Vol. 174, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 15, 1991 Pages 222-227

GENOMIC ORGANIZATION OF THE α CHAIN OF THE HUMAN C4b-BINDING PROTEIN GENE

Teijiro ASO[a], Seiichi OKAMURA[b,*], Tetsuya MATSUGUCHI[a], Norihiro SAKAMOTO[c], Teizo SATA[d] and Yoshiyuki NIHO[a]

[a]The First Department of Internal Medicine, [b]Cancer Center and [c]Research Laboratory for Genetic Information, Faculty of Medicine, Kyushu University, [d]Sata Clinic, Fukuoka, Japan

Received November 22, 1990

SUMMARY: C4b-binding protein (C4bp) is a serum glycoprotein that is one of the regulators of the complement activation (RCA) family. This protein is composed of structurally related 70-kDa (α chain) and 45-kDa (β chain) polypeptides. The α chain of C4bp (C4bpα) consists of eight short consensus repeats (SCR), which constitute the amino-terminal 491 residues. Human C4bp is also one of the acute-phase reactants. In order to clarify the genetic basis of the SCR and to understand the regulatory mechanisms of C4bp synthesis, we isolated 6 genomic DNA clones covering all of the human C4bpα gene. This gene consists of 12 exons and spans about 40 kb. Each of the SCRs is encoded by a single exon, except for the second SCR (SCR II), which is encoded by two separate exons, demonstrating that human C4bpα has a split SCR at the genomic level. The 5' flanking region was sequenced up to 380 bases upstream from the putative transcription initiation site. Several possible binding sites for transcription factors were identified.

Human C4b-binding protein (C4bp) is a serum glycoprotein of Mr approximately 550,000, which regulates the C3-convertase (C4bC2a, EC3.4.21.43) of the classical pathway of complement (1,2). C4bp is composed of eight chains, seven identical 70-kDa polypeptides (α chain) and one 45-kDa polypeptide (β chain), that are linked covalently by their C-terminal regions to give the molecule a spider-like structure (3,4,5). The complement regulatory functions of C4bp involve the α chain, whereas the β chain appears to be the binding site for the anticoagulant vitamin K-dependent protein S (4,6). The α chain of C4bp (C4bpα) contains eight internally homologous repeat protein domains, defined as short consensus repeats (SCRs) (7), which are approximately 60 amino acids long. These SCRs are also found in other C3b/C4b-binding proteins including C3b/C4b receptor (CR1), C3dg receptor (CR2), decay-accelerating factor (DAF), membrane cofactor protein (MCP) and factor H (8). The genes encoding these six proteins are tightly linked and form a gene cluster, defined as regulators of the complement activation (RCA) locus,

* To whom correspondence should be addressed at Cancer Center, Kyushu University Hospital, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812, Japan.

Abbreviations: C4bp, C4b-binding protein; RCA, regulator of complement activation; SCR, short consensus repeat.

0006-291X/91 $1.50

which maps to band q32 of chromosome 1 (9). All of these proteins serve to down-regulate the complement system at the level of C3 convertase. Thus, the RCA gene family consists of structurally homologous, functionally related, and genetically linked proteins.

Recently it was reported that a human hepatoma cell line, Hep G2 (10), was capable of synthesizing C4bp, and that the secretion of C4bp was enhanced by interleukin-6 and tumor necrosis factor (11), which are known to be modulators of acute-phase reactants (12,13). In addition, the plasma concentration of C4bp was found to be increased in patients with acute inflammation (11,14), indicating that C4bp is an acute-phase reactant.

A preliminary characterization of the human C4bpα gene was recently presented (15), but details are not clear. In this report, we describe that the human C4bpα gene has a split SCR at the genomic level and has several possible regulatory signals in the 5' promoter region.

MATERIALS AND METHODS

Isolation of genomic clones and DNA sequencing. Human genomic libraries constructed from peripheral blood lymphocyte DNA (16) were screened by plaque hybridization using a human C4bpα cDNA (17) as a probe. The restricted DNA fragments carrying the exons were identified by Southern blotting (18), subcloned into PUC 118 or PUC 119 and sequenced by the dideoxy chain-termination method (19). The positions of the exons were determined by comparing our sequence data with the cDNA sequence (17,20).

Primer extension analysis. Poly(A)(PA) enriched (6 μg) or total (TO) RNA (35 μg) from surgically resected human liver was incubated at 30°C for 12 h with 200 ng (20 pmol) of the ^{32}P-labeled primer oligodeoxynucleotide in a solution (20 μl) containing 40 mM Pipes (pH 6.4), 0.4 M NaCl and 1 mM EDTA. As a primer, a 30-mer, 5'-TTTTTCTATGAAGAGCCCCAGATGGAGTTT-3', was used. The 5' terminus of the primer was labeled with ^{32}P using T4 polynucleotide kinase and [γ-^{32}P]ATP. After incubation, RNA and primer were precipitated by ethanol and resuspended in a solution (30 μl) containing 50 mM Tris-HCl (pH 8.3), 50 mM KCl, 8 mM $MgCl_2$, 4 mM dithiothreitol and 30 units of AMV reverse transcriptase (Life Sciences). The reaction mixture was incubated at 42°C for 60 min, and then subjected to 8% polyacrylamide gel electrophoresis. To determine the length of the extension products, a dideoxynucleotide (A, G, C, T) chain-termination reaction (19) with a known DNA fragment and the end-labeled pBR322/HpaII fragments (MA) were coordinately run.

RESULTS AND DISCUSSION

Cloning and structural organization of the human C4bpα gene

A human genomic library was screened using the plaque hybridization technique and a total of 6 overlapping clones were obtained. These clones were characterized by restriction endonuclease mapping and Southern blotting, using ^{32}P-labeled C4bpα cDNA as a probe, and by sequencing of the fragments carrying the exons. The restriction map and the location of the exons are shown in Fig. 1. The cloned C4bpα gene consists of 12 exons and spans about 40 kb, which are consistent with earlier data (15).

Exon-intron boundaries were then determined by comparing genomic sequences with C4bpα cDNA (17,20). The sequences of exon-intron junctions and the relation between the exon and the protein domains are shown in

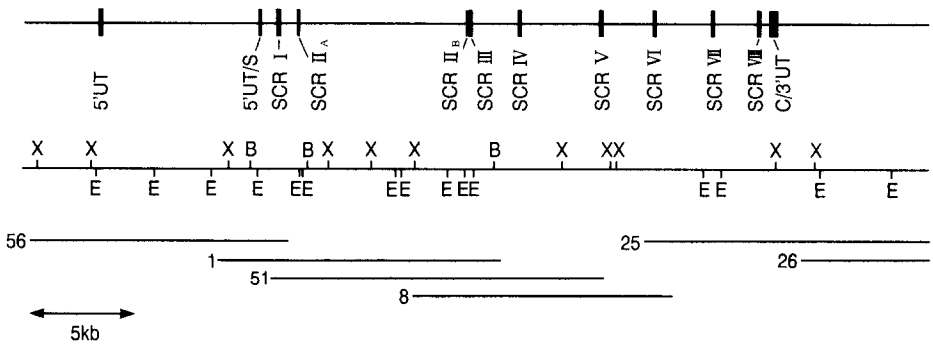

Fig.1. Restriction map and structural organization of the human C4bpα gene. The top line denotes the gene from 5' to 3' (left to right). The positions and sizes of the exons are drawn approximately to scale; the solid boxes show the positions of the exons. Abbreviations for exons are 5'UT/S: 5' untranslated region / signal peptide, SCR: short consensus repeat, C/3'UT: carboxy-terminal domain / 3' untranslated region.

Table I. All exon-intron junctions conform to the GT/AG sequence rule (21). The first two exons encode the 5' untranslated region and signal peptide. The C-terminal 57 amino acids of the protein, which have no similarities to the SCR (20), and the 3' untranslated region are encoded by a separate exon.

The SCRs of C4bpα are each encoded by single exons of approximately equal size, except for the second SCR (SCR II), which is encoded by two separate exons 8 kb apart. The split within SCR II occurs at the codon position number 34 glycine of the SCR consensus structure after the first two guanines of the boundary codon, whereas the other SCR splice junctions occur after the first nucleotide of a codon (Table I).

Split SCRs have also been found in other members of the RCA gene family, CR1 (22), CR2 (23), DAF (24), MCP (25) and murine factor H (26). The splits in these SCRs invariably occurs after the second nucleotide of the consensus number 34 glycine.

These findings suggest that this split SCR occurred very early in the evolutionary history of these proteins and provide strong evidence that the

Table I. Exon-intron boundaries of the human C4bpα gene. Each exon is shown with the encoded protein domain. Exon sequences are in capital letters; introns are in lower case letters. Abbreviations are as in Fig. 1.

Exon number	Protein domain	5' Intron	Exon	3' Intron	Exon size(bp)
1	5'UT		...AAA G	gtcggtgg	265
2	5'UT/S	ttcagcag	AA AAA...CTT G	gtgagtag	167
3	SCR I	gtgtccag	GC AAT...ATC T	gtaagtat	186
4	SCR IIA	tctcccag	AC AAA...GAA GG	gtgagtgt	100
5	SCR IIB	ttattcag	A TTT...GAA A	gtaagtaa	86
6	SCR III	tacctcag	TT GTC...GAA A	gtaagtca	192
7	SCR IV	atctttag	AA ATC...CCC A	gtaagtat	183
8	SCR V	atgagtag	AT AGT...GAG G	gtgagttt	195
9	SCR VI	tttttcag	CG TTA...GAC A	gtaagagt	189
10	SCR VII	ccctcaag	TT TGC...AAA G	gtaactta	171
11	SCR VIII	cattacag	CT CTG...TGG	gtaagtgg	176
12	C/3'UT	tcttttag	GAG...TAT end		420

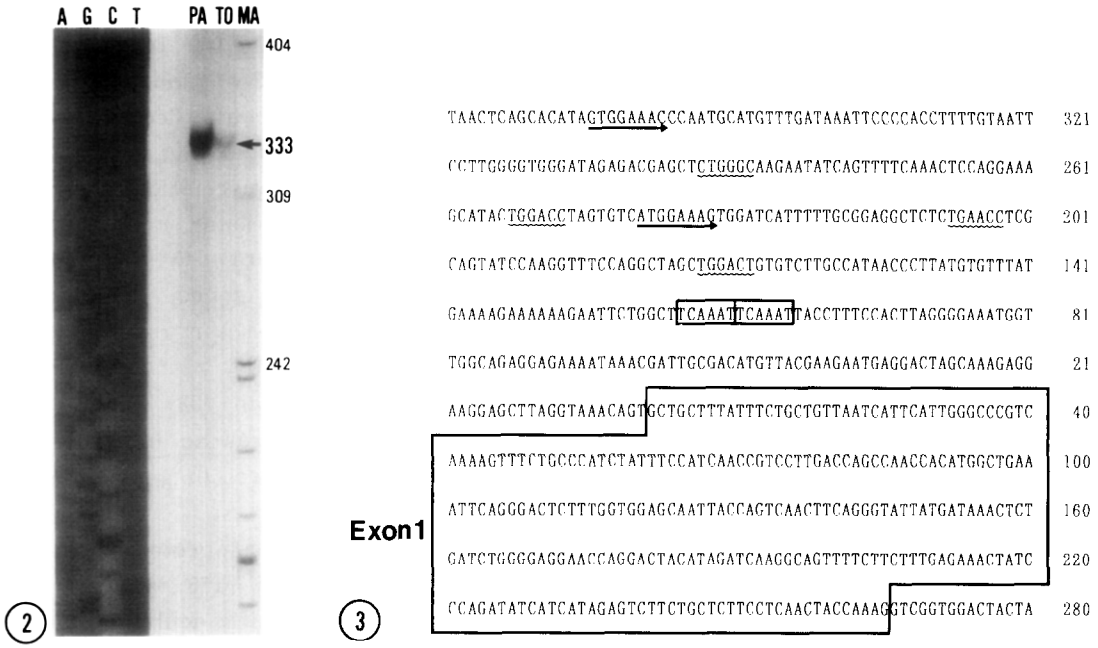

Fig.2. Primer extension analysis of the human C4bpα gene. The ^{32}P-labeled primer oligodeoxynucleotide (30-mer) was incubated with Poly(A)(PA) RNA or total (TO) RNA from surgically resected human liver. After primer extension by reverse transcriptase, products were separated on an 8% polyacrylamide gel. As size markers, a dideoxynucleotide (A,G,C,T) chain-termination reaction (19) with a known DNA fragment and the end-labeled pBR322/HpaII fragments (MA) were coordinately run.

Fig.3. Nucleotide sequence of the 5' flanking region of the human C4bpα gene. Nucleotides are numbered from the G of the tentative transcription start point. The exons and the putative CAAT-box are boxed. The consensus sequence found in various acute-phase reactant genes and those homologous to the enhancer core sequence are indicated by a wavy line and horizontal arrows, respectively.

RCA family arose by multiple duplication of an ancestral SCR unit.

Characterization of the 5' end of the C4bpα gene

To determine the 5' end of C4bpα mRNA, a primer extension experiment was carried out. Poly(A) RNA or total RNA from human liver was hybridized with the oligodeoxynucleotide primer, the 5' end of which was labeled with ^{32}P, and subjected to the reverse transcriptase reaction. The products were then analyzed by polyacrylamide gel electrophoresis. For reference, sequence ladders of a known DNA fragment and the end-labeled pBR322/HpaII fragments were coordinately run. The extended DNA products were 333 nucleotides in length (Fig. 2). This indicates that the transcription start site of the C4bpα gene is located at guanine +1 (Fig. 3).

Usually two consensus sequences are present in the proximal 5' flanking region of eukaryotic genes, a TATA-box and a CAAT-box about 20-30 and about 40-110 bp upstream, respectively, from the cap site of the gene (21). In the case of the human C4bpα gene we did not find these canonical sequences, but a CAAT-box equivalent sequence: TCAAAT. It should be noted that Kunz et al. (27) found a similar sequence (TCAAAA) in the rat α_2-macroglobulin gene.

As the 5' flanking region of genes contains regulatory signals, we compared the sequences with known motifs that are binding sites for DNA-binding proteins, especially hepatic binding factors. Several sequence similarities between the 5' flanking region and these motifs were identified (Table I). Of particular interest was the presence of the hexanucleotide CTGGGA homology sequence, CTGGGC, 295 bp upstream from the transcription start point. The sequence CTGGGA has been found in the 5' promoter region of several acute-phase protein genes. Fowlkes et al. (28) were the first to describe this sequence motif in the rat α, $\beta$and γfibrinogen genes. Subsequently this consensus sequence was discovered in the promoter region of human C-reactive protein (29,30), murine serum amyloid A (31), rat α_1-acid glycoprotein (32), human transferrin (33), rat α_2-macroglobulin (34) and rat prealbumin (35). Recently, this hexanucleotide was identified as an interleukin-6-responsive cis-acting element in the rat α_1-acid glycoprotein gene (36) and rat α_2-macroglobulin gene (37).

The sequence TG(G/A)AC(C/T), corresponding to the nuclear protein LF-A1 (38) binding site, at nucleotide positions -254, -209 and -175 was also frequently found over the 5' flanking region of the liver specific acute-phase reactant gene. Some enhancer-like sequences were found in the sequenced region. The sequences at -366 to -359 and -241 to -234 resembled the enhancer core sequence of SV 40 (39), GTGGAAAG.

Functional studies will be required to define the precise role of these putative binding sites in the regulation of transcription. Comparison of the 5' flanking region with other acute-phase reactant genes did not show any significant homology.

ACKNOWLEDGMENTS

We wish to thank Drs. Y. Sakaki and M. Hattori (Research Laboratory for Genetic Information, Kyushu University) for generous provision of a human genomic library and helpful discussion. We also thank Drs. Y. Takihara and Y. Kobayashi for technical advice. This work was supported in part by Grants-in-Aid for Cancer Research and Scientific Research (Nos. 60770949, 61015074, 63015063 and 02256102) from the Ministry of Education, Science and Culture of Japan, and from the Fukuoka Anti-Cancer Society.

REFERENCES

1. Gigli, I., Fujita, T. and Nussenzweig, V. (1979) Proc. Natl. Acad. Sci. USA 76, 6596-6600.
2. Nagasawa, S., Ichihara, C. and Stroud, R.M. (1980) J. Immunol. 125, 578-582.
3. Dahlback, B., Smith, C.A. and Muller-Eberhard, H.J. (1983) Proc. Natl. Acad. Sci. USA 80, 3461-3465.
4. Hillarp, A. and Dahlback, B. (1988) J. Biol. Chem. 263, 12759-12764.
5. Scharfstein, J., Ferreira, A., Gigli, I. and Nussenzweig, V. (1978) J. Exp. Med. 148, 207-222.
6. Hillarp, A., Hessing, M. and Dahlback, B. (1989) FEBS Lett. 259, 53-56.
7. Klickstein, L.B., Wong, W.W., Smith, J.A., Weis, J.H., Wilson, J.G. and Fearon, D.T. (1987) J. Exp. Med. 165, 1095-1112.
8. Kristensen, T., D'Eustachio, P., Ogata, R.T., Chung, L.P., Reid, K.B.M. and Tack, B.F. (1987) Federation Proc. 46, 2463-2469.
9. Lublin, D.M., Liszewski, M.K., Post, T.W., Arce, M.A., Le Beau, M.M., Rebentisch, M.B., Lemons, R.S., Seya, T. and Atkinson, J.P. (1988) J. Exp. Med. 168, 181-194.

10. Morris, K.M., Aden, D.P., Knowles, B.B. and Colten, H.R. (1982) J. Clin. Invest. 70, 906-913.
11. Saeki, T., Hirose, S., Nukatsuka, M., Kusunoki, Y. and Nagasawa, S. (1989) Biochem. Biophys. Res. Commun. 164, 1446-1451.
12. Perlmutter, D.H., Dinarello, C.A., Punsal, P.I. and Colten, H.R. (1986) J. Clin. Invest. 78, 1349-1354.
13. Castell, J.V., Gomez-Lechon, M.J., David, M., Hirano, T., Kishimoto, T. and Heinrich, P.C. (1988) FEBS Lett. 232, 347-350.
14. Funakoshi, M., Sasaki, J. and Arakawa, K. (1988) Biochim. Biophys. Acta 963, 98-108.
15. Rey-Campos, J., Marshall, P., Rubinstein, P. and Rodriguez de Cordoba, S. (1989) Complement Inflammation 6, 393-394 (Abstr.).
16. Tomatsu, S., Kobayashi, Y., Fukumaki, Y., Yubisui, T., Orii, T. and Sakaki, Y. (1989) Gene 80, 353-361.
17. Matsuguchi, T., Okamura, S., Aso, T., Sata, T. and Niho, Y. (1989) Biochem. Biophys. Res. Commun. 165, 138-144.
18. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74,5463-5467.
20. Chung, L.P., Bentley, D.R. and Reid, K.B.M. (1985) Biochem. J. 230, 133-141.
21. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
22. Hourcade, D., Miesner, D.R., Atkinson, J.P. and Holers, V.M. (1988) J. Exp. Med. 168, 1255-1270.
23. Fujisaku, A., Harley, J.B., Frank, M.B., Gruner, B.A., Frazier, B. and Holers, V.M. (1989) J. Biol. Chem. 264, 2118-2125.
24. Post, T.W., Arce, M.A., Liszewski, M.K., Thompson, E.S., Atkinson, J.P. and Lublin, D.M. (1990) J. Immunol. 144, 740-744.
25. Post, T.W. and Atkinson, J.P. (1989) FASEB J. 3, A368 (Abstr.)
26. Vik, D.P., Keeney, J.B., Munoz-Canoves, P., Chaplin, D.D. and Tack, B.F. (1988) J. Biol. Chem. 263, 16720-16724.
27. Kunz, D., Zimmermann, R., Heisig, M. and Heinrich, P.C. (1989) Nucleic Acids Res. 17, 1121-1138.
28. Fowlkes, D.M., Mullis, N.T., Comeau, C.M. and Crabtree, G.R. (1984) Proc. Natl. Acad. Sci.USA 81, 2313-2316.
29. Woo, P., Korenberg, J.R. and Whitehead, A.S. (1985) J. Biol. Chem. 260, 13384-13388.
30. Lei, K-J., Liu, T., Zon, G., Soravia, E., Liu, T-Y. and Goldman, N.D. (1985) J. Biol. Chem. 260, 13377-13383.
31. Lowell, C.A., Potter, D.A., Stearman, R.S. and Morrow, J.F. (1986) J. Biol. Chem. 261, 8442-8452.
32. Reinke, R. and Feigelson, P. (1985) J. Biol. Chem. 260, 4397-4403.
33. Adrian, G.S., Korinek, B.W., Bowman, B.H. and Yang, F. (1986) Gene 49, 167-175.
34. Tsuchiya, Y., Hattori, M., Hayashida, K., Ishibashi, H., Okubo, H. and Sakaki, Y. (1987) Gene 57, 73-80.
35. Fung, W-P., Thomas, T., Dickson, P.W., Aldred, A.R., Milland, J., Dziadek, M., Power, B., Hudson, P. and Schreiber, G. (1988) J. Biol. Chem. 263, 480- 488.
36. Prowse, K.R. and Baumann, H. (1988) Mol. Cell. Biol. 8, 42-51.
37. Hattori, M., Abraham, L.J., Northemann, W. and Fey, G.H. (1990) Proc. Natl.Acad. Sci. USA 87, 2364-2368.
38. Hardon, E.M., Frain, M., Paonessa, G. and Cortese, R. (1988) EMBO J. 7, 1711-1719.
39. Weiher, H., Konig, M. and Gruss, P. (1983) Science 219, 626-631.