Internal homologies of the Ba fragment from human complement component Factor B, a class III MHC antigen

Bernard J.Morley and R.Duncan Campbell*

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

*To whom reprint requests should be sent

Communicated by R.R.Porter

The amino acid sequence of Ba, a fragment of the complement protein Factor B, has been determined from the sequence of its corresponding cDNA. Ba is composed of 234 amino acids and from the sequence two striking regions of internal homology are apparent which are related to a third less homologous region. Analysis of cloned genomic DNA using an 81-bp cDNA probe containing coding information for part of a leader peptide and nine amino acids at the N terminus of Ba has established the extent of the 5' end of the Factor B gene and shown that the region of the gene encoding Ba is ~1.6 kb in length.

Key words: Ba sequence/internal homology/Factor B gene/MHC class III antigen

Introduction

Factor B, a component of the alternative pathway of complement activation (Reid and Porter, 1981), is a class III major histocompatibility complex (MHC) antigen encoded between the HLA-B and HLA-D loci on human chromosome 6 (Alper, 1981). It is a serum glycoprotein and consists of a single polypeptide chain of ~90 000 mol. wt. (Curman et al., 1977; Kerr and Porter, 1978). During activation of the alternative pathway Factor B associates with C3b in a Mg^{2+}-dependent reaction (Vogt et al., 1975) and is then cleaved by Factor D into Ba of ~30 000 mol. wt., derived from the N terminus of the molecule, and Bb of ~60 000 mol. wt. (Curman et al., 1977; Kerr, 1979). The Bb fragment is the catalytically active component of the complex proteinases of the alternative pathway, C3 convertase C3bBb, and C5 convertase (C3b)$_n$Bb (Medicus et al., 1976). Sequence analysis of the Bb fragment has shown Factor B to be an unusual serine proteinase with a catalytic chain of approximately twice the size of that in other serine proteinases (Christie and Gagnon, 1983). Factor B is polymorphic and at least 11 genetic variants have been identified on the basis of differences in electrophoretic mobility (Alper et al., 1972; Mauff et al., 1978). It has been suggested that the majority of the charge heterogeneity may reside in the Ba fragment (Alper et al., 1972).

We have determined the structure of Ba from the sequence of the corresponding cDNA. Ba is composed of 234 amino acids and from the sequence a striking internal homology is apparent. Isolation and subsequent sequence analysis of a 600-bp BamHI genomic fragment detected using an 81-bp cDNA probe containing part of the leader peptide and nine amino acids at the N terminus of Ba, has enabled us to determine the extent of the 5' end of the Factor B gene.

Results

Isolation of cDNA clones

Approximately 75 000 clones of a full-length cDNA library prepared against human liver mRNA (Bentley and Porter, in preparation) were screened for those containing inserts corresponding to Factor B using two regions of human genomic DNA (Campbell and Porter, 1983) as probes. They were 750-bp and 400-bp in length and contained coding information for the C-terminal regions of Ba and Bb, respectively (Figure 4). In total 11 clones, containing inserts ranging in size from 1.5 to 2.4 kb in length, were isolated which screened positively with both probes. Sequence analysis of the longest clones established that the inserts contained sequences coding for Factor B, but showed that the whole of Ba had been reversed with respect to Bb such that the amino termini of Ba and Bb were adjacent (Figure 1). This situation is not without precedent as it has been suggested (Williams, 1981; Maniatis et al., 1982) that the looped-back, self-priming reaction, for second strand synthesis may lead to inversion of a region of cDNA if a sufficiently strong homology exists between the 3' end of the molecule and an internal site, and the S1 nuclease cuts near this region. However, since the amino- and carboxy-terminal sequences of Ba are known it has been possible to derive the complete amino acid sequence of Ba from sequence analysis of a 660-bp fragment produced using an internal BamHI site, situated near the 5' end of Ba and the ClaI site of the pAT153/PvuII/8 vector (Choo et al., 1982).

The sequence of Ba

The restriction map of the 660-bp BamHI/ClaI fragment together with the approach used to determine the nucleotide sequence, is shown in Figure 1. The entire sequence of both DNA strands was determined by the Maxam and Gilbert (1980) procedure and the sequence at all restriction sites overlapped. The complete nucleotide sequence and the derived
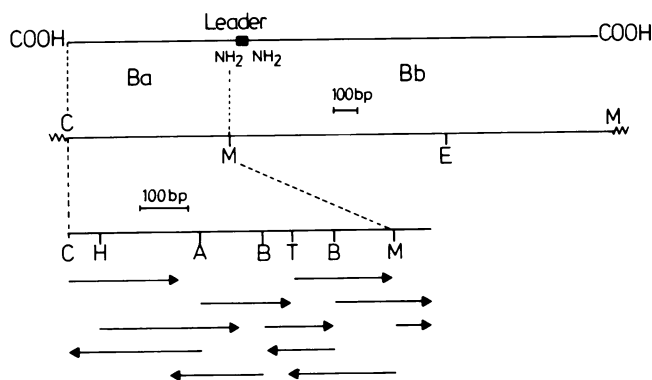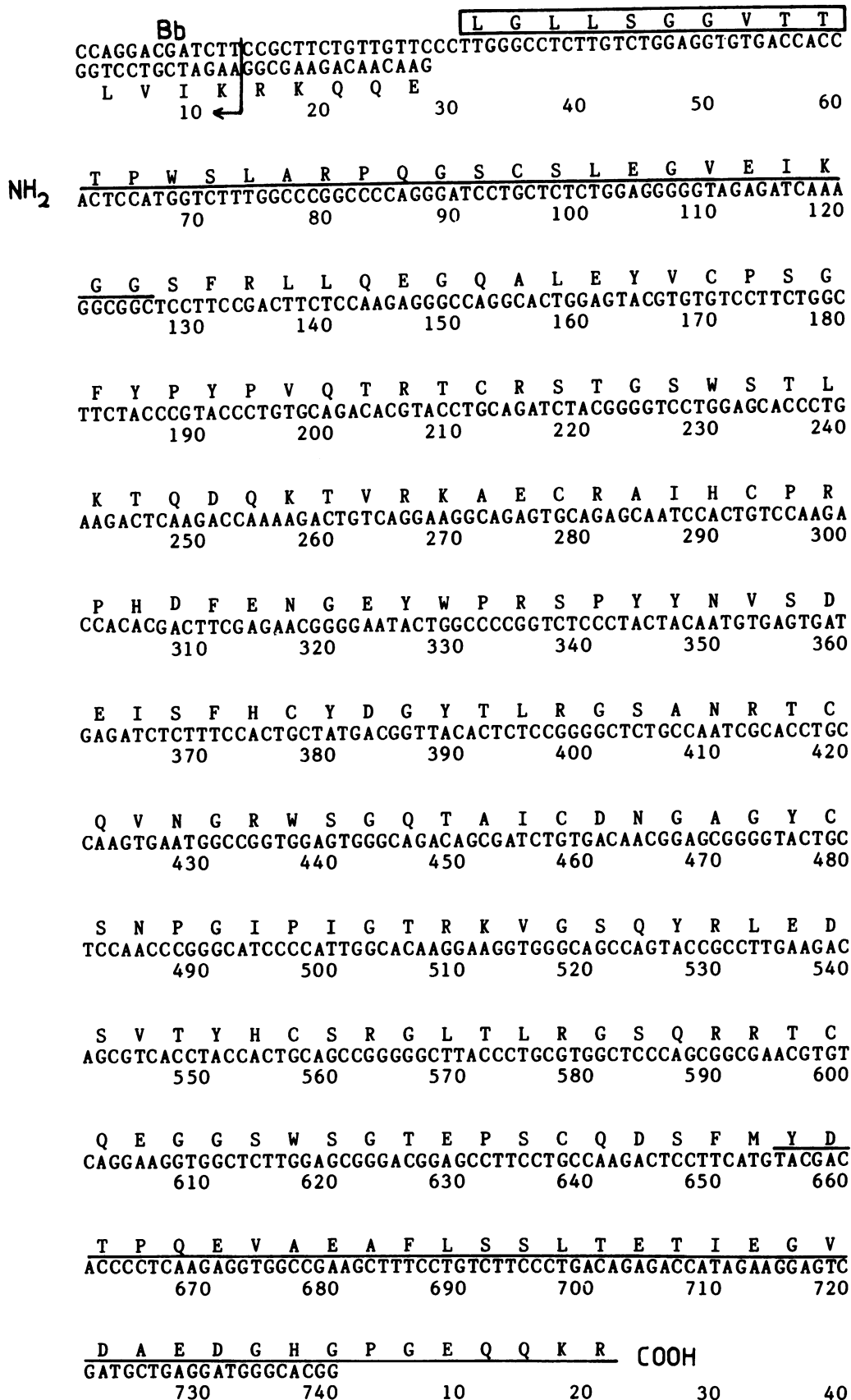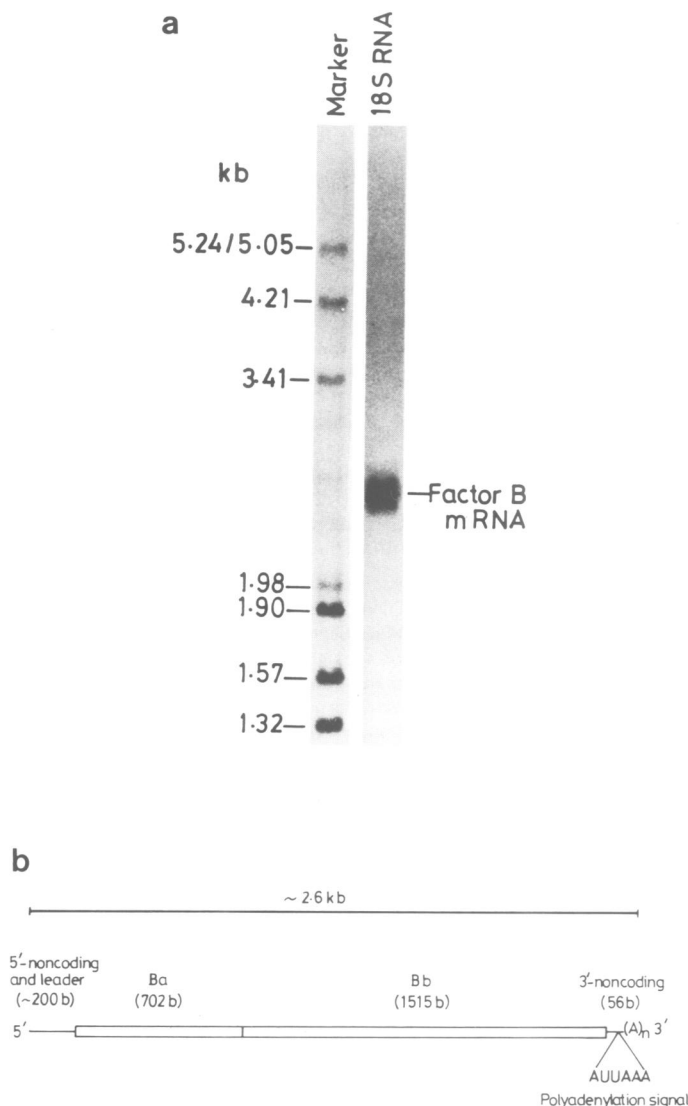

Fig. 1. Sequencing strategy and restriction map for the Ba region of the full-length Factor B cDNA clone, pFB3b, together with a schematic representation of the orientation of Ba and Bb within the clone. C, ClaI; H, HindIII; A, AvaI; B, BglII; T, TaqI; M, BamHI; E, EcoRI.

```
                                            ┌─────────────────────────────┐
         Bb                                 │ L  G  L  L  S  G  G  V  T  T │
      CCAGGACGATCTT CCGCTTCTGTTGTTCCCTTGGGCCTCTTGTCTGGAGGTGTGACCACC
      GGTCCTGCTAGAA GGCGAAGACAACAAG
        L  V  I  K │ R  K  Q  Q  E
              10 ◄─┘    20          30        40        50        60
```

```
        T  P  W  S  L  A  R  P  Q  G  S  C  S  L  E  G  V  E  I  K
NH₂   ACTCCATGGTCTTTGGCCCGGCCCCAGGGATCCTGCTCTCTGGAGGGGGTAGAGATCAAA
        70        80        90        100       110       120
```

```
        G  G  S  F  R  L  L  Q  E  G  Q  A  L  E  Y  V  C  P  S  G
      GGCGGCTCCTTCCGACTTCTCCAAGAGGGCCAGGCACTGGAGTACGTGTGTCCTTCTGGC
        130       140       150       160       170       180
```

```
        F  Y  P  Y  P  V  Q  T  R  T  C  R  S  T  G  S  W  S  T  L
      TTCTACCCGTACCCTGTGCAGACACGTACCTGCAGATCTACGGGGTCCTGGAGCACCCTG
        190       200       210       220       230       240
```

```
        K  T  Q  D  Q  K  T  V  R  K  A  E  C  R  A  I  H  C  P  R
      AAGACTCAAGACCAAAAGACTGTCAGGAAGGCAGAGTGCAGAGCAATCCACTGTCCAAGA
        250       260       270       280       290       300
```

```
        P  H  D  F  E  N  G  E  Y  W  P  R  S  P  Y  Y  N  V  S  D
      CCACACGACTTCGAGAACGGGGAATACTGGCCCCGGTCTCCCTACTACAATGTGAGTGAT
        310       320       330       340       350       360
```

```
        E  I  S  F  H  C  Y  D  G  Y  T  L  R  G  S  A  N  R  T  C
      GAGATCTCTTTCCACTGCTATGACGGTTACACTCTCCGGGGCTCTGCCAATCGCACCTGC
        370       380       390       400       410       420
```

```
        Q  V  N  G  R  W  S  G  Q  T  A  I  C  D  N  G  A  G  Y  C
      CAAGTGAATGGCCGGTGGAGTGGGCAGACAGCGATCTGTGACAACGGAGCGGGGTACTGC
        430       440       450       460       470       480
```

```
        S  N  P  G  I  P  I  G  T  R  K  V  G  S  Q  Y  R  L  E  D
      TCCAACCCGGGCATCCCCATTGGCACAAGGAAGGTGGGCAGCCAGTACCGCCTTGAAGAC
        490       500       510       520       530       540
```

```
        S  V  T  Y  H  C  S  R  G  L  T  L  R  G  S  Q  R  R  T  C
      AGCGTCACCTACCACTGCAGCCGGGGGCTTACCCTGCGTGGCTCCCAGCGGCGAACGTGT
        550       560       570       580       590       600
```

```
        Q  E  G  G  S  W  S  G  T  E  P  S  C  Q  D  S  F  M  Y  D
      CAGGAAGGTGGCTCTTGGAGCGGGACGGAGCCTTCCTGCCAAGACTCCTTCATGTACGAC
        610       620       630       640       650       660
```

```
        T  P  Q  E  V  A  E  A  F  L  S  S  L  T  E  T  I  E  G  V
      ACCCCTCAAGAGGTGGCCGAAGCTTTCCTGTCTTCCCTGACAGAGACCATAGAAGGAGTC
        670       680       690       700       710       720
```

```
        D  A  E  D  G  H  G  P  G  E  Q  Q  K  R   COOH
      GATGCTGAGGATGGGCACGG
        730       740       10        20        30        40
```

Fig. 2. The cDNA sequence and derived protein sequence of Ba including 10 amino acids of a putative leader sequence (boxed). The amino- and carboxy-terminal residues of Ba are designated by NH₂ and COOH, respectively. The known protein sequence at the amino- and carboxy-termini (Christie and Gagnon, 1982) is underlined. Also shown is the region of cDNA encoding the five carboxy-terminal residues of Ba and the N-terminal region of Bb.
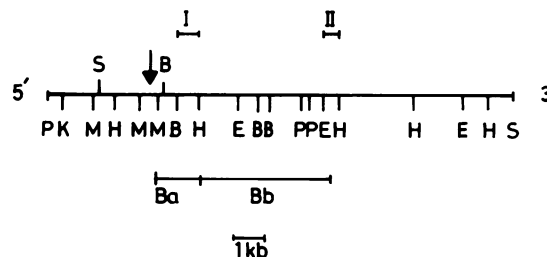
Fig. 4. Partial restriction map of overlapping 8.4-kb *Pvu*II and 11-kb *Sma*I restriction fragments which contain the Factor B gene. The map was prepared from restriction digests of the DNA and separation of the fragments on agarose gels. Enzymes used were: B, *Bgl*II; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; M, *Bam*HI; P, *Pvu*II; S, *Sma*I.

gel and hybridisation with the FB1 cDNA probe indicated that the Factor B mRNA was ~2.6 kb in length (Figure 3). The structure of the mRNA was deduced from the known protein sequence data of Bb and the cDNA sequence of clone pFB3b shown in Figure 2. The coding region is 2217 bases of which 702 bases account for Ba and 1515 bases account for Bb. Sequence analysis at the 3' end of the cloned insert has established that the 3'-non-coding region including the stop codon (UAA) is 56 bases, in agreement with Woods *et al.* (1982), and contains one polyadenylation signal, AUUAAA, beginning 20 bases before the poly(A) tract. The sequence of the 3'-non-coding region agrees exactly with that found in the gene (Campbell and Porter, 1983). The 5'-untranslated region and the leader peptide are contained within 100−200 bases assuming an average poly(A) tract of 100−150 bases (Brawerman, 1976).

The limits of the Factor B gene

A restriction map of overlapping 8.4-kb *Pvu*II and 11-kb *Sma*I genomic DNA fragments from the genomic clone, cosA2, found to hybridise with the FB1 cDNA probe (Campbell and Porter, 1983) is shown in Figure 4. Southern blot analysis of restriction digests of the cloned DNA showed that the pFB3b cDNA probe hybridised to adjacent 2.7-kb and 4.4-kb *Hind*III fragments. Previous work has defined the 3' end of the gene and established that the Bb portion of the gene is contained within the 4.4-kb fragment and is ~4.1 kb in length (Campbell and Porter, 1983).

To characterise the 5' end of the gene an 81-bp *Bam*HI/*Sau*3A restriction fragment derived from the pFB3b cDNA clone and containing coding information for part of the leader peptide and nine amino acids at the N terminus of Ba was used as a hybridisation probe on Southern blots of restriction digests of the cloned genomic DNA. The probe hybridised strongly to a 600-bp *Bam*HI fragment (arrowed in Figure 4), and sequence analysis has shown that the *Bam*HI site defining the 3' end of the fragment corresponds to the *Bam*HI site found in the cDNA. The sequence analysis also showed that an intron occurs within the leader peptide and results in the four C-terminal amino acids of the leader (GVTT) being encoded in the same exon as the N-terminal region of Ba (results not shown). Thus the region of the gene encoding Ba is ~1.6 kb in length (Figure 4) and the complete gene spans greater than 5.7 kb of DNA.

Fig. 3. (a) Northern blot analysis of 18S human liver RNA hybridised with the 515-bp Factor B cDNA probe, FB1. (b) A schematic diagram of the Factor B mRNA illustrating the coding (boxed) and non-coding regions, together with the polyadenylation (AUUAAA) signal.

amino acid sequence of Ba is shown in Figure 2. The amino acid sequence agrees exactly with the known N-terminal sequence of Ba and continues in an open reading frame to seven amino acids short of the carboxy terminus, overlapping by 29 amino acids the known C-terminal sequence. Five (EQQKR) of the remaining seven carboxy-terminal amino acids are in the correct orientation with respect to Bb and are consequently coded for in the region of DNA immediately before the amino-terminus of Ba (Figure 2). The derived amino acid sequence indicates that Ba is composed of 234 amino acids and has two potential glycosylation sites, conforming to the sequence Asn-X-Thr or Ser (Neuberger *et al.*, 1972) lying close to one another at amino acid residues 97 and 117.

Sequence analysis across the *Bam*HI site shows that immediately adjacent to the N-terminal Thr residue of Ba the cDNA sequence codes for 10 amino acids which are predominantly non-polar in nature and may represent part of a leader peptide (Figure 2).

Northern blot analysis

Electrophoresis of 18S adult human liver RNA in an agarose

Discussion

The amino acid sequence of Ba reported here together with the sequence of Bb determined by Christie and Gagnon
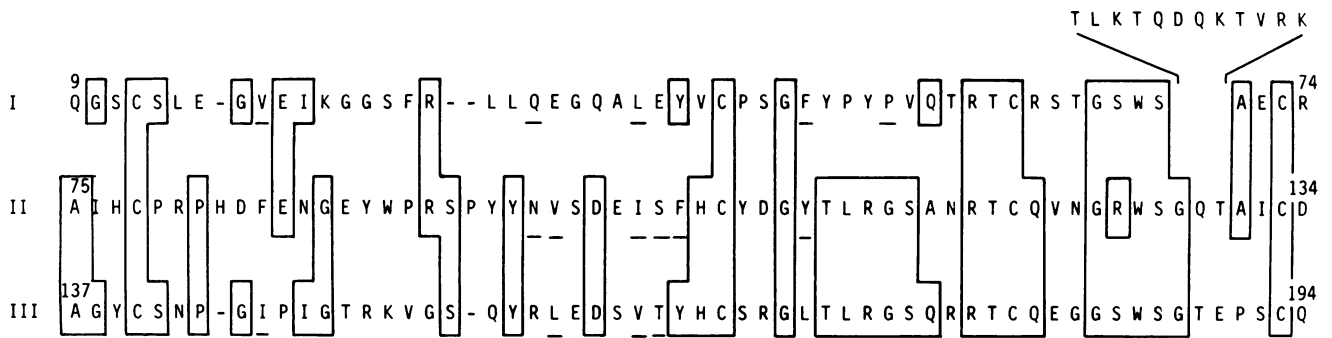
Fig. 5. Comparison of three regions of the Ba amino acid sequence illustrating the high degree of internal homology. The three regions correspond to amino acids 9–74 (I), 75–134 (II) and 137–194 (III). Boxed areas represent identity and underlined regions represent functionally conserved amino acids, designated by Dayhoff *et al.* (1972) from chemical similarity and accepted point-mutation data and consist of (A,P,G), (N,Q), (D,E), (S,T), (C), (V,I,M,L), (K,R,H) and (F,Y,W).

(1983) completes the sequence of Factor B and shows that the zymogen is composed of 739 amino acids. The complete sequence of the zymogen has also been presented by Mole *et al.* (1983).

The Ba fragment is composed of 234 amino acids and from the sequence a striking internal homology is apparent. Three regions of sequence, corresponding to amino acids 9–74 (I), 75–134 (II) and 137–194 (III), are aligned in Figure 5. It has only been necessary to insert two small gaps in the N-terminal halves of regions I and III and a small insertion in the C-terminal half of region I in order to align the four cysteine residues contained in each region. Region I appears to be as related to region II as it is to region III with 27% and 29% homology, respectively. Regions II and III show greater similarity to one another than to region I with an overall homology of 47%. Taking the C-terminal 30 amino acids this homology increases to 57% identity. Although the number of cysteine residues is high there is no significant homology between Ba and the kringle structure of the blood clotting proteins (Dayhoff, 1978). Nor is there any obvious homology between Ba and the amino acid sequences of immunoglobulins or MHC class I and class II antigens (Kabat *et al.*, 1983).

The strong homology that exists between the sequences in regions II and III suggests that they may have arisen from a DNA duplication event. This mechanism may also have been responsible for the emergence of the sequence in region I, but because this region appears more distantly related to regions II and III the duplication event which gave rise to this region may have occurred earlier in the evolution of the Ba fragment. Preliminary sequence analysis of the Ba portion of the Factor B gene (Morley and Campbell, unpublished data) suggests that the three regions shown in Figure 5 are contained within separate exons.

The Factor B gene has been shown to span >5.7 kb of DNA. Due to the very close linkage of the Factor B and C2 genes (Carroll *et al.*, 1983) it is likely that the 600-bp *Bam*HI fragment defining the 5' end of the Ba portion of the gene (Figure 4) contains the sequences responsible for control of Factor B expression, suggesting that the gene is ~6 kb in length. Work is continuing to determine the complete structure of the gene.

Materials and methods

Enzymes

Restriction endonucleases were purchased from New England Biolabs and

Amersham International. Nick translation 'kit' and all radiolabelled nucleotides were also from Amersham International.

Clone isolation and sequencing

75 000 clones of a full-length cDNA library, constructed from human liver mRNA by Drs. A.Connolly and D.R.Bentley (Bentley and Porter in preparation), were plated onto nitrocellulose filters at a colony density of ~5000 clones per plate. Replica filters were prepared and processed for colony hybridisation as described by Grosveld *et al.* (1981). Filters were prehybridised and hybridised at 42°C in buffer containing 50% formamide, according to the procedure of Bernards and Flavell (1980). Two regions of genomic DNA, 750 bp and 400 bp in length and containing coding information for the C-terminal regions of Ba and Bb, respectively, were excised from *Pvu*II and *Sma*I subclones of cosA2 (Campbell and Porter, 1983) using *Bgl*II/*Hind*III (750 bp) and *Eco*RI/*Hind*III (400 bp) double digests (regions I and II, respectively, in Figure 4). These fragments were nick-translated to a specific activity of ~10⁸ c.p.m./μg of DNA according to Rigby *et al.* (1977), to be used independently as hybridisation probes for the replica filters. Colonies which screened positively with both genomic probes were rescreened using the FB1 cDNA probe (Campbell and Porter, 1983). Plasmid DNA was extracted from bacterial colonies by the alkaline-SDS method (Birnboim and Doly, 1979). End-labelled fragments to be sequenced by the Maxam and Gilbert (1980) procedure were prepared by strand separation or secondary restriction enzyme cleavage.

Northern blot analysis

18 μg of 18S human liver RNA were denatured in 50% formamide/16% formaldehyde at 60°C for 15 min, electrophoresed in a 1% agarose-formaldehyde gel (Lehrach *et al.*, 1977), and washed consecutively, at room temperature, in water, 0.05 M NaOH, 0.1 M Tris/HCl, pH 7.5 and 1 x SSC (0.15 M NaCl, 0.015 M Na₃ citrate, pH 7). Transfer to nitrocellulose was as described by Wahl *et al.* (1979) and hybridisation to the nick-translated FB1 cDNA probe was carried out according to Campbell and Porter (1983).

Subcloning

*Pvu*II and *Sma*I digests of a cosmid clone of genomic DNA, cosA2, which contains the Factor B gene (Campbell and Porter, 1983) were ligated into the *Pvu*II site of the plasmid vector pAT153/*Pvu*II/8. Clones which contained inserts corresponding to the 8.4-kb *Pvu*II and 11-kb *Sma*I fragments were detected by colony hybridisation using the FB1 cDNA probe.

Acknowledgements

References

Alper,C.A. (1981) in Dorf,M.E. (ed.), *The Role of the Major Histocompatibility Complex in Immunobiology*, Garland, NY, pp. 173-220.
Alper,C.A., Boenisch,T. and Watson,L. (1972) *J. Exp. Med.*, **135**, 68-80.
Bernards,R. and Flavell,R.A. (1980) *Nucleic Acids Res.*, **8**, 1521-1534.
Birnboim,H.C. and Doly,T. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.
Brawerman,G. (1976) *Prog. Nucleic Acid Res. Mol. Biol.*, **17**, 117-148.
Campbell,R.D. and Porter,R.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4464-4468.

Carroll,M.C., Campbell,R.D., Bentley,D.R. and Porter,R.R. (1983) *Nature,* in press.

Choo,K.H., Gould,K.G., Rees,D.J.G. and Brownlee,G.G. (1982) *Nature,* **299**, 178-180.

Christie,D.L. and Gagnon,J. (1982) *Biochem. J.,* **201**, 555-567.

Christie,D.L. and Gagnon,J. (1983) *Biochem. J.,* **209**, 61-70.

Curman,B., Sandberg-Tragardh,L. and Peterson,P.A. (1977) *Biochemistry (Wash.),* **16**, 5368-5375.

Dayhoff,M.O., ed. (1978) *Atlas of Protein Sequence and Structure* 5, supplement 3, published by National Biomedical Research Foundation, pp. 84-86.

Dayhoff,M.O., Eck,R.V. and Park,C.M. (1972) in Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure,* National Biomedical Research Foundation, pp. 89-99.

Grosveld,F.G., Dahl,H.H.M., de Boer,R. and Flavell,R.A. (1981) *Gene,* **13**, 227-237.

Kabat,E.A., Wu,T.T., Bilofsky,M., Reid-Miller,M. and Perry,H. (1983) in *Sequences of Proteins of Immunological Interest,* U.S. Dept. Health and Human Services.

Kerr,M.A. (1979) *Biochem. J.,* **183**, 615-622.

Kerr,M.A. and Porter,R.R. (1978) *Biochem. J.,* **171**, 99-107.

Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry (Wash.),* **16**, 4743-4751.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, a Laboratory Manual,* published by Cold Spring Harbor Laboratory Press, NY, pp. 214-216.

Mauff,G., Hauptmann,G., Hitzeroth,H.W., Gauchel,F. and Scherz,R. (1978) *Z. Immunitforsch. Immunobiol.,* **154**, 115-119.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.

Medicus,R.G., Gotze,O. and Muller-Eberhard,H.J. (1976) *Scand. J. Immunol.,* **5**, 1049-1055.

Mole,J.E., Woods,D., Colten,H. and Anderson,J.K. (1983) *Immunobiology,* **164**, 279.

Neuberger,A., Gottschalk,A., Marshall,R.D. and Spiro,R.G. (1972) in Gottschalk,A. (ed.), *The Glycoproteins: Their Composition, Structure and Function,* Elsevier, Amsterdam, pp. 450-490.

Reid,K.B.M. and Porter,R.R. (1981) *Annu. Rev. Biochem.,* **50**, 433-464.

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.,* **113**, 237-251.

Vogt,W., Schmidt,G., Dieminger,L. and Lynen,R. (1975) *Z. Immunforsch. Exp. Ther.,* **149**, 440-447.

Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA,* **76**, 3683-3687.

Williams,J.G. (1981) in Williamson,R. (ed.), *Genetic Engineering 1,* Academic Press, pp. 1-59.

Woods,D.E., Markham,A.F., Ricker,A.T., Goldberger,G. and Colten,H. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 5661-5665.