The catalytic chain of human complement subcomponent C1r

Purification and *N*-terminal amino acid sequences of the major cyanogen bromide-cleavage fragments

Gérard J. ARLAUD,* Jean GAGNON and Rodney R. PORTER

Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford,
South Parks Road, Oxford OX1 3QU, U.K.

(Received 1 June 1981/Accepted 12 August 1981)

1. The a- and b-chains of reduced and alkylated human complement subcomponent C1r were separated by high-pressure gel-permeation chromatography and isolated in good yield and in pure form. 2. CNBr cleavage of C1r b-chain yielded eight major peptides, which were purified by gel filtration and high-pressure reversed-phase chromatography. As determined from the sum of their amino acid compositions, these peptides accounted for a minimum molecular weight of 28 000, close to the value 29 100 calculated from the whole b-chain. 3. *N*-Terminal sequence determinations of C1r b-chain and its CNBr-cleavage peptides allowed the identification of about two-thirds of the amino acids of C1r b-chain. From our results, and on the basis of homology with other serine proteinases, an alignment of the eight CNBr-cleavage peptides from C1r b-chain is proposed. 4. The residues forming the 'charge-relay' system of the active site of serine proteinases (His-57, Asp-102 and Ser-195 in the chymotrypsinogen numbering) are found in the corresponding regions of C1r b-chain, and the amino acid sequence around these residues has been determined. 5. The *N*-terminal sequence of C1r b-chain has been extended to residue 60 and reveals that C1r b-chain lacks the 'histidine loop', a disulphide bond that is present in all other known serine proteinases.

Component C1, the first component of the classical complement pathway, is a Ca²⁺-dependent complex consisting of three distinct glycoproteins, subcomponents C1q, C1r and C1s (Lepow *et al.*, 1963). Binding of C1 to activators such as antibody-antigen aggregates is mediated mainly by C1q and induces activation of proenzyme C1r to its proteinase form subcomponent C1r, which in turn cleaves and activates proenzyme C1s. Subcomponent C1s is the proteinase responsible for specific cleavage by C1 of complement components C2 and C4 (see review by Porter, 1977).

Abbreviations used: the nomenclature of complement components is that recommended by the World Health Organisation (1968); activated components are indicated by a bar, e.g. C1r; iPr₂P-F, di-isopropyl phosphorofluoridate; iPr₂P-, di-isopropyl phosphoryl-; SDS, sodium dodecyl sulphate; dansyl, 5-dimethylaminonaphthalene-1-sulphonyl.

* Present address: D.R.F./Biologie Moléculaire et Cellulaire, C.E.N.G., 85X, 38041 Grenoble Cedex, France.

Subcomponent C1r is a dimer of apparent mol.wt. 166 000–188 000, each monomer being, in its proenzyme form, a single polypeptide chain of apparent mol.wt. 83 000–95 000, cleaved on activation into two disulphide-linked polypeptide chains a and b, of respective apparent mol.wts. 56 000–60 000 and 27 000–36 000 (de Bracco & Stroud, 1971; Valet & Cooper, 1974; Ziccardi & Cooper, 1976*a,b*; Sim *et al.*, 1977). Both C1r and C1r dimers are dissociated into monomers at low pH (Arlaud *et al.*, 1980*a*) or in denaturing conditions. Both a- and b-chains contain carbohydrate (Sim *et al.*, 1977; Arlaud *et al.*, 1980*a*), and the b-chain, which is derived from the C-terminal part of C1r (Sim *et al.*, 1977), contains the iPr₂P-F-reactive site (Takahashi *et al.*, 1975; Sim & Porter, 1976). The a-chain has a blocked *N*-terminal amino acid, and the *N*-terminal sequence of the b-chain has been obtained, showing homology with the corresponding part of serine proteinases (Sim *et al.*, 1977).

The first proteolytic event of C1 activation is currently attributed to C1r itself. This is supported by the experiments of Dodds *et al.* (1978), who

proposed that C1r is autocatalytically activated when incorporated within the C1 complex bound to antibody-antigen aggregates. An intramolecular autocatalytic mechanism has also been proposed to account for activation of isolated C1r (Arlaud *et al.*, 1980b). However, there are conflicting reports that isolated C1r by itself has (Ziccardi & Cooper, 1976b; Assimeh *et al.*, 1978; Arlaud *et al.*, 1980b) or has not (Dodds *et al.*, 1978; Bauer *et al.*, 1980) the capacity to autoactivate. Activated C1r is a very specific proteinase, as its biological function in C1 is thought to be restricted to a single cleavage in proenzyme C1s. Its esterolytic activity also appears to be very limited, as only two synthetic esters are unanimously recognized as C1r substrates (Sim, 1981).

In order to elucidate the structural basis of the unique specificity of C1r, the *N*-terminal sequences of the major peptides generated by CNBr treatment of the reduced and alkylated C1r b-chain have been investigated. The results obtained extend the *N*-terminal sequence of C1r b-chain to residue 60 and allow the location of the histidine, aspartic acid and serine residues involved in the charge-relay system.

Materials and methods

Materials

Sephadex G-50 (superfine grade) was obtained from Pharmacia Fine Chemicals, Uppsala, Sweden. Iodoacetic acid was from Sigma Chemical Co., Poole, Dorset, U.K. Dithiothreitol was obtained from Calbiochem, Hereford, U.K., and CNBr was from Aldrich Chemical Co., Gillingham, Dorset, U.K.

Quadrol [*NNN'*-tetrakis-(2-hydroxypropyl)-ethenediamine trifluoroacetate] was purchased from Beckman Instruments, Palo Alto, CA, U.S.A. Polybrene (1,5-diaza-1,5-dimethylundecamethylene polymethobromide) was from Aldrich. All other chemicals used in sequencing were obtained from Rathburn Chemicals, Walkerburn, Peeblesshire, Scotland, U.K.

Iodo[2-¹⁴C]acetic acid (54 Ci/mol) and [1,3-³H]iPr₂P-F (3.4 Ci/mmol) were purchased from The Radiochemical Centre, Amersham, Bucks., U.K.

Purification of subcomponent C1r

C1r was purified from human serum as described previously (Arlaud *et al.*, 1979). The isolated protein (0.5–0.6 mg/ml) in 5 mM-triethanolamine/HCl/145 mM-NaCl buffer, pH 7.4, was incubated at 30°C for 30 min in the presence either of 10 mM-iPr₂P-F or of 2 mM-[1,3-³H]iPr₂P-F (97 Ci/mol). iPr₂P-F-treated C1r was exhaustively dialysed at 4°C against 1% (v/v) acetic acid and freeze-dried.

Reduction and alkylation of subcomponent C1r

C1r (270 nmol) was dissolved in 6 M-guanidine/

HCl/0.4 M-Tris/HCl/2 mM-EDTA buffer, pH 8.0 (5 ml), and reduced with 20 mM-dithiothreitol, then alkylated by iodo[2-¹⁴C]acetic acid (250 μCi) as described by Johnson *et al.* (1980). The labelled protein was dialysed at 4°C against four changes of 1% (v/v) acetic acid (2 litres) and freeze-dried.

Separation of subcomponent-C1r chains

Reduced and alkylated C1r (260 nmol) was dissolved in 6 M-urea/0.2 M-formic acid (2.6 ml). The C1r a- and b-chains were isolated by high-pressure gel-permeation chromatography, by repetitive loading of 100 μl fractions on a 7.5 mm × 600 mm column of TSK-G 3000 SW (Toyo Soda Manufacturing, Tokyo, Japan) equilibrated in 6 M-urea/0.2 M-formic acid and pumped at 0.5 ml/min by a Waters Associates system (model 6000 A). Isolated a- and b-chains were thoroughly dialysed at 4°C against 1% (v/v) acetic acid, then freeze-dried.

SDS/polyacrylamide-gel electrophoresis

The purity of C1r and its chains were assessed by electrophoresis as described by Fairbanks *et al.* (1971) or by Laemmli (1970), by using 7.5% (w/v) polyacrylamide gels in the presence of 0.1% SDS.

CNBr cleavage of subcomponent-C1r b-chain and purification of CNBr-cleavage peptides

Reduced and alkylated C1r b-chain (200 nmol) was dissolved in 70% (v/v) formic acid (1.3 ml) containing CNBr (1.6 mmol) and kept in the dark for 24 h at 4°C. The mixture was diluted 1:10 with water and freeze-dried, then redissolved in 20% (v/v) acetic acid (1.0 ml) and applied to a column (2.2 cm × 90 cm) of Sephadex G-50 (superfine grade), equilibrated with 10% (v/v) acetic acid, and run at a flow rate of 12 ml/h. Five pools, denoted CB-I to CB-V, were collected and freeze-dried. CNBr-cleavage peptides were further purified by high-pressure reversed-phase chromatography, on a Waters μBondapak C-18 column (3.9 mm × 300 mm) with one of three solvent systems. In the first system, used for purification of peptides from pools CB-I and CB-III, the column was equilibrated with a mixture composed of 0.1% (w/v) NH₄HCO₃ and CH₃CN in the ratio 95:5 (v/v), then eluted with a linear gradient to give a final ratio of 45:55 (v/v). In the second system, used for purification of peptides from pools CB-IV and CB-V, the column was equilibrated with a mixture composed of solutions A [0.1% (v/v) trifluoroacetic acid] and B (CH₃CN/methanol/propan-1-ol, 1:1:1, by vol.) in the ratio 95:5 (v/v), then eluted with a linear gradient to give a final ratio of 45:55 (v/v). Further purification of peptide CB-II was achieved by using a third system, in which the column was initially equilibrated with a mixture of solutions A and B in the ratio 70:30 (v/v) and eluted with a

linear gradient to give a final ratio of 20:80 (v/v). In each case the μ Bondapak C-18 column was pumped at 1.0 ml/min by a Waters Associates system (model 6000 A), and peptides were detected by recording of the absorption at 206 nm.

Amino acid analysis

Reduced and alkylated samples were hydrolysed under reduced pressure at 110°C in constant-boiling HCl containing 0.1% (v/v) 2-mercaptoethanol and 4 mM-phenol. Cystine was determined as *S*-carboxymethylcysteine, and methionine was measured in C1r and its chains as methionine sulphone after performic acid oxidation (Hirs, 1967). Tryptophan was not determined. Analyses were performed on a Durrum D500 analyser.

N-Terminal determination and automatic sequencing

N-Terminal amino acids were identified by the dansyl procedure (Gray, 1972), and dansyl-amino acids were resolved as described by Woods & Wang (1967).

N-Terminal amino acid sequences of the CNBr-cleavage peptides from C1r b-chain were determined in a Beckman 890C sequencer as described by Johnson *et al.* (1980). Amino acid phenylthiohydantoin derivatives were identified by high-pressure liquid chromatography as described by Bridgen *et al.* (1976). Assignment of *S*-carboxymethylcysteine residues was confirmed by liquid-scintillation counting of the radioactivity of a 200 μ l portion of the butyl chloride wash.

Results

Isolation of subcomponent-C1r chains

A typical separation of the reduced and alkylated C1r chains by high-pressure gel-permeation chromatography is shown in Fig. 1. Three peaks were eluted successively and pooled as indicated (Fig. 1). Analysis by SDS/polyacrylamide-gel electrophoresis (Laemmli, 1970) showed that the second and the third peaks contained respectively the α -chain (apparent mol.wt. 58 000) and the β -chain (apparent mol.wt. 35 000). The two chains were isolated in pure form, with an average recovery of 80–90%. Both chains were present in the first peak (Fig. 1), which might represent undissociated C1r or aggregated material. The amino acid compositions of C1r and its isolated chains are given in Table 1.

CNBr digestion of subcomponent-C1r b-chain and purification of CNBr-cleavage peptides

The CNBr digest was initially fractionated on Sephadex G-50 (superfine grade) (Fig. 2). The early fractions, eluted at 111–140 ml, contained partially

Fig. 1. Separation of subcomponent-C1r chains. Reduced and alkylated C1r (10 nmol) dissolved in 6 M-urea/0.2 M-formic acid (100 μ l) was loaded on a TSK G-3000SW column (7.5 mm $\times$ 600 mm) equilibrated in the same buffer and pumped at 0.5 ml/min by a Waters Associates system (model 6000 A), as described in the text. Three pools were collected as indicated.

cleaved material accounting for 18% of the total radioactivity eluted from the column. Three major peaks, denoted CB-I, CB-II and CB-III, each containing *S*-carboxy[14 C]methionine radioactivity, were eluted at 143–158 ml, 161–181 ml and 184–210 ml respectively. The radioactivity corresponding to [1,3- 3 H]iPr $_2$ P- was found associated with the second peak (CB-II). Two other pools, denoted CB-IV and CB-V, were collected at 231–251 ml and 263–289 ml respectively. No significant peptide material was detected in pools made between CB-III and CB-IV (213–228 ml), between CB-IV and CB-V (254–260 ml) and after CB-V (292–321 ml). The average recovery of peptide material from the Sephadex column was estimated to be 80%.

Peptides from the five pools CB-I to CB-V were further purified by high-pressure liquid chromatography on a μ Bondapak C-18 column, with three different systems as described in the Materials and

Table 1. Amino acid compositions of subcomponent C̄r and its polypeptide chains

Amino acid compositions were calculated from triplicate 24 h, 48 h and 72 h HCl hydrolysates, except for methionine, which was determined as methionine sulphone from triplicate 24 h HCl hydrolysates of samples after performic acid oxidation. Corrections were made for destruction of serine and threonine. Half-cystine was estimated as *S*-carboxymethylcysteine. Tryptophan was not determined.

Amino acid	Amino acid composition (residues/molecule)		
	C̄r a-chain	C̄r b-chain	C̄r
Asx	47.1	33.0	80.6
Thr	28.4	10.4	40.6
Ser	31.3	13.7	44.6
Glx	69.4	26.2	95.3
Pro	34.7	11.1	47.3
Gly	40.4	28.8	69.7
Ala	17.8	14.3	33.3
Val	18.7	18.9	41.6
Cys	23.4	5.8	28.9
Met	6.0	6.9	13.0
Ile	19.6	11.0	32.2
Leu	34.3	25.1	62.6
Tyr	28.3	8.8	37.2
Phe	24.9	13.7	42.3
His	11.8	8.7	21.4
Lys	27.9	11.8	41.3
Arg	25.8	14.4	44.3

methods section. With the first system, peptide CB-III was recovered as a single major peak containing *S*-carboxyl[¹⁴C]methylcysteine radioactivity, and pool CB-I was fractionated into four peaks (Fig. 3a). On the basis of *N*-terminal determination and amino acid analysis, it was concluded that the first and second peaks corresponded to different forms of the same peptide CB-Ia, whereas the third and fourth peaks were two forms of another peptide, CB-Ib. It is likely that in both cases one of the forms had its *C*-terminal homoserine in the lactone form, as similar doublets have been observed for other CNBr-cleavage peptides (Johnson *et al.*, 1980; Kerlavage & Taylor, 1980). Peptide CB-Ia contained *S*-carboxyl[¹⁴C]methylcysteine radioactivity, whereas peptide CB-Ib did not.

With the second system (see the Materials and methods section), peptide CB-IV was recovered as two major peaks, corresponding to different forms of a unique peptide, as both had the same amino acid composition. The same system allowed the resolution of pool CB-V into three peaks, corresponding to distinct peptides denoted CB-Va, CB-Vb and CB-Vc (Fig. 3b).

By means of a third system (see the Materials and methods section), peptide CB-II was separated from a contaminating fraction that, from amino acid composition, was identified as peptide CB-Ib. Peptide CB-II contained both *S*-carboxyl[¹⁴C]methyl-

Fig. 2. Initial fractionation on Sephadex G-50 of peptides obtained by CNBr digestion of subcomponent-C̄r b-chain. The CNBr digest from 200 nmol of C̄r b-chain was freeze-dried, redissolved in 20% (v/v) acetic acid, and fractionated on Sephadex G-50 (superfine grade) as described in the text. Five pools, denoted CB-I to CB-V, were collected as indicated. ----, A_{280} ; O, radioactivity of *S*-carboxyl[¹⁴C]methylcysteine; ●, radioactivity of [1,3-³H]IPr₂P-.

Fig. 3. Further purification by high-pressure liquid chromatography of peptides from pools CB-I and CB-V (a) Pool CB-I was freeze-dried, redissolved in 0.1% (w/v) NH_4HCO_3 , and one-third was loaded on to a μ Bondapak C-18 column (3.9 mm $\times$ 300 mm) equilibrated with a mixture composed of 0.1% NH_4HCO_3 and CH_3CN in the ratio 95:5 (v/v). The column was eluted with a linear gradient to give a final ratio of 45:55 (v/v), as described in the text. —, A_{206} ; ----, relative percentage of CH_3CN . (b) Pool CB-V was freeze-dried, redissolved in 0.1% (v/v) trifluoroacetic acid, and one-half was loaded on to the μ Bondapak C-18 column equilibrated with a mixture composed of solution A (0.1% trifluoroacetic acid) and solution B (CH_3CN /methanol/propan-1-ol, 1:1:1, by vol.) in the ratio 95:5 (v/v). The column was eluted with a linear gradient to give a final solution A/solution B ratio of 45:55 (v/v), as described in the text. —, A_{206} ; ----, relative percentage of solution B. In both cases the μ Bondapak C-18 column was pumped at 1.0 ml/min by means of a Waters Associates system (model 6000 A). Pools were made as indicated.

cysteine radioactivity and $[1,3\text{-}^3\text{H}]\text{Pr}_2\text{P}$ - radioactivity. The average recovery of CNBr-cleavage peptides from the μ Bondapak C18 column was estimated to be 50–60%.

Amino acid compositions and N-terminal amino acid sequences of subcomponent-C1r b-chain and its CNBr-cleavage peptides

The amino acid compositions of the eight major peptides yielded by CNBr cleavage of C1r b-chain are shown in Table 2. Although small discrepancies can be noted, particularly for glycine and phenylalanine, the sum of the amino acid compositions of individual peptides is in close agreement with that of the whole C1r b-chain.

The N-terminal amino acid sequences of C1r b-chain and its CNBr-cleavage peptides Cb-Ia, Cb-Ib, CB-II, CB-III, CB-Va, CB-Vb and CB-Vc are shown in Fig. 4. The characteristics of each N-terminal analysis are detailed in Table 3. The N-terminal sequence of the whole C1r b-chain was determined for 20 residues and provided the overlap between the N-terminal CNBr-cleavage peptide (identified from its amino acid composition as peptide CB-IV) and peptide CB-Ib. Peptide CB-Ib

was estimated to be 65 amino acid residues long, and an N-terminal sequence extending for 51 residues was obtained, with two undetermined residues at positions 42 and 44 of this peptide, corresponding to residues 58 and 60 in overall numbering (Fig. 4). Peptide CB-Ia was estimated to be 66 residues long, and its N-terminal sequence was determined for 42 residues.

The sequence of the first eight N-terminal amino acid residues of peptide CB-Vc was obtained, although the sixth residue was only tentatively identified as valine and its provisional identification was based on the amino acid composition of peptide CB-Vc (Table 2). Peptide CB-III was estimated to be 33 amino acid residues long, and N-terminal sequence analysis gave sequence of the first 25 residues, with the exception of residue at position 24. Peptide CB-Vb was found to be seven residues long, and its first six residues were identified. Peptide CB-II was estimated to be 64 amino acid residues long, and the first 29 amino acid residues were determined, except for residues at positions 17, 23, 24 and 28. By scintillation counting of the radioactivity of 3% of the butyl chloride extracts, the radioactivity associated with $[1,3\text{-}^3\text{H}]\text{Pr}_2\text{P}$ - was

Table 2. *Amino acid compositions of CNBr-cleavage peptides from subcomponent-C1r b-chain*

The amino acid compositions of peptides were calculated from duplicate 24 h HCl hydrolysates. Homoserine was quantified as the sum of homoserine and homoserine lactone values. Serine and threonine yields were corrected assuming the same rate of destruction as observed for C1r b-chain. Half-cystine was estimated as *S*-carboxymethyl-cysteine, and tryptophan was not determined. All amino acid compositions are given as residues/molecule. The numbers in parentheses represent the values extrapolated to the nearest integer.

Amino acid composition (residues/molecule)										
Amino acid	Peptide CB-Ia	Peptide CB-Ib	Peptide CB-II	Peptide CB-III	Peptide CB-IV	Peptide CB-Va	Peptide CB-Vb	Peptide CB-Vc	Sum of CNBr-cleavage peptides	C1r b-chain
Asx	11.7 (12)	7.0 (7)	6.5 (7)	4.2 (4)	0.1 (0)	1.0 (1)	2.1 (2)	0.1 (0)	33	33.0 (33)
Thr	2.3 (2)	4.4 (4)	3.1 (3)	0.3 (0)	—	—	—	—	9	10.4 (10)
Ser	3.6 (4)	3.1 (3)	4.0 (4)	0.9 (1)	0.2 (0)	0.2 (0)	0.9 (1)	0.9 (1)	14	13.7 (14)
Hse	0.9 (1)	0.9 (1)	1.0 (1)	1.0 (1)	0.9 (1)	—	0.9 (1)	0.9 (1)	7	—
Glx	6.1 (6)	7.2 (7)	3.7 (4)	4.3 (4)	1.1 (1)	2.9 (3)	1.1 (1)	0.1 (0)	26	26.2 (26)
Pro	5.0 (5)	2.1 (2)	1.9 (2)	1.5 (2)	—	—	—	—	11	11.1 (11)
Gly	5.2 (5)	9.0 (9)	9.8 (10)	2.0 (2)	2.1 (2)	0.3 (0)	0.2 (0)	3.2 (3)	31	28.8 (29)
Ala	1.3 (1)	5.6 (6)	4.2 (4)	2.9 (3)	1.0 (1)	—	—	—	15	14.3 (14)
Val	3.5 (4)	3.6 (4)	5.8 (6)	2.6 (3)	—	—	1.1 (1)	2.1 (2)	20	18.9 (19)
Cys	1.2 (1)	—	2.9 (3)	1.0 (1)	—	—	—	—	5	5.8 (6)
Met	—	—	—	—	—	—	—	—	—	6.9 (7)
Ile	3.1 (3)	2.2 (2)	2.9 (3)	1.0 (1)	1.5 (2)	—	—	—	11	11.0 (11)
Leu	10.9 (11)	7.7 (8)	2.9 (3)	3.0 (3)	—	—	—	—	25	25.1 (25)
Tyr	3.0 (3)	1.2 (1)	3.1 (3)	0.2 (0)	—	—	—	1.0 (1)	8	8.8 (9)
Phe	1.9 (2)	3.2 (3)	2.9 (3)	1.3 (1)	—	—	0.9 (1)	0.9 (1)	11	13.7 (14)
His	2.2 (2)	4.4 (4)	1.2 (1)	1.0 (1)	—	—	—	—	8	8.7 (9)
Lys	1.3 (1)	1.6 (2)	3.9 (4)	2.0 (2)	1.9 (2)	0.1 (0)	—	—	11	11.8 (12)
Arg	3.4 (3)	2.4 (2)	3.2 (3)	3.8 (4)	—	—	—	—	12	14.4 (14)
Total	66	65	64	33	9	4	7	9	257	263

Table 3. *Technical characteristics of N-terminal amino acid sequence analyses performed on subcomponent-C1r b-chain and its CNBr-cleavage peptides CB-Ia, CB-Ib, CB-II, CB-III, CB-Va, CB-Vb and CB-Vc*

	C1r b-chain	Peptide CB-Ia	Peptide CB-Ib	Peptide CB-II	Peptide CB-III	Peptide CB-Va	Peptide CB-Vb	Peptide CB-Vc
Amount used in sequencer (nmol)	30	60	47	56	31	56	51	52
Recovery at first step (nmol)	13	28	29	29	17	38	45	42
Stepwise yield (%)*	97	92	93	92	89	68	89	57

* Calculated by regression analysis of all steps.

located at position 17. The absence of positive identification of this residue as serine was expected, since the intact C1r used to prepare CNBr-cleavage peptides has been treated by $[1,3\text{-}^3\text{H}]\text{iPr}_2\text{P-F}$ under conditions (see the Materials and methods section) where most of the serine residues at the active site were expected to be blocked by the reagent. Peptide CB-II was found associated with a contaminating peptide (about 10% of the total material) of the following *N*-terminal sequence:

By superimposition of this sequence with those determined for peptide CB-Vb and CB-II, it became apparent that this contaminating peptide resulted from incomplete digestion of the methionyl bond at position 7 in peptide CB-Vb and was composed of peptides CB-Vb and CB-II.

Among the eight major peptides generated by CNBr cleavage, peptide CB-Va was the only one lacking homoserine (Table 2), and was therefore concluded to be the *C*-terminal peptide of C1r b-chain.

Asp-Val-Phe-()-Gln-Asn-()-Phe-Cys-Ala()-()-Pro-()-Leu-Lys

Fig. 4. *N*-Terminal amino acid sequences of the major CNBr-cleavage peptides from subcomponent-C1r b-chain: homology with sequences of other serine proteinases. Sequence data for elastase, chymotrypsin, trypsin, plasmin, Factor X and thrombin are from Young *et al.* (1978). — denotes that a gap was left to give maximum homology on alignment of the amino acid residues. * denotes the position of the [1,3-³H]Pr₂P-labelled serine residue in peptide CB-II. The residue numbering indicated is arbitrary, and takes into account the maximum number of positions in all the proteinases listed. The conserved residues shown are from Young *et al.* (1978), those underlined being invariant in the known sequences (Young *et al.*, 1978). The single-letter code of amino acid residues has been used: A, Ala; B, Asx; 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; X, unknown; Y, Tyr; Z, Glx.

Fig. 4. *N*-Terminal amino acid sequences of the major CNBr-cleavage peptides from subcomponent-C1r b-chain: homology with sequences of other serine proteinases. Sequence data for elastase, chymotrypsin, trypsin, plasmin, Factor X and thrombin are from Young *et al.* (1978). — denotes that a gap was left to give maximum homology on alignment of the amino acid residues. * denotes the position of the [1,3-³H]Pr₂P-labelled serine residue in peptide CB-II. The residue numbering indicated is arbitrary, and takes into account the maximum number of positions in all the proteinases listed. The conserved residues shown are from Young *et al.* (1978), those underlined being invariant in the known sequences (Young *et al.*, 1978). The single-letter code of amino acid residues has been used: A, Ala; B, Asx; 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; X, unknown; Y, Tyr; Z, Glx.

Elastase (pig)	N T D D V A A	100	110	120	130
	- - - - -	- - - - -	G Y D I A L L R L A Q S V T L N S Y V Q L G V L P R A G T -		
Chymotrypsin A (cow)	N S L T I N N	- - - - -	- - - - -	D I T L L K L S T A A S F S Q T V S A V C L P S A S D -	
Trypsin (cow)	N S N T L N N	- - - - -	- - - - -	D I M L I K L K S A A S L N S R V A S I S L P T S C A -	
Plasmin (human)	- - - T R K	- - - - -	- - - - -	D I A L L K L S S P A V I T D K V I P A C L P S P N Y -	
Factor X (cow)	V - K E T Y D	- - - - -	- - - - -	F D I A V L R L K T P I R F - R N V A P A C L P E K D W A	
Thrombin (cow)	N W K E N L D	- - - - -	- - - - -	R D I A L L K L K R P I E L S D Y I H P V C L P D K Q T A	
Conserved residues	N			D I A L L K L P A	V P C L P
Clr b-chain	R V S V H P D Y R Q D E S Y N F E G D I A L L E L E N S V T L G P N				

CB-Ia

Elastase (pig)	- - I L A N N S P C Y I T G W G L	140	150	160	170	180
	- - - - -	- - - - -	- - - - -	T R - T N G Q L A Q T L Q Q A Y L P T V D Y A I		
Chymotrypsin A (cow)	- - D F A A G T T C V T T G W G L	- - - - -	- - - - -	T R Y T N A N T P D R L Q Q A S L P L S N T N		
Trypsin (cow)	- - - S A G T Q C L I S G W G N	- - - - -	- - - - -	T K S S G T S Y P D V L K C L K A P I L S N S S		
Plasmin (human)	- - V V A D R T E C F I T G W G E	- - - - -	- - - - -	T Q - - G T F G A G L K E A Q L P V I E N K V		
Factor X (cow)	A E T L Q T K T - G I V S G F G R	- - - - -	- - - - -	T H - E K G R L S S T L K M L E V P Y V D R S T		
Thrombin (cow)	A K L L H A G F K G R V T G W G N R R R E T W T T S V A E V Q P S V L Q V V N L P L V E R P V					
Conserved residues	G T C T G W G			T	L Q A L P V	
Clr b-chain	M G Y V S G V G F			M E E K I A H D L R F V R L P V A N P Q A		

CB-Ia

CB-Vc

CB-III

	190	200	210	220	230
Elastase (pig)	C S S S Y W G S T V K N S M V C A G - - - G N G V R - S G C Q G D S G G P L H - - C L V N				
Chymotrypsin A (cow)	C K K - - Y W G T K I K D A M I C A G - - - A S G V - - S S C M G D S G G P L V - - C K K N				
Trypsin (cow)	C K S A - Y P G - Q I T S N M F C A G Y - - L E G G K - D S C Q G D S G G P V V - - C S G K				
Plasmin (human)	C N R Y E F L N G R V Q S T E L C A G H - - L A G G T - D S C Q G D S G G P L V - - C F E K				
Factor X (cow)	C K L S - S S F - T I T P N M F C A G Y - - D T Q P E - D A C Q G D S G G P H V - - T R F K				
Thrombin (cow)	C K A S - T R I - R I T N D M F C A G Y K P G E G K R G D A C E G D S G G P F V M K S P Y N				
Conserved residues	C Y V T M C A G Y G G D C Q G D S G G P L V C				
C1r b-chain	C E N X L M D V F S Q N M F C A G H - - P S L K Q - D A C Q G D S G G V F A X X D P N X				
	-CB-III-	-CB-Vb-	-CB-II-		

	240	250	260	270
Elastase (pig)	G Q Y A V H G V T S F V S R R L G C N V T R K F T V F T R V S A Y I S W I N N V I A S N			
Chymotrypsin A (cow)	G A W T L V G I V S W G S S - T C S - T S T P G V Y A R V T A L V N W V Q Q T L A A N			
Trypsin (cow)	- - - L Q G I V S W G S - - G C A Q K N K P G V Y T K V C N Y V S W I K Q T I A S N			
Plasmin (human)	D K Y I L Q G V T S W G L - - G C A R P N K P G V Y V R V S R F V T W I E G V M R N N			
Factor X (cow)	D T Y F V T G I V S W G E - - G C A R K G K R G V Y T K V S N F L K W I D K I M K A R A G			
Thrombin (cow)	N R W Y Q M G I V S W G E - - G C D R N G K Y G F Y T H V F R L K K W I Q K V I D R L G S			
Conserved residues	G I V S W G G C A R K P G V Y T V S W I I N			
C1r b-chain	D			M E E E D
		-CB-II-		-CB-Va

Discussion

High-pressure gel-permeation chromatography permitted the complete separation of C1r a- and b-chains. Compared with conventional gel-filtration

methods, this technique is faster, one separation being achieved within 15–30 min, depending on the flow rate. This is of importance, as long stages in 6 M-urea/0.2 M-formic acid can lead to acid cleavage of the protein. Amino acid analyses performed on C1r and its isolated chains (Table 1) are in agreement with those reported previously (Sim & Porter, 1976; Ziccardi & Cooper, 1976b; Sim *et al.*, 1977). The methionine value obtained in the present study for C1r b-chain (2.63 residues/100 residues) indicated 7 methionine residues/molecule of b-chain, and eight peptides were therefore expected from CNBr cleavage. This treatment produced peptides CB-Ia, CB-Ib, CB-II, CB-III, CB-IV, CB-Va, CB-Vb and CB-Vc, which were obtained in yields of 38%, 33%, 31%, 34%, 52%, 31%, 20% and 21% respectively. The minimum molecular weight of C1r b-chain, as calculated from its composition or from the sum of amino acid composition of its CNBr-cleavage peptides, was 29 100 and 28 000 respectively. The eight CNBr-cleavage peptides purified in the present study can therefore account for the whole C1r b-chain, although the possibility of a ninth, very small, peptide cannot be entirely excluded.

The amino acids identified from *N*-terminal sequence analyses of C1r b-chain and its CNBr-cleavage peptides represent two-thirds of the total amino acids of the C1r b-chain. Comparison of the *N*-terminal sequences of C1r b-chain and peptide CB-Ib (Fig. 4) shows that the expected *N*-terminal peptide is peptide CB-IV (see Table 2), and allows the alignment of peptides CB-IV and CB-Ib, thus extending the *N*-terminal sequence of C1r b-chain to 60 residues.

Our results agree exactly with the identification of residues at positions 1–20 reported by Sim *et al.* (1977), with the exception of residue 20, which had been tentatively identified as glutamic acid/glutamine, whereas we found isoleucine.

The alignment of peptides CB-Vb and CB-II was provided by the sequence of the overlapping peptide, and peptide CB-Va was identified as the *C*-terminal peptide of C1r b-chain (see the Results section). Therefore the above findings allow the relative alignment of five peptides in the order:

From homology with other serine proteinases

(Young *et al.*, 1978), peptides CB-Ia, CB-Vc and CB-III were located in this order between peptides CB-Ib and CB-Vb (Fig. 4). Thus the complete alignment proposed for the eight major CNBr-cleavage peptides of C1r b-chain is the following:

Positioning of the eight peptides in this order shows extensive homology with the conserved residues of the catalytic chain of serine proteinases (Fig. 4). The three active residues common to all serine proteinases (His-57, Asp-102 and Ser-195, according to the chymotrypsinogen numbering) are found at position 39 of C1r b-chain, position 27 of peptide CB-Ia and position 17 of peptide CB-II, in a linear distribution that is in accordance with that found in other serine proteinases (de Haën *et al.*, 1975; Young *et al.*, 1978).

The presence of an aspartic acid residue at position 11 of peptide CB-II is likely to indicate that C1r belongs to the family of 'trypsin-like' enzymes. This would be consistent with the proteolytic specificity of C1r, which probably cleaves a single Arg-Ile or Lys-Ile bond in proenzyme C1s (Sim, 1981), and with the esterolytic specificity of C1r, which cleaves the synthetic substrates acetylglucyl-Lysine methyl ester and acetylarginine methyl ester (Naff & Ratnoff, 1968; Volanakis *et al.*, 1977; Sim *et al.*, 1977; Andrews & Baillie, 1979).

The sequence found at positions 15–20 of peptide CB-II (-Gly-Asp-Ser-Gly-Gly-Val) is homologous with the primary binding site of serine proteinases (Young *et al.*, 1978), with the exception of position 20, which is invariably a proline residue in other known serine proteinases (Fig. 4). More important appears to be the lack in C1r b-chain structure of two half-cystine residues invariant in other serine proteinases, in which they are linked to form a disulphide bond called the 'histidine loop' (Young *et al.*, 1978). These residues, located at positions 30 and 46 in Fig. 4, are unequivocally missing in C1r b-chain structure. In view of the close proximity of this bond to the histidine residue involved in the active site, its absence from C1r b-chain is likely to imply important structural and functional consequences. The cysteine residues identified at position 21 of peptide CB-III and position 2 of peptide CB-II (Fig. 4) are equivalent to those forming the 'methionine loop' in other serine proteinases, and the cysteine residue at position 13 of peptide CB-II is likely to represent half of the other invariant disulphide bridge linking the primary and secondary binding sites of serine proteinases (Young *et al.*, 1978). At least two other cysteine residues are present in C1r b-chain, one being found in peptide CB-Ia, the other in peptide CB-II.

Peptide CB-Va, the C-terminal CNBr-cleavage peptide of C1r b-chain, contains four acidic amino acid residues, which is an unusual feature, compared with the C-terminal part of other serine proteinases (de Haën *et al.*, 1975; Young *et al.*, 1978; Jackson, 1980). For this reason, the precise location of peptide CB-Va by homology with other serine proteinases (Fig. 4) must be considered as tentative. Overall comparison of all the sequence determined in the present study with the corresponding sequences of other serine proteinases (Fig. 4) indicates 23%, 26%, 27%, 29%, 29% and 35% homology of C1r b-chain with chymotrypsin, trypsin, elastase, plasmin, Factor X and thrombin respectively, thus indicating a closer relationship between C1r b-chain and thrombin.

We are very grateful to Professor M. G. Colomb, Dr. R. B. Sim, Dr. K. B. M. Reid and Dr. D. L. Christie for advice and discussion. We thank Mr. A. Gascoygne for performing the amino acid analyses, Mr. A. C. Willis for sequence analysis and Mrs. M. Lacroix for excellent technical assistance. G. J. A. held a N.A.T.O. Research Fellowship.

References

- Andrews, J. M. & Baillie, R. D. (1979) *J. Immunol.* **123**, 1403–1408
- Arlaud, G. J., Sim, R. B., Duplaa, A.-M. & Colomb, M. G. (1979) *Mol. Immunol.* **16**, 445–450
- Arlaud, G. J., Chesne, S., Villiers, C. L. & Colomb, M. G. (1980a) *Biochim. Biophys. Acta* **616**, 105–115
- Arlaud, G. J., Villiers, C. L., Chesne, S. & Colomb, M. G. (1980b) *Biochim. Biophys. Acta* **616**, 116–129
- Assimeh, S. N., Chapuis, R. M. & Isliker, H. (1978) *Immunochemistry* **15**, 13–17
- Bauer, J., Weiner, E. & Valet, G. (1980) *J. Immunol.* **124**, 1514 (Abstr.)
- Bridgen, P. J., Cross, G. A. M. & Bridgen, J. (1976) *Nature (London)* **263**, 613–614
- de Bracco, M. M. E. & Stroud, R. M. (1971) *J. Clin. Invest.* **50**, 838–848
- de Haën, C., Neurath, H. & Teller, D. C. (1975) *J. Mol. Biol.* **92**, 225–259
- Dodds, A. W., Sim, R. B., Porter, R. R. & Kerr, M. A. (1978) *Biochem. J.* **175**, 383–390
- Fairbanks, G., Stech, T. L. & Wallach, D. F. H. (1971) *Biochemistry* **10**, 2606–2617
- Gray, W. R. (1972) *Methods Enzymol.* **25**, 121–138
- Hirs, C. H. W. (1967) *Methods Enzymol.* **11**, 197–199
- Jackson, C. M. (1980) *Annu. Rev. Biochem.* **49**, 765–811
- Johnson, D. M. A., Gagnon, J. & Reid, K. B. M. (1980) *Biochem. J.* **187**, 863–874
- Kerlavage, A. R. & Taylor, S. S. (1980) *J. Biol. Chem.* **255**, 8483–8488
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lepow, I. H., Naff, G. B., Todd, E. W., Pensky, T. & Hinz, C. F., Jr. (1963) *J. Exp. Med.* **117**, 983–1008
- Naff, G. B. & Ratnoff, O. D. (1968) *J. Exp. Med.* **128**, 571–593
- Porter, R. R. (1977) *Biochem. Soc. Trans.* **5**, 1659–1674
- Sim, R. B. (1981) *Methods Enzymol.* **80**, in the press
- Sim, R. B. & Porter, R. R. (1976) *Biochem. Soc. Trans.* **4**, 127–129
- Sim, R. B., Porter, R. R., Reid, K. B. M. & Gigli, I. (1977) *Biochem. J.* **163**, 219–227
- Takahashi, K., Nagasawa, S. & Koyama, J. (1975) *FEBS Lett.* **55**, 156–160
- Valet, G. & Cooper, N. R. (1974) *J. Immunol.* **112**, 1667–1673
- Volanakis, J. E., Schrohenloher, R. E. & Stroud, R. H. (1977) *J. Immunol.* **119**, 337–342
- Woods, K. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **133**, 369–370
- World Health Organisation (1968) *Bull. W.H.O.* **39**, 935–938
- 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
- Ziccardi, R. J. & Cooper, N. R. (1976a) *J. Immunol.* **116**, 496–503
- Ziccardi, R. J. & Cooper, N. R. (1976b) *J. Immunol.* **116**, 504–509