Volume 232, number 2, 328–332 FEB 05819 May 1988

Derivation of the sequence of the signal peptide in human C4b-binding protein and interspecies cross-hybridisation of the C4bp cDNA sequence

S.J. Lintin*, A.R. Lewin+ and K.B.M. Reid

MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU and +*Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, England*

Received 19 February 1988

A 5′ cDNA clone coding for human C4b-binding protein (C4bp) was isolated, characterised and sequenced to complete the cDNA sequence coding for residues 1–32 thus confirming the protein sequence data of Chung et al. [(1985) Biochem. J. 230, 133–141]. The sequence extended to allow derivation of the putative leader peptide sequence which was 32 residues in length and showed a high of hydrophobicity typical of other documented leader sequences. Cross hybridisation was detected between the human C4bp cDNA probes and genomic DNA isolated from various species on Southern blots suggesting that genomic sequence homologous to that coding for C4bp has been conserved during evolution.

Signal peptide; C4b-binding protein; Cross-hybridization; Complement protein; Evolution

1. INTRODUCTION

C4b-binding protein (C4bp) is one of a group of proteins which collectively regulate the two pathways of complement activation and are now known to be genetically linked (reviews [1,2]). It is a serum glycoprotein of 550 kDa and is thought to be composed of 7 identical disulphide-bond linked subunits [3]. Each subunit is composed of eight internally homologous repeat units, defined as short consensus repeats (SCR) [4], which are approximately 60 amino acids long, and constitute the amino-terminal 491 residues [5]. The carboxy-terminal 58 residues show no similarity in sequence to the remainder of the chain and are thought to be involved in the disulphide-bonded 'core' of the molecule [5]. Other complement and non-complement proteins have been found to contain varying numbers of SCRs suggesting that C4bp is a member of the recently defined SCR superfamily [6,7].

The complete amino acid sequence of C4bp was elucidated by a combination of cDNA and peptide sequencing [5]. Overlapping cDNA clones, isolated by Chung et al. [5] included the cDNA coding for amino acid residues 32–549 and N-terminal protein sequence data included amino acid residues 1–40. This paper reports the subsequent isolation of a cDNA clone which confirmed and extended the N-terminal protein data allowing derivation of the signal peptide sequence.

C4bp has also been isolated from mouse [8,9] and guinea pig [10] and in both cases is thought to be composed of multiple subunits, similar in size to those found in human C4bp. The complete amino acid sequence of mouse C4bp was derived from cDNA data and shown to have an overall identity of 51%, at the amino acid level, with the human sequence [11]. As in the human sequence the amino terminus of the mouse sequence contains contiguous SCRs. There are, however, only 6 SCRs within each mouse C4bp monomer com-

Correspondence address: K.B.M. Reid, MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England

**Present address:* Department of Pharmacology, University of Oxford, South Parks Road, Oxford OX1 3QS, England

Published by Elsevier Science Publishers B.V. (Biomedical Division)

pared with 8 in the human sequence. The cDNA sequence coding for human C4bp has been used as a probe to demonstrate the presence of a region of conserved genomic sequence in a variety of different species.

2. MATERIALS AND METHODS

2.1. *Enzymes*

Restriction endonucleases were purchased from New England Biolabs, Boehringer Mannheim and Amersham International. The nick translation 'kit' and all radiolabelled nucleotides were also from Amersham International.

2.2. *Isolation and characterisation of 5′ cDNA clones*

Approximately 30000 colonies of a cDNA library constructed from human liver poly(A^+) RNA [12] were screened with the C4bp cDNA probe, A400 [5] using the nitrocellulose replica filter preparation techniques [12]. Filters were hybridised for 16 h at 42°C in a buffer containing 50% formamide [14]. The probe was radiolabelled using nick translation [15] to a specific activity of $\sim 10^8$ cpm/μg and used at a final concentration of 6×10^5 cpm per ml hybridisation buffer.

Plasmid DNA was prepared by the alkaline-SDS method [16]. Initial characterisation involved restriction site mapping and sequencing using the Maxam and Gilbert technique [17]. More extensive sequencing of the cDNA to derive the signal peptide and N-terminal amino acid sequence was performed using the dideoxy chain termination method [18].

2.3. *Southern blotting*

Genomic DNA was prepared from whole blood [19], all samples were supplied through the Department of Biochemistry (Oxford) with the exception of shark blood which was a gift from Dr Sylvia Smith (Florida International University); for analysis, 8–10 μg was digested with 20 units of enzyme for 20 h, electrophoresed in 0.7% (w/v) agarose and transferred to nitrocellulose (Amersham) [20]. The filters were then hybridised using a nick translated cDNA probe as described above, the probe used was C4bp S1 [21]. The filters were then washed three times in 2 × SSC + 0.1% SDS at room temperature for 20 min; twice in 5 × SSC at 65°C for 15 min and twice in 3 × SSC at 65°C for 15 min.

3. RESULTS AND DISCUSSION

3.1. *Isolation and characterisation of 5′ cDNA clones*

Eight positive clones were identified and the sizes of cloned inserts assessed by agarose gel electrophoresis of *Bam*HI/*Hin*dIII restriction enzyme digests (the library was constructed using the pAT/*Pvu*II/8 plasmid [22] which has unique *Bam*HI and *Hin*dIII sites flanking the cloning site). The two clones, designated pBP51 and pBP62, containing the largest inserts of approx. 1.7 kb were selected for further analysis. Comprehensive restriction maps together with limited nucleotide sequencing demonstrated that both clones overlapped extensively with known cDNA data and that pBP51 extended into the 5′ untranslated region (fig.1).

Thus the *Bgl*I/*Hin*fI fragment shown in fig.1 was sequenced, the results of which are shown in fig.2. The use of the self-priming mechanism in the synthesis of the second cDNA strand, during library construction [12], introduced 'loop backs' into the double-stranded DNA and as a result the coding sequence in the 5′ region of pBP51 was inverted and on the opposite strand (fig.1). This did not cause any ambiguity in this study because of the availability of a detailed restriction map for the known cDNA and peptide data for the N-terminus region of the protein. The extent of the 5′ untranslated region can only be derived by the use of S1 nuclease mapping and cannot be derived accurately from cDNA clones.

The sequence derived for residue 1 (asparagine) through to residue 33 (cysteine) confirmed the N-terminal sequence determined by conventional protein sequencing [23] and included two cysteines and a proline residue involved in the consensus sequence defined for the SCR [6]. When translated the sequence immediately 5′ to the first codon in the coding region allows derivation of the putative leader peptide sequence. This region of the amino acid sequence shows a high proportion of hydrophobic residues with two universal translation initiation codons (ATG), one at position −13 and the other at −32, the latter being preceded by four polar residues. The ATG at position −32 lies closest to the CAP site and is considered to be the point of initiation of translation. Thus the putative leader peptide is 32 residues in length for human C4bp. The mouse C4bp leader peptide was also shown to include two universal translation initiation codons at position −13 and −56 and is hence considered to be 56 residues in length [11].

Leader peptides show no well-defined consensus sequence of amino acids but the presence of polar residues followed by a region of relatively high hydrophobicity are well-documented features of other leader peptides [24]. The hydrophobic nature of the putative leader peptide reflects its role in anchoring the lengthening protein in the membrane during translation, thus mediating secretion of the

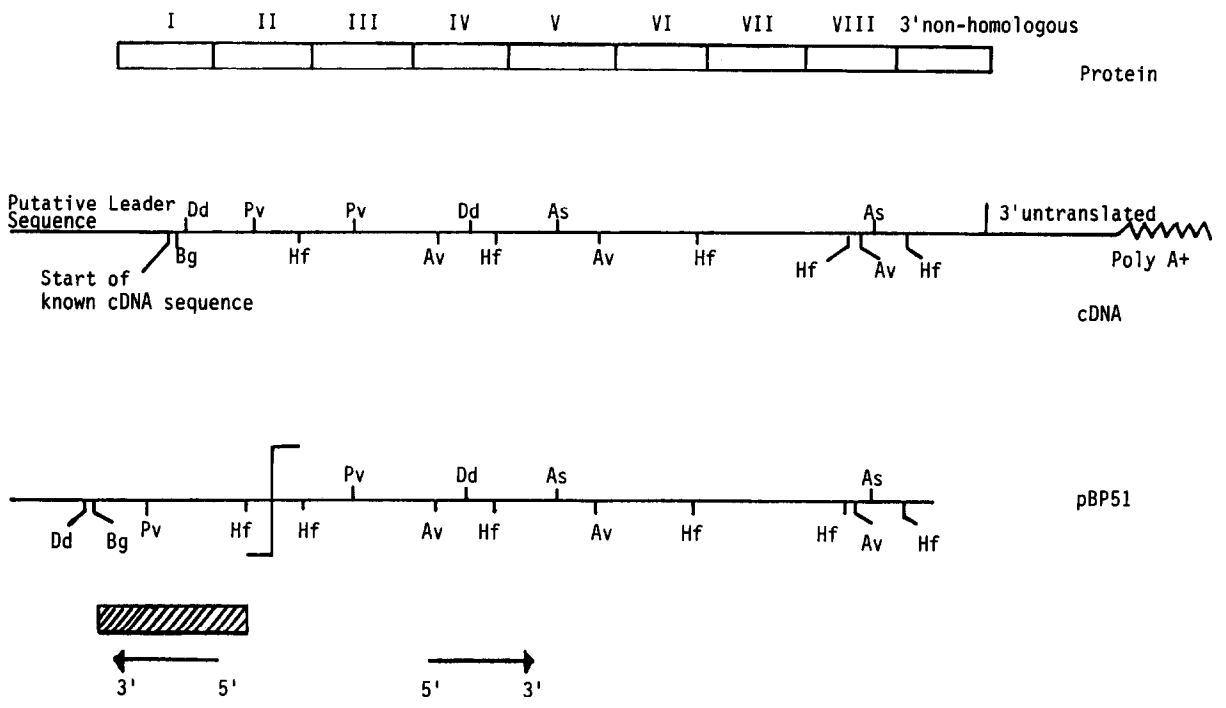

Fig.1. Restriction map constructed for pBP51 compared with the known cDNA restriction map and primary protein structure showing diagrammatically the short consensus repeats. The 'loop back' in the new clone is marked with ⌠ and the resulting inversion of the coding sequence is shown with appropriate arrows. The hatched bar indicates the *Hin*fI/*Bgl*I fragment purified for sequencing. The restriction sites in the map are marked As, *Asp*718; Av, *Ava*II; Bg, *Bgl*I; Dd, *Dde*I; Hf, *Hin*fI; Pv, *Pvu*II.

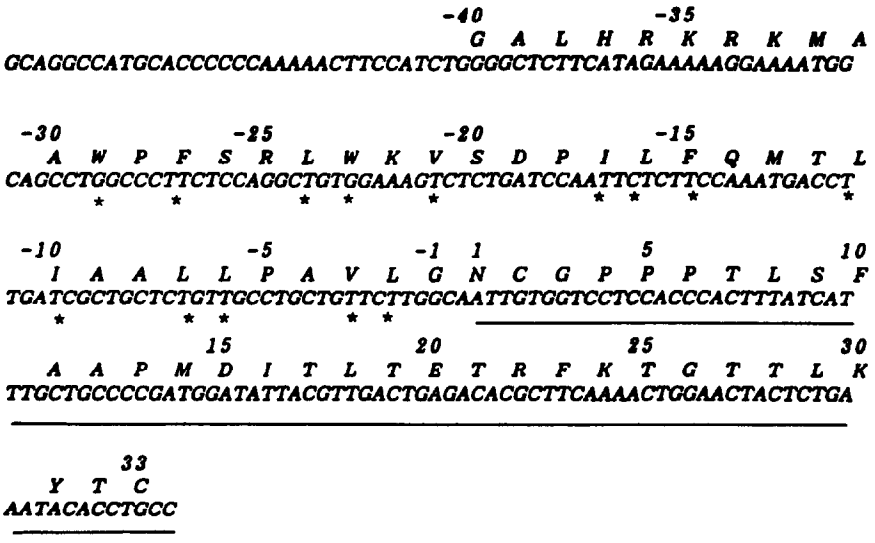

Fig.2. cDNA sequence and derived amino acid sequence for the *Hin*fI/*Bgl*I fragment from clone pBP51. The underlined sequence corresponds to the N-terminus peptide sequence [23]. The two universal translation initiation codons (ATG) are boxed and the hydrophobic residues in the leader peptide are marked *. Sequence 5′ to the codon for residue −32 (ATG) corresponds to the 5′ untranslated region, the complete sequence obtained in this region is not shown as it was obtained on one strand only and appeared to contain cloning artefacts.

mature protein from the cell and into the fluid phase [25]. After completion of translation the leader peptide is cleaved from the mature protein. As a result of this cleavage the leader peptide sequence is available only by derivation from the cDNA sequence.

3.2. *Interspecies cross hybridisation of the C4bp cDNA sequence*

Genomic DNA was prepared from the white blood cells of various mammalian and non-mammalian species as described. The restriction endonuclease *Bgl*II was then used to digest the various DNA samples and the digests analysed using the technique of Southern blotting as described in section 2.3. Similar amounts of genomic DNA were digested for each track and all the digests appeared complete on ethidium bromide stained gels. The hybridisation patterns seen are demonstrated in fig.3 and suggest the presence of sequences related to human C4bp cDNA in most species with the exception of shark. The hybridisation pattern in the track containing human genomic DNA was considerably more complex than is detected under the normally more stringent washing conditions [21], suggesting a possible cross-hybridisation with other related sequences.

The results show that genomic sequence homologous to human C4bp has been conserved during evolution; it does not demonstrate conclusively the presence of an active C4bp gene in these species. Many of the hybridisation patterns on the blots are complex, reflecting either the presence of more than one related sequence or the splitting of homologous regions of sequence by long regions of non-homologous intervening sequence. Results from analysis of cloned genomic DNA in human [21] and mouse [26] suggest that in both cases the coding sequence is split into a number of relatively short exons following closely the SCR organisation of the protein sequences, separated by introns of up to 6 kb which is reflected in complex hybridisation patterns on Southern blots.

The absence of cross-hybridisation between the human cDNA probe and genomic DNA isolated from shark blood suggest that the sequence has diverged considerably during evolution from the more primitive species. This correlates with a functional study [27] in which human C4b and C3b

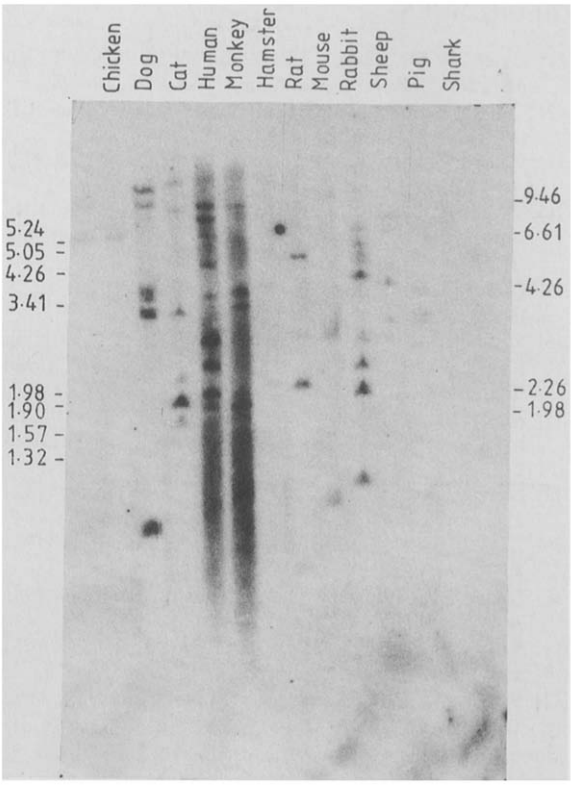

Fig.3. Results of the interspecies cross-hybridisation using the human C4bp cDNA probe, SI(21) to hybridise genomic DNA isolated from various mammalian and non-mammalian species.

were used as substrates to evaluate the presence of molecules serving the function of complement regulatory proteins in the sera of a number of different species. Serum samples from the more primitive species including shark failed to participate in the regulatory functions normally attributed to C4bp [27].

Thus a cDNA clone was isolated which allowed derivation of the putative leader peptide sequence of C4bp and confirmation, at the cDNA level, of the N-terminal peptide sequence [23]. The newly isolated clone has subsequently been used to prepare a cDNA probe and the sequence has been used to synthesize an oligonucleotide probe, both of which are being used to characterise genomic clones covering the 5′ region of the human C4bp gene. Use of cDNA probes in Southern blot analysis has demonstrated the presence of sequence homologous to human C4bp in many mammalian and non-mammalian species.

REFERENCES

[1] Holers, V.M., Cole, J.L., Lublin, D.M., Seya, T. and Atkinson, J.P. (1985) Immunol. Today 6, 188–192.

[2] Campbell, R.D., Law, S.K., Reid, K.B.M. and Sim, R.B. (1987) Annu. Rev. Immunol. 6, 161–195.

[3] Dahlbäck, B., Smith, C.A. and Müller-Eberhard, H.J. (1983) Proc. Natl. Acad. Sci. USA 80, 3461–3465.

[4] 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.

[5] Chung, L.P., Bentley, D.R. and Reid, K.B.M. (1985) Biochem. J. 230, 133–141.

[6] Reid, K.B.M., Bentley, D.R., Campbell, R.D., Chung, L.P., Sim, R.B., Kristensen, T. and Tack, B.F. (1986) Immunol. Today 7, 230–233.

[7] Kristensen, T., D'Eustachio, P., Ogata, R.T., Chung, L.P., Reid, K.B.M. and Tack, B.F. (1987) Fed. Proc. 46, 2463–2469.

[8] Ferreira, A., Takahashi, M. and Nussenzweig, V. (1977) J. Exp. Med. 146, 1001–1018.

[9] Kaidoh, T., Natsume-Sakai, S. and Takahashi, M. (1981) J. Immunol. 126, 463–467.

[10] Burge, J., Nicholson-Weller, A. and Austen, K.F. (1981) J. Immunol. 126, 232–235.

[11] Kristensen, T., Ogata, R.T., Chung, L.P., Reid, K.B.M. and Tack, B.F. (1987) Biochemistry 26, 4668–4674.

[12] Bentley, D.R. and Porter, R.R. (1984) Proc. Natl. Acad. Sci. USA 81, 1212–1215.

[13] Grosveld, F.G., Lund, T., Murray, E.J., Mellor, P.L., Dahl, H. and Flavell, R.A. (1982) Nucleic Acids Res. 10, 6715–6731.

[14] Bernards, R. and Flavell, R.A. (1980) Nucleic Acids Res. 8, 1521–1534.

[15] Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.

[16] Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513–1523.

[17] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.

[18] Messing, J. (1983) Methods Enzymol. 101, 20–78.

[19] Bell, G.I., Karam, J.H. and Rutter, W.J. (1981) Proc. Natl. Acad. Sci. USA 78, 5759–5763.

[20] Southern, E. (1979) Methods Enzymol. 68, 152–176.

[21] Lintin, S.J. and Reid, K.B.M. (1986) FEBS Lett. 204, 77–81.

[22] Anson, D.S., Choo, K.H., Rees, D.J.G., Gianelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1984) EMBO J. 3, 1053–1060.

[23] Chung, L.P., Gagnon, J. and Reid, K.B.M. (1985) Mol. Immunol. 22, 427–435.

[24] Watson, M.E.E. (1984) Nucleic Acids Res. 12, 5145–5164.

[25] Blobel, G. (1980) Proc. Natl. Acad. Sci. USA 77, 1496–1500.

[26] Barnum, S., Kenney, J., Kristensen, T., Noack, D., Seldon, M., D'Eustachio, P., Chaplin, D. and Tack, B. (1987) Complement 4, 131.

[27] Kaidoh, T. and Gigli, I. (1987) J. Immunol. 139, 194–201.