Structure of the Gene Coding for the \( \alpha \) Polypeptide Chain of the Human Complement Component C4b-binding Protein

By Santiago Rodriguez de Cordoba, Pilar Sanchez-Corral, and Javier Rey-Campos

From the Unidad de Immunología, Centro de Investigaciones Biológicas (CSIC), 28006 Madrid, Spain

Summary

The human gene coding for the 70-kD polypeptide of the complement regulatory component C4b-binding protein (C4BPa) spans over 40 kb of DNA and is composed of twelve exons. Upon transcription in liver, or in Hep-G2 cells, this gene produces a single transcript of 2,262 nucleotides, excepting the poly A tail, that presents an unusually long 5' untranslated region (5' UTR) of 223 nucleotides. The C4BPa gene is organized as follows: the first exon codes for the first 198 nucleotides of the 5' UTR. It is separated by a large intron from the second exon including the remaining of the 5' UTR and the coding region for the signal peptide. Each of the eight 60-amino acid repeats (short consensus repeats [SCRs]) that compose the C4BPa polypeptide chain is encoded by a single exon, except for the second SCR, which is split in two exons. At the 3' end of the C4BPa gene, the twelfth exon codes for the COOH-terminal 57 amino acids of the mature protein, which have no similarities to the SCRs, and the 245 nucleotides of the 3' UTR. Examination of the nucleotide sequence of the first exon revealed an interesting characteristic, strongly suggesting that this exon may specify a functional domain of the C4BPa transcript. It includes two in-phase ATG codons, in a different frame respect to that coding the C4BPa polypeptide, followed by an in-frame termination codon, also within the first exon.

Comparison between mouse and human C4BPa transcripts indicates conservation of this structure within the 5' UTR. C4BP is expressed in the liver and is an acute phase protein. A computer search of the genomic sequences upstream the transcription start site demonstrates the presence of potential cis-acting regulatory elements similar to those found in the promoters of other liver-expressed and/or acute phase genes.

Complement is a major defense and clearance system in the bloodstream which can be activated by two different routes, the classical and the alternative pathways. Activation by each pathway leads to the formation of complex enzymes, the C3-convertases, which catalyze the crucial step of complement activation, i.e., the cleavage and activation of the complement component C3 (reviewed in reference 1).

C4b-binding protein (C4BP)\(^1\) is a regulator of complement activation. It binds to C4b, accelerates the decay of the classical pathway C3-convertase (C4b,2a), and functions as a cofactor in the factor I-mediated proteolytic inactivation of C4b (2–6). In addition, C4BP binds to the anticoagulant vitamin K-dependent protein S and renders it inactive as a cofactor to activated protein C in the inactivation of coagulation factors Va and VIIIa (7–10).

\(^1\) Abbreviations used in this paper: C4BP, C4b-binding protein; DAF, decay accelerating factor; FXIII B, coagulation factor XIIB subunit; H, factor H; MCP, membrane cofactor protein; RCA, regulator of complement activation; SCR, short consensus repeat; uORF, upstream open reading frame; UTR, untranslated region.

C4BP is an abundant protein. It is present in plasma at a concentration of 150 mg/liter (11) and, as many other complement components, is probably synthesized primarily by the liver. Recently, data from different laboratories (11, 12) indicate that C4BP is an acute phase protein.

C4BP shows a characteristic molecular heterogeneity. The major molecular form of C4BP is composed of eight chains, seven identical 70-kD polypeptides (\( \alpha \) chain) and one 45-kD polypeptide (\( \beta \) chain), that are covalently linked by their COOH-terminal regions to give the molecule a spider-like structure (2, 13–15). Both \( \alpha \) and \( \beta \) chains are evolutionarily related. They probably originated as result of a gene duplication event and then diverged to perform different functions (16). The complement regulatory functions of C4BP involve the \( \alpha \) chain, whereas the \( \beta \) chain appears to be the binding site for the anticoagulant vitamin K-dependent protein S (13, 17).

The complete amino acid sequence of the human C4BPa polypeptide chain has been deduced from nucleotide sequencing of cDNA clones (18, 19). It belongs to a family
of proteins with a characteristic structural organization based on a repetitive unit of 60 amino acids denominated as short consensus repeat (SCR). Because most of the proteins containing SCRs bind to C3b and/or C4b, this family is referred to as the superfamily of the C3b/C4b-binding proteins. The typical SCR framework of conserved residues includes four cysteines, two prolines, one tryptophan, and several other partially conserved glycine and hydrophobic residues. Starting at the NH₂-terminal end, the C4BPα polypeptide chain is composed of eight SCRs (18). The sequence ends with a COOH-terminal nonrepeat region containing two cysteines that are thought to be involved in the interchain disulfide linkage.

The human C4BPα gene is very closely linked to that coding for the C4BPβ polypeptide (16) and probably this is also the case in rodents (21). In humans, the C4BPα and C4BPβ genes are arranged in tandem with the 3' end of the C4BPβ gene located 3.5–5 kb from the 5' end of the C4BPα gene. The C4BPα and C4BPβ genes have been linked to the genes coding for the proteins membrane cofactor protein (MCP), CR1, CR2, decay accelerating factor (DAF), factor H (H), and coagulation factor XIII B subunit (FXIII B) (22–26). All these proteins share the same structural organization based in the presence of multiple SCR units. In addition, most of them are complement-regulatory components.

This group of genes, known as the regulator of complement activation (RCA) gene cluster (22), is located on the long arm of human chromosome 1 (1q32) (27, 28).

The present report describes the structural organization of the human C4BPα gene and analyzes its relationship to other genes of the RCA gene cluster. It also provides the structural basis for the subsequent analysis of the regulatory elements controlling the expression of C4BPα. In this respect, our results demonstrate a peculiar organization of the 5' untranslated region (5' UTR) of the C4BPα mRNA which may have important consequences in the control of the expression of this polypeptide.

Materials and Methods

Probes and Libraries. The C4BPα-cDNA clone used in these studies has been described previously (16). It is a 2,190-bp-long cDNA clone isolated from human cDNA library (HL 1001b) purchased from Clontech Laboratories, Inc. (Palo Alto, CA). It includes a portion of the 5' end UTR, the entire coding regions for the signal peptide and the mature secreted protein, and the 3' UTR with a 13 nucleotide poly A tail. It is shown in this report that this cDNA clone only lacks the first 86 nucleotides of the 5' end region of the C4BPα transcript. This C4BPα-cDNA clone is apparently identical to that reported recently by Matsuguchi et al. (29). This cDNA clone, or fragments of it, was used to screen two human genomic libraries constructed in the EMBL3 vector (HL 1006d and HL 1067f; Clontech Laboratories, Inc.). Additional probes were prepared from cloned restriction fragments of the human genomic DNA isolated from the EMBL3 clones.

Primer Extension Analysis. A 35mer oligonucleotide (GAAGAAAACGCTCTGATATGATGATTCTCTGTTC) corresponding to nucleotides 19–54 of the noncoding strand of the C4BPα-cDNA was end labeled with T4 polynucleotide kinase to a specific activity of ~10⁶ cpm/μg. 10 ng of labeled primer was annealed to 50 μg of human liver total RNA in 50 mM PIPES buffer, pH 6.5, containing 400 mM NaCl, 1 mM EDTA, and 50% formamide overnight at 40°C. Poly(A)⁺ RNA and annealed primer were then isolated using oligo-dT cellulose and precipitated in ethanol. Reverse transcription was carried out by adding 20 μl of 50 mM TRIS-HCl buffer, pH 8.5, containing 1 mM DTT, 10 mM MgCl₂, 4 mM KCl, 1 mM of each dNTP, 40 U of RNASin-ribonuclease inhibitor (Promega Corp., Madison, WI), and 10 U of AMV reverse transcriptase (Promega Corp.) for 1 h at 42°C. Samples were made 0.4 M NaOH and incubated overnight at room temperature to eliminate RNA. After neutralization with acetic acid, samples were precipitated in ethanol, redissolved in 6 μl of sequencing loading buffer, and analyzed in 6% acrylamide/urea sequencing gels.

SI Nuclelease Protection Assays. SI analysis of mRNA using single-stranded DNA probes were performed as described (30) with slight modifications. Briefly, single-stranded DNA probes were prepared using T7 DNA polymerase to extend 5'-end radiolabeled oligonucleotides annealed to denatured linearized-plasmid DNA templates. Labeled probes were separated from the DNA template using 6% acrylamide/urea sequencing gels, extracted from the acrylamide in 100 μl of 0.5 M Ammonium acetate, 1 mM EDTA, 1% SDS, and further purified with sephadex G-50 spin columns equilibrated with H₂O.

5 × 10⁵ cpm of labeled single-stranded probe were hybridized with 50 μg of total RNA from human liver, or with 50 μg of tRNA, in 20 μl of 50 mM PIPES buffer, pH 6.5, containing 400 mM NaCl, 1 mM EDTA, and 80% formamide overnight at 30°C. After hybridization, 300 μl of 50 mM sodium acetate buffer, pH 4.5, containing 280 mM NaCl, 4.5 mM ZnSO₄, 20 μg/ml salmon sperm DNA, and 120 U of SI nuclease (Boehringer Mannheim GmbH, Mannheim, West Germany) was added to each sample, and the tubes were incubated for 1 h at 30°C. Digestions were stopped by the addition of 80 μl of 4 M ammonium acetate/20 mM EDTA and nucleic acids precipitated in ethanol. Redissolved samples were treated overnight at room temperature with 0.4 M NaOH to remove RNA, neutralized with acetic acid, and precipitated again in ethanol. Finally, samples were redissolved in 6 μl of sequencing loading buffer and analyzed in 6% acrylamide/urea sequencing gels.

DNA Sequence Analysis. Restriction fragments of the genomic clones hybridizing with the C4BPα-cDNA probes were subcloned into plasmid vectors for DNA sequencing by the dideoxy chain termination technique with either Sequenase (United States Biochemical Corp., Cleveland, OH) or T7 DNA polymerase (Pharmacia LKB Biotechnology, Uppsala, Sweden) using 3²S-radiolabeled dATP. Primers for DNA sequencing were mostly oligonucleotides derived from the C4BPα-cDNA sequence. Programs included in the DNASTAR software package (DNASTAR, Inc., Madison, WI) were used for the sequence analysis and to compare our sequences with the National Biomedical Research Foundation (NBRF) protein sequence database and the Genebank nucleic acid database.

Results

Genomic Organization of Human C4BPα. We have used a 2,190-bp-long human C4BPα-cDNA clone (16) to screen two different EMRL3 human genomic libraries. Several overlapping clones hybridizing to this probe were identified, spanning a region of ~60 kb. Five of these clones, G562, G211, G912, G105-2, and G17, were selected and characterized by restriction endonuclease digestion and Southern blot hybridization with different C4BPα-cDNA-derived probes. The
Figure 1. Map of the human C4BPκ gene. The first line represents the complete EcoR I (R), BamH I (B), Hind III (H), and Sal I (S) restriction map of the C4BPκ gene. It resulted from the alignment of the restriction maps of five selected overlapping genomic clones, G562, G211, G912, G105-2, and G17. The position of these clones is indicated below the map. The localization of the twelve exons that compose C4BPκ is shown with vertical bars labeled from 1 to 12. The black solid circle (●) is to indicate the position of the 3' end of the C4BPκ gene.

| EXON (bp) | DOMAIN | INTRON (-kb) | 3' INTRON | EXON | 5' INTRON |
|-----------|--------|--------------|-----------|------|----------|
| I 198 | 5' UT | 1 9 | AACCGTCTCTACCAAG | gtcgggt |
| II 167 | 5' UT/L | 2 1 | cagcag AAAAACACCTCTT G | gtcaggt |
| III 186 | SCR 1 | 3 0.9 | gttcag GC ATG AATAG T | tgaagt |
| IV 100 | SCR 2A | 4 8 | tcccag AC AAA GAAG G | gtcaggt |
| V 86 | SCR 2B | 5 0.167 | attcag A TTT GAA A | tgaagt |
| VI 192 | SCR 3 | 6 2 | cctcag TT GTC GAA A | tgaagt |
| VII 183 | SCR 4 | 7 5 | cttaag AA ATG CCCC A | tgaagt |
| VIII 195 | SCR 5 | 8 4 | gatcag AT AGT GAG G | tgaagt |
| IX 189 | SCR 6 | 9 6 | tttcag CG TTA GAG A | tgaaga |
| X 171 | SCR 7 | 10 1.5 | cttaag TT TGG AAA G | tgaact |
| XI 176 | SCR 8 | 11 1 | ttcag CT CTG GAG TGG | tgaagt |
| XII 425 | CT/3'UT | 11 1 | tttcag GAG ACC TAT |

Figure 2. Exon/intron organization of the C4BPκ gene. (a) The precise size of each of the C4BPκ exons and the sequences corresponding to their 3' and 5' exon/intron boundaries are shown. Sizes of the introns, excepting intron 5 are approximate and were determined by agarose gel electrophoresis. (b) Diagramatic representation of the genomic organization of the C4BPκ transcript. Solid bars (-----) are to indicate sequences of the mRNA corresponding to the 5' and 3' untranslated regions (UT). Coding regions are boxed. The signal peptide is referred to as L and the 57 amino acid COOH-terminal region as CT.
alignment of these genomic clones is presented in Fig. 1 to show the overall organization of the human C4BPα gene. C4BPα is composed of 12 exons spanning 40 kb of DNA.

The precise size of each of these 12 exons and the results of the analysis of all exon/intron junctions are described in Fig. 2. The data were obtained by comparison of the C4BPα-cDNA sequence to those sequences obtained from selected genomic subclones. Fig. 2 also presents a description of the structural domains associated with each of the exons and estimates of the length of the different introns based on agarose gel electrophoresis analyses. These results fully confirm and extend preliminary data on the organization of the human C4BPα gene (31, 32).

The 5' UTR of the C4BPα transcript is split in two exons. The first exon includes the initial 198 nucleotides of the 5' UTR (see below) and it is separated by a ~9-kb-long intervening sequence from the second exon (167 bp), coding for the remainder of the 5' UTR and the predicted coding region for the signal peptide. The third exon codes for the first of the SCRs and starts the coding region for the mature secreted protein. Each of the eight SCRs that compose the C4BPα polypeptide chain is encoded by a single exon, except for the second SCR, which is split in two exons. Finally, the twelfth exon of the C4BPα gene codes for the COOH-terminal 57-amino acid region, which has no similarities to the SCRs, and the 245 nucleotides of the 3' UTR (Fig. 2). Characterization of the precise position to which the poly A is added could not be determined since the first six A nucleotides of the poly A tail found in the cDNA clones are also present in genomic DNA.

A partial structure for the gene coding the murine homologue to human C4BPα has been reported (33). It includes six exons that are homologous to the fifth, sixth, seventh, tenth, eleventh, and twelfth exons of human C4BPα. The human and murine C4BPα transcripts, as will be shown in the Discussion, present very similar 5' UTRs. Thus, excepting for the absence of exons 8 and 9, the organization of the murine C4BPα gene can be anticipated to be similar to that of the human C4BPα gene.

It should be mentioned that the screening of the human genomic libraries with the C4BPα-cDNA-derived probes also yield a separated set of overlapping clones with exon-like regions 80% homologous to the C4BPα exons. Pulsed field gel electrophoresis analyses of human genomic DNA, using specific probes, demonstrated that the two sets of C4BPα-specific overlapping clones map to very closely linked, but distinct, genomic regions (16). Thus, the second set of genomic clones probably identifies a "recent" duplication of the C4BPα gene, referred to as C4BPα-like gene, which will be described elsewhere (P. Sanchez-Corral, F. Pardo-Manuel, and S. Rodriguez de Cordoba, manuscript in preparation).

Characterization of the Transcriptional Unit of the Human C4BPα Gene. The 2,190-bp-long cDNA-C4BPα clone is the one among several cDNA clones that contains the longest sequence upstream the putative translational start site for the C4BPα polypeptide. Genomic sequences upstream the cDNA start site were obtained by DNA sequencing the G562RV0.3 subclone, a 352-bp-long EcoR I/EcoR V restriction fragment of the phage clone G562 including the first 98 bp of the C4BPα-cDNA, but no canonical TATA or CAAT boxes were encountered that could help to localize the cap site of the C4BPα-mRNA.

To characterize the transcription start site of the human C4BPα gene and to rule out the possibility of additional exons, we used a combination of primer extension and S1 nuclease protection assays (Fig. 3). A 32P-labeled 35mer oligonucleotide, corresponding to nucleotides 19–54 of the noncoding strand of the C4BPα-cDNA, was annealed to total RNA from human liver, or from the human hepatoma cell line Hep-G2, and extended with AMV reverse transcriptase. As shown in Fig. 3a the antisense 35mer oligonucleotide extended up to a single labeled fragment of 139 nucleotides, which suggests that the mRNA start site is located 104 nucleotides upstream the 3' end of the primer. Identical results were obtained using RNA from the human hepatoma cell line Hep-G2 (data not shown).

The same 35mer oligonucleotide was also used to generate a 333-bp-long single-stranded DNA probe using the G562RV0.3 restriction fragment as the template (Fig. 3c). In agreement with the results of the primer extension experiments, the S1 nuclease protection assays of this probe with total liver RNA showed a major "protected" band of nearly identical size to that found by primer extension analysis (Fig. 3b). This finding further supports that the C4BPα gene transcription is initiated at a single site located 86 nucleotides upstream the 5' most proximal nucleotide found in our longest C4BPα-cDNA clone. Taking into account these 86 nucleotides, the predicted length of the entire C4BPα transcript, excepting the poly A tail, is estimated in 2,252 nucleotides.

Close examination of the 5' UTR sequence (198 bp) included within the first exon revealed the presence, in a different frame with respect to that encoding the C4BPα polypeptide, of two in-phase ATG codons at positions 25 and 82, which potentially could initiate translation of a small "upstream open reading frame" (uORF). This unexpected finding together with the unusual genomic organization of the 5' UTR of the human C4BPα-cDNA prompted us to confirm
the 2,262-bp-long transcript, including both the first and second exons, identifies the majority of the C4BPα-mRNA in liver.

Total RNA from human liver was used this time to protect a 324-nucleotide-long 5'-end ³²P-labeled single-stranded DNA fragment including 270 nucleotides corresponding to the sequence from position 145 to position 415 of the non-coding strand of the C4BPα-mRNA; i.e., spanning the last 54 bp of exon I, exon II, and the first 49 bp of exon III. As shown in Figure 4, a “protected” band was detected that exactly corresponds to the length of the C4BPα-mRNA sequence included in the input probe. More important, no additional bands were observed (particularly in the vicinity of the boundaries between the third and second exons or between the second and the first exons), thus indicating that most, if not all, C4BPα transcripts in liver include the same 223-bp-long 5' UTR preceding the first ATG in-frame with the coding sequence for C4BPα.

Genomic Sequences Upstream the Transcription Initiation Site. Over 5 kb of genomic DNA flanking the 5' end of the C4BPα gene were sequenced using various subclones of the G562 λ genomic clone. As a first result, the analysis of this genomic sequence demonstrates that the 3' end of the C4BPα gene, which we have previously shown to lie adjacent to the 5' end of the C4BPα gene (16), is exactly located 4,178 bp from the transcription start site of the C4BPα gene.

C4BP is expressed in the liver and is an acute phase protein (12, 13). Analysis of the 4,178-bp nucleotide sequence separating both genes demonstrates the presence of several sequences that share identity with the consensus sequence (TT/GNNGNAA/T/G) recognized by NF-IL6, a nuclear factor implicated in the gene regulation of acute-phase genes (34). Some of these elements and their positions are shown in Fig. 5. Although no data on functional assays for these potential regulatory elements are yet available, it is suspected that some of them could be responsible for the observed enhanced secretion of C4BP by Hep-G2 cells upon stimulation by IL-6 (13). The human C4BPα gene does not present a conventional TATA promoter. However, a search of this sequence to identify potential cis-acting regulatory elements demonstrated the presence of the motif GTTAAATCATTCAT at position −50 (Fig. 5). This 13-bp sequence matches the proposed consensus sequence for the binding site of HNF1 in the promoter region of other liver-expressed genes (35).

Discussion

The C4BPα polypeptide belongs to a family of proteins characterized by a structural organization based on the presence of internal repeat units of 60 amino acids (SCR) that share a framework of highly conserved residues (20). Human C4BPα is composed of eight of these repeats followed by a 57-amino acid nonrepeat region (18).
The results of the comparison between the genomic organization of the C4BPα gene with other RCA genes are in agreement with the current understanding of the evolution of the RCA gene cluster. Genes encoded at this location most likely originated as a result of the duplication of an ancient gene organized by SCRs and then diverged to present distinct COOH-terminal coding regions as well as different 5'-noncoding regions. These genes maintained, however, the characteristic genetic organization of the SCR exons and it can be speculated that this fact facilitated the appearance of the successive generations of RCA genes. The peculiar uniformity of the splice junctions in this region, which permits insertion or deletion of SCR-coding exons without changing the correct reading frame, provides a particularly suitable structural organization for the evolution of these genes involving genetic mechanisms, like unequal recombination, which are normally deleterious for a gene without this feature. This model for the evolution of the RCA genes is supported by their physical localization within the RCA gene cluster. Thus, RCA genes presenting similar “non-SCR regions” map together within specific subregions of the RCA gene cluster.

The C4BPα gene shows a peculiar 5’ end organization. It presents an unusually long 5’ UTR that is encoded by two exons separated by a large ~9-kb intron. The first exon codes for the initial 198 nucleotides of the 5’ UTR of the C4BPα mRNA and includes two ATG codons that lie in a sequence context suitable for being the translational start site of a small mRNA and includes two ATG codons that lie in a sequence context suitable for being the translational start site of a small mRNA. Upstream from the predicted start site of the C4BPα mRNA and includes two ATG codons that lie in a sequence context suitable for being the translational start site of a small mRNA and includes two ATG codons that lie in a sequence context suitable for being the translational start site of a small mRNA.

The complete amino acid sequence of the murine homologue of human C4BPα has been deduced from nucleotide sequencing of a 1,889-bp-long cDNA clone isolated from a mouse-liver-cDNA library using human C4BPα-cDNA probes (46). The predicted mouse C4BPα polypeptide chain is 51% identical at the amino acid level to its human homologue and lacks the residues corresponding to the fifth and sixth SCRs of human C4BPα. Upstream from the predicted exon starting the coding region for the mature secreted protein, the murine cDNA clone includes 370 bp with two in-phase ATG codons, positions 203 and 332, which yield putative signal peptides of 56 and 13 residues, respectively (46). Prompted by the peculiarities found in the 5’ end region of the human C4BPα transcript, we have reexamined the 5’ end region sequence of this 1,889-bp-long murine C4BPα-cDNA clone and found that further upstream of the two ATG codons in-phase with the C4BPα polypeptide there are three additional ATG codons, two of which (positions 32 and 107) could be translation start sites for two small uORFs. These uORFs lie within a region of this clone that was earlier suggested to correspond to the 3’ end of an intron sequence (46). Alignment of the murine cDNA-C4BPα sequence and the completed human C4BPα sequence, however, shows homology between both 5’ UTRs (data not shown), strongly suggesting
that the sequence including the two uORFs belongs to the murine C4BPα mRNA and supporting the view that the 1,889-bp-long cDNA clone reported in reference 46 represents the entire sequence of the murine C4BPα transcript.

To identify unusually conserved sequences between human and mouse C4BPα in the region upstream from the predicted codon corresponding to the NH₂ terminus of the mature secreted protein, we have performed dot matrix comparisons at the nucleotide level using high stringent conditions (window = 20 n; identity = 80%). This analysis demonstrates the existence of two of such conserved regions (Fig. 6 a). One of these regions corresponds to that containing the uORFs and the other includes the coding sequences for the signal peptide.

Close examination of the nucleotide sequences surrounding the 5' most proximal ATG codons starting either the upstream or the signal peptide coding frames indicates, however, that these ATG codons have not been conserved; i.e., human and mouse C4BPα present these ATG codons at distinct sequence positions. Interestingly, when we compared the deduced human and mouse amino acid sequences encoded at these locations, we found that, in spite of using different ATG codons, the reading frames have been maintained. Alignment of the deduced sequences for the signal peptide, assuming
that the translation start site is located at the most 5'-proximal
in-phase ATG codon, is shown in Fig. 6 b. Except for nine
extra residues at the beginning of the murine sequence, both
signal peptides were 56% identical. In addition, they share
other positions with chemically similar residues. These results
strongly argue in favor of the ATG positions 203 and 224,
mouse and human, respectively, as the authentic translation
start sites of the C4BPα polypeptides.

The comparison of the deduced amino acid sequences of
the uORFs present in the 5' UTR of the human and mouse
C4BPα shows a degree of conservation between these se-
quences comparable to that found between those coding for
the signal peptides (Fig 6 b) or between those coding for the
mature secreted protein (46). The striking conservation of
these 5' UTRs sequences between human and mouse C4BPα
suggests a functional role for these uORFs. Experiments are
in progress to test this hypothesis by investigating a possible
role of these sequences in controlling the expression of the
C4BPα polypeptide.

C4BP is unique among RCA proteins in that it is orga-
nized by two functionally distinct polypeptide chains (C4BPα
and C4BPβ). They share similar COOH-terminal regions
that contain the two cysteines that establish the C4BP inter-
chain disulfide linkages (15, 18). These polypeptides are en-
coded by genes arranged in tandem that, as discussed above,
probably originated by the duplication of a common ancestor
at the C4BP-subregion of the RCA gene cluster. Because these
genes code for subunits of the same molecule, it is conceiv-
able that they also share regulatory sequences at their 5'
flanking sequences which allow their coordinated expression.

As a first approach to the characterization of these regula-
tory sequences in the human C4BPα gene, we have sequenced
>5 kb of DNA upstream the transcription start site and
sought for the presence of potential cis-acting regulatory ele-
ments. Although the human C4BPα gene does not present
a conventional TATA promoter, a number of sequences similar
to known regulatory elements were identified. These include
sequences that share identity with the consensus sequence
for the binding sites of HNF1 or NF-IL6 (Fig. 5). The pre-
ence of these sequences is particularly suggestive because C4BP
is expressed in liver and is an acute phase reactant. Whether
these structural elements are functional regulatory elements
involved in the control of C4BPα expression is currently under
investigation in our laboratory.

We thank Drs. J. Paz-Ares, M. A. PenaAlva, and A. Puyet for the stimulating discussions concerning the
5' untranslated region of the C4BPα gene.

This work was supported by the grants from the Spanish Direccicon General de Investigacion Cientifica
y Tecnica PM88-0002 and PM89-0013.

Address correspondence to Santiago Rodriguez de Cordoba, Unidad de Inmunologia, Centro de Inves-
tigaciones Biologicas (CSIC), Velazquez 144, 28006 Madrid, Spain.

Received for publication 24 December 1990 and in revised form 4 February 1991.

References

1. Law, S.K.A., and K.B.M. Reid. 1988. Complement. IRL Press,
Oxford. 1-72.

2. Schafsstein, J., A. Ferreira, I. Gigli, and V. Nussenzweig. 1978.
Human C4-binding protein I. Isolation and characterization.
J. Exp. Med. 148:207.

3. Fujita, T., I. Gigli, and V. Nussenzweig. 1978. Human C4-
binding protein. II. Role in proteolysis of C4b by C3b-
inactivator. J. Exp Med. 148:1044.

4. Fujita, T., and V. Nussenzweig. 1979. The role of C4-binding
protein and 1H in proteolysis of C4b and C3b. J. Exp. Med.
150:267.

5. Nagasawa, S., and R.M. Stroud. 1980. Purification and charac-
terization of a macromolecular weight cofactor for C3b-
inactivator, C4b/C3INA-cofactor, of the human plasma. Mol.
Immunol. 17:1365.

6. Gigli, I., T. Fujita, and V. Nussenzweig. 1979. Modulation of
the classical pathway C3 convertase by plasma proteins C4-
binding proteins and C3b inactivator. Proc. Natl. Acad. Sci.
USA. 76:6596.

7. Dahlbäck, B., and J. Stenflo. 1981. High molecular weight com-
plex in human plasma between vitamin K-dependent protein
S and complement component C4B-binding protein. Proc. Natl.
Acad. Sci. USA. 78:2512.

8. Dahlbäck, B. 1986. Inhibition of protein Ca cofactor function
of human and bovine protein S by C4b-binding protein. J. Biol.
Chem. 261:12022.

9. Comp, P.C., R.R. Nixon, M.R. Cooper, and C.T. Esmon.
1984. Familial protein S deficiency is associated with recurrent
trombosis. J. Clin. Invest. 74:2082.

10. Walker, F.J. 1981. Regulation of activated protein C by pro-
tein S. J. Biol. Chem. 256:11128.

11. Barnum, S.R., and B. Dahlbäck. 1990. C4b-binding protein, a
regulatory component of the classical pathway of comple-
ment, is an acute-phase protein and is elevated in systemic lupus
erythematosus. Complement Inflammation. 7:71.

12. Sack, T., S. Hirose, M. Nukatsuka, Y. Kusunoki, and S.
Nagasawa. 1989. Evidence that C4b-binding protein is an acute
phase protein. Biochem. Biophys. Res. Commun. 164:1446.

13. Hillarp, A., and B. Dahlbäck. 1988. Novel subunit in C4b/
binding protein required for protein S binding. J. Biol.
Chem. 263:12759.

14. Dahlbäck, B., C.A. Smith, and H.J. Muller-Eberhard. 1983.
Visualization of human C4b-binding protein and its complexes
with vitamin K-dependent protein S and complement protein
C4b. Proc. Natl. Acad. Sci. USA. 80:3461.

15. Hillarp, A., and B. Dahlbäck. 1990. Cloning of cDNA coding
for the β-chain of human C4b-binding protein, sequence ho-
mology with the α-chain. Proc. Natl. Acad. Sci. USA. 87:1183.
16. Pardo-Manuel, F., J. Rey-Campos, A. Hillaire, B. Dahlbäck, and S. Rodríguez de Córdoba. 1990. Human genes for the α and β chains of complement C4b-binding protein are closely linked in a head-to-tail arrangement. Proc. Natl. Acad. Sci. USA. 87:4529.
17. Hillarp, A., M. Hegg, and B. Dahlbäck. 1989. Protein S binding in relation to the subunit composition of human C4b-binding protein. FEBS (Fed. Eur. Biochem. Soc.) Lett. 259:53.
18. Chung, L.P., D.R. Bentley, and K.B.M. Reid. 1985. Molecular cloning and characterization of the cDNA coding for C4b-binding protein, a regulatory protein of the classical pathway of the human complement system. Biochem. J. 230:133.
19. Lintin, S.J., A. Lewin, and K.B.M. Reid. 1988. Derivation of the sequence of the signal peptide in human C4b-binding protein and interspecies cross-hybridization of the C4bp cDNA sequence. FEBS (Fed. Eur. Biochem. Soc.) Lett. 232:328.
20. Reid, K.B.M., D.R. Bentley, R.D. Campbell, L.P. Chung, R.B. Sim, T. Kristensen, and B.F. Tack. 1986. Complement system proteins which interact with C3b or C4b. A superfamily of structurally related proteins. Immunol. Today. 7:230.
21. Andersson, A., B. Dahlbäck, C. Hanson, A. Hillaire, G. Levan, J. Szpirer, and C. Szpirer. 1990. Genes for C4b-binding protein α- and β-chains (C4BPA and C4BPP) are located on chromosome 1, band q32, in humans and on chromosome 13 in rats. Somat. Cell. Mol. Genet. 16:493.
22. Rodríguez de Córdoba, S., D. Lublin, P. Rubinstein, and J.P. Atkinson. 1985. Human genes for three complement components that regulate the activation of C3 are tightly linked. J. Exp. Med. 161:1189.
23. Rey-Campos, J., P. Rubinstein, and S. Rodríguez de Córdoba. 1987. Decay accelerating factor: genetic polymorphism and linkage to the RCA (regulator of the complement activation) gene cluster in humans. J. Exp. Med. 166:246.
24. Rey-Campos, J., P. Rubinstein, and S. Rodríguez de Córdoba. 1988. A physical map of the human regulator of complement activation gene cluster linking the complement genes CR1, CR2, DAF, and C4BP. J. Exp. Med. 167:664.
25. Bora, N.S., D.L. Lublin, B.V. Kumar, R.D. Hockett, V.M. Holers, and J.P. Atkinson. 1989. Structural gene for human membrane cofactor protein (MCP) of complement maps to within 100 kb of the 3' end of the C3b/C4b receptor gene. J. Exp. Med. 169:597.
26. Rey-Campos, J., S. Baeza-Sanz, and S. Rodríguez de Córdoba. 1990. Physical linkage of the human genes coding for complement factor H and coagulation Factor XIII B subunit. Genomics. 7:644.
27. Weis, J.H., C.C. Morton, G.A. Bruns, J.J. Weis, L.B. Klickstein, W.W. Wong, and D.T. Fearon. 1987. A complement receptor locus genes encoding C3b/C4b receptor and C3d/Epstein-Barr virus receptor map to q32. J. Immunol. 138:312.
28. Lublin, D., R.S. Lemons, M.M. Le Beau, V.M. Holers, M.L. Tykocinski, M.E. Medof, and J.P. Atkinson. 1987. The gene encoding decay accelerating factor (DAF) is located in the complement regulatory locus on the long arm of chromosome 1. J. Exp. Med. 165:1731.
29. Matsuguchi, T., S. Okamura, T. Aso, T. Sata, and Y. Niho. 1989. Molecular cloning of the cDNA coding for Proline-Rich Protein (PRP): Identity of PRP as C4-binding protein. Biochem. Biophys. Res. Commun. 165:138.
30. Aurubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl. 1989. Current Protocols in Molecular Biology. John Wiley & Sons, New York. pp. 4.6.1.
31. Rey-Campos, J., P. Marshall, P. Rubinstein, and S. Rodríguez de Córdoba. 1989. Structure of the human C4BP gene. Complement Inflammation. 6:393 (Abstr.).
32. Lintin, S.J., and K.B.M. Reid. 1986. Studies on the structure of the human C4-binding-protein gene. FEBS (Fed. Eur. Biochem. Soc.) Lett. 204:77.
33. Barnum, S.R., T. Kristensen, D.D. Chaplin, M.F. Seldin, and B.F. Tack. 1989. Molecular analysis of the murine C4b-binding protein gene. Chromosome assignment and partial gene organization. Biochemistry. 28:8312.
34. Akira, S., H. Ishikii, T. Sugita, O. Tanabe, S. Kinoshita, Y. Nishio, T. Nakajima, T. Hirano, and T. Kishimoto. 1990. A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. EMBO (Eur. Mol. Biol. Organ.) J. 9:1897.
35. Courtois, G., S. Baumhuetter, and G.R. Crabtree. 1988. Purified hepatocyte nuclear factor 1 interacts with a family of hepatocyte-specific promoters. Proc. Natl. Acad. Sci. USA. 85:7937.
36. Lintin, S.J., and K.B.M. Reid. 1986. Studies on the structure of the human C4bp binding-protein gene. FEBS (Fed. Eur. Biochem. Soc.) Lett. 204:77.
37. Courtice, D., D.R. Miesner, J.P. Atkinson, and V.M. Holers. 1988. Identification of an alternative polyadenylation site in the human C3b/C4b receptor (complement receptor type 1) transcriptional unit and prediction of a secreted form of complement receptor type 1. J. Exp. Med. 168:1255.
38. Post, T.W., M.A. Arce, M.K. Lisszewski, E.S. Thompson, J.P. Atkinson, and D.M. Lublin. 1990. Structure of the gene for human complement protein decay accelerating factor. J. Immunol. 144:740.
39. Courtice, D., D.R. Miesner, J.P. Atkinson, and V.M. Holers. 1988. Identification of an alternative polyadenylation site in the human C3b/C4b receptor (complement receptor type 1) transcriptional unit and prediction of a secreted form of complement receptor type 1. J. Exp. Med. 168:1255.
40. Post, T.W., and J.P. Atkinson. 1989. The structure and organization of the MCP gene. PASEB (Fed. Am. Soc. Exp. Biol.) J. 3:368a (Abstr.).
41. Fujisaku, A., J.B. Harley, M. Barton Frank, B.A. Gruner, B. Frazier, and V.M. Holers. 1989. Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus receptor. J. Biol. Chem. 264:2118.
42. Vik, D.P., J.B. Keeney, P. Muñoz-Cánoves, D.D. Chaplin, and B.F. Tack. 1988. Structure of the murine complement factor H gene. J. Biol. Chem. 263:16720.
43. Wong, W.W., J.M. Cahill, M.D. Rosen, C. Kennedy, E.T. Bonaccio, M.J. Morris, J.G. Wilson, L.B. Klickstein, and D.T. Fearon. 1989. Structure of human CR1 gene. Molecular basis of structural and quantitative polymorphism and identification of a new CR1-like allele. J. Exp. Med. 169:847.
44. Marth, J.D., R.W. Overell, K.E. Meier, E.G. Krebs, and R.M. Perlmutter. 1988. Translational activation of the Ick proto-oncogene. Nature (Lond.). 332:171.
45. Futerer, J., K. Gordon, H. Sanfàcon, J.-M. Bonneville, and V.M. Holers. 1988. Identification of an alternative polyadenylation site in the human C3b/C4b receptor (complement receptor type 1) transcriptional unit and prediction of a secreted form of complement receptor type 1. J. Exp. Med. 168:1255.
46. Kristensen, T., R.T. Ogata, L.P. Chung, K.B.M. Reid, and B.F. Tack. 1987. cDNA cloning of decay accelerating factor indicates novel use of splicing to generate two protein forms. Nature (Lond.). 325:545.
47. Mueller, P.P., and A.G. Hinnebusch. 1986. Multiple upstream AUG codons mediate translational control of GCNA. Cell. 45:201.
48. Marth, J.D., R.W. Overell, K.E. Meier, E.G. Krebs, and R.M. Perlmutter. 1988. Translational activation of the Ick proto-oncogene. Nature (Lond.). 332:171.
49. Futere, J., K. Gordon, H. Sanfàcon, J.-M. Bonneville, and V.M. Holers. 1990. Positive and negative control of translation by the leader sequence of cauliflower mosaic virus pregenomic 35S RNA. EMBO (Eur. Mol. Biol. Organ.) J. 9:1697.
50. Kristensen, T., R.T. Ogata, L.P. Chung, K.B.M. Reid, and B.F. Tack. 1987. cDNA structure of Murine C4b-binding protein, a regulatory component of the serum complement system. Biochemistry. 26:4668.