translation initiation site. Therefore, the molecular basis for translocation of ϵBP across the membrane has not yet been elucidated.

Examination of the ϵBP genomic sequence also revealed the mechanism for generating two mRNA species with varying 5′ UT sequences. Two potential donor sites occur in exon I separated by 27 bp. Theoretically, both sites can be spliced to a common acceptor site located 4 bp 5′ of the ATG translation initiation codon in exon II (Figure 6). This splice acceptor site consists of a $(Y)_{10}$TTAG/G preceded by another octamer polypyrimidine track separated by two purines (GG). Both splice donor sites are identical, consisting of the sequence AG/GTGAGC. In addition, there is a small genomic duplication of the sequence GAGCGGC$(X)_5$GAGCACTA occurring 3 or 4 bp 5′ of each donor site.

Analysis of Alternative 5′-End Utilization in mRNA Using PCR. We examined poly(A)-containing RNA isolated from a variety of murine cell lines and tissues to see if preferential usage of either of the 5′ UT regions correlated with the cell

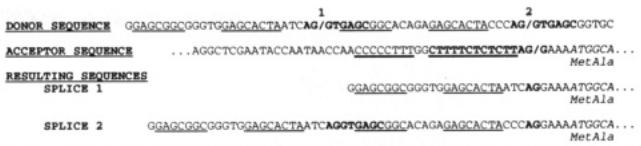

FIGURE 6: Alternative splicing to generate two ϵBP mRNAs varying only in 5′ UT sequence. The donor sequences in the 5′ UT exon I region and the accepter sequence of the exon II region are shown. Boldface letters show the donor and accepter sequences with a slash to indicate the splice junctions. Two donor sites present in ϵBP mRNA are indicated by 1 and 2 above the slash marks. The coding sequence is shown in italic with the corresponding amino acid abbreviation underneath. The pyrimidine-rich sequence in the accepter region is double-underlined, and the repeated sequences GAGCGGC-$(X)_5$GAGCACTA present in donor region are single-underlined.

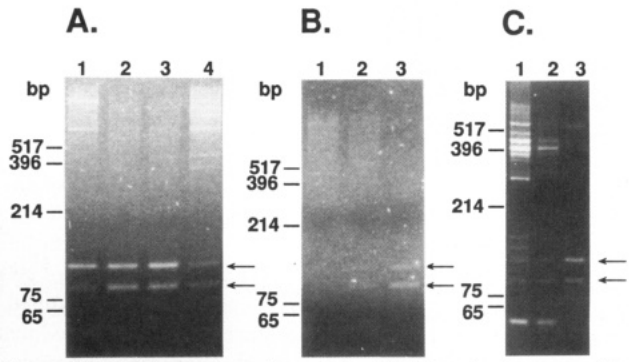

FIGURE 7: Gel electrophoresis analysis of PCR products representing two ϵBP mRNAs differing in the 5′ UT sequence. Poly(A) mRNA from various murine cell lines and tissues was used as template in PCR. The 5′ primer used corresponds to a sequence present on both Mac-2 cDNAs [pBC 7 and pBC 12 (Cherayil et al., 1989)]; the 3′ primer used corresponds to a sequence in exon III. The reaction products were analyzed on 1% agarose/1.3% Nu-sieve gels and revealed by ethidium bromide staining. (Panel A) Lane 1, SBV6; lane 2, P388D1; lane 3, WEHI-3; lane 4, 3T3 (for each, first-strand template, 1 μL; number of PCR cycles, 25). (Panel B) Lane 1, mouse heart; lane 2, mouse brain; lane 3, mouse liver (for each, first-strand template, 5 μL; number of PCR cycles, 25). (Panel C) Lane 1, mouse brain; lane 2, mouse heart (for each, first-strand template, 20 μL; number of PCR cycles, 40); lane 3, WEHI-3 included as a reference (first-strand template, 1 μL; number of PCR cycles, 40). Positions of size markers are indicated on the left of each panel. The positions of expected PCR products (121 and 94 bp) corresponding to two different 5′ UT sequences are indicated by arrows.

types or functions. We isolated poly(A)-containing RNA from murine cell lines that represent macrophages (P388D1 and WEHI-3), hepatocytes (SVB6), and fibroblasts (3T3) and from murine liver, brain, and heart tissues. The presence of both potential 5′ mRNA ends was detected as two DNA fragments of 94 and 121 bp that were generated using first-strand cDNA synthesis followed by PCR amplification of the DNA. The 94 bp fragment represents mRNAs that have been spliced using the 5′ donor splice site of the UT exon I (donor site 1, Figure 6), whereas the 121 bp fragment represents mRNA spliced at the 3′ donor site of the UT exon I (donor site 2, Figure 6).

From all four cell lines, both PCR products were easily detected on sieving agarose gels as shown in Figure 7A. The corresponding products were not detected when mRNAs generated from various mouse tissues (liver, brain, and heart) were amplified using similar amounts of the template and PCR conditions (data not shown). However, when the amounts of the templates obtained from tissue samples were increased 5-fold, both products were generated from the liver mRNA (Figure 7B). When the number of cycles in the PCR reactions was increased from 25 to 40, the two bands were clearly detected in the products amplified from heart mRNA (Figure 7C). However, under these conditions, the background

increased so much that the two bands were obscured in the products from brain mRNA (Figure 7C). The identity of the PCR products was confirmed by DNA (Southern) blot hybridization using oligomer probes that correspond to the terminal 5′ sequence and to an internal sequence present only in products of splicing at donor position 2 (data not shown). Both the 121 and 94 bp fragments contain the same 5′ region and hybridize to the 5′ probe, whereas only the 121 bp fragment contains the internal sequence and hybridizes to that probe. In addition, the longer PCR product was also sequenced, confirming that the amplified DNA contains the UT exon I.

In most PCR reactions, both the 94 and 121 bp products were present in nearly equal amounts as detected by gel analysis. However, for SVB6, there is significantly more of the larger PCR fragment as compared to the smaller one. This pattern was seen consistently in three separate experiments. Because the same primers were used to amplify both products, there should be no technical preference for amplification of either product (Wang et al., 1989). Therefore, we conclude that there is no preferential usage of either of the 5′ UT ends in most of the cell lines and tissues examined. Whether in some cells, such as SVB6 cells, the preferential usage of one of the 5′ UT ends is a function of the rate of cell proliferation, among other factors, remains to be determined. In addition, there was significantly more of both PCR products detected in the cell lines analyzed compared to the mouse tissues using identical conditions for cDNA synthesis and PCR amplification. Although the results would require confirmation by more quantiative methods, they correlate with the observations of ourselves and others that more ϵBP is present in transformed cells or tumors lines compared to normal cells (Moutsatsos et al., 1987; Raz et al., 1989). The difference in the amounts of the PCR products from various tissues most likely reflects the variation in the level of ϵBP mRNA in these tissues. We have previously shown by Northern blot analysis of rat tissues that there is much less ϵBP transcript in the brain as compared to the liver (Gritzmacher et al., 1988).

In conclusion, we confirmed that alternative splicing occurs at the 5′ UT end of ϵBP mRNA. However, there does not seem to be a correlation between 5′ UT splicing and ϵBP gene expression in macrophages, where the alternative splicing was first detected (Cherayil et al., 1989), or in any other cell type. Whether a biological function for alternative RNA splicing in the ϵBP message exists is yet to be determined.

ACKNOWLEDGMENT

We thank Drs. Dan Hsu and Mark Richards for critical review of the manuscript and Velda Comstock for excellent assistance in preparation of the manuscript. All oligonucleotide primers were synthesized in the department DNA Core Lab at The Scripps Research Institute supported by a grant from the Sam and Rose Stein Charitable Foundation.

REFERENCES

Abbott, W. M., & Feizi, T. (1989) *Biochem. J. 259*, 291.
Albrandt, K., Orida, N. K., & Liu, F.-T. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6859.
Barondes, S. H. (1988) *Trends Biochem. Sci. 13*, 480
Cherayil, B. J., Weiner, S. J., & Pillai, S. (1989) *J. Exp. Med. 170*, 1959.
Cherayil, B. J., Chaitovitz, S., Wong, C., & Pillai, S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 7324.
Clerch, L. B., Whitney, P., Hass, M., Brew, K., Miller, T., Werner, R., & Massaro, D. (1988) *Biochemistry 27*, 692.
Crittenden, S. L., Roff, C. F., & Wang, J. L. (1984) *Mol. Cell. Biol. 4*, 1252.
Drickamer, K. (1988) *J. Biol. Chem. 263*, 9557.
Frigeri, L. G., & Liu, F.-T. (1992) *J. Immunol. 148*, 861.
Gitt, M. A., & Barondes, S. H. (1991) *Biochemistry 30*, 82.
Gritzmacher, C. A., Robertson, M. W., & Liu, F.-T. (1988) *J. Immunol. 141*, 2801.
Henikoff, S. (1984) *Gene 28*, 357.
Hirabayashi, J., & Kasai, K.-I. (1988) *J. Biochem. 104*, 1.
Ho, M.-K., & Springer, T. A. (1982) *J. Immunol. 128*, 1221.
Jia, S., & Wang, J. L. (1988) *J. Biol. Chem. 263*, 6009.
Leffler, H., & Barondes, S. H. (1986) *J. Biol. Chem. 261*, 10119.
Leffler, H., Masiarz, F. R,. & Barondes, S. H. (1989) *Biochemistry 28*, 9222.
Liu, F.-T. (1990) *CRC Crit. Rev. Immunol. 10*, 289.
Liu, F.-T., Albrandt, K., Mendel, E., Kulczycki, A., Jr., & Orida, N. K. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4100.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Moutsatsos, I. K., Davis, J. M., & Wang, J. L. (1986) *J. Cell Biol. 102*, 477.
Moutsatsos, I. K., Wade, M., Schindler, M., & Wang, J. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6452.
Oda, Y., Leffler, H., Sakakura, Y., Kasai, K., & Barondes, S. H. (1991) *Gene 99*, 279.
Ohyama, Y., & Kasai, K. (1988) *J. Biochem. 104*, 173.
Paroutaud, P., Levi, G., Teichberg, V. I., & Strosberg, A. D. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6345.
Raz, A., Meromsky, L., Zvibel, I., & Lotan, R. (1987) *Int. J. Cancer 39*, 353.
Raz, A., Pazerini, G., & Carmi, P. (1989) *Cancer Res. 49*, 3489.
Raz, A., Carmi, P., Raz, T., Hogan, V., Mohamed, A., & Wolman, S. R. (1991a) *Cancer Res. 51*, 2173.
Raz, A., Zhu, D., Hogan, V., Shah, N., Raz, T., Karkash, R., Pazerini, G., & Carmi, P. (1991b) *Int. J. Cancer 46*, 871.
Robertson, M. W., Albrandt, K., Keller, D., & Liu, F.-T. (1990) *Biochemistry 29*, 8093.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.
Smith, M. W. (1988) *J. Mol. Evol. 27*, 45.
Southern, E. (1975) *J. Mol. Biol. 98*, 503.
Wang, A. M., Doyle, M. V., & Mark, D. F. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9717.
Wilson, T. J., Firth, M. N., Powell, J. T., & Harrison, F. L. (1989) *Biochem. J. 261*, 847.
Woo, H. J., Shaw, L. M., Messier, J. M., & Mercurio, A. M. (1990) *J. Biol. Chem. 265*, 7097.