Studies on the structure of the human C4b-binding protein gene

Susan J. Lintin and K.B.M. Reid+

MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England

Received 30 May 1986

Protein and cDNA sequence data have shown human C4b-binding protein to contain eight internally homologous repeat units, each approx. 61 amino acids in length. Repeat units conforming to the same consensus sequence have been found in other complement and non-complement proteins. Southern blot analysis together with isolation, characterisation and sequencing of genomic clones has allowed the study of intron/exon organisation in the human C4b-binding protein gene and the identification of a *Bgl*II restriction fragment length polymorphism.

Complement *C4b-binding protein* *Gene structure* *Polymorphism*

1. INTRODUCTION

C4b-binding protein (C4bp) is one of a group of regulatory proteins, which includes complement receptor type 1 (CR1), factor H and gp45–70 [1], within the complement cascade which maintain homeostasis of C3 levels. C4bp performs this regulation in the classical pathway of complement activation [2,3] by accelerating the decay of C3 convertase (C4bC2a), through binding to C4b thus displacing C2a and acting as a cofactor for factor I which proteolytically splits C4b [4,5]. These functions rely upon the ability of C4bp to bind C4b; similar dependence of C4b/C3b binding is noted in the regulatory function of the plasma protein factor H of the alternative pathway and the cell surface molecules CR1, DAF and gp45–70 [1].

C4bp is a serum glycoprotein of 550 kDa and is thought to be composed of 7 identical disulphide-bond-linked chains [6]. Each chain is composed of 8 contiguous repeating homology units, each approx. 61 amino acids long, making up the N-terminal 491 amino acids while the C-terminal 58 amino acids show no homology to the rest of the chain [7]. The repeating units conform to a consensus sequence having a framework of highly conserved residues consisting of one tryptophan, two proline and four half-cystine residues and a number of other partially conserved positions involving glycine or hydrophobic residues. These repeating units are widespread being found in a number of complement proteins (C2, factor B, C4bp, factor H and CR1) and non-complement proteins (β_2-glycoprotein, IL-2 receptor; blood clotting factor XIII, haptoglobin-2 precursor) (review [8]).

Two phenotypes of human C4bp have been identified [9] and have been closely linked to the loci for CR1 and factor H and thus form a recently described gene cluster of complement regulatory proteins [10] which has been mapped to chromosome 1 ([11] and Solomon, E., unpublished). Use of the cDNA probes and Southern blot analysis has allowed the mapping of the C4bp gene and identification of a *Bgl*II restriction fragment length polymorphism. The isolation of genomic clones has allowed the study of the intron/exon

+ To whom correspondence and reprint requests should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

organisation in the gene which will be informative with respect to understanding the evolution of this group of regulatory proteins.

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. *Southern blot analysis*

Genomic DNA was prepared from whole blood [12]; for analysis 8–10 μg was digested with 20 units of enzyme for 20 h, electrophoresed in 0.7% agarose and transferred to nitrocellulose (Amersham) [13]. These were then hybridised with cDNA probes (fig.1a), nick-translated to a specific activity of $\simeq 10^8$ cpm/μg of DNA [14] in buffer containing 50% formamide and 6×10^6 cpm/ml for 30 h. The blots were then washed three times at room temperature in 2 × SSC + 0.1% SDS and three times at 68°C in 0.2 × SSC.

2.3. *Genomic clone isolation and characterisation*

Approx. 10^6 plaques from a λ phage library (constructed by A. Bentley) were screened by standard techniques [15] using the nick-translated cDNA probes. λDNA was prepared [15] and partially characterised by Southern blotting. Restriction fragments from a *Pvu*II digest of λDNA was subcloned into pAT153/*Pvu*II/8 [16] and detected

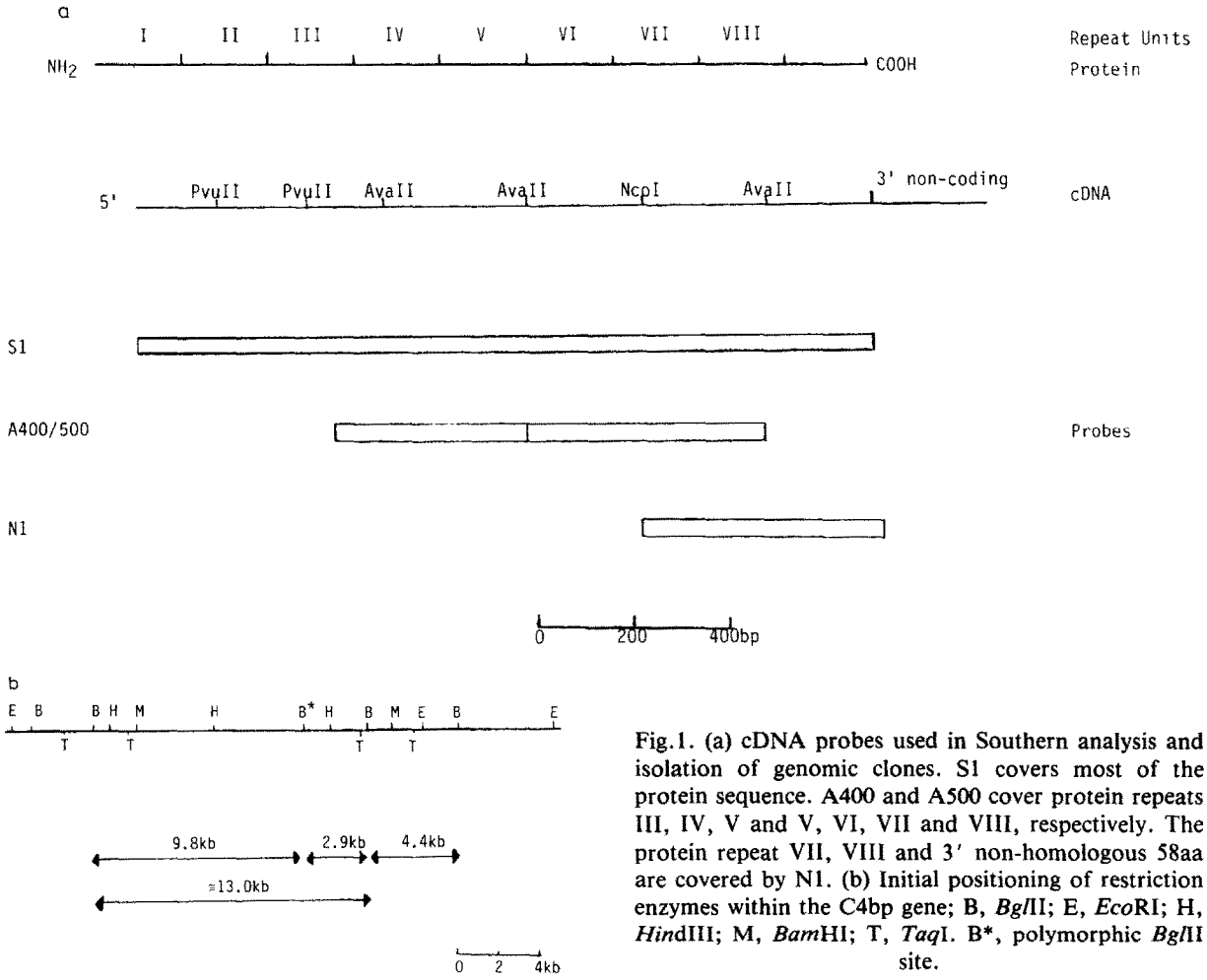

Fig.1. (a) cDNA probes used in Southern analysis and isolation of genomic clones. S1 covers most of the protein sequence. A400 and A500 cover protein repeats III, IV, V and V, VI, VII and VIII, respectively. The protein repeat VII, VIII and 3′ non-homologous 58aa are covered by N1. (b) Initial positioning of restriction enzymes within the C4bp gene; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; M, *Bam*HI; T, *Taq*I. B*, polymorphic *Bgl*II site.

by colony hybridisation with the A400 and S1 cDNA probes. Plasmid DNA was prepared by the alkaline-SDS method [17] and DNA sequencing carried out using the Maxam and Gilbert method.

3. RESULTS AND DISCUSSION

From initial genomic DNA blotting results it has been possible to construct a restriction map (fig.1b) using the four available cDNA probes which together cover most of the cDNA sequence. The overall size of the gene is approx. 30 kb as estimated by Southern blotting using 5′ and 3′ cDNA probes. From Northern blots the C4bp mRNA is estimated to be 2.5 kb long [7] and a large proportion of the gene must therefore be composed of intron sequences.

A restriction fragment length polymorphism was detected using the restriction endonuclease *Bgl*II and the S1 and A500 probes. A total of 18 genomic DNA samples (15 of which were a gift from A. Palsdottir) were then analysed with the same enzyme and probes (fig.2). Using the cDNA probes available the *Bgl*II polymorphism was mapped as shown in fig.1b and is due to the presence or absence of a *Bgl*II site 2.9 kb upstream of the 5′-end of the 4.4. kb *Bgl*II fragment.

Of the 18 samples 3 were members of a family, in which the father is homozygous for the smaller fragments and the mother and daughter are heterozygous. Thus the polymorphism is inherited and in a Mendelian fashion.

The polymorphism has been detected in 50% of the 18 samples tested and is therefore unlikely to show a simple correlation with the protein polymorphism which has gene frequencies of 0.981 and 0.018 [9]. During this study no other restriction fragment length polymorphisms using 10 other common restriction enzymes have been detected in the C4bp gene which suggests the C4bp gene, unlike the C4 and related MHC associated genes, may not be highly polymorphic although a

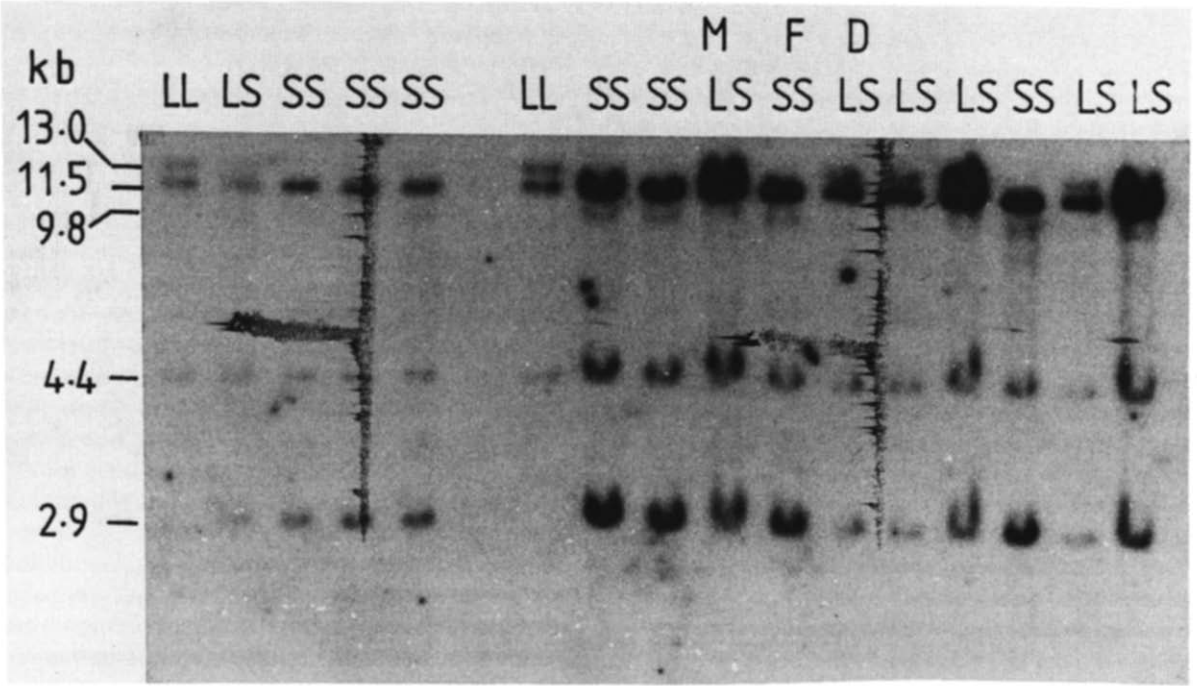

Fig.2. Southern analysis of the *Bgl*II restriction fragment length polymorphism using the A500 probe. Three different patterns can be seen, the 11.5 kb and 4.4 kb fragments being common in all. Samples marked SS have two extra bands at 9.8 kb and 2.9 kb (and are homozygous for the small fragments); LL, have one extra band at approx. 13.0 kb (and are homozygous for the large fragment). LS have all five bands (and are heterozygous). The samples marked M (mother), F (father), D (daughter) are members of one family.

number of differences between the derived protein sequence from the cDNA of an individual and the protein sequence of C4bp isolated from pooled plasma have been reported [7].

Preliminary sequencing data from one of the subclones of the 4 isolated genomic clones, clone BP/Pv.a, have determined the location of two intron/exon boundaries (fig.3) which from comparison with the published amino acid sequence of C4bp [7] corresponds to the C-terminal of the third internal repeat, and N-terminal of the fourth. From these early data it can be speculated that each of the 60 amino acid repeats will be precisely encoded within one exon, as has been found for the three internally homologous repeats in the complement protein factor B [18].

A number of complement proteins contain these internal repeating units [8], the N-terminal Ba and C2b portions of factor B and C2, both contain 3 and the chains of C4bp and factor H [19] contain 8 and 20 repeating units respectively; the precise number in CR1 is not known yet [20] but there could be as many as 36. Evidence for the encoding of the protein repeat within one exon in the three complement genes so far examined, C4bp, factor B and C2 [8], would suggest that a mechanism of tandem duplication of a single exon has occurred. In factor B and C2 the repeat structure has been combined with a catalytic chain containing a serine proteinase domain [18]. The evolution of these proteins can now be discussed in terms of exon shuffling as described for the LDL receptor [21].

The complement proteins containing the repeating units are all functionally, as well as structurally, related since they all interact with C3b or C4b [8]. However, the discovery of the repeat in non-complement proteins would imply that the presence of repeating units may be a general structural feature of a superfamily of related proteins.

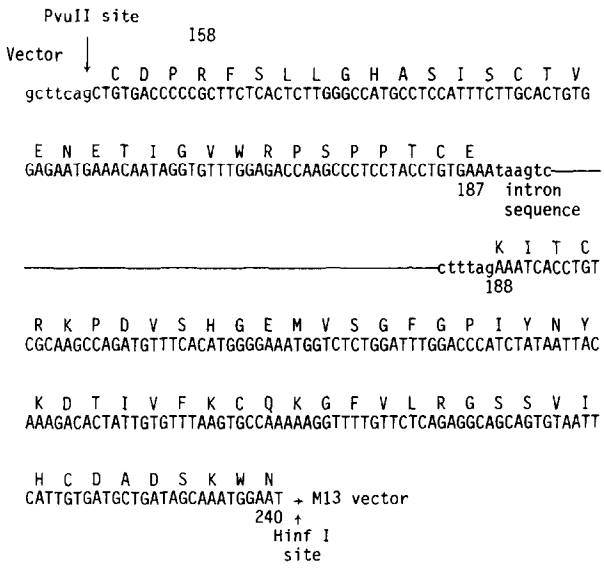

Fig.3. The nucleic acid sequence of the C4bp gene showing the coding region from Cys 155 to Glu 187 and some intron sequence and coding region from lysine 188 to asparagine 240. Amino acid 187 marks the end of the third protein repeat in C4bp.

REFERENCES

[1] Holers, V.M., Cole, J.L., Lublin, D., Seya, T. and Atkinson, J.P. (1985) Immunol. Today 6, 188–192.
[2] Gigli, I., Fujita, T. and Nussenzweig, V. (1979) Proc. Natl. Acad. Sci. USA 76, 6596–6660.
[3] Nagasawa, S., Ischihara, C. and Stroud, R.M. (1980) J. Immunol. 125, p578–582.
[4] Fujita, T., Gigli, I. and Nussenzweig, V. (1978) J. Exp. Med. 148, 1044–1051.
[5] Fujita, T. and Nussenzweig, V. (1979) J. Exp. Med. 150, 267–276.
[6] Dahlbäck, B., Smith, C.A. and Müller-Eberhard, H.J. (1983) Proc. Natl. Acad. Sci. USA 80, 3461–3465.
[7] Chung, L.P., Bentley, D.R. and Reid, K.B.M. (1985) Biochem. J. 230, 133–141.
[8] 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, in press.
[9] De Cordoba, S.R., Ferreira, D., Nussenzweig, V. and Rubinstein, P. (1983) J. Immunol. 131, 1565–1569.
[10] De Cordoba, S.R., Lublin, D., Rubinstein, P. and Atkinson, J. (1985) J. Exp. Med. 161, 1189–1195.
[11] Klickstein, L.D., Wong, W.W., Smith, J.A., Morton, C., Fearon, D.T. and Weiss, J.H. (1985) Complement 2, 44.
[12] Sykes, B.C. (1983) Lancet i, 787–788.
[13] Wahl, G.M., Stem, M. and Start, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683–3687.
[14] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Borg, P. (1977) J. Mol. Biol. 113, 237–251.
[15] Maniatis, T., Fritsch, G.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[16] Anson, D.S., Choo, K.H., Rees, D.J.G., Giannelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1986) EMBO J. 3, 1053.
[17] Birnboim, H.C. and Daly, J. (1979) Nucleic Acids Res. 7, 1513–1523.
[18] Campbell, R.D., Bentley, D.R. and Morley, B.J. (1986) Phil. Trans. R. Soc. Lond. 306, 367–378.
[19] Kristensen, T. and Tack, B.F. (1986) Proc. Natl. Acad. Sci. USA, in press.
[20] Wong, W.W., Klickstein, L.D., Smith, J.A., Weis, J.H. and Fearon, D.T. (1985) Proc. Natl. Acad. Sci. USA 82, 7711–7715.
[21] Sudhof, T.C., Goldstein, J.L., Brown, M.S. and Russell, D.W. (1985) Science 228, 815–822.