GENOMICS 16, 63–68 (1993)

Cloning and Structure of the Human Corticotrophin Releasing Factor-Binding Protein Gene (CRHBP)

DOMINIC P. BEHAN,*,[1] ELLEN POTTER,* KATHY A. LEWIS,* NANCY A. JENKINS,† NEAL COPELAND,† PHILIP J. LOWRY,‡ AND WYLIE W. VALE*

*Clayton Foundation Laboratories for Peptide Biology, The Salk Institute, La Jolla, San Diego, California 92037; ‡Department of Biochemistry and Physiology, Reading University, Reading RG6 2AJ, England; †Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702-1201

Received April 13, 1992; revised December 15, 1992

The human CRF-binding protein gene has been cloned and mapped to the distal region of chromosome 13 and loci 5q in the mouse and human genomes, respectively. The gene consists of 7 exons and 6 introns. The mature protein has 10 cysteines and 5 tandem disulfide bridges 4 of which are contained within exons 3, 5, 6, and 7. One bridge is shared by exons 3 and 4. The signal peptide and the first 3 amino acids of the mature protein were coded for by an extreme 5′ exon. Primer extension analyses revealed the transcriptional initiation site to be located 32 bp downstream from a consensus TATA box. The promoter sequence contained a number of putative promoter elements including an AP-1 site, three ER-half sites, the immunoglobulin enhancer elements NF-κB and INF-1, and the liver-specific enhancers LFA1 and LFB1. © 1993 Academic Press, Inc.

INTRODUCTION

A 37-kDa protein (CRF-BP) capable of binding the neuropeptide corticotrophin releasing factor (CRF) with high affinity (Linton and Lowry, 1986; Orth and Mount, 1987; Linton *et al.*, 1988, 1989; Suda *et al.*, 1988) has been purified from human plasma (Behan *et al.*, 1989), partially sequenced and cloned (Potter *et al.*, 1991) from both human liver and rat brain libraries. We have recently found that the protein is widely distributed throughout the rat brain and have been successful in localizing its expression in many brain areas including the cortex and pituitary corticotropes (Potter *et al.*, 1992). We have also determined that the protein is expressed in primate brain and in human liver (Potter *et al.*, 1991). The hCRF-BP primary amino acid sequence contains 11 cysteines, thus predicting the presence of 5 disulfide bonds in the mature protein after cleavage of the signal peptide, which contains 1 cysteine (Potter *et al.*, 1991). Recently, we reported that the secondary structure of the hCRF-BP consisted of five loops formed by the sequential bonding of adjacent cysteines (Fischer *et al.*, 1991). To understand more about the CRF-BP's structure and its tissue-specific expression, we have cloned the human CRF-BP gene and identified 949 bp of its extreme 5′ sequence that is likely to be responsible for the regulation of CRF-BP gene expression.

[1] To whom correspondence should be addressed. Telephone: (619) 453-4100, ext. 510. Fax: (619) 552-1546.

MATERIALS AND METHODS

Construction of the human genomic library. A number of commercial genomic libraries were first screened for the CRF-BP gene without success; therefore, we decided to construct our own library from human liver DNA. Fifteen micrograms of human liver DNA was subjected to partial digestion with 7.5 units of *Bam*HI for 1 h at 37°C. The DNA was then ethanol precipitated and phosphatased with 10 units of alkaline phosphatase for 30 min at 37°C. SDS and EDTA were then added to final concentrations of 0.5% and 5 m*M*, respectively. The reaction was then mixed and incubated for a further 30 min with proteinase K at 56°C. The digested DNA was ethanol precipitated and reconstituted in 20 μl of sterile water. Approximately 2 μg (3 μl) of this DNA was ligated to 600 ng of *Bam*HI-digested λ phage arms (Stratagene, La Jolla, CA) in a total reaction volume of 15 μl. The ligated DNA was then packaged using the Gigapack II gold packaging extract (Stratagene). Briefly, 15 μl of the ligated DNA was mixed with 18 μl of packaging mix and 30 μl of sonic extract. This was incubated at 22°C for 2 h and subsequently diluted in 1 ml SM buffer. The estimated titer of the unamplified library (using P2PLK-17 as the host bacterial strain) was 2×10^6 PFU/ml.

Gene cloning and mapping. One million plaque-forming units of a human genomic library were screened using a ^{32}P randomly labeled (Amersham multiprime labeling kit) *Pst*I DNA fragment that contained the first 500 bp of the human cDNA coding sequence. Duplicate nitrocellulose filters were hybridized in 5× SSC, 1× Denhart's solution, 0.1% SDS, 100 μg/ml salmon sperm DNA, and 1×10^6 cpm/ml of ^{32}P-labeled cDNA probe. Filters were washed in 2× SSC, 0.1% SDS for 2 × 15 min at 65°C. One clone obtained from the primary screen was plaque purified and used to make pure phage DNA. The DNA was digested with *Bam*HI and ligated directly into *Bam*HI-digested pBluescript KS (Stratagene). After transformation of XL1-blues, two subclones that harbored 10- and 8-kb *Bam*HI fragments upon digestion with *Bam*HI were obtained. The entire gene was subcloned as an 18-kb *Sal*I fragment (utilizing the *Sal*I sites on either side of the *Bam*HI sites in the polylinker of the λ phage clone) into *Sal*I-digested pBluescript in the same manner. Sense and antisense primers, made to correspond to evenly spaced cDNA coding sequences, were used to prime and obtain sequence from the three subclones using the Sanger

0888-7543/93 $5.00

dideoxy termination method (IBI). When intron/exon junctions were identified, oligonucleotide primers were made to the exonic side of the junction and sequence was then primed across the exon to the opposing intron/exon junction. Thus, sense oligonucleotides were used to prime and obtain sequence at the 3′ end of an exon and/or 5′ end of an intron and antisense oligonucleotides were used to prime and obtain sequence at the 3′ end of an intron and/or at the 5′ end of an exon. With this strategy we could account for all of the coding sequence, and all intron/exon junctions were located. The extreme 5′ putative promoter stretch was sequenced on both strands and all intron/exon junctions were sequenced on one strand.

Restriction mapping of the gene was performed by exposing the three different subclones to various restriction enzymes and analyzing the products after separation on 0.5–1% agarose gels. The DNA from the gels was transferred onto nitrocellulose (as described by Maniatis) and then probed with ^{32}P-end-labeled oligonucleotides corresponding to exons 2, 4, 6, and 7. The oligonucleotides were end labeled by the kinase method described for oligonucleotide 48 in the primer extension method below. The Southern blot of human genomic DNA was probed using a ^{32}P randomly labeled PCR fragment corresponding to the first 377 bp of coding sequence, which spanned exons 1–4.

Chromosomal mapping. Interspecific backcross progeny were generated by mating [C57BL/6J. × *Mus spretus*] F_1 females and C57BL/6J males as described (Copeland and Jenkins, 1991). A total of 205 N_2 mice were used to map the CRF-BP locus (see text for details). DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer, and hybridization were performed essentially as described (Jenkins *et al.*, 1982). All blots were prepared with Zetabind nylon membrane (AMF-Cuno). The probe, a 630-bp rat cDNA clone, was labeled with [α-^{32}P]dCTP using a random prime labeling kit (Amersham): washing was done to a final stringency of 0.8× SSCP, 0.1% SDS, 65°C. Major fragments of 3.8, 1.1, and 0.6 kb were detected in *Taq*I-digested *M. spretus* DNA. The presence or absence of the 2.8- and 1.2-kb *M. spretus*-specific *Taq*I fragments, which cosegregated, was followed in backcross mice.

A description of the probes and RFLPs for the loci linked CRF-BP including neuroendocrine convertase-1 (NEC-1), RAS p21 protein activator (RASA), dihydrofolate reductase (DHFR), and cytotoxic T lymphocyte-associated protein-3 (CTLA-3) has been given previously (Copeland *et al.*, 1992). Recombination distances were calculated as described (Green, 1981) using the computer program SPRETUS MADNESS. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

The human CRF-BP chromosomal location was obtained by probing a Southern blot of DNA obtained from somatic cell hybrids. Briefly, a panel of DNA was isolated from 24 human/rodent somatic cell hybrids retaining one or two human chromosomes. All but 2 of the hybrids retained a single intact human chromosome. They are either human/mouse or human/Chinese hamster ovary cell (CHO) hybrids. The DNAs (Coriell Cell Repositories, NJ) were digested with *Eco*RI, separated on a 1% agarose gel, and then Southern blotted. The blot was then probed with a randomly ^{32}P-labeled PCR fragment to the first 377 bp of the hCRF-BP cDNA. The Southern blot was kindly provided by Dr. Makel Djabali of the Salk Institute.

Primer extension analysis. Antisense oligonucleotide DOM 48 (100 μg) was purified from free label on a 12% urea sequencing gel. The purified oligonucleotide (400 ng) was then kinased in a mixture containing 10 μl of [γ-^{32}P]ATP and 1 μl polynucleotide kinase for 30 min at 37°C. Free label was then separated from labeled oligonucleotide by Sephadex G-50 chromatography (0.75 × 5-cm column) in TE buffer, collecting 300-μl fractions throughout.

Human fetal liver RNA (100 μg of each) was ethanol precipitated and resuspended in 7 μl sterile TE, 2 μl 5× annealing buffer (1× TE containing 1.25 *M* KCl) and 1 μl of ^{32}P-end-labeled primer (total volume = 10 μl). The oligonucleotide was annealed at 55°C for 1 h and primer extension was subsequently performed by the addition of 23 μl of extension mix (20 m*M* Tris–HCl, pH 8.7, 10 m*M* $MgCl_2$, 100 μg/ml actinomycin D, 5 m*M* DTT, 0.33 m*M* dNTPs) and 10 units of AMV reverse transcriptase followed by further incubated at 50°C for 1 h. The RNA was then ethanol precipitated and resuspended in 4 μl of sterile water and 4 μl of sequenase stop solution. Two controls were used for this analysis: first yeast tRNA [100 μg] was subjected to primer extension in the same manner and second an identical reaction was performed without the addition of any RNA. Reaction products were then analyzed on an 8% sequencing gel.

RESULTS

The human CRF-BP gene was found to consist of 7 exons and 6 introns that spanned 16 kb of the cloned 18-kb *Sal*I fragment (Fig. 1). The intron/exon boundaries all contained consensus splice donor and acceptor sequences following the donor and acceptor rules previously reported (Breathnach *et al.*, 1987; Table 1). To generate the full-length genomic clone (*Sal*I BP), *Sal*I sites external to the *Bam*HI sites of the phage polylinker were used; i.e., the *Sal*I sites do not represent sites in genomic DNA (see gene cloning section under Materials and Methods). The signal peptide was located on a separate exon on the extreme 5′ end of the gene, and the polyadenylation signal sequence AATAAA was found to be contained on the 3′ untranslated region at the end of exon 7 (Fig. 1). The restriction map of the purified genomic clone shown in Fig. 1 predicted all the *Eco*RI, *Pst*I, and *Hin*dIII fragments shown on the Southern blot of human genomic DNA (Fig. 2). Interestingly, exons 5, 6, and 7 each contained two cysteines which we have previously reported to be linked to form sequential disulfide loops in the purified recombinant CRF-BP molecule (Fischer *et al.*, 1991; Fig. 1). Three cysteines are provided by exon 3 and one by exon 4 such that disulfide bonded loops 1 and 2 of the CRF-BP sequence are coded for by exons 3 and 4, respectively (Fig. 1).

Chromosomal mapping revealed the human gene to be present on chromosome 5 and the mouse locus to be located in the distal region of chromosome 13 (Fig. 3), which is indeed synonymous with human chromosome 5q.

Examination of the CRF-BP promoter sequence revealed a conserved TATA box 122 bp upstream from the first methionine of the N-terminal signal peptide (Fig. 4). To ensure that the 849-bp putative promoter sequence was not simply an extreme 5′ intron, primer extension analyses were performed on human fetal liver RNA. One major extended fragment was generated from human fetal liver RNA but not from yeast tRNA or when the reaction proceeded without the addition of RNA (Fig. 5). This identified a putative transcriptional start site to be the adenosine of the cap site found 32 bp downstream from the conserved TATA box (Fig. 4). The first methionine of the protein sequence was thus located at +91 bp and the N-terminal tyrosine codon of the reported CRF-BP sequence (Potter *et al.*, 1991) was positioned at +163 bp (Fig. 4). The extreme 5′ exon thus coded for the signal peptide plus YLE of the extreme amino terminus of the mature protein, with the first exon/intron junction being located at +171 bp.

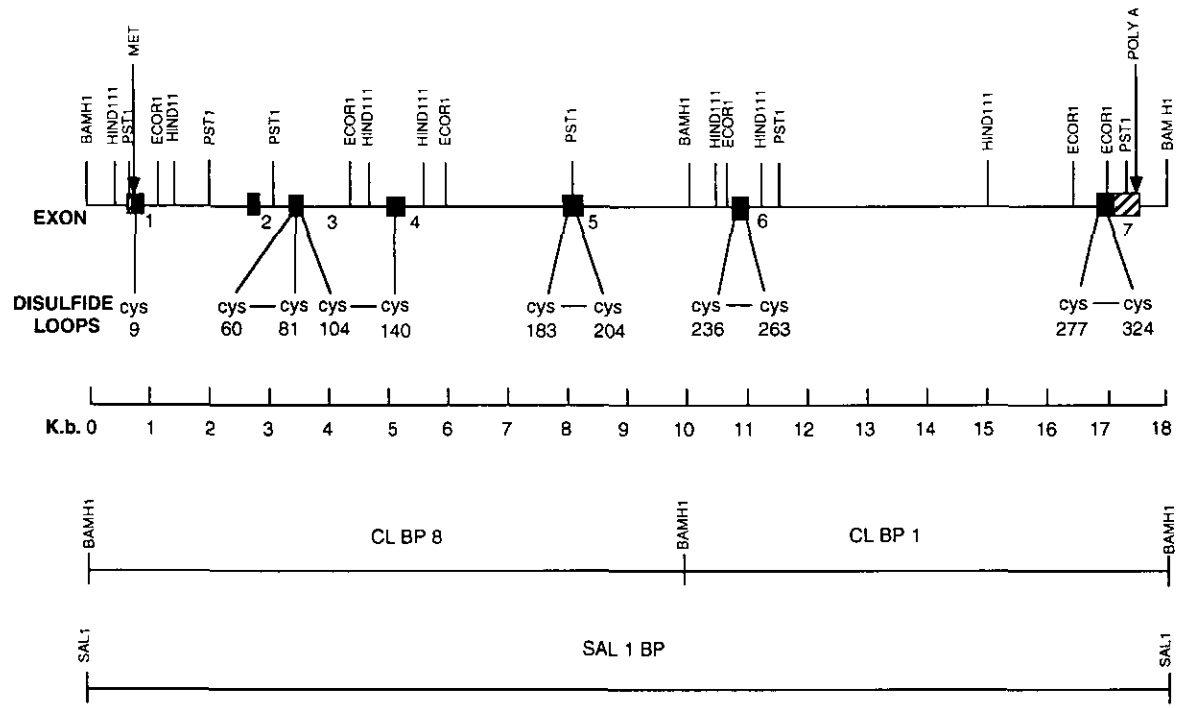

FIG. 1. Diagram showing the complete restriction map of the human CRF-BP gene. The introns are represented by solid lines and the exons are represented by black boxes. The exon number is marked next to each black box. Restriction mapping was performed by digesting the three subclones (*Sal*I BP, CL8 BP, CL1 BP) with various restriction enzymes (*Pst*I, *Hin*dIII, and *Eco*RI) and analyzing the fragments on a Southern blot after probing with ^{32}P-labeled oligonucleotides directed to each exon. The disulfide-bonded loops, positioned relative to their respective exons, are indicated by cys–cys and the residue number of each cysteine in the protein sequence is indicated by the number underneath. The hatched bars represent both 5′ and 3′ untranslated sequences present in the CRF-BP mRNA.

DISCUSSION

In this study we present the genomic structure and extreme 5′ nucleotide sequence of the human CRF-BP gene. The gene was found to have a multiple intronic/exonic structure consisting of seven exons and six introns. Exons 5, 6, and 7 each contained two cysteines, which are bonded to form disulfide loops in the purified recombinant CRF-BP molecule (Fischer *et al.*, 1991; Fig. 1). The four cysteines which form the first and second disulfide loops in the sequence are provided by exons 3 and 4, with three of the cysteines being provided by exon 3 and one by exon 4. Exon 1 codes for one cysteine in the signal peptide, which is cleaved in the mature secreted protein. Since exons can code for functional domains within proteins (Gilbert, 1978; Tonegawa, 1983) it is possible that these disulfide loops form domains in the mature CRF-BP. This structure was of particular interest since it resembled the genomic organization of genes from the IgG family, many of which have exons that code for disulfide-bonded loops (Barclay *et al.*, 1988). Furthermore, exons coding for domains in the IgG molecule have been reported to undergo somatic rearrangement and thus to generate antibody diversity by exon shuffling (Tonegawa, 1983). It is, therefore, possible that alternate splicing of the CRF-BP primary mRNA transcript may result in the expression of a different molecule(s) from the same gene by a reorganization of the disulfide loop domains.

TABLE 1

Sequences for Each Intron/Exon Junction in the CRF-BP Gene

Exon	Codon interupted	Exon–intron junction	
		5′ Donor	3′ Acceptor
1	Glu^{27}	CTAGAGgtgagc	.....tgccagCTGAGG
2	Arg^{59}	CTCTGCgtgagt	.....ccccagGGTGCC
3	Val^{112}	CTGAAGgtgagg	.....tcaaagGTATTT
4	Pro^{182}	TCTTTCgtaagg	.....acaaagCTTGCA
5	Lys^{231}	TTAAAGgtgagt	.....ccttagAAATCC
6	Ala^{271}	GCCCGGgtgagg	.....ttatagCCCAGA

Note. Capital letters indicate exons and lowercase letters indicate introns.

In this study we provide evidence for only one human CRF-BP gene since all the sizes of restricted bands identified on a Southern blot of human genomic DNA were identical to those generated from the human CRF-BP genomic clones (Fig. 2).

The mouse chromosomal location of CRF-BP was determined by interspecific backcross analysis using progeny derived from matings of [(CR7BL/6J × *M. spretus*) F_1 × C575BL/6J] mice. This interspecific backcross mapping panel has been typed for over 1100 loci that are well distributed among all autosomes as well as the X chromosome (Copeland and Jenkins, 1991). C57BL/6J and *M. spretus* DNAs were digested with several en-

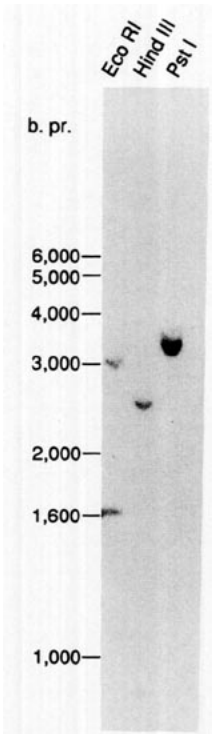

FIG. 2. Diagram showing a Southern blot of human genomic DNA after digestion with *Pst*I, *Hin*dIII, and *Eco*RI. The blot was probed with a 377-bp PCR fragment (1×10^6 cpm/ml) spanning exons 2, 3, and 4 of the genomic map. Prehybridizations and hybridizations were carried out in a solution of 5× SSC, 1× Denhart's solution, 0.1% SDS, 100 μg/ml salmon sperm DNA. All the bands seen on the Southern blot were predicted from the genomic restriction map shown in Fig. 1.

zymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLPs) using a rat cDNA CRF-BP probe. The 2.8- and 1.2-kb *M. spetus Taq*I RFLPs (see Materials and Methods) were used to follow the segregation of the CRF-BP locus in backcross mice. The mapping results indicated that CRF-BP is located in the distal region of mouse chromosome 13 linked to NEC-1, RASA, DHFR, and CTLA-3. Although 98 mice were analyzed for every marker and are shown in the segregation analysis (Fig. 3), up to 161 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the most likely gene order are centromere–NEC-1–2/161–RASA–3/116–DHFR–0/115–CRF-BP–14/137–CTLA–3. The recombinants were detected between DHFR and CRF-BP in 115 animals typed in common, suggesting that the two loci are within 2.6 cM of each other (upper 95% confidence limit).

We have compared our interspecific map of chromosome 13 with a composite mouse linkage map that reports the map location of many uncloned mouse mutations (compiled by M. T. Davisson, T. H. Roderick, A. L. Hillyard, and D. P. Doolittle and provided from GBASE, a computerized database maintained at The Jackson Laboratory, Bar Harbor, ME). CRF-BP mapped in the region of the composite map that lacks mouse mutations with a phenotype that might be expected for an alteration in this locus (data not shown).

The distal region of mouse chromosome 13 shares homology with human chromosome 5q (summarized in Fig. 3). In particular, DHFR has been placed on human 5q11.2–5q13.3. The tight linkage between DHFR and CRF-BP suggests that CRF-BP will reside on 5q as well.

To test this hypothesis we also determined the human chromosomal location of CRF-BP, which was indeed found to be present on human chromosome 5 by Southern blotting DNA from human–rodent hybrid cells and subsequently probing for the human CRF-BP gene with a labeled CRF-BP cDNA fragment.

Interestingly, the previously reported rat cDNA sequence (Potter *et al.*, 1991) started only 3 bp from the consensus TATA box of the human CRF-BP gene instead of the expected 32 bp downstream as predicted from the primer extension analyses (Fig. 4). This may suggest that there is some translational regulation of the CRF-BP gene expression so that the published 5′ rat cDNA sequence may be accounted for by the utilization of TATA boxes further upstream in the CRF-BP promoter (e.g., at positions −286 and −556; Fig. 4). Indeed, since the rat CRF-BP cDNA was cloned from a brain

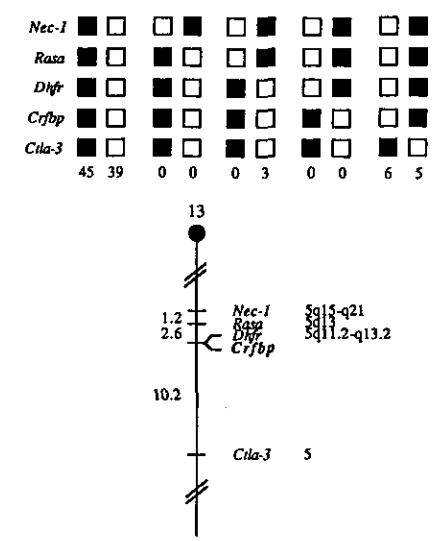

FIG. 3. CRF-BP maps in the distal region of mouse chromosome 13. CRF-BP was placed on mouse chromosome 13 by interspecific backcross analysis and the segregation patterns of CRF-BP and flanking genes in 98 backcross animals that were typed (see text). Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M. spretus*) F_1 parent. The shaded boxes represent the presence of a C57BL/6J allele and white boxes represent the presence of a *M. spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 13 linkage map showing the location of CRF-BP in relation to linked genes is shown at the bottom of the figure. Recombination distances between loci in centimorgans are shown to the left of the chromosome and the positions of loci mapped in this study can be obtained from GDB (Genome Data Base), a computerized database of human linkage information maintained by The William H. Welch Medical Library of The Johns Hopkins University (Baltimore, MD).

```
GGATCCCTAAGTCATCTTTTTAGCCCTGCAAATTTGCTGGTGAGCAGTTA  -849
BAM H1 SITE
CAATAAAGCTTCCTGGCAGTTCCTTCTACCAGAGTGTTAAAGTAATCCAG  -799

GTATTTGAAGGCACGAAATGGAATAGCTTAATCTCATCTCTCAGTGGTTC  -749

AGATGGATGCGAAGGTCATGGCTACAGCTACCTGGTCTAGGAGTGTGTAC  -699
           ER HALF SI>
ATCATGTAAACATCCAGGTCAAGTGACAGAGCTCAGCTAAAATCCAGAAG  -649
              ER HALF SI> INF-1>
TCCCCTGACACCAGAGTTAGTTTTCCATGAAAGAAATGCCTATAAGAAAC  -599

TGGACCAGGATAAAGCAAAGAGTAAGGGGCAGGAAGAAAGCTTATAATTG  -549
LF-A1>                                    TATA BOX>
CATTACTGCAATCTTCTGCAAAATCAGACTTCTGCTTCTGTTCCACCTCT  -499

TTAAAGCTGAAAATGCCTCAAGTTTCTTACTGAGAGACATGAGAAATTCC  -449

ATGAATACTAAATTCTTTGACAATTTCTAGCATGGTTTCATCCAAGGTTA  -399

TGTGATGAGGCTGCATAATATAGGCTCTATATGTTTGGGTATTTTATGCG  -349

TGTATGTAACTTTGTCACTTTGATTTGCAAAATGGAAAATTCCCTGAATG  -299
                                  <NF-kB CS1
CCCACAGAAAAGTATAATGTAGCATTTAGCAGGTCAACAAAATGCACTCC  -249
        TATA BOX>          ER HALF SI>
CTTATTCTTTGTCTCCGCCTCATTCTCCAACACTTTAGTGCACTAAGTTG  -199

TAGTGACTGAAATCATTCTTGGTGAATAATGCTGGGCCACGCTGAAAATT  -149
                       AP1 CS3>
TGTGGCTGAGAGCTGGACCCTCGTCATCGCCACGTACTCTGTCAATGAGA  -99
             LF-A1>
CAGGGTAACCCATGGTTACTGAGCTCCGATCAGTTAAGGGGCAGAGAGCC  -49
                                                 : :
TCCGCTCCCCACTGCTCTATAAAAGAGACCCAGCAAAGGGACCCTACCAG  +2
                 TATA BOX>                 CAP SITE>
CTTCTAGCTCTCAGTCTGCGCGAGGGTGTAGGAAGGAAAGCCCAGGACCT  +52
                                        MET
CCGGAGCAGAGCACAGCAGCTGCAGAGGCAAGGCCAGCATGTCGCCCAAC  +102
                  <OLIGO DOM 48
TTCAAACTTCAGTGTCACTTCATTCTCATCTTCCTGACGGCTCTAAGAGG  +152

          SPLICE JUNCTION
GGAAAGCCGGTACCTAGAGGTGAGCCACCCCTGGACTGACCCATCTCAC   +203
            Y  L  E
N-TERMINAL AMINO ACIDS
```

FIG. 4. Diagram showing the extreme 5′ sequence of the human CRF-BP gene up to the 5′ *Bam*HI site of CL BP 8. The position of the conserved TATA box is shown underlined as TATA and the transcriptional initiation site is indicated by +1. The first methionine is indicated by MET and the first three N-terminal amino acids of the mature protein are indicated by YLE. The first exon/intron splice junction and the oligonucleotide, Dom 48, used for the primer extension analyses are also marked. All putative promoter elements are underlined.

library it is possible that the CRF-BP message starts from different places depending on if it is expressed in the brain or in the liver. It still must be noted, however, that the rat promoter sequence may differ in this respect, which would account for the longer 5′ untranslated region of the rat cDNA (Potter *et al.*, 1991). Clarification of this awaits the cloning and characterization of the rat CRF-BP promoter sequence.

The promoter sequence was screened for promoter elements using the computer program Macvector. A number of putative promoter elements were found, which may explain to some degree the proteins tissue distribution and give clues to its transcriptional regulation. There is an AP1 site at −176 bp, which is known to bind the general transcription factors fos and jun (Lee *et al.*, 1987), but there was no evidence for a cyclic AMP response element (CRE) in the extreme 5′ sequence (Fig. 4). We have previously reported that the human CRF-BP is expressed in the liver (Potter *et al.*, 1992) and that this tissue is likely to be the main source of the circulating protein in normal human plasma. In our search for tissue-specific elements we were therefore interested in finding sequences corresponding to the core motif of TG[GA][AC]CC at −135 and −598 bp, which has been reported to bind two distinct liver-specific enhancer proteins LFA1 and LFB1 (Hardon *et al.*, 1988; Fig. 4). In light of the structural similarities between the hCRF-BP gene and the immunoglobulin family of genes it was intriguing to find a number of immunoglobulin-like enhancer elements in the extreme 5′ flanking region. The immunoglobulin enhancer element for NF-κB, a transcription factor known to regulate immunoglobulins and interleukins, was located at −305 bp (Lenardo and Baltimore, 1989; Fig. 4). This was of particular interest since NF-κB has been shown to bind to the acute-phase response element (AFPE) of the angiotensin gene and is thought to be important in activating liver-specific genes in response to inflammation or injury (Ron *et al.*, 1989). Indeed one other immune-type enhancer element

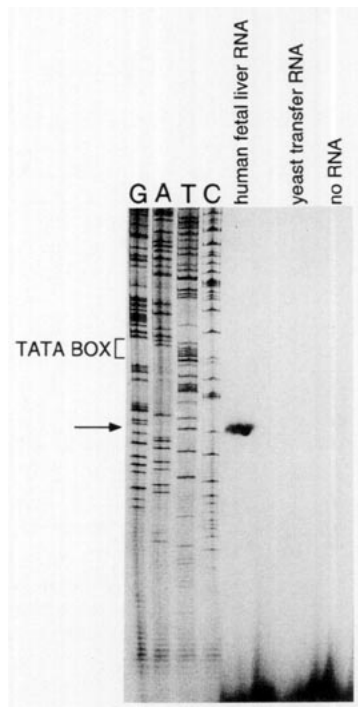

FIG. 5. Diagram showing primer extension analyses of human fetal liver RNA. Human fetal liver RNA (100 μg) was ethanol precipitated and resuspended in 7 μl sterile TE, 2 μl 5× annealing buffer (1× TE containing 1.25 *M* KCl) and 1 μl of ^{32}P-end-labeled primer Dom 48 (total volume = 10 μl). The reaction was annealed at 55°C for 1 h after which time 23 μl of extension mix (20 m*M* Tris–HCl, pH 8.7, 10 m*M* $MgCl_2$, 100 μg/ml actinomycin D, 5 m*M* DTT, 0.33 m*M* dNTPs) and 10 units of AMV reverse transcriptase were added. The reaction was then further incubated at 50°C for 1 h and the RNA was precipitated and resuspended in 4 μl of Sequenase stop solution. Two controls were used for this analysis: first yeast tRNA was subjected to primer extension in the same manner and second the reaction was allowed to proceed without the addition of RNA. All reaction products were then analyzed on 8% urea sequencing gels. Sequence was primed from subclone CL BP8 using the oligonucleotide Dom 48, which identified the first base in the primary transcript to be the thyamine adjacent to the arrowhead. In the sense orientation this identified the transcriptional initiation site to be the adenosine of the putative CAP site at +1 (see Fig. 4).

that binds INF-1, a transcription factor known to regulate the interferon gene (Fujita *et al.*, 1987), was located at −676 bp (Fig. 4). Finally, a number of estrogen receptor half-sites (Vaccaro *et al.*, 1990; Fig. 4) were also identified at positions −735, −682, and 267 bp. Further studies are needed to ascertain if any of these putative promoter elements within the 849-bp flanking region are important in the control CRF-BP gene expression.

In summary, we have cloned the human CRF-BP gene, mapped its location to human chromosome 5, and isolated and characterized 898 bp of its extreme 5′ sequence. Primer extension analyses located the position of a potential transcriptional start site to be 32 bp from a conserved TATA box. The gene consists of 7 exons and 6 introns. The mature protein has 10 cysteines and 5 tandem disulfide bridges 4 of which are contained within exons 3, 5, 6, and 7. One bridge is shared by exons 3 and 4. Furthermore, this multiple-intronic/exonic structure of the human CRF-BP gene may provide the potential for an alternate splicing event.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grant DK 26741 and NIH DK 07044 (E.P.). This work was also supported in part by the Adler (D.B.) and Kleberg Foundations. We acknowledge the Wellcome Trust (England, P.L.) and the Reading University (England, P.L.) Endowment Fund for their support. We also thank D. J. Gilbert for excellent technical assistance in the mouse chromosomal mapping studies, which were supported, in part, by the National Cancer Institute, under Contract NO1-74101.

REFERENCES

Barclay, N. A., Johnson, P., McCaughan, G. W., and Williams, A. F. (1988). Immunoglobulin-related structures associated with vertebrate cell surfaces. *In* "The T-Cell Receptors," Chap. 3, pp. 53–85.

Behan, D. P., Linton, E. A., and Lowry, P. J. (1989). Isolation of the human plasma corticotrophin-releasing factor binding protein. *J. Clin. Endocrinol. Metab.* **70:** 1574–1580.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon–intron boundaries. *Proc. Natl. Acad. Sci. USA* **75:** 4853–4857.

Copeland, N. G., and Jenkins, N. A. (1991). Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* **7:** 113–118.

Copeland, N. G., Gilbert, D. J., Chretien, M., Seidah, N. G., and Jenkins, N. A. (1992). Regional localization of three convertases, PC1 (*Nec-1*), PC2 (*Nec-2*), and (*Fur*), on mouse chromosomes. *Genomics* **13:** 1356–1358.

Fischer, W. H., Behan, D. P., Potter, E., Park, M., Lowry, P. J., and Vale, W. (1991). "Fifth Symposium of the Protein Society," June 22–26, Baltimore, MD, Poster T70.

Fujita, T., Shibuya, H., Hotta, H., Yamanishi, K., and Taniguchi, T. (1987). Interferon-B gene regulation: Tandemly repeated sequences of a synthetic 6bp oligomer function as a virus-inducible enhancer. *Cell* **49:** 357–367.

Gilbert, W. (1978). Why genes in pieces. *Nature* **271:** 501.

Gobout, R., Ingram, S. R., and Tilghman, S. M. (1987). Fine-structure mapping of the three mouse a-fetoprotein gene enhances. *Mol. Cell. Biol.* **8:** 1169–1178.

Green, E. L. (1981). Linkage, recombination, and mapping. *In* "Genetics and Probability in Animal Breeding Experiments," pp. 77–113, Oxford Univ. Press, New York.

Hardon, E. M., Frain, M., Paonessa, G., and Corteses, R. (1988). Two distinct factors interact with the promoter regions of several liver-specific genes. *EMBO J.* **7:** 1711–1719.

Jenkins, N. A., Copeland, N. G., Taylor, B. A., and Lee, B. K. (1982). Organization, distribution, and stability of endogenous ectropic murine leukemia virus DNA sequences in chromosomes of *Mus musculus*. *J. Virol.* **43:** 26–36.

Lee, W., Mitchell, P., and Tjian, R. (1987). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49:** 741–752.

Lefevre, C., Imagawa, M., Dana, S., Grindly, J., Bodner, M., and Karin, M. (1987). Tissue-specific expression of the human growth hormone gene is conferred in part by the binding of a specific transacting factor. *EMBO J.* **6:** 971–981.

Lenardo, J. M., and Baltimore, D. (1989). NF-kB: A pleotrophic mediator of inducible and tissue-specific gene control. *Cell* **58:** 227–229.

Linton, E. A., Behan, D. P., Saphier, P. W., and Lowry, P. J. (1989). Corticotrophin-releasing hormone binding protein: Reduction in the ACTH-releasing activity of placental but not hypothalamic CRH. *J. Clin. Endocrinol. Metab.* **70:** 1574–1580.

Linton, E. A., and Lowry, P. J. (1986). A Large molecular weight carrier for CRF-41 in human plasma. *J. Endocrinol.* **111** (suppl.). [Abstract 150]

Linton, E. A., Wolfe, C. D. A., Behan, D. P., and Lowry, P. J. (1988). A specific carrier substance for human corticotrophin releasing factor in late gestational maternal plasma which could mask the ACTH-releasing activity. *Clin. Endocrinol.* **28:** 315–324.

McGillis, J. P., Park, A., Rubin-Fletter, P., Turck, C., Dallman, M. F., and Payan, D. G. (1989). Stimulation of rat B-lymphocyte proliferation by corticotrophin-releasing factor. *J. Neurosci. Res.* **23,** 346–352.

Orth, D. N., and Mount (1987). Specific high affinity binding protein for human corticotrophin-releasing hormone in normal human plasma. *Biochem. Biophys. Res. Commun.* **143:** 411–417.

Potter, E., Behan, D. P., Fischer, W. H., Linton, E. A., Lowry, P. J., and Vale, W. (1991). Cloning and characterization of cDNAs for the human and rat corticotrophin-releasing factor binding proteins. *Nature* **349:** 423–426.

Potter, E., Behan, D. P., Linton, E. A., Lowry, P. J., Sawchenko, P. E., and Vale, W. (1992). The central distribution of a CRF-binding protein predicts multiple sites and modes of interaction with CRF. *Proc. Natl. Acad. Sci.* **89:** 4192–4196.

Ron, D., Brasier, A. R., Wright, K. A., Tate, J. E., and Habener, J. F. (1989). An inducible 50-kilodalton NFkB-like protein and a constitutive protein both bind the acute phase response element of the angiotensinogen gene. *Mol. Cell. Biol.* 1023–1032.

Suda, T., Iwashita, M., Tozawa, F., Ushiyama, T., Tamori, N., Sumitomo, T., Nakagami, Y., Demura, H., and Shizuma, K. (1988). Characterization of CRH binding protein in human plasma by chemical cross-linking and its binding during pregnancy. *J. Clin. Endocrinol. Metab.* **67:** 1278–1283.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302:** 575–581.

Vaccaro, M., Pawlak, A., and Jost, JP. (1990). Positive and negative regulatory elements of chicken vitellogenin II gene characterized by in vitro transcription competition assays in a homologous system. *Proc. Natl. Acad. Sci. USA* **87:** 3047–3051.