*Biochemical J.* (1983) **209**, 61–70
*Printed in Great Britain*



# Amino acid sequence of the Bb fragment from complement Factor B

## Sequence of the major cyanogen bromide-cleavage peptide (CB-II) and completion of the sequence of the Bb fragment

David L. CHRISTIE and Jean GAGNON
*M.R.C. Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.*

(*Received 7 April 1982/Accepted 30 September 1982*)

The amino acid sequence of peptide CB-II, the major product (mol.wt. 30000) of CNBr cleavage of fragment Bb from human complement Factor B, is given. The sequence was obtained from peptides derived by trypsin cleavage of peptide CB-II and clostripain digestion of fragment Bb. Cleavage of two Asn–Gly bonds in peptide CB-II was also found useful. These results, along with those presented in the preceding paper [Gagnon & Christie (1983) *Biochem. J.* **209**, 51–60], yield the complete sequence of the 505 amino acid residues of fragment Bb. The *C*-terminal half of the molecule shows strong homology of sequence with serine proteinases. Factor B has a catalytic chain (fragment Bb) with a molecular weight twice that of proteinases previously described, suggesting that it is a novel type of serine proteinase, probably with a different activation mechanism.

The preceding paper (Gagnon & Christie, 1983) described the alignment of the CNBr-cleavage peptides from the Bb fragment of Factor B. This provided the sequence of 252 residues from the *N*-terminus of the Bb fragment, leaving only the sequence of the *C*-terminal CNBr-cleavage peptide CB-II (mol.wt. approx. 30000) to be completed.

Partial sequence studies have previously shown that the CB-II portion of Factor B is homologous to serine proteinases (Christie *et al.*, 1980; Mole & Niemann, 1980). The complete sequence of peptide CB-II is described in the present paper, and this provides the complete sequence of the active-site-containing Bb fragment from human complement Factor B.

The complement proteins Factor B, component C2 and component C4 are each coded for by genes within the major histocompatibility locus (Barnstaple *et al.*, 1979). Factor B is the first of these proteins for which there are extensive sequence data available.

## Materials and methods

### *Materials*

Materials were as described previously (Christie & Gagnon, 1982) unless otherwise stated. Hydroxylamine hydrochloride was obtained from Cambrian Chemicals, Croydon, Surrey, U.K. LiChrospher SI 500 (10 $\mu$m, 50nm pore size) was from Merck and supplied by BDH Chemicals, Poole, Dorset, U.K. n-Octyltrichlorosilane and trimethylchlorosilane were from Silar Laboratories and supplied by Field Instrument Co., Surbiton, Surrey, U.K. Carboxypeptidase Y (20 units/mg) was from Boehringer Mannheim, Lewes, E. Sussex, U.K.

Abbreviations used: h.p.l.c., high-pressure liquid chromatography. The nomenclature of complement components is that recommended by the World Health Organisation (1968).

### *General methods*

Unless stated otherwise, all procedures were as described previously (Christie & Gagnon, 1982).

### *Preparation of peptides Bb,Cl-4-5 and Bb,Cl-5-9*

These peptides were isolated from the digest of 43mg of the Bb fragment by clostripain and separated by gel filtration on a Sephadex G-75 column (2cm × 90cm) in a manner identical with that described in the preceding paper (Gagnon & Christie, 1983). Fractions corresponding to elution volumes of 208–232ml and 244–270ml were designated Bb,Cl-4 and Bb,Cl-5 respectively. Peptide Bb,Cl-4-5 and peptide Bb,Cl-5-9 were subsequently isolated from the Bb,Cl-4 and Bb,Cl-5 pools respectively by h.p.l.c. on a $\mu$-Bondapak $C_{18}$ column with the 0.1% $NH_4HCO_3$/acetonitrile solvent system

### *Preparation of hydroxylamine-cleavage peptides from peptide CB-II*

Hydroxylamine cleavage was performed as



0306-3275/83/010061-10$02.00 © 1983 The Biochemical Society

described by Balian *et al.* (1971). Guanidinium chloride (5.73g) and hydroxylamine hydrochloride (1.39g) were weighed into a 20ml vial and dissolved in water (2–3ml) with vigorous stirring in an ice bath. The pH was adjusted to 6–7 with 12.5M-NaOH, and 1.0M-$K_2CO_3$ (2ml) was added. Then the solution was adjusted to pH9.0, diluted to 10ml with water and warmed to 45°C. Fully reduced and *S*-carboxymethylated peptide CB-II (240nmol) was dissolved by the addition of 5ml of the hydroxylamine reagent (6M-guanidinium chloride/2.0M-hydroxylamine hydrochloride/0.2M-$K_2CO_3$, pH9.0), and the mixture was incubated for 4h at 45°C. Then the reaction was stopped by lowering the pH to 3.0 with acetic acid. Vigorous stirring and the addition of propan-1-ol (1ml) was required to prevent excessive foaming. The digest was dialysed extensively against 10% (v/v) acetic acid and then freeze-dried. The digest was redissolved in 10% (v/v) acetic acid/10% (v/v) propan-1-ol (1.1ml), and portions (0.1ml) were injected on to an LKB Ultropac TSK G-3000 SW gel-permeation column (7.5mm × 600mm) equilibrated with 10% acetic acid/10% propan-1-ol and pumped at a flow rate of 0.7ml/min. Peptides were further purified by reversed-phase h.p.l.c. on a $C_8$ 50nm-pore-size column as described in the Results section. The $C_8$ column was prepared with LiChrospher SI 500 (10μm, 50nm pore size) by the procedure of Lewis *et al.* (1980). The silica particles to which octyl groups were covalently bound were packed into a column (3.9mm × 30mm) by courtesy of Shandon Southern Products, Runcorn, Cheshire, U.K., with a Shandon column-packing instrument.

*Preparation of peptide CB-II,SP-5-2*

Non-reduced peptide CB-II (8.8mg) in 50mM-ammonium acetate/2mM-EDTA, pH4.0, was incubated with V8 proteinase (0.18mg) for 18h at 37°C and the digest was freeze-dried. The digest was redissolved in 0.1M-$NH_4HCO_3$ (1.0ml) and centrifuged, and the pellet was extracted twice with 0.1M-$NH_4HCO_3$ (0.5ml). The combined supernatants were applied to a column (2cm × 90cm) of Sephadex G-50 (superfine grade) equilibrated with 0.1M-$NH_4HCO_3$. The material eluted between 248ml and 260ml was designated CB-II,SP-5. This material was freeze-dried and further purified by h.p.l.c. on a μ-Bondapak $C_{18}$ column equilibrated with 0.1% trifluoroacetic acid/solvent B (95:5, v/v); solvent B was methanol/acetonitrile/propan-2-ol (1:1:1, by vol.). A linear gradient of 60min was applied to the column to give a final solvent composition 0.1% trifluoroacetic acid/solvent B (50:50, v/v). Peptide CB-II,SP-5-2 was eluted 10min from the time of injection.

*Carboxypeptidase-Y digest of tryptic peptide Bb,T-3-9b*

Peptide Bb,T-3-9b (40nmol) was dissolved in 50mM-pyridine/acetate buffer, pH5.5 (0.3ml), containing norleucine (150nmol). Carboxypeptidase Y (25μg) was added and the sample was incubated at 37°C. Portions (25μl) were taken at appropriate time intervals and placed in a boiling-water bath for 2min. These samples were subsequently freeze-dried after addition of 0.4ml of water before amino acid analysis. Amino acid analyses were performed with a Durrum D500 amino acid analyser and results were corrected from the recovery of norleucine. No amino acids were detected from control incubations in which carboxypeptidase Y was incubated alone for 60min at 37°C.

## Results

*Sequence of the C-terminal CNBr-cleavage peptide CB-II from the Bb fragment of Factor B*

Partial sequence studies have previously shown that the CB-II portion of Factor B is homologous to serine proteinases (Christie *et al.*, 1980). The *N*-terminal 78 residues were established by extending the *N*-terminal sequence of peptide CB-II with that of an overlapping peptide CB-II,T-1 derived from a tryptic digest of succinoylated peptide CB-II. The amino acid at position 278 was found to be isoleucine, instead of lysine as previously reported (Christie & Gagnon, 1982). The sequences of other tryptic peptides from succinoylated CB-II, namely CB-II,T-5a, CB-II,T-5b, CB-II,T-4 and CB-II,T-6c, were aligned from the sequence of a peptide Bb, AC-4, which was derived from cleavage of the Bb fragment with dilute acid under conditions selective for the cleavage of Asp–Pro bonds (Christie *et al.*, 1980) This provided the sequence of a segment of 60 residues near the *C*-terminus of the molecule. The complete sequence of peptide CB-II has now been established as shown in Fig. 1. The residues are numbered according to their position in the Bb sequence.

A key factor in establishing the complete sequence of peptide CB-II has been the isolation and sequence of two peptides generated by the cleavage of Asn–Gly bonds in peptide CB-II with hydroxylamine. This enabled the alignment of peptides derived from cleavage at arginine residues by trypsin and clostripain. The sequence of the CNBr-cleavage peptide CB-II completes the sequence of the active-site-containing Bb fragment from Factor B.

*Sequence analysis of the tryptic peptides from the Bb fragment*

Details of the isolation of the tryptic peptides from the succinoylated Bb fragment were described in the

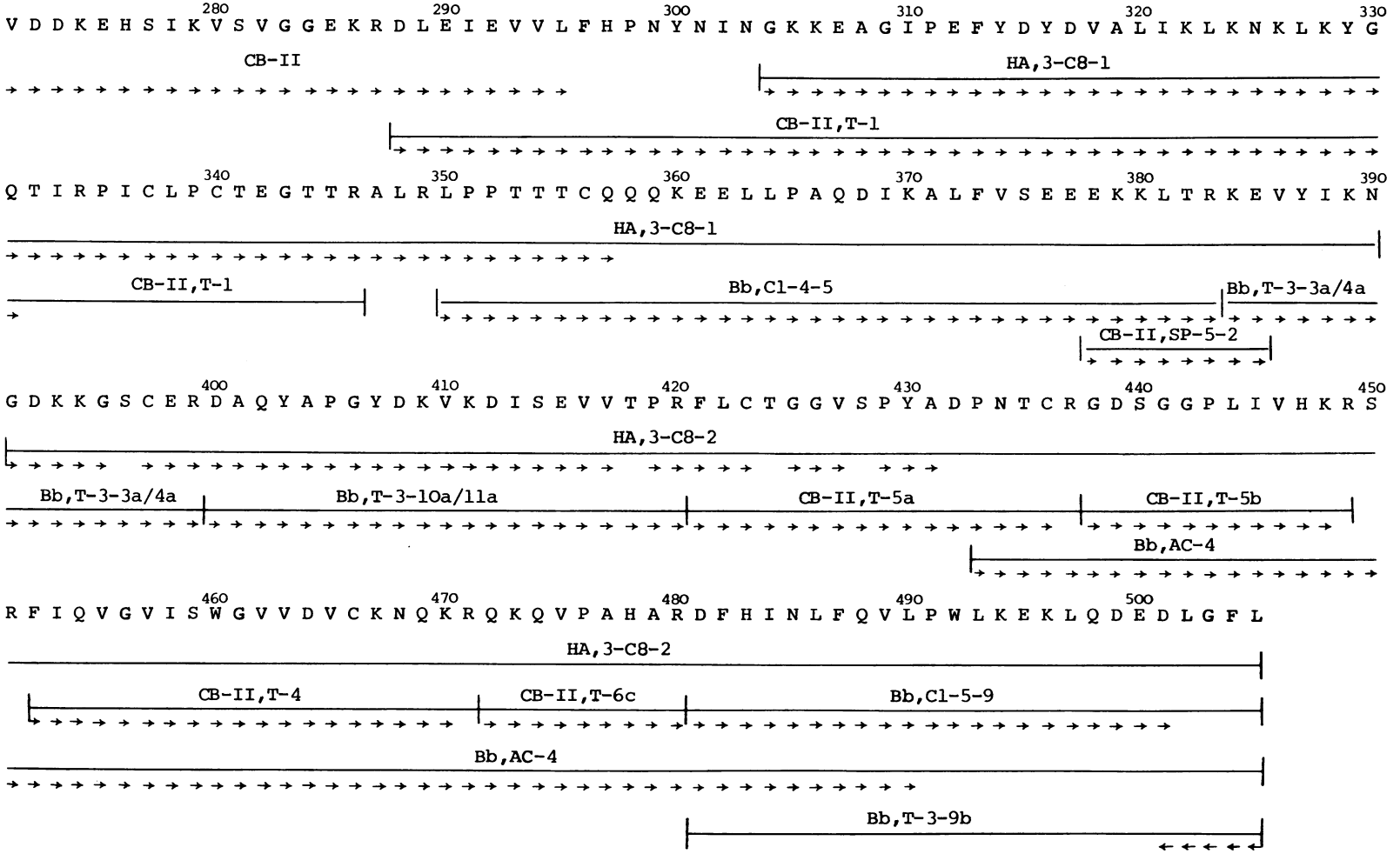


Fig. 1. *Complete sequence of the C-terminal CNBr-cleavage peptide CB-II*

Residues are numbered according to their position in the Bb sequence. The *N*-terminal sequences of peptides CB-II, CB-II,T-1, CB-II,T-5a, CB-II,T-5b. CB-II,T-4 and Bb,AC-4 have been presented previously (Christie *et al.*, 1980). This provided the *N*-terminal 78 residues and the sequence of a segment of 60 residues from the *C*-terminal portion of the molecule. Residues identified by automated Edman degradation or after carboxypeptidase-Y digestion are indicated → and ← respectively. The following peptide nomenclature was used: Ha, hydroxylamine-cleavage peptide; CB, CNBr-cleavage peptide; AC, dilute-acid-cleavage peptide; T, tryptic-digest peptide; Cl, clostripain-digest peptide; Sp, V8-proteinase-digest peptide. Amino acids are in the single-letter notation: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Table 1. *Amino acid compositions and details of automated Edman degradation of peptides used to complete the sequence of peptide CB-II*

Compositions are given as mol of residues/mol of peptide and were calculated from 24h-hydrolysis values. Values less than 0.3 have been omitted. Values for any particular peptide are normalized with respect to the residue marked with an asterisk (*). Cysteine was determined as *S*-carboxymethylcysteine. Abbreviation: N.D., not determined.

| | Amino acid composition (mol of residue/mol) | | | | | | |
|---|---|---|---|---|---|---|---|
| Peptide ... | Bb,T-3-3a/4a | Bb,T-3-10a/11a | Bb,Cl-4-5 | Bb,Cl-5-9 | HA,3-C8-1 | HA,3-C8-2 | CB-II,SP-5-2 |
| Asx | 2.0 | 3.1 | 1.4 | 4.1 | 5.4 | 12.6 | — |
| Thr | — | 0.9 | 3.2 | — | 7.2 | 3.5 | 1.0 |
| Ser | 0.8 | 0.8 | 1.1 | — | 1.7 | 5.3 | — |
| Glx | 2.0 | 2.1 | 9.4 | 3.8 | 15.0 | 12.8 | 2.6 |
| Pro | — | 1.8 | 3.1 | 0.9 | 7.0 | 8.4 | — |
| Gly | 2.1 | 1.1 | — | 1.3 | 4.7 | 11.0 | — |
| Ala | — | 2.0 | 1.8 | — | 5.0 | 5.6 | — |
| Cys | 1.0 | — | 1.0* | — | 3.0 | 3.7 | — |
| Val | 1.0 | 2.9 | 1.1 | 1.4 | 3.8 | 10.7 | — |
| Met | — | — | — | — | — | — | — |
| Ile | 1.0 | 1.0 | 1.2 | 1.0* | 5.9 | 4.8 | — |
| Leu | — | — | 4.8 | 5.9 | 9.5 | 8.5 | 1.3 |
| Tyr | 1.0 | 1.9 | — | — | 3.9 | 3.1 | — |
| Phe | — | — | 0.9 | 2.7 | 2.0 | 5.0 | — |
| His | — | — | — | 1.0 | — | 3.1 | — |
| Lys | 3.7 | 2.0 | 3.7 | — | 11.4 | 10.0* | 2.9 |
| Arg | 1.0* | 1.0* | N.D. | 2.1 | 4.0* | 6.3 | 1.0* |
| Total residues | 16 | 21 | 33 | 24 | 90 | 114 | 8 |
| Yield of peptide (%) | 53 | 46 | 18 | 26 | 11 | 11 | 22 |
| Amount used in automated sequence analysis (nmol) | 70 | 70 | 138 | 196 | 26 | 30 | 55 |
| Initial yield (%) | 69 | 66 | 40 | 55 | 81 | 87 | 87 |
| Repetitive yield (%) | 90 | 87 | 92 | 93 | 93 | 93 | 86 |

preceding paper (Gagnon & Christie, 1983). The amino acid compositions of the peptides sequenced are presented in Table 1.

*Peptide Bb,T-3-10a/11a.* The entire sequence of 16 residues of peptide Bb,T-3-3a/4a was established from automated sequence analysis of 140nmol of peptide (Fig. 1 and Table 1).

*Peptide Bb,T-3-10a/11a.* The entire sequence of 21 residues of this peptide was established from automated sequence analysis of 73nmol of peptide (Fig. 1 and Table 1).

### *Sequence analysis of peptides from a clostripain digest of the Bb fragment*

*Peptide Bb,Cl-4-5.* The entire sequence of 34 residues was determined by automated sequence analysis (Fig. 1 and Table 1).

*Peptide Bb,Cl-5-9.* Amino acid analysis of peptide Bb, Cl-5-9 showed it to contain no arginine, indicating that it is the *C*-terminal peptide of the Bb fragment. The *N*-terminal 21 of the 25 residues of this peptide were determined by automated sequence analysis (Fig. 1 and Table 1). Although in a previous publication (Christie *et al.*, 1980) the sequence of peptide Bb,AC-4 was shown only up to the arginine residue corresponding to position 480 of the Bb sequence (Fig. 1), automated Edman degradation had been continued past this point. The additional sequence of peptide Bb,Ac-4 was identical with the *N*-terminal sequence of peptide Bb,Cl-5-9, which provided a continuous sequence corresponding to positions 421–501 from the *C*-terminal sequence of the Bb fragment (Fig. 1).

### *Isolation of hydroxylamine-cleavage peptides from peptide CB-II*

Peptide CB-II (240nmol) was treated with hydroxylamine as described in the Materials and methods section and subjected to initial fractionation by high-pressure gel-permeation chromatography (Fig. 2*a*). Three main peaks of material were obtained, and each was characterized by sodium dodecyl sulphate / polyacrylamide - gel electrophoresis. The first two peaks contained products of incomplete cleavage, and the third peak, which was pooled as indicated, contained two peptides with

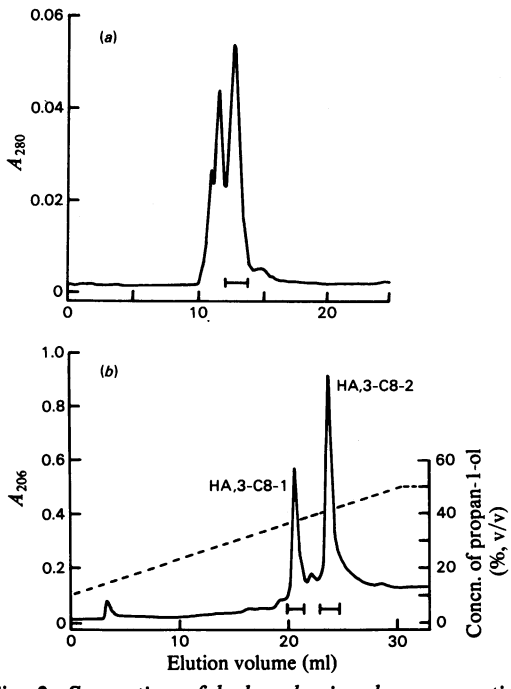


Fig. 2. *Separation of hydroxylamine-cleavage peptides from peptide CB-II*

(*a*) The freeze-dried hydroxylamine digest of peptide CB-II (240nmol) was dissolved in 10% acetic acid/10% propan-1-ol (1.1ml), and portions (0.1ml) were injected on to an LKB Ultropac TSK G-3000 SW gel-permeation column equilibrated with 10% acetic-acid/10% propan-1-ol and pumped at 1ml/min. The third peak was pooled as indicated and freeze-dried. ——, $A_{280}$. (b) The freeze-dried material from (*a*) was dissolved in 0.1% trifluoroacetic acid/10% propan-1-ol (3.3ml) and portions (0.4ml) were injected on to a $C_8$ 50nm-pore-size h.p.l.c. column (3.9mm × 300mm). A linear gradient of 60min was pumped to the column at 0.5ml/min to give a final solvent composition of 0.1% trifluoroacetic acid/50% propan-1-ol. Peptides HA,3-C8-1 and HA,3-C8-2 were pooled as indicated. ——, $A_{280}$; – – – –, concn. of propan-1-ol.

molecular weights approx. 12000 and 10000 respectively. After being freeze-dried, this material was redissolved in 0.1% trifluoroacetic acid/10% propan-1-ol (3.3ml), and portions (0.4 to 1.4ml) were injected on to a $C_8$ 50nm-pore-size h.p.l.c. column (3.9mm × 300mm) equilibrated with 0.1% trifluoroacetic acid/10% propan-1-ol (Fig. 2*b*). A linear gradient of 60min was pumped to the column at 0.5ml/min to give a final solvent composition of 0.1% trifluoroacetic acid/50% propan-1-ol. The peptides HA,3-C8-1 and HA,3-C8-2 were eluted 41min and 47min from the time of injection respectively. The amino acid compositions and yields of these peptides are presented in Table 1.

### *Automated sequence analysis of hydroxylamine-cleavage peptides from peptide CB-II*

*Peptide HA,3-C8-1.* The *N*-terminal 54 residues of this peptide were determined by automated sequence analysis (Fig. 1 and Table 1). It is clear that this peptide results from the cleavage of an Asn–Gly bond corresponding to position 303 in the Bb sequence. The sequence of peptide HA,3-C8-1 provided an overlap between the *C*-terminal sequence of peptide CB-II,T-1 and the *N*-terminal sequence of peptide Bb,Cl-4-5 (Fig. 1). This extended the *N*-terminal sequence of the Bb fragment to 383 residues. The sequence of residues 318–342 was confirmed by the sequence of a peptide Bb,T-1b,SP-6, isolated from a V8-proteinase digest of peptide Bb,T-1b (results not shown).

*Peptide HA,3-C8-2.* Automated sequence analysis of peptide HA,3-C8-2 enabled identification of 37 of the *N*-terminal 41 residues (Fig. 1 and Table 1). The sequence provided an overlap with the sequence of peptides Bb,T-3-3a/4a, Bb,T-3-10a/11a and CB-II,T-5a, and thus the 60-residue segment corresponding to residue 421–480 of the Bb fragment that had been published previously (Christie *et al.*, 1980). This provided a continuous segment of 118 residues corresponding to positions 384–501 of the Bb fragment (Fig. 1). There were four gaps in the sequence of peptide HA,3-C8-2, corresponding to positions 396, 418, 424 and 428 of the Bb fragment. The gap at position 396 resulted from the complete loss of a sample due to a cracked centrifuge tube. A serine residue was assigned to this position from the sequence of peptide Bb,T-3-3a/4a. A threonine residue was assigned to position 418 from the sequence of peptide Bb,T-3-10a/11a. Similarly, threonine and serine residues were assigned to positions 424 and 428 of the Bb fragment respectively from the sequence of peptide CB-II-T-5a.

### *Alignment of the N-terminal 383 residues with residues 384–501 of the Bb fragment*

The entire sequence of eight residues of a peptide CB-II,SP-5-2, which was isolated from a V8-proteinase digest of non-reduced peptide CB-II as described in the Materials and methods section, was determined by automated sequence analysis (Fig. 1 and Table 1). The sequence Glu-Lys-Lys-Leu-Thr-Arg-Lys-Glu provides an overlap between peptide Bb,Cl-4.5 and peptide Bb,T-3-3a/4a. This sequence is sufficient to align residues 1–383 and 384–501 of the Bb fragment, as peptide Bb,T-3-3a/4a is the only tryptic peptide with an *N*-terminal sequence commencing Lys-Glu- (Fig. 1). This establishes a continuous sequence corresponding to residues 1–501 of the Bb fragment.

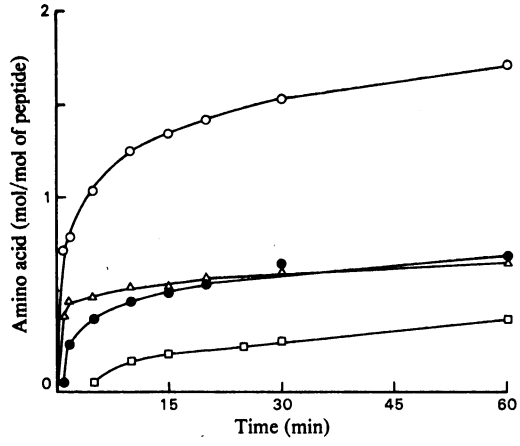


Fig. 3. *Carboxypeptidase-Y digestion of the Bb fragment from Factor B*

Peptide Bb,T-3-9b (40nmol) in 50mM-pyridine/acetate buffer, pH5.5 (0.3ml), was digested with carboxypeptidase Y (25$\mu$g) at 37°C. Samples (25$\mu$l) were taken at various time intervals, freeze-dried and subjected to amino acid analysis. Norleucine was used as an internal standard. Symbols: ○, leucine; △, phenylalanine; ●, glycine; □, aspartic acid.

*Carboxypeptidase-Y digest of the C-terminal tryptic peptide from the Bb fragment*

Comparison of the *N*-terminal sequence of the *C*-terminal peptide Bb,Cl-5-9 with its amino acid composition indicates that the unidentified sequence has the composition $Gly_1$,$Phe_1$,$Leu_2$. The hydrophobic nature of these residues prevents the identification of the remaining sequence by automated Edman degradation owing to the wash-out of peptide from the sequencer cup. A peptide identical with Bb,Cl-5-9, namely peptide Bb,T-3-9b, was isolated from the tryptic digest of the succinoylated Bb fragment as described in the preceding paper (Gagnon & Christie, 1983). This peptide was digested with carboxypeptidase Y, and the results are presented in Fig. 3. The kinetics suggest a *C*-terminal sequence Asp-Leu-Gly-Phe-Leu. A *C*-terminal sequence for the Bb fragment has been reported as -Phe-Leu (Lesavre *et al.*, 1979; Christie *et al.*, 1980). The *C*-terminal sequence of peptide Bb,T-3-9b provides an overlap of one residue with the *N*-terminal sequence of peptide Bb,Cl-5-9, and accounts for all the residues unidentified after *N*-terminal sequence analysis. Furthermore, a peptide with the amino acid composition $Asp_1$,$Gly_1$,$Leu_2$,$Phe_1$, isolated from a V8-proteinase digest of peptide CB-II, gave the sequence Asp-Leu-Gly. This completes the sequence of peptide CB-II and thus the entire sequence of the Bb fragment of 505 residues.

*Complete sequence of the Bb fragment of Factor B*

The amino acid sequence and the amino acid composition based on the sequence of the Bb fragment are presented in Fig. 4. The polypeptide portion of the Bb fragment has a calculated molecular weight of 57000. The amino acid sequence is derived from the present work and that reported by Christie *et al.* (1980) and Christie & Gagnon (1982). The sequence of 22 *N*-terminal and 36 *C*-terminal amino acid residues of the smaller Ba fragment have also been described elsewhere (Christie & Gagnon, 1982).

## Discussion

The complete sequence of 505 amino acid residues of the Bb fragment from Factor B has been determined. The amino acid composition based on the sequence is in good agreement with results of amino acid analysis (Christie & Gagnon, 1982). The polypeptide portion of the Bb fragment has a calculated molecular weight of 57000. This compares with a value of approx. 60000 obtained by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis (Kerr, 1979; Lesavre *et al.*, 1979). The difference between these two values provides an estimate of the contribution of asparagine-linked carbohydrate at positions 26 and 119 of the Bb fragment (Christie & Gagnon, 1982; Gagnon & Christie, 1983).

A key factor in estimating the complete sequence of large *C*-terminal peptide CB-II was the isolation and sequence of two peptides derived from cleavage of Asn–Gly bonds in peptide CB-II with hydroxylamine. The combination of high-pressure gel-permeation chromatography followed by reverse-phase chromatography on a $C_8$ 50nm-pore-size column is well suited to the task of isolating relatively small amounts of large denatured peptides after chemical cleavage procedures. An overall yield of 11% for the two peptides was considered acceptable, taking into account the incomplete cleavage of Asn–Gly bonds by hydroxylamine. which is generally in the range 30–80% (Bornstein & Balian, 1977). In the present study the yield of cleavage appears to be about 50% (Fig. 2*a*).

Partial amino acid sequence studies have previously shown Factor B to be an unusual type of serine proteinase of approximately twice the size of other known serine proteinases (Christie *et al.*, 1980; Mole & Niemann, 1980). Residues 222–505 of the Bb fragment are compared with both bovine prothrombin and bovine chymotrypsinogen A in Fig. 5 in accordance with the alignment reported by Young *et al.* (1978). Prothrombin was chosen, as this is the longest serine proteinase characterized thus far. The numbering of such alignments is normally based on the sequence of chymotryp-

```
                10                  20                  30                  40                  50                  60
K I V L D P S G S M N I Y L V L D G S D S I G A S(N)F T G A K K C L V N L T E K V A S Y G V K P R Y G L V T Y A T Y P K

                70                  80                  90                  100                 110                 120
I W V K V S E A D S S N A D W V T K Q L N E I N Y E 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

                130                 140                 150                 160                 170                 180
T R H V I I L M T D G L H N M G G D P I T V I D E I R D L L Y I G K D R K N P R E D Y L D V Y V F G V G P L V N Q V N I

                190                 200                 210                 220                 230                 240
N A L A S K K D N E Q H V F K V K D M E N L E D V F Y Q M I 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 K I

                250                 260                 270                 280                 290                 300
S V I 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 D K E H S I K V S V G G E K R D L E I E V V L F H P N Y

                310                 320                 330                 340                 350                 360
N I N G K K E A G I P E F Y D Y D V A L I K L K N K L K Y G Q T I R P I C L P C T E G T T R A L R L P P T T T C Q Q Q K

                370                 380                 390                 400                 410                 420
E E L L P A Q D I K A L F V S E E E K K L T R K E V Y I K N G D K K G S C E R D A Q Y A P G Y D K V K D I S E V V T P R

                430                 440                 450                 460                 470                 480
F L C T G G V S P Y A D P N T C R G D S G G P L I V H K R S R F I Q V G V I S W G V V D V C K N Q K R Q K Q V P A H A R

                490                 500
D F H I N L F Q V L P W L K E K L Q D E D L G F L
```

Fig. 4. *Complete sequence of fragment Bb from human complement Factor B*

The total number of residues is 505, and the molecular weight of the unmodified polypeptide chain is 57000. The amino acid composition is $Ala_{24}$,$Arg_{20}$,$Asn_{23}$,$Asp_{36}$,$Cys_{11}$,$Gln_{20}$,$Glu_{31}$,$Gly_{35}$,$His_{13}$,$Ile_{29}$,$Leu_{40}$,$Lys_{47}$,$Met_{9}$,$Phe_{14}$,$Pro_{25}$, $Ser_{29}$,$Thr_{24}$$Trp_{8}$,$Tyr_{21}$,$Val_{46}$. Amino acids are in the single-letter notation indicated in Fig. 1 legend. (N) indicates a likely carbohydrate-attachment site. A free thiol group is located at position 33 of the sequence.

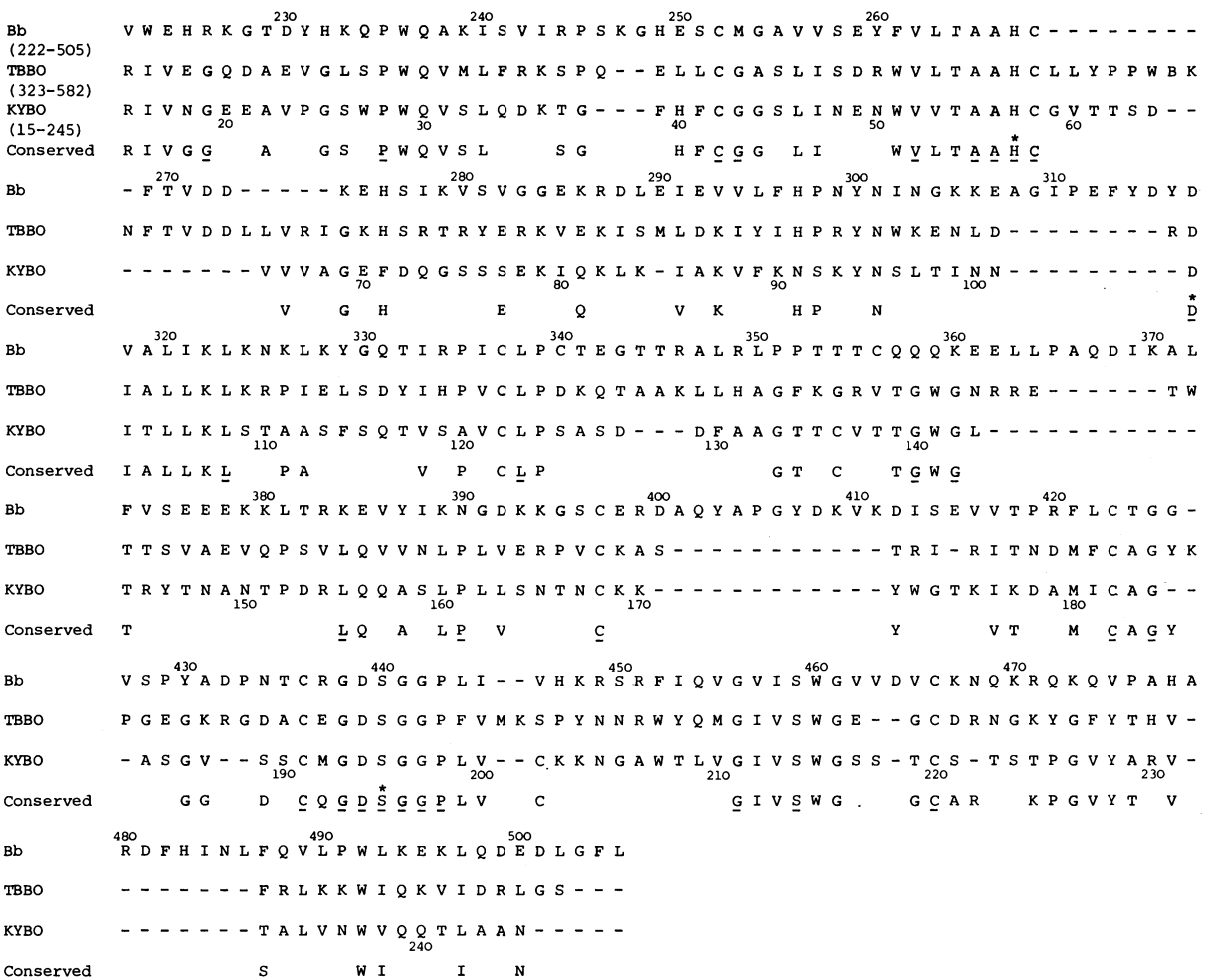


Fig. 5. *Alignment of the serine proteinase portion of the human Factor B with those of bovine prothrombin and chymotrypsinogen*

Sequence data and alignment for prothrombin (TBBO) and chymotrypsinogen (KYBO) are from Young *et al.* (1978), but with the insertion of four blocks of gaps in the prothrombin and chymotrypsinogen sequences to accommodate the Bb sequence, between positions 308–315, 365–370, 401–411 and 479–486. To maximize homology, four blocks of gaps are introduced in the Bb sequence at positions 268–269, 273–274, 426–427 and 445–446. The residue numbering indicated above is that of fragment Bb (Fig. 4), the one below is that of chymotrypsinogen. The conserved residues shown are from Young *et al.* (1978), those underlined being invariant in the known sequences (Young *et al.*, 1978). – denotes a gap introduced in the sequence; * denotes the position of active-site residues. The single-letter code of amino acid residues is the same as in Fig. 1 legend.

sinogen A. To accommodate the longer Bb sequence, we have inserted four blocks of gaps in the prothrombin and chymotrypsinogen sequences. These gaps correspond to positions 308–315, 365–370, 401–411 and 479–486 of the Bb fragment sequence. The homology between the alignment reported by Young *et al.* (1978) and the Bb fragment sequence was increased by introducing breaks in four positions of our sequence between residues 268–269, 273–274, 426–427 and 445–446.

The three active-site residues corresponding to His-57, Asp-102 and Ser-193 (chymotrypsinogen numbering) are found at positions 267, 317 and 440 of the Bb sequence respectively. The primary binding site at residues 438–443 (Gly-Asp-Ser-Gly-Gly-Pro) is absolutely conserved. This is to be expected, as even minor substitutions would alter the conformation of the binding pocket (Stroud *et al.*, 1971).

The secondary binding site corresponding to residues 439–441 (Ser-Trp-Gly) is also conserved. This sequence is thought to form hydrogen bonds with the P2 and P3 residues of the substrates in an anti-parallel $\beta$-sheet arrangement (Segal *et al.*, 1971).

Serine proteinases generally contain three conserved disulphide bonds linking Cys-142 to Cys-158 (the 'histidine loop'), Cys-168 to Cys-182 (the 'methionine loop') and Cys-191 to Cys-220, which links the primary and secondary binding sites (chymotrypsinogen numbering). Notable exceptions are C$\bar{1}$r and C$\bar{1}$s, which are missing the 'histidine loop' (Arlaud & Gagnon, 1981), and rat intestinal muscle group-specific proteinase, which is missing the disulphide bond linking the primary and secondary binding sites (Woodburg *et al.*, 1978).

The Bb fragment contains 11 cysteine residues. The cysteine residue at position 33 carries a free thiol group, and the cysteine residues at positions 219 and 252 are both disulphide-linked to cysteine residues within the CB-II portion of the molecule (Christie & Gagnon, 1982). The disulphide bonds of the Bb fragment have yet to be assigned, although it is likely that the three commonly conserved disulphide bonds are present and are represented by disulphide bridges linking Cys-252 to Cys-286, Cys-396 to Cys-423 and Cys-436 to Cys-466 (Fig. 5). The Bb fragment has two other disulphide bonds: from homology with thrombin it is likely that Cys-337 is disulphide-linked to the cysteine residue present in peptide CB-VII-I, corresponding to position 219 to the Bb fragment. The remaining disulphide bridge probably links Cys-340 and Cys-356.

A noteworthy feature of the sequence of fragment Bb is that the sequence containing the catalytic residues is considerably longer than that of other serine proteinases. Although the spacing between the histidine and the aspartic acid residues of 49 residues compares favourably with that for chymotrypsin (44 residues), the spacing between the aspartic acid and the serine residues (122 residues) is longer than those found for thrombin (103 residues) and chymotrypsin (92 residues). These additional residues make the comparison with other serine proteinases of the sequence in this region difficult. The alignment (Fig. 5) in this region is based on minimizing the number of gaps rather than maximizing the number of gaps rather than maximizing homology.

When compared with thrombin and chymotrypsinogen, it can be seen that there has been both deletion and insertion of residues between the active-site histidine and aspartic acid residues.

In addition to the insertions and deletions present in Factor B when compared with thrombin and chymotrypsinogen, six of the 29 invariant residues of serine proteinases (Young *et al.*, 1978) are missing. These correspond to Gly-19, Gly-43, Gly-140, Gly-142, Leu-155 and Pro-161 of chymotrypsinogen. The glycine residues at positions 19 and 43 of chymotrypsinogen have been substituted by an arginine and a methionine residue in the Bb fragment respectively (Fig. 5). It is difficult to comment on the exact substitutions of the other invariant residues. These residues are found between the active-site aspartic acid and serine residues, and this region of the Bb fragment differs considerably from those of thrombin and chymotrypsin. However, no sequence equivalent to Gly-Trp-Gly (residues 140–142 of chymotrypsinogen) is found in this region of the sequence of fragment Bb. Similarly, leucine and proline residues are not found within this region with a spacing equivalent to residues Leu-155 and Pro-161 of chymotrypsinogen (Fig. 5).

A number of the conserved residues of serine proteinases, based on the alignment reported by Young *et al.* (1978), are not present in the Bb fragment (Fig. 5). The position equivalent to residue 192 of chymotrypsinogen is usually occupied by a neutral amino acid. In the Bb fragment an arginine residue is found at this position. Two other complement proteinases also have basic amino acids at this position, arginine in C$\bar{1}$r and lysine in Factor D (Johnson *et al.*, 1980; Arlaud *et al.*, 1982). The residue at position 189 of chymotrypsinogen is thought to confer substrate-specificity. Factor B cleaves an Arg–Ser bond of C3 (Tack *et al.*, 1979), and thus an aspartic acid residue would have been expected, as is the case for trypsin, instead of an asparagine (residue 434 of the Bb fragment). However, an aspartic acid residue is present at position 432, which may function in this capacity.

From the alignment presented in Fig. 5, 53 and 46 residues of fragment Bb are found in equivalent positions in thrombin and chymotrypsin respectively. This is equivalent to approx. 20% and 16% identity between the serine proteinase portion of the Bb fragment and those of thrombin and chymo-

trypsin respectively. By way of comparison, 33% of the residues of chymotrypsin and thrombin are indentical according to the alignment reported by Young *et al.* (1978).

Although the Bb fragment of Factor B shows considerable homology to serine proteinases, the number of insertions, deletions and absence of some invariant and conserved residues prevents any evolutionary relationship being demonstrated to any particular serine proteinases. Also, the zymogens of other serine proteinases share a common activation mechanism, which involves cleavage of an Arg–Ile bond at a position equivalent to residue 15 of chymotrypsinogen. Activation of Factor B by Factor D involves the cleavage of an Arg–Lys bond 222 residues nearer the *N*-terminus of the molecule (Christie & Gagnon, 1982). Moreover, the *N*-terminal sequence of the Bb fragment shows no homology with the conserved *N*-terminal sequences of other serine proteinases.

These factors suggest that Factor B represents a novel class of serine proteinase. It is possible that C2, a proteinase of analogous function from the classical pathway of complement and with similar physicochemical properties to Factor *B* (Kerr & Porter, 1978; Kerr, 1979), also belongs to this class of serine proteinase.

It is tempting to speculate on a domain structure for Factor B. Firstly, Factor B is cleaved to an *N*-terminal Ba fragment (mol.wt. 30000) and a *C*-terminal Bb fragment (mol.wt. 60000) by Factor D in the presence of C3b or by the action of trypsin alone (Curman *et al.*, 1977; Kerr, 1979; Lesavre *et al.*, 1979). The *C*-terminal half of the Bb fragment contains the essential active-site residues characteristic of serine proteinases. Limited proteolysis of the Bb fragment by clostripain results in an *N*-terminal fragment representing residues 1–226 (Gagnon & Christie, 1983). This fragment contains a free thiol group and two sites bearing *N*-linked carbohydrate. Whether this portion of the molecule or the Ba fragment is involved in the initial interaction of Factor B with C3b cannot be deduced at present. It is, however, an essential feature of both Factor B and C2 that their role in the activation of the complement system, i.e. proteolytic activation of C3, occurs only when they are associated with a second protein, C3 or C4. Activation of C5 depends on the interaction of at least one additional C3b molecule. Parts of the Factor B molecule must be involved in the interactions with these other proteins to give the complex C3 and C5 convertases.

We are grateful to Professor R. R. Porter for his continued help and encouragement. We acknowledge Dr. A. Fallon for his help with regard to the $C_8$ column for the h.p.l.c. We thank Mr. P. Fisher for technical assistance in the purification of Factor B, and Mr. A. C. Willis and Mr. T. Gascoyne for their assistance in amino acid sequence determination and amino acid analysis. D. L. C. is a Nuffield Dominions Demonstrator.

## References

Arlaud, G. & Gagnon, J. (1981) *Biosci. Rep.* **1**, 779–787

Arlaud, G., Gagnon, J. & Porter, R. R. (1982) *Biochem. J.* **210**, 49–59

Balian, G., Click, E. M. & Bornstein, P. (1971) *Biochemistry* **10**, 4470–4478

Barnstaple, C. J., Jones, E. A. & Bodmer, W. F. (1979) in *Defence and Recognition IIA* (Lennox, E. S., ed.), vol. 22, pp. 151–225, University Park Press, Baltimore

Bornstein, P. & Balian, G. (1977) *Methods Enzymol.* **47**, 132–145

Christie, D. L. & Gagnon, J. (1982) *Biochem. J.* **201**, 555–567

Christie, D. L., Gagnon, J. & Porter, R. R. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4923–4927

Curman, B., Sandberg-Trägårdh, L. & Peterson, P. A. (1977) *Biochemistry* **16**, 5368–5375

Gagnon, J. & Christie, D. L. (1983) *Biochem. J.* **209**, 51–60

Johnson, D. M. A., Gagnon, J. & Reid, K. B. M. (1980) *Biochem. J.* **187**, 863–874

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

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

Lesavre, P. H., Hugli, T. E., Esser, A. F. & Müller-Eberhard, H. J. (1979) *J. Immunol.* **123**, 529–534

Lewis, R. V., Fallon, A., Stein, S., Gibson, K. D. & Udenfriend, S. (1980) *Anal. Biochem.* **104**, 153–159

Mole, J. E. & Niemann, M. A. (1980) *J. Biol. Chem.* **255**, 8472–8476

Segal, D. M., Powers, J. C., Cohen, G. H., Davies, D. R. & Wilcox, P. E. (1971) *Biochemistry* **10**, 3728–3738

Stroud, R. M., Kay, L. M. & Dickerson, R. E. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 125–140

Tack, B. F., Morris, S. C. & Prahl, J. W. (1979) *Biochemistry* **18**, 1497–1503

Woodburg, R. G., Katunuma, N., Kobayashi, K., Titani, K. & Neurath, H. (1978) *Biochemistry* **17**, 811–819

World Health Organisation (1968) *Bull. W.H.O.* **39**, 935–936

Young, C. L., Barker, W. C., Tomaselli, C. M. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), vol. 5, suppl. 3, pp. 73–93, National Biomedical Research Foundation, Washington