Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 4464–4468, July 1983
Immunology

Molecular cloning and characterization of the gene coding for human complement protein factor B

(HLA class III antigen/serine proteinase/DNA sequence analysis/cosmid library/synthetic oligonucleotides)

R. DUNCAN CAMPBELL AND RODNEY R. PORTER

Medical Research Council Immunochemistry Unit, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, United Kingdom

Contributed by R. R. Porter, March 21, 1983

ABSTRACT **Four cosmid clones, each with an average insert size of 40 kilobase pairs and containing the factor B gene, were isolated from a human genomic DNA library. The clones were identified by hybridization with a 515-base-pair cDNA probe isolated by using a unique 17-base synthetic oligonucleotide probe from a human liver cDNA library. The cosmid clones were characterized by restriction endonuclease digestion and Southern blotting, and a partial restriction map of the DNA represented in the cosmids was constructed. The *Bb* portion of the factor B gene is about 4 kb in length. DNA sequence analysis has resulted in the determination of 3.3 kb of sequence at the 3′ end of the gene. This region codes for amino acids 87–505 of Bb and includes the whole of the serine proteinase domain of the protein. The three active site residues of histidine, aspartic acid, and serine found at positions 267, 317, and 440 of the Bb sequence, respectively, lie on separate exons. Other functional regions within the serine proteinase domain are separated also by intervening sequences.**

The major histocompatibility complex on human chromosome 6 is a highly polymorphic multigene system that encodes proteins known to have key roles in cell–cell recognition, especially those governing the functions of the immune system. The gene products fall into three major classes of antigen: class I and class II molecules are both integral cell-surface glycoproteins that are specified, respectively, by the loci for the major histocompatibility antigens HLA-A, -B, and -C and by the locus for HLA-DR, whereas class III molecules are components of the serum complement system—namely, factor B and the second and fourth components of complement, C2 and C4 (reviewed in refs. 1 and 2).

Factor B, a glycoprotein of M_r 90,000, is a component of the alternative pathway of complement activation (3). It is cleaved by factor $\overline{D}$ in the presence of C3b into two fragments, Ba of M_r 30,000 and Bb of M_r 60,000 (4). The Bb fragment is the catalytically active component of the complex proteinases of the alternative pathway, C3 convertase $\overline{C3bBb}$, and C5 convertase $\overline{C3b_nBb}$ (5). Amino acid sequence studies have shown factor B to be a novel type of serine proteinase with a catalytic chain of approximately twice the size of other known serine proteinases (6). Factor B is polymorphic, and at least 11 genetic variants have been identified on the basis of differences in electrophoretic mobility (7).

The availability of cDNA probes to class I (8–10) and class II (11–13) antigens has led to a better understanding of the molecular organization and genetic structure of the genes for these molecules. Although cDNA clones containing sequences coding for factor B have been isolated (14), little is known about the detailed molecular structure of the gene for this protein. To this end we have used a human factor B cDNA clone, isolated by using a synthetic oligonucleotide as probe, to screen human genomic DNA cosmid libraries. Here we report the isolation and partial characterization of the factor B gene and the complete structure and sequence of that part of the gene specifying the serine proteinase domain of Bb.

MATERIALS AND METHODS

Synthesis of Oligonucleotide Probes. The 14-base mixture of eight oligonucleotides and the unique 17-base oligonucleotide were synthesized by the solid-phase phosphotriester technique (15).

Synthesis of cDNA. The 14-base oligonucleotide mix was 5′-labeled by using [γ-^{32}P]ATP and T4 polynucleotide kinase (16). Synthesis of single-stranded cDNA was based on the method of Wickens *et al.* (17) and was carried out by incubating 500 ng of the labeled oligonucleotide mix and 180 μg of 18S RNA with 50 units of avian myeloblastosis virus reverse transcriptase (J. W. Beard) in 100 μl at 42°C for 60 min. The mixture was incubated at 70°C for 20 min in 0.1 M NaOH/1 mM EDTA to degrade the RNA, precipitated with ethanol, and desalted on a Sephacryl S-300 column equilibrated in 0.2 M NaCl/1 mM EDTA/10 mM Tris, pH 7.4. The products of the reaction were analyzed under denaturing conditions on a 3% polyacrylamide gel.

Isolation of cDNA Clones. Approximately 30,000 clones of a human liver cDNA fragment library (18) were plated onto ampicillin plates, then transferred to Whatman 541 filter paper (19), and processed for hybridization as described by Gergen *et al.* (20). Filters were treated prior to hybridization (prehybridized) as described by Carroll and Porter (18), then hybridized for 16 hr at 46°C in solution containing the end-labeled unique 17-base oligonucleotide probe (2 ng/ml; 8×10^5 cpm/ml). After successive washes at 4°C (1 hr) and 46°C (30 min), the filters were dried in air and autoradiographed at −70°C overnight.

Isolation of Genomic Clones. Approximately 200,000 clones from each of two cosmid libraries constructed by F. G. Grosveld *et al.* (21) using placental DNA partially digested with *Mbo* I were plated onto nitrocellulose filters. Replicas were prepared and processed for colony hybridization as described by Grosveld *et al.* (22). Filters were prehybridized and hybridized at 42°C in buffer containing 50% formamide (23).

The 515-base-pair (bp) insert of the factor B cDNA clone, FB1, was nick-translated to a specific activity of $\approx 10^8$ cpm/μg of DNA (24). The hybridization mix contained $2–5 \times 10^5$ cpm/ml. After hybridization for 20 hr, the filters were washed as described by Carroll and Porter (18).

Preparation and Analysis of Cloned DNA. Plasmid (25) and cosmid (22) DNA were extracted from bacterial colonies by standard procedures. Partial characterization of the cosmids was achieved by a series of single, double, and triple digests and by separation of the fragments on agarose gels. DNA fragments

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: C2, C3, and C4, second, third, and fourth components of complement; bp, base pair; kb, kilobase pair.

were blotted onto nitrocellulose filters as described by Wahl *et al.* (26) and hybridized with either the factor B cDNA probe or the cosmid vector as probe.

Restriction fragments from the cosmid DNA found to hybridize with the cDNA probe were subcloned into the pAT153/*Pvu* II/8 plasmid vector as follows. A *Bgl* II digest and a *Sma* I digest were ligated into the *Bam*HI and *Pvu* II sites of the vector, respectively, and clones containing inserts corresponding to the 0.4-kilobase (kb) and 2.6-kb *Bgl* II fragments and the 11-kb *Sma* I fragment were detected by colony hybridization. To facilitate sequence analysis, two *Eco*RI fragments of 1.6 kb and 1.1 kb from the 11-kb *Sma* I subclone were filled in by using the Klenow fragment of DNA polymerase I, purified by polyacrylamide gel electrophoresis, and blunt-end-ligated into the *Pvu* II site of the vector. Several clones were tested for the presence of 1.6-kb and 1.1-kb inserts by restriction analysis of isolated plasmids.

DNA sequence analysis was carried out as described by Maxam and Gilbert (16).

RESULTS AND DISCUSSION

Isolation and Characterization of Factor B cDNA Clones. Synthetic oligonucleotides have been used successfully as primers of cDNA synthesis (27) and as specific hybridization probes in the detection and characterization of recombinant DNA molecules (14, 18, 19, 28). This approach was adopted to isolate cDNA clones corresponding to factor B and an oligonucleotide was synthesized against a region of the factor B sequence with the least ambiguity in codon usage. The sequence chosen was Met-Val-Trp-Glu-His, corresponding to positions 221–225 of Bb (6), and a 14-base mixture of eight synthetic oligonucleotides,

$$5'\ d\left(T\text{-}G\text{-}\begin{matrix}T\\C\end{matrix}\text{-}T\text{-}C\text{-}C\text{-}C\text{-}A\text{-}\begin{matrix}A\\T\\C\\G\end{matrix}\text{-}A\text{-}C\text{-}C\text{-}A\text{-}T\right)\ 3',$$

was synthesized. The oligonucleotide mix was phosphorylated at the 5′ OH and was used as a primer to synthesize cDNA from the 18S fraction of human liver RNA. The reaction products were analyzed under denaturing conditions on a 3% polyacrylamide gel, and a prominent band at ≈1,560 bases was visualized (Fig. 1). This band appeared to be a strong candidate for factor B cDNA, as the expected product based on the known protein sequence data would be at least 1,425 bases long (Fig. 1). Sufficient radioactivity (≈3,000 cpm) was contained in the band to allow elution and sequence analysis of the 5′ end by the Maxam and Gilbert technique. The sequence of 30 bases was determined, enough to confirm that this band did indeed

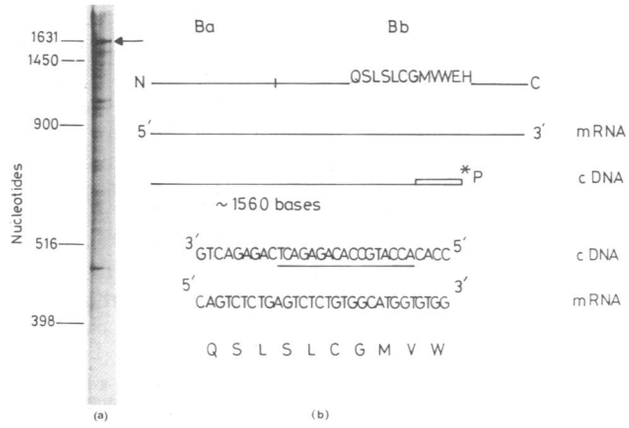


FIG. 1. Analysis of cDNA products synthesized from 18S RNA by using the 14-base oligonucleotide mix as primer. (*a*) The products of the reaction were run under denaturing conditions on a 3% polyacrylamide/urea gel, and the band indicated by the arrow was recovered for sequence analysis. Numbers represent size standards. (*b*) Schematic representation of the priming reaction and the sequence of 30 bases determined from the 5′ end of the cDNA product running at ≈1,560 bases. The derived amino acid sequence agrees with the known amino acid sequence of this portion of Bb (6). The sequence underlined was used to synthesize a unique 17-base oligonucleotide probe.

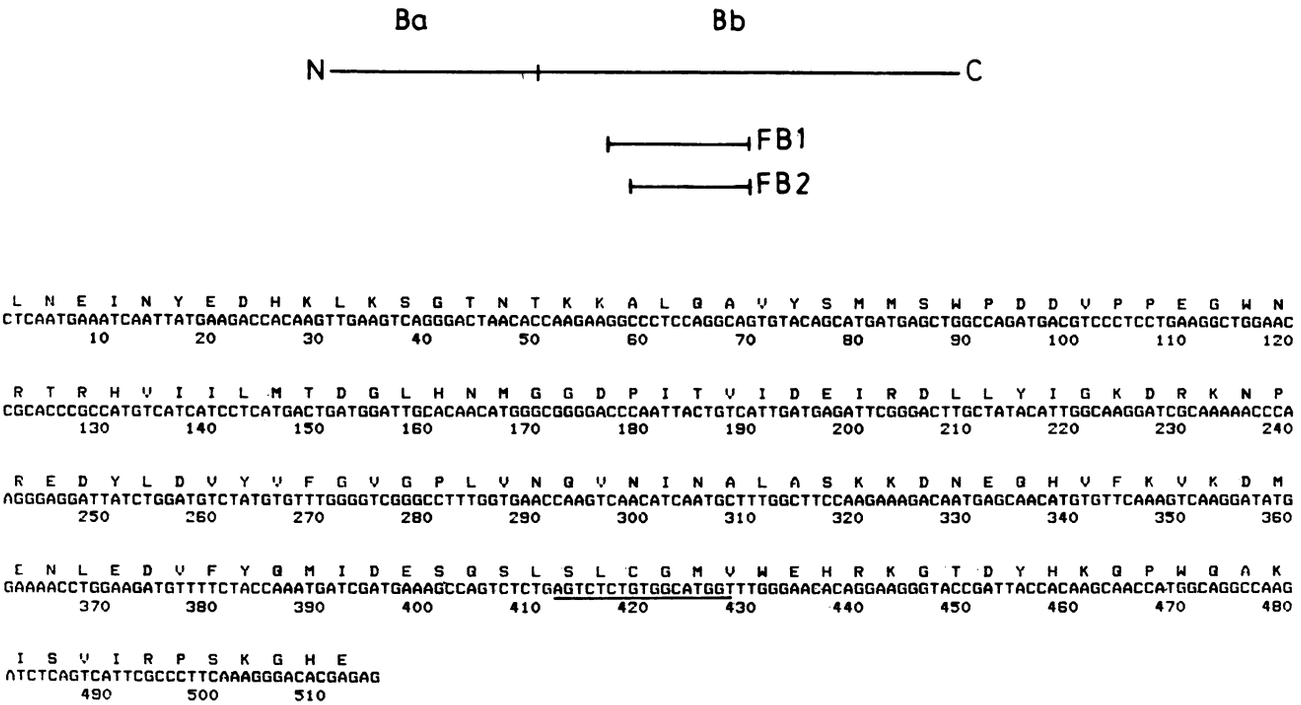


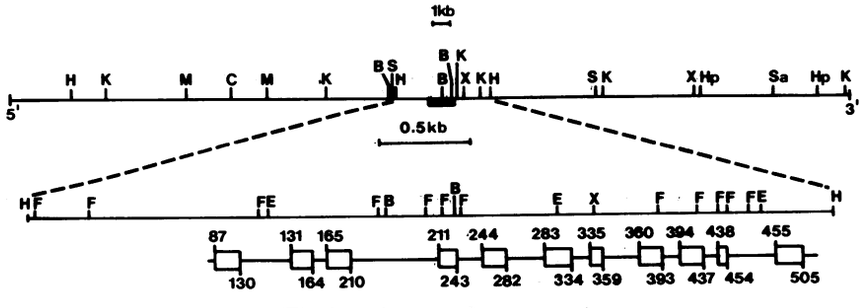
FIG. 2. Nucleotide sequence of the FB1 cDNA insert. (*Upper*) Position of the cloned cDNA inserts FB1 and FB2 in relation to factor B. (*Lower*) The nucleotide sequence was determined on end-labeled restriction fragments isolated from polyacrylamide gels (16). The sequence complementary to the 17-base oligonucleotide probe is underlined.

correspond to factor B cDNA primed by using the 14-base oligonucleotide mix (Fig. 1).

The sequence underlined in Fig. 1 was used to synthesize a unique 17-base oligonucleotide probe. The probe was end-labeled and used to screen, by colony hybridization, ≈30,000 clones from an *Alu* I restriction fragment library of human liver cDNA cloned in the *Pvu* II site of the plasmid vector pAT153/*Pvu* II/8 (18). Two positive clones, FB1 and FB2, were identified over background after washes at 46°C and 55°C. The sequences of the cloned inserts were determined by the Maxam and Gilbert procedure. Clone FB1 contains an insert of 515 bp and codes for amino acids 80–250 of Bb, whereas clone FB2 contains an insert of 429 bp and codes for amino acids 109–250 (Fig. 2). The insert of clone FB1 is a partial digest product having an *Alu* I site at position 86 and contains within it the complete sequence of clone FB2. The amino acid sequence derived from the nucleotide sequence agrees exactly with the published amino acid sequence of Bb (6). It also confirms the assignment of an asparagine residue at position 119 of the protein sequence.

Isolation and Characterization of Factor B Genomic Clones. Initial screening of ≈200,000 clones from a human genomic DNA cosmid library constructed by F. G. Grosveld (21) using the nick-translated FB1 cDNA insert as probe resulted in the

```
                D  H  K  L  K  S  G  T  N  T  K  K  A  L  Q  A  V  Y  S  M  M  S  W  P  D  D  V  P  P  E  G  W  N  R  T
CTCTTCCCTCTCCACAGACCACAAGTTGAAGTCAGGGACTAACACCAAGAAGGCCCTCCAGGCAGTGTACAGCATGATGAGCTGGCCAGATGACGTCCCTCCTGAAGGCTGGAACCGCAC
      10        20        30        40        50        60        70        80        90        100       110       120

  R  H  V  I  I  L  M  T  D
CCGCCATGTCATCATCCTCATGACTGATGGTCAGAAGGGACCTCTCTCCTGTCCCAGCCTCCCCACCTTCTCAGACCAGCATGTGGCCCTTAAGTCCACTTGTAACACTATACCCATGGT
      130       140       150       160       170       180       190       200       210       220       230       240

TGGGGCCCTGAATGTGACTCATAGCTGGCTGTTCATCTCTCCTGTGACCCTTCATAAGGAATTCTTCCTAAGCCCTGTGATCAACTATCTCTAACCCTTCCTCAACTTGCTCACCCTGCC
      250       260       270       280       290       300       310       320       330       340       350       360

                                                                                G  L  H  N  M  G  G  D  P  I  T  V  I  D  E  I  R
ATGTGTATCCCTGCCTTTAGCCAGTTTATCTTCCTTATCTCCTACCCTCATGGTCCTGTCTCTTCTGCAGGATTGCACAACATGGGCGGGGACCCAATTACTGTCATTGATGAGATCCGG
      370       380       390       400       410       420       430       440       450       460       470       480

  D  L  L  Y  I  G  K  D  R  K  N  P  R  E  D  Y  L
GACTTGCTATACATTGGCAAGGATCGCAAAAACCCAAGGGAGGATTATCTGGGTGAGTAACCTGCCTAGGACCCAGCACCCCACTTCCTCAGGGCTTGGACCCTCATCCTTCCTTTTTAT
      490       500       510       520       530       540       550       560       570       580       590       600

        D  V  Y  V  F  G  V  G  P  L  V  N  G  V  N  I  N  A  L  A  S  K  K  D  N  E  G  H  V  F  K  V  K  D  M  E  N  L
CCCTCAGATGTCTATGTGTTTGGGGTCGGGCCTTTGGTGAACCAAGTGAACATCAATGCTTTGGCTTCCAAGAAAGACAATGAGCAACATGTGTTCAAAGTCAAGGATATGGAAAACCTG
      610       620       630       640       650       660       670       680       690       700       710       720

  E  D  V  F  Y  G  M  I
GAAGATGTTTTCTACCAAATGATCGGTAGGGAGATACAAGGGAATAAAGAACACAACTCTCCTCAGGTTCCCCTGAAGTAATTCATTCTTCCTCTACACCTGAAGCTCTAGTTGCCTGGA
      730       740       750       760       770       780       790       800       810       820       830       840

AAGCCTTCTTCATTCCTCCTTCTCTACCTCAGTGTCACTATTCTTGTTTCCTGGCACTGTTCACTTAACCTTAGAATCACAGAGCTCTGAGCACTTCAGAGATCTTTCTACAGTCCTACA
      850       860       870       880       890       900       910       920       930       940       950       960

TTTGACACGTGGAAACAGAAGCCAAAGGAGGTCAAGGGACAGCAAGTTAGCAACAAGGGTGGGCTTGAAAACAGCCAGGCCTCTGACAGCTTGATCCCAAGTTCTTTCCCTTTTCAGTCC
      970       980       990       1000      1010      1020      1030      1040      1050      1060      1070      1080

ACCATAGCAGTTTTCTCCTAACACGAGGAAACAAATACCCGTGGTCTTTCCCTTTCTCCTTTTGGGCCTTTGCTCCCCATAGACTCCTACCCAAAGGCTGCTGCCATTTGGGAATGAAGT
      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200

                                  D  E  S  Q  S  L  S  L  C  G  M  V  W  E  H  R  K  G  T  D  Y  H  K  Q  P  W  Q  A
GTTCCGAGTTTTCAGCACATTCTCCTTCTCTGCCAGATGAAAGCCAGTCTCTGAGTCTCTGTGGCATGGTTTGGGAACACAGGAAGGGTACCGATTACCACAAGCAACCATGGCAGGCCA
      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320

K  I  S  V  I
AGATCTCAGTCATTGTAAGCACAGAATCCCAGTAGTGGGGACTTGGGGGAGGTGAGGTCAAGGTGAAATGGGAGTAGGGGAAGGAAAAAATGGCCATAAGAGATGGTGGTTTGTGAAAGT
      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440

                              R  P  S  K  G  H  E  S  C  M  G  A  V  V  S  E  Y  F  V  L  T  A  A  H  C  F  T  V  D
TGAGCTTTCCCTCTCTACTGTTGTGTCCCCAGCGCCCTTCAAAGGGACACGAGAGCTGTATGGGGGCTGTGGTGTCTGAGTACTTTGTGCTGACAGCAGCACATTGTTTCACTGTGGATG
      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1580

D  K  E  H  S  I  K  V  S  V
ACAAGGAACACTCAATCAAGGTCAGCGTAGGTAAGGATGCAACTGAAGGTCCTGGGCTGCACCTATGCTCTCCAGGCAACACCTCCCACTTTCTACAGATCCTACACTCCACCCATCCTC
      1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680
```



(Fig. 3 continues on the next page.)

isolation of three positive clones. Comparison of the patterns of restriction fragments showed that two of the three cosmids, which had inserts of ≈40 kb, were identical. The third cosmid had common restriction enzyme sites with the first two but appeared to have a 2-kb deletion in the internal 4.3-kb *Bam*HI fragment spanning the *Cla* I site (Fig. 3). Screening of ≈200,000 clones from a second cosmid library constructed by F. G. Grosveld resulted in the isolation of one further positive clone. This clone, which had an insert of ≈35 kb, extended the other three by 7 kb at the 3′ end. Thus, the total length of genomic DNA represented in the cosmids is ≈47 kb (Fig. 3). Southern blot analysis shows that the factor B cDNA probe hybridizes to restriction fragments derived from the middle of this region (Fig. 3). No other sequences corresponding to the FB1 cDNA probe were detected in this analysis, suggesting the presence of only one factor B gene.

In order to study the gene in greater detail, the 0.4-kb and 2.6-kb *Bgl* II fragments and the 11-kb *Sma* I fragment found to hybridize with the FB1 cDNA probe were subcloned and characterized by DNA sequence analysis. Various restriction enzyme digests of the subclones were used to generate fragments for sequence determination, and fragments labeled at both ends were either digested with another restriction enzyme or subjected to strand separation. In all cases, overlapping sequences were obtained for restriction sites used to generate sequence. This has allowed the almost complete structure of the *Bb* portion of the factor B gene to be determined. The 3.3 kb of continuous sequence presented in Fig. 3 comprises the entire 3′ end of the gene coding for amino acids 87–505 of Bb and the complete 3′-untranslated region. The sequence of the 3′-untranslated region agrees with that found by Woods *et al.* (14) in one of their cDNA clones except at positions 3,223 and 3,230. The locations of the intron/exon boundaries were determined from comparison of the possible translation products and the published amino acid sequence of Bb (6). In all cases, the nucleotide sequence at the 5′ end of an intron showed the splice

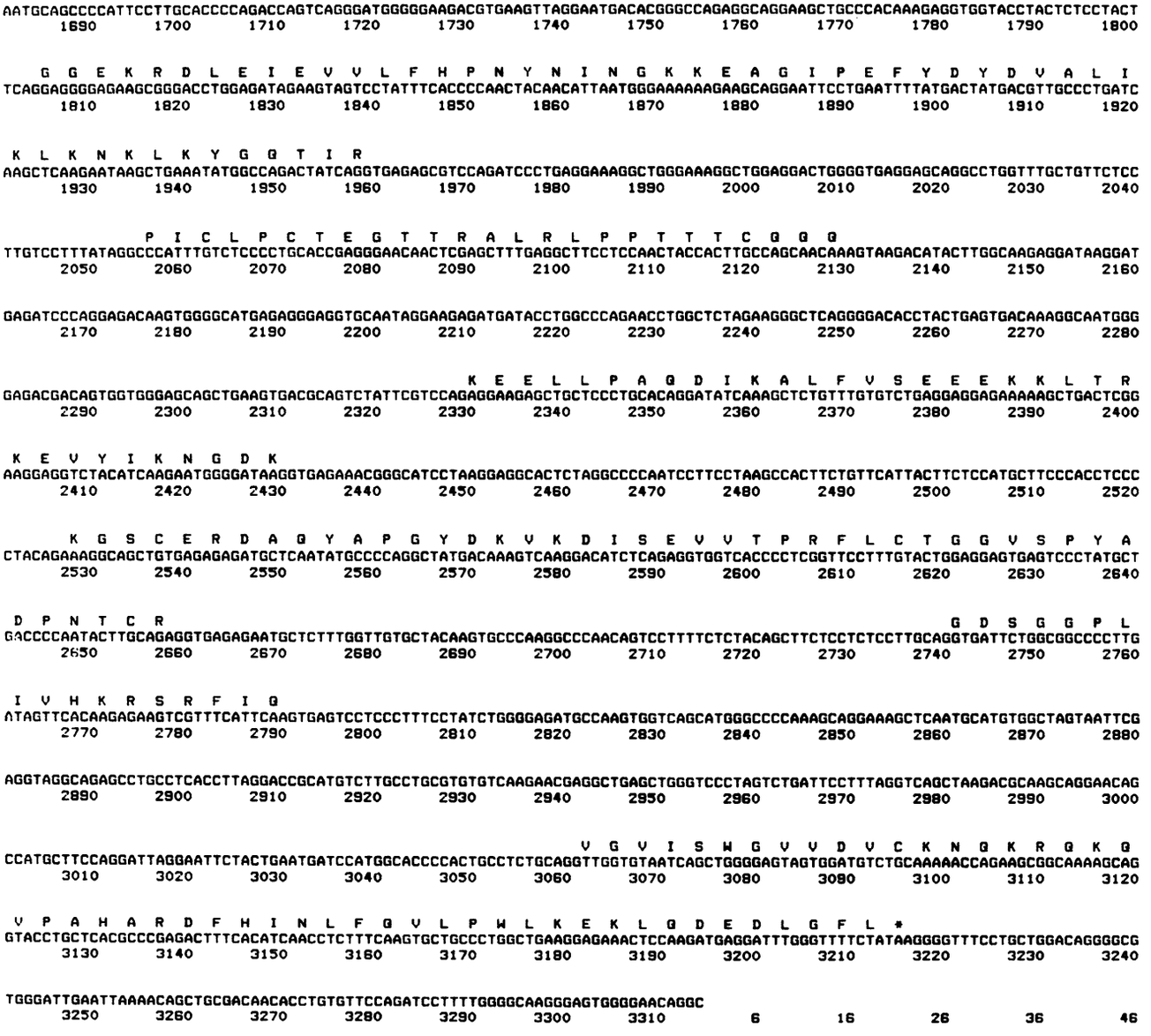


FIG. 3. (*Upper Left and Right*) Nucleotide sequence of 3.3 kb of DNA that encodes residues 87–505 of Bb and the 3′ untranslated region of the gene. (*Lower*) Partial restriction map of the genomic DNA represented in the cosmid clones. The map was prepared by restriction digests of the DNA and separation of the fragments on agarose gels. Southern blot analysis was carried out as described. The region of the DNA to which the FB1 cDNA probe hybridizes is indicated by a solid box. The expanded region of the map shows the 4.4-kb *Hin*dIII fragment that contains the *Bb* portion of the gene. The intron/exon organization was determined by DNA sequence analysis. The numbering for the amino acids that lie at either end of an exon is taken from the published sequence of Bb (6). Restriction enzyme sites: B, *Bgl* II; C, *Cla* I; E, *Eco*RI; F, *Hin*fI; H, *Hin*dIII; Hp, *Hpa* I; K, *Kpn* I; M, *Bam*HI; Sa, *Sal* I; S, *Sma* I; and X, *Xho* I.

signal G-T, and at the 3′ end, the splice signal A-G (29). No discrepancies were found between the amino acid sequence derived from the nucleotide sequence and the published amino acid sequence (6).

Sequence analysis around the *Hin*dIII site (Fig. 3) in the 2.6-kb *Bgl* II subclone has shown that this restriction enzyme site lies about 60 bp from the 3′ end of the exon coding for the COOH-terminal region of the Ba fragment (result not shown). Thus, the *Bb* portion of the factor B gene is contained in the 4.4-kb *Hin*dIII fragment (Fig. 3) and is about 4 kb in length. Because the *Hin*dIII site lies about 21 kb from the 5′ end of the cosmid DNA, it is likely that the complete gene is represented in the DNA isolated.

Structure of the Gene. Factor B is an unusual serine proteinase with a catalytic chain of approximately twice the size of other serine proteinases (6). The 3.3 kb of DNA sequence obtained (Fig. 3) encodes 11 exons corresponding to amino acids 87–505 of Bb. The first three exons are of similar size coding for 44, 34, and 46 amino acids, respectively, which lie in the NH_2-terminal half of Bb. The fourth exon, encoding 33 amino acids, specifies a region of the molecule that, in other serine proteinases, contains the peptide bond that is cleaved during activation.

Residues 222–505 of Bb correspond to that portion of the molecule homologous to the catalytic chain of other serine proteinases. The structure of this region of the gene, which is 2 kb in length, is shown schematically in Fig. 4. It is split into eight exons separated by introns of various lengths and with the codon for residue 222 lying 33 bases from the 5′ end of the first exon. The catalytically important residues histidine-267, aspartic acid-317, and serine-440 are found in separate exons (Fig. 4). The exon specifying the primary binding site, residues 438–443 (Gly-Asp-Ser-Gly-Gly-Pro), and containing serine-440 is only 50 bp long. This exon is separated by a 270-nucleotide-long intron from the exon containing the secondary binding site corresponding to residues 459–461 (Ser-Trp-Gly) and by an 81-nucleotide-long intron from the exon that contains the amino acid residue determining specificity in the binding pocket, probably aspartic acid-432 (6) (Fig. 4).

Each functionally important part of the active site is contained in a separate exon, and, because none are functionally independent, they must have been conserved in a definite relationship to each other throughout evolution. This also appears to be the case with chymotrypsinogen and trypsinogen, as the three catalytically important residues His, Asp, and Ser and the amino acid conferring specificity at the active site are all contained in separate exons (30). It has been suggested that the intron/exon structure of a gene may be a mechanism for increasing the rate of evolution and that recombination within introns could reassort protein functions to produce novel proteins from parts of existing ones (31, 32). If the functionally important parts of the catalytic sites of other serine proteinases are also found to lie on separate exons, such an arrangement may have facilitated the interchange of specificity-determining sequences and played a role in the emergence of the many serine proteinases which exist.

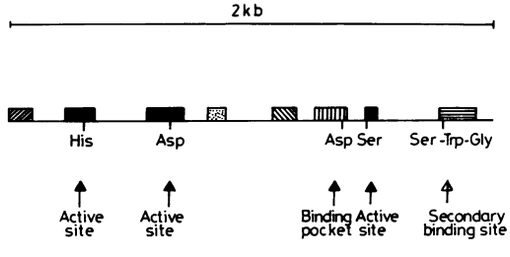


FIG. 4. Organization of exons in the serine proteinase domain of *Bb*. Residues 222–505 of Bb correspond to the catalytic chain of other serine proteinases (6). The region of the gene encoding these amino acids is 2 kb in length and comprises eight exons, with the codon for residue 222 lying 33 bp from the 5′ end of the first exon. Exons that contain the catalytically important amino acids histidine-267, aspartic acid-317, and serine-440 are shown, as is the exon that contains the amino acid conferring specificity in the binding pocket and the exon containing the secondary binding site.

We thank Professor G. G. Brownlee for generous provision of facilities and for his help and advice throughout this work. We also thank Dr. M. C. Carroll for providing the human liver cDNA library; Dr. F. G. Grosveld for providing the cosmid libraries; Mr. Keith Gould for help in synthesis of oligonucleotides; Drs. C. Blake, J. Gagnon, M. C. Carroll, and D. H. Bentley for helpful discussion; and Miss N. J. Janjua for excellent technical assistance.

1. Barnstaple, C. J., Jones, E. A. & Bodmer, W. F. (1979) in *Defense and Recognition II B, Cellular Aspects*, MTP International Review of Biochemistry, ed. Lennox, E. S. (University Park Press, Baltimore), Vol. 22, pp. 151–224.
2. Alper, C. A. (1981) in *The Role of the Major Histocompatibility Complex in Immunobiology*, ed. Dorf, M. E. (Garland, New York), pp. 173–220.
3. Reid, K. B. M. & Porter, R. R. (1981) *Annu. Rev. Biochem.* **50**, 433–464.
4. Kerr, M. A. (1979) *Biochem. J.* **183**, 615–622.
5. Medicus, R. G., Götze, O. & Müller-Eberhard, H. J. (1976) *Scand. J. Immunol.* **5**, 1049–1055.
6. Christie, D. L. & Gagnon, J. (1983) *Biochem. J.* **209**, 61–70.
7. Alper, C. A., Boenisch, T. & Watson, L. (1972) *J. Exp. Med.* **135**, 68–80.
8. Ploegh, H. L., Orr, H. T. & Strominger, J. L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6081–6085.
9. Sood, A. K., Pereira, D. & Weissman, S. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 616–620.
10. Malissen, M., Malissen, B. & Jordan, B. R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 893–897.
11. Lee, J. S., Trowsdale, J. & Bodmer, W. F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 545–549.
12. Lee, J. S., Trowsdale, J., Travers, P. J., Carey, J., Grosveld, F., Jenkins, J. & Bodmer, W. F. (1982) *Nature (London)* **299**, 750–752.
13. Korman, A. J., Auffrey, C., Schamboeck, A. & Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6013–6017.
14. Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colten, H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5661–5665.
15. Duckworth, M. L., Gait, M. J., Goelet, P., Hong, G. F., Singh, M. & Titmas, R. C. (1981) *Nucleic Acids Res.* **9**, 1691–1706.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253**, 2483–2495.
18. Carroll, M. C. & Porter, R. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 264–267.
19. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6613–6617.
20. Gergen, J. P., Stern, R. H. & Websink, P. C. (1979) *Nucleic Acids Res.* **7**, 2115–2136.
21. Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10**, 6715–6732.
22. Grosveld, F. G., Dahl, H. H. M., de Boer, E. & Flavell, R. A. (1981) *Gene* **13**, 227–237.
23. Bernards, R. & Flavell, R. A. (1980) *Nucleic Acids Res.* **8**, 1521–1534.
24. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
25. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
26. Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683–3687.
27. Agarwal, K. L., Brunstedt, J. & Noyes, B. E. (1981) *J. Biol. Chem.* **256**, 1023–1028.
28. Choo, K. H., Gould, K. G., Rees, D. J. G. & Brownlee, G. G. (1982) *Nature (London)* **299**, 178–180.
29. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
30. Craik, C. S., Sprang, S., Fletterick, R. & Rutter, W. J. (1982) *Nature (London)* **299**, 180–182.
31. Gilbert, W. (1978) *Nature (London)* **271**, 501.
32. Blake, C. (1978) *Nature (London)* **273**, 267.