Biochem. J. (1985) **225**, 135–142
Printed in Great Britain

Primary structure of the A chain of human complement-classical-pathway enzyme C$\bar{1}$r

N-Terminal sequences and alignment of autolytic fragments and CNBr-cleavage peptides

Jean GAGNON* and Gerard J. ARLAUD†
**M.R.C. Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and †D.R.F. Laboratoire de Biologie Moléculaire et Cellulaire, Centre d'Études Nucléaires de Grenoble, B.P. 85X, 38041 Grenoble Cedex, France*

(*Received 12 June 1984/Accepted 11 September 1984*)

Activated human complement-classical-pathway enzyme C$\bar{1}$r has previously been shown to undergo autolytic cleavages occurring in the A chain [Arlaud, Villiers, Chesne & Colomb (1980) *Biochim. Biophys. Acta* **616**, 116–129]. Chemical analysis of the autolytic products confirms that the A chain undergoes two major cleavages, generating three fragments, which have now been isolated and characterized. The *N*-terminal α fragment (approx. 210 residues long) has a blocked *N*-terminus, as does the whole A chain, whereas *N*-terminal sequences of fragments β and γ (approx. 66 and 176 residues long respectively) do not, and their *N*-terminal sequences were determined. Fragments α, β and γ, which are not interconnected by disulphide bridges, are located in this order within C$\bar{1}$r A chain. Fragment γ is disulphide-linked to the B chain of C$\bar{1}$r, which is *C*-terminal in the single polypeptide chain of precursor C1r. CNBr cleavage of C$\bar{1}$r A chain yields seven major peptides, CN1b, CN4a, CN2a, CN1a, CN3, CN4b and CN2b, which were positioned in that order, on the basis of *N*-terminal sequences of the methionine-containing peptides generated from tryptic cleavage of the succinylated (3-carboxypropionylated) C$\bar{1}$r A chain. About 60% of the sequence of C$\bar{1}$r A chain (440–460 residues long) was determined, including the complete sequence of the *C*-terminal 95 residues. This region shows homology with the corresponding parts of plasminogen and chymotrypsinogen and, more surprisingly, with the α1 chain of human haptoglobin 1-1, a serine proteinase homologue.

The classical pathway of complement (Reid & Porter, 1981) is triggered by the activation of C1, a Ca^{2+}-dependent complex consisting of a recognition unit, C1q, and two zymogens of serine proteinases, C1r, and C1s (Sim, 1981). C1 activation is a two-step process involving: (i) autocatalytic activation of C1r to its proteinase form C$\bar{1}$r; (ii) C$\bar{1}$r-mediated activation of C1s to its proteinase form C$\bar{1}$s, responsible for the proteolytic activity of the whole C$\bar{1}$ complex. The key event of C1 activation is thus the activation of the C1r subunit by itself. This basic autoactivation property is expressed in the whole C1 complex, but modulated by the other subcomponents (Dodds *et al.*, 1978; Arlaud *et al.*, 1980*b*; Ziccardi, 1982).

Human proenzyme C1r is a dimeric glycoprotein consisting of two apparently identical single-chain polypeptides of M_r approx. 90000. During activation, each monomer undergoes a limited proteolytic cleavage, which generates two disulphide-linked polypeptide chains A (M_r 56000–60000) and B (M_r 30000–36000) (Valet & Cooper, 1974; Ziccardi & Cooper, 1976*a*,*b*; Sim *et al.*, 1977; Arlaud & Gagnon, 1983). On prolonged incubation at 37°C, activated C$\bar{1}$r undergoes

Abbreviations used: h.p.l.c., high-pressure liquid chromatography; iPr_2P-F, di-isopropyl phosphorofluoridate; the nomenclature of complement components is that recommended by the World Health Organisation (1968, 1981); activated components are indicated by a superscript bar, e.g. C$\bar{1}$r.

further proteolytic cleavages, occurring in the *N*-terminal A chain, which splits into three fragments (Assimeh *et al.*, 1978; Okamura & Fujii, 1978; Arlaud *et al.*, 1980*b*). Fragments α (M_r 35000) and β (M_r 7000–11000) are removed from each monomer, leaving a molecule of smaller size (M_r 110000) that is still a dimer, each monomer now consisting of fragment γ disulphide-linked to the original B chain (Arlaud *et al.*, 1980*b*). The final product, $C\bar{1}r$ II, retains the original antigenic properties of $C\bar{1}r$ and a functional active site (Arlaud *et al.*, 1980*b*), but has lost the capacity to bind $C\bar{1}s$ (Arlaud *et al.*, 1980*b*) and Ca^{2+} ions (Villiers *et al.*, 1980).

The sequence of the catalytic B chain of human $C\bar{1}r$ (*C*-terminal in C1r) has already been completed (Arlaud & Gagnon, 1983). The chain, which lacks the 'histidine loop', a disulphide bridge that is conserved in other mammalian serine proteinases, contains 242 amino acid residues and two carbohydrate moieties, both attached via asparagine residues (Arlaud & Gagnon, 1983).

The *N*-terminal A chain, which is thought to be involved in intermonomer and intersubunit interactions (Villiers *et al.*, 1980; Arlaud *et al.*, 1980*a*), also contains carbohydrate, and has a blocked *N*-terminus (Sim *et al.*, 1977). The present study is the first step in the elucidation of the structure of $C\bar{1}r$ A chain. In addition to CNBr and tryptic cleavages, we have taken advantage of the autolytic capacity of $C\bar{1}r$, which provides a convenient means of performing a limited proteolysis of the A chain.

Experimental

Materials

Trypsin [treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one ('TPCK')] was obtained from Worthington Biochemical Corp. Iodo[2-^{3}H]-acetic acid (61 Ci/mol) was from The Radiochemical Centre. Reagents for automated amino acid-sequence analysis were obtained as described by Christie & Gagnon (1982). Methanol, acetonitrile, propan-2-ol and all reagents used for h.p.l.c. were purchased from Merck, Rathburn Chemical or Fluka. Spectrapor 6 dialysis tubing (M_r cut-off 1000) was obtained from Spectrum Medical Industries.

Methods

$C\bar{1}r$ was purified from human serum as described previously (Arlaud *et al.*, 1979). Reduction of $C\bar{1}r$, alkylation by iodo[2-^{3}H]acetic acid and separation of $C\bar{1}r$ A and B chains by high-pressure gel-permeation chromatography were performed as described previously (Arlaud *et al.*, 1982).

$C\bar{1}r$ autolytic cleavage

$C\bar{1}r$ (0.46 mg/ml) was incubated for 7 h at 37°C in 145 mM-NaCl/5 mM-triethanolamine/HCl buffer, pH 7.4. After exhaustive dialysis in Spectrapor 6 dialysis tubing against 0.5% (v/v) acetic acid, and freeze-drying, the autolytic-cleavage products were dissolved in 6 M-urea/0.2 M-formic acid and separated by high-pressure gel-permeation chromatography, by repetitive loading of 100 μl fractions on an LKB Blue column (TSK G-3000 SW) (see Fig. 1). Pools 1, 2 and 3 contained fragments $C\bar{1}r$ II, α and β respectively. These were dialysed against 0.5% (v/v) acetic acid, freeze-dried, and then reduced and *S*-[^{3}H]carboxymethylated as described for $C\bar{1}r$ chains (Arlaud *et al.*, 1982). Reduction of fragment $C\bar{1}r$ II released $C\bar{1}r$ B chain and fragment γ, which were separated by high-pressure gel-permeation chromatography on TSK G-3000 SW as described above.

Peptide separation by reversed-phase h.p.l.c.

All separations were performed on a μBondapak C18 column (Waters Associates) with two solvent systems as described previously (Arlaud & Gagnon, 1983). Briefly, system 1 consisted of 0.1% NH_4HCO_3 and acetonitrile, and system 2 consisted of 0.1% trifluoroacetic acid and acetonitrile/methanol/propan-2-ol (1:1:1, by vol.). Peptides were detected from the absorbance at 206 nm or 215 nm.

CNBr cleavage of $C\bar{1}r$ A chain

Reduced and *S*-[^{3}H]carboxymethylated $C\bar{1}r$ A chain (250 nmol) was dissolved in 70% (v/v) trifluoroacetic acid (1 ml). CNBr (4 mmol) was added, and the mixture was kept in the dark for 24 h at 4°C, then diluted 1:10 with water and freeze-dried. The CNBr digest was redissolved in 0.1 M-NH_4HCO_3 (5 ml) and fractionated on a Sephadex G-75 column equilibrated in the same buffer (see Fig. 3). Four pools were collected and freeze-dried. Pool 1 was further analysed by high-pressure gel-permeation chromatography on TSK G3000 SW as described for $C\bar{1}r$ autolytic fragments, whereas pools 2, 3 and 4 were submitted to reversed-phase h.p.l.c. with respectively solvent systems 1, 2 and 2 (Arlaud & Gagnon, 1983).

Protein succinylation (3-carboxypropionylation) and tryptic cleavage

Reduced and *S*-[^{3}H]carboxymethylated $C\bar{1}r$ A chain (150 nmol) and fragment α (170 nmol) were succinylated by the procedure of Koide *et al.* (1978), as previously described for $C\bar{1}r$ B chain (Arlaud & Gagnon, 1983). Reduced and *S*-[^{3}H]-carboxymethylated succinylated $C\bar{1}r$ A chain and fragment α were cleaved by trypsin, as described previously for $C\bar{1}r$ B chain (Arlaud & Gagnon,

1983). The tryptic digest of C$\bar{1}$r A chain was fractionated by gel filtration on Sephadex G-50 (see Fig. 4). Ten pools were collected, and the methionine-containing peptides were further purified by reversed-phase h.p.l.c. with solvent system 1 (Arlaud & Gagnon, 1983). The tryptic digest of fragment α was fractionated by gel filtration on a Sephadex G-50 (fine grade) column (2.5cm × 100cm) equilibrated in 0.1M-NH_4HCO_3. Six pools were collected, and peptide α ST1 was obtained from pool 1 without further purification.

Amino acid analysis

Peptides were hydrolysed for 24h under reduced pressure at 110°C in constant-boiling HCl containing 0.1% (v/v) 2-mercaptoethanol and 4mM-phenol. Cysteine was determined as *S*-(carboxymethyl)cysteine. Hexosamines were positively identified and quantified in the same hydrolysates as amino acids, although no correction was made for destruction during acid hydrolysis. Analyses were performed on a Durrum D-500 or a LKB 4400 analyser. Tryptophan was not determined.

Automated sequence analysis

Automated Edman degradation was performed in a Beckman 890C Sequenator with the 0.3M-Quadrol programme of Hunkapiller & Hood (1978). Identification of amino acid phenylthiohydantoin derivatives by reversed-phase h.p.l.c. was performed as described previously (Christie & Gagnon, 1982).

Peptide nomenclature

Autolytic fragments are designated α, β and γ. Fragments obtained from CNBr cleavage and tryptic cleavage after succinylation are designated CN and ST respectively. The next number indicates the elution position of the peptide in the initial separation used (i.e. gel filtration), and it is followed by a letter indicating its position in the second separation (i.e. h.p.l.c.).

Results

Characterization of C$\bar{1}$r A-chain autolytic fragments

C$\bar{1}$r autolytic cleavage yielded fragments α, β and C$\bar{1}$r II, which were isolated by high-pressure gel-permeation chromatography before reduction of disulphide bridges (Fig. 1). C$\bar{1}$r II consisted of fragment γ disulphide-linked to C$\bar{1}$r B chain, which were separated after reduction and alkylation. There are no disulphide bridges between the three α, β and γ fragments from C$\bar{1}$r A chain, whereas, as previously shown (Arlaud *et al.*, 1980*b*), fragment γ is disulphide-bonded to C$\bar{1}$r B chain.

Amino acid compositions of fragments α, β and γ are indicated in Table 1. The sum of the amino acid compositions of individual fragments is in close agreement with that of the whole C$\bar{1}$r A chain, and fragment β lacks methionine, whereas fragments α and γ each contain three methionine residues. *N*-Terminal sequence analysis of fragment α showed that this fragment has a blocked *N*-terminus, whereas *N*-terminal sequences of fragments β and γ, determined for 50 and 62 cycles respectively, indicated that both fragments have an *N*-terminal glycine residue (Fig. 2). Fragment β contains two half-cystine residues (Table 1), located at positions 22 and 40 of this fragment (Fig. 2). Since, as shown above, fragment β is released from C$\bar{1}$r A chain without reduction, it can be concluded that these residues form a disulphide bridge in C$\bar{1}$r. *N*-Terminal sequence analysis of fragment γ revealed that the major peptide Gly-Trp-Lys-Leu-Arg-Tyr-Thr-Thr-Glu... was contaminated with a minor peptide (6% on a molar basis) commencing Tyr-Thr-Thr-Glu..., thus indi-

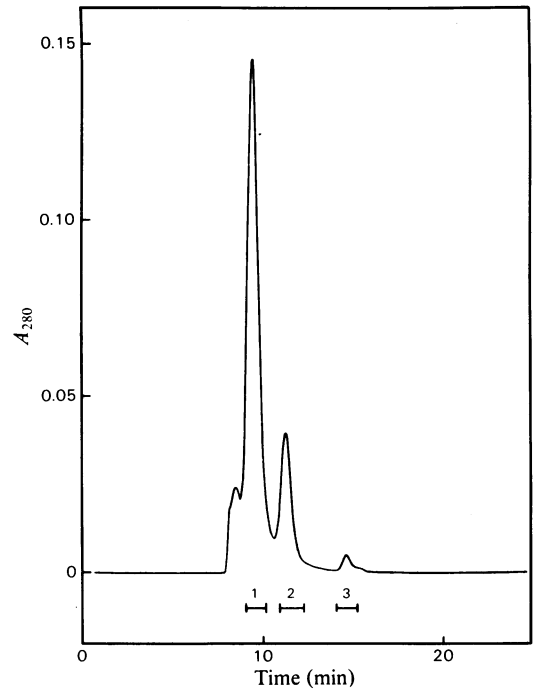

Fig. 1. *Separation of C$\bar{1}$r autolytic fragments by high-pressure gel-permeation chromatography on TSK G3000 SW*

C$\bar{1}$r autolytic fragments (10nmol) were dissolved in 6M-urea/0.2M-formic acid (100μl) and loaded on a TSK G-3000 SW column (7.5mm × 600mm) equilibrated in the same buffer and eluted at 1.0ml/min. Pools 1, 2 and 3 were collected as indicated.

Table 1. *Amino acid compositions of autolytic fragments and CNBr-cleavage peptides from C1r A chain*
Amino acid composition of C1r A chain was taken from Arlaud *et al.* (1982) and re-evaluated to give 17 valine residues. Amino acid compositions of fragments and peptides were calculated from duplicate 24h HCl hydrolysates, and no correction was done. Cysteine was estimated as *S*-(carboxymethyl)cysteine. Methionine from CNBr-cleavage peptides was quantified as the sum of homoserine and homoserine lactone values. Results are expressed in mol of residues/mol, and values less than 0.2mol/mol are omitted. Values obtained from the sequence are indicated in parentheses.

Amino acid	Proteolytic fragments				C1r chain	CNBr-cleavage peptides							
	α	β	γ	Total	A	Total	CN1a	CN1b	CN2a	CN2b	CN3	CN4a	CN4b
Asx	17.1	11.2	15.3	43	43.0	43	27.5	5.0	2.7 (3)	2.0 (2)	3.9 (4)	1.0 (1)	1.1 (1)
Thr	8.7	2.6	12.1	24	30.0	26	11.6	3.7	2.5 (3)	2.7 (3)	2.5 (3)		0.8 (1)
Ser	16.0	3.6	4.8	25	28.6	24	16.2	3.9	1.0 (1)	1.8 (2)		0.8 (1)	
Glx	28.2	7.8	26.5	63	63.3	62	37.9	6.9	2.0 (2)	11.3 (11)	1.2 (1)	1.0 (1)	2.0 (2)
Pro	16.3	4.1	11.7	32	31.7	32	16.7	6.9		3.9 (4)	2.8 (3)		1.0 (1)
Gly	17.8	5.3	14.4	37	36.9	35	19.1	6.4	1.1 (1)	4.7 (5)	2.3 (2)	1.2 (1)	1.2 (1)
Ala	7.0	2.1	7.6	17	16.2	16	11.6	1.4		1.9 (2)			1.1 (1)
Cys	10.4	1.9	8.7	21	21.3	22	14.0	1.9		3.0 (3)	1.9 (2)		0.9 (1)
Val	9.9	2.4	9.2	21	17.0	17	8.4	4.1		3.6 (4)			1.0 (1)
Met	2.9		3.0	6	5.5	6	1.0	0.9	1.1 (1)		0.9 (1)	0.8 (1)	0.8 (1)
Ile	6.5	2.9	9.6	20	17.9	18	9.9	2.5	0.8 (1)	1.8 (2)	0.8 (1)		0.8 (1)
Leu	16.2	7.4	8.7	32	31.3	32	22.8	4.8	1.8 (2)	1.0 (1)	1.0 (1)		
Tyr	14.7	2.0	10.9	28	25.9	28	18.1	2.9		1.0 (1)	1.0 (1)		4.7 (5)
Phe	12.6	4.6	5.7	24	22.8	22	13.5	5.1	2.2 (2)		1.0 (1)		
His	4.5	2.9	3.6	12	10.7	10	8.0	0.3	1.0 (1)				0.9 (1)
Lys	11.5	3.1	12.4	26	25.4	24	9.5	5.3		3.8 (4)	1.8 (2)	0.9 (1)	1.7 (2)
Arg	9.2	2.0	11.0	22	23.6	25	13.6	2.0		4.7 (5)	3.0 (3)		1.0 (1)
Trp										(1)			
Total	210	66	176	453	452	442	261	64	17 (17)	49 (50)	25 (25)	6 (6)	20 (20)

cating a partial cleavage of the Arg–Tyr bond at positions 5 of fragment γ.

Purification and N-terminal sequence analysis of the CNBr-cleavage peptides from C1r A chain

Amino acid analysis of the whole C1r A chain indicated five or six methionine residues (Table 1), and six or seven peptides were therefore expected from CNBr cleavage. The CNBr digest was initially fractionated on Sephadex G-75 (Fig. 3), and four pools were collected and further analysed by high-pressure gel-permeation chromatography or reversed-phase h.p.l.c. Pool 1 from Sephadex G-75 contained two peptides, CN1a and CN1b, separated by high-pressure gel-permeation chromatography. The second pool from Sephadex G-75 also contained two major peptides, CN2a and CN2b, purified by reversed-phase h.p.l.c. Peptide CN2b appeared as a single peak, whereas peptide CN2a was resolved into four peaks of identical amino acid compositions. A single major peptide CN3 was found in pool 3 from Sephadex G-75, appearing as a double peak on reversed-phase h.p.l.c., whereas two peptides, CN4a and CN4b, were recovered from pool 4.

In summary, seven major peptides, CN1a, CN1b, CN2a, CN2b, CN3, CN4a and CN4b, resulting from CNBr cleavage of C1r A chain were purified. The sum of their individual amino acid compositions, shown in Table 1, is in agreement with the amino acid composition of the whole C1r A chain. *N*-Terminal sequence analyses of the peptides are indicated in Fig. 2.

The *N*-terminal amino acid sequence of peptide CN1a was determined for 32 cycles, whereas peptide CN1b was found to have a blocked *N*-terminus. It is noteworthy that peptide CN1b (approx. 64 residues) was eluted on Sephadex G-75 at the same position as peptide CN1a (approx. 261 residues), a behaviour due to the high tendency of peptide CN1b to aggregate, which also explains the low purification yield of this peptide. Peptide CN1a was found to contain *N*-acetylglucosamine, estimated as 2.1 mol/mol of peptide.

The *N*-terminal sequence of peptide CN2a was determined for 15 cycles, with the exception of the residue at position 13. As threonine was identified at position 15, it is likely that position 13 is occupied by an asparagine-linked carbohydrate moiety, which is in agreement with the characteristic Asn-Xaa-Thr sequence required for attachment of that type of carbohydrate (Neuberger *et al.*, 1972). The presence of asparagine at position 13 is consistent with the amino acid composition (Table 1) and complete sequence (see below) of peptide CN2a, which was also found to contain *N*-

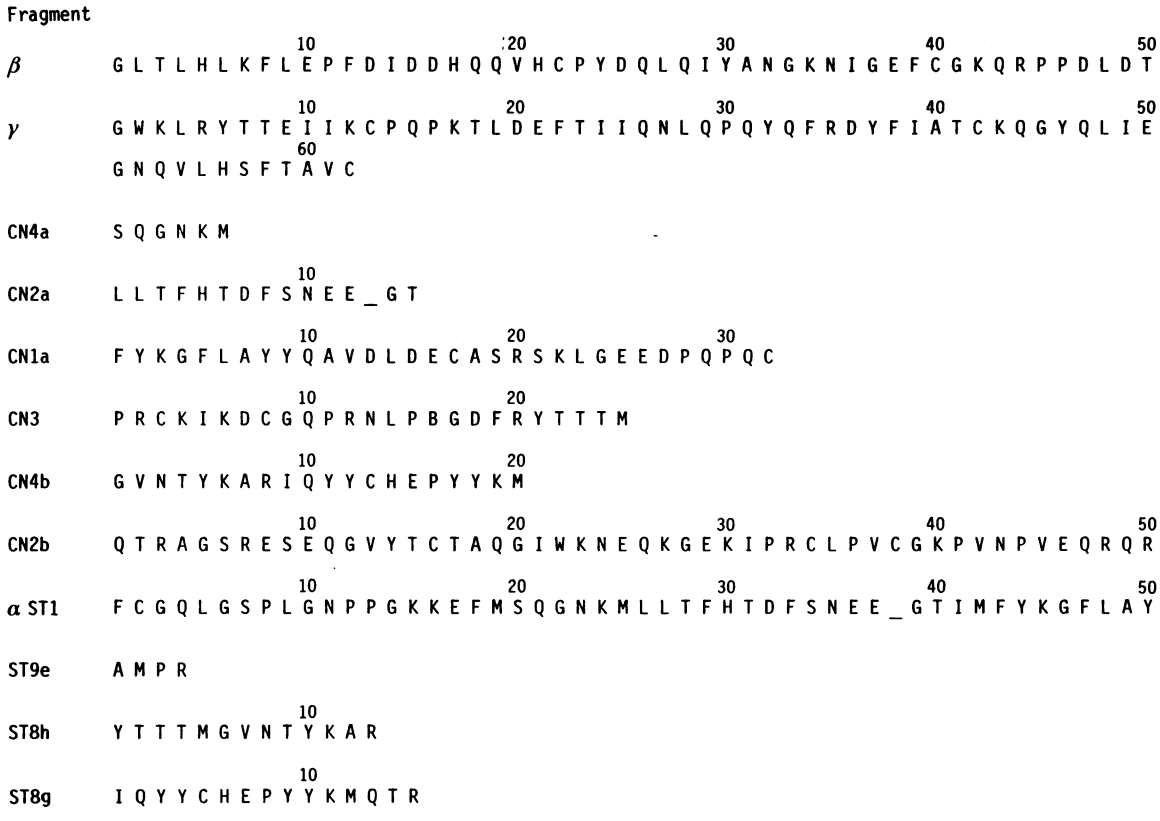

Fig. 2. *N-Terminal sequences of autolytic fragments β and γ, CNBr-cleavage peptides and methionine-containing tryptic-digest peptides from C1r A chain*

The single-letter notation is 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. _ indicates the presence of an Asn-linked carbohydrate moiety.

acetylglucosamine (1.4mol/mol of peptide). The heterogeneity of peptide CN2a on reversed-phase h.p.l.c. might arise partly from some heterogeneity of its carbohydrate moiety, although h.p.l.c. has already been reported to discriminate the homoserine and homoserine lactone forms of other CNBr-cleavage peptides (Johnson *et al.*, 1980; Kerlavage & Taylor, 1980; Arlaud *et al.*, 1982).

Peptide CN2b was the only CNBr-cleavage peptide lacking homoserine (Table 1), and was therefore identified as the *C*-terminal peptide. On the basis of the complete correlation between the residues identified by sequence (Fig. 2) and by amino acid analysis (Table 1), it can be assumed that the entire amino acid sequence of peptide CN2b has been determined.

The complete amino acid sequence of peptide CN3 (25 residues) was determined (Fig. 2), although equivalent amounts of asparagine and aspartic acid were found at position 16, which was therefore provisionally assigned as Asx. Complete amino acid sequences of peptides CN4a (6 residues) and CN4b (20 residues) were also obtained (Fig. 2). It is noteworthy that the recovery of peptide CN4a was readily improved when the CNBr cleavage was performed in trifluoroacetic acid, which indicates a more complete cleavage of the methionine–serine bond under these conditions.

Purification of the methionine-containing tryptic peptides from C1r A chain

The tryptic digest of reduced and *S*-[^{3}H]carboxymethylated succinylated C1r A chain was initially fractionated on Sephadex G-50 (Fig. 4), and methionine-containing pools 2, 8 and 9 were further analysed by reversed-phase h.p.l.c. From pool 8, two methionine-containing peptides ST8g and ST8h were isolated, the latter being also recovered from pool 9, together with peptide ST9e. The fourth methionine-containing tryptic peptide, αST1, was primarily found in pool 2, but could not

be separated from a peptide with the same *N*-terminal amino acid sequence as fragment β. Tryptic cleavage of reduced and *S*-[^{3}H]carboxymethylated succinylated fragment α also yielded peptide αST1, which was obtained in pure form from pool 1 of the fractionation on Sephadex G-50.

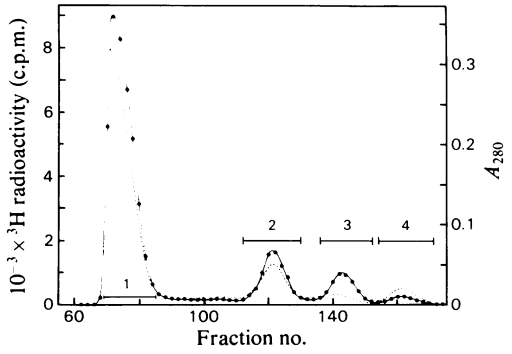

Fig. 3. *Initial fractionation of CNBr-cleavage peptides from reduced and S-[^{3}H]carboxymethylated CIr A chain by gel filtration on Sephadex G-75*

The CNBr digest of 250nmol of CIr A chain was loaded on a Sephadex G-75 column (2.5cm × 110cm) equilibrated in 0.1M-NH_4HCO_3 and eluted at a flow rate of 12.5ml/h. Fractions of volume 3ml were collected, and *S*-[^{3}H]carboxymethylcysteine radioactivity was determined on 20μl samples. Pools 1, 2, 3 and 4 were collected as indicated. ------, A_{280}; ●, radioactivity.

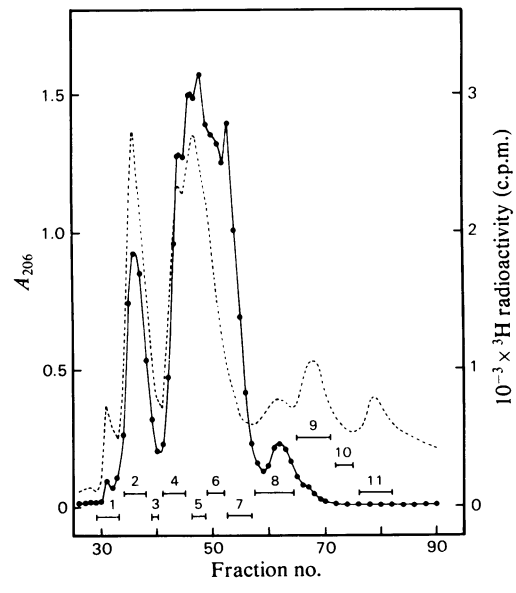

Fig. 4. *Initial fractionation of a tryptic digest of reduced and S-[^{3}H]carboxymethylated succinylated CIr A chain by gel filtration on Sephadex G-50*

The tryptic digest of 150nmol of CIr A chain was loaded on a Sephadex G-50 (superfine grade) column (2.0cm × 90cm) equilibrated in 0.1M-NH_4HCO_3 and eluted at a flow rate of 28ml/h. Fractions of volume 5ml were collected, and *S*-[^{3}H]-carboxymethylcysteine radioactivity was determined on 50μl samples. Pools were collected as indicated. ------, A_{206}; ●, radioactivity.

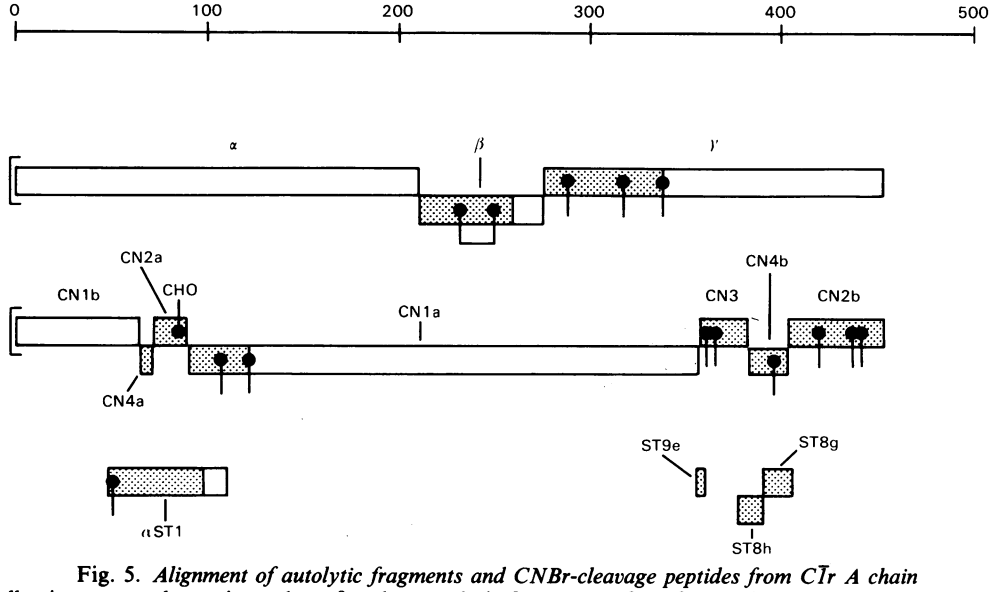

Fig. 5. *Alignment of autolytic fragments and CNBr-cleavage peptides from CIr A chain*

The following nomenclature is used: α, β and γ, autolytic fragments; CN, CNBr-cleavage peptides; ST, peptides from tryptic cleavage of succinylated CIr A chain. The known carbohydrate-attachment site is indicated by -CHO. Those half-cystine residues that have been identified are shown, as is the disulphide bridge in fragment β. Stippling denotes the portions of CIr A chain that have been sequenced. [indicates a blocked *N*-terminus.

Alignment of CNBr-cleavage peptides

The complete amino acid sequences of peptides ST8h and ST8g were obtained (Fig. 1), and provided overlaps between CNBr-cleavage peptides CN3 and CN4b on the one hand, and between peptides CN4b and CN2b on the other hand, thus establishing the alignment CN3–CN4b–CN2b (Figs. 2 and 5). The short peptide ST9e was clearly located at the junction between CN3 and the preceding CNBr-cleavage peptide (Figs. 2 and 5), whereas the *N*-terminal sequence of peptide αST1, determined for 50 cycles, overlapped four CNBr-cleavage peptides, among them CN4a, CN2a and CN1a, which were thus aligned in this order (Figs. 2 and 5). By taking into account that C$\bar{1}$r A chain has a blocked *N*-terminus (Sim *et al.*, 1977) and that CN1b is the only CNBr-cleavage peptide resistant to Edman degradation, it can therefore be postulated that peptide CN1b originates from the *N*-terminal end of C$\bar{1}$r A chain. As, on the other hand, peptide CN2b has been identified as the *C*-terminal CNBr-cleavage peptide (see above), it becomes apparent that C$\bar{1}$r A chain CNBr-cleavage peptides are aligned in the order (Fig. 5):

CN1b–CN4a–CN2a–CN1a–CN3–CN4b–CN2b

Alignment of autolytic fragments

Fragment α, which is the only autolytic fragment having a blocked *N*-terminus, can be assigned as the *N*-terminal fragment of C$\bar{1}$r A chain. Whereas fragments α and γ each contain three methionine residues, fragment β does not contain methionine (Table 1). Furthermore, its *N*-terminal sequence cannot be located within the sequence of the 95 *C*-terminal residues of C$\bar{1}$r A chain, known from the sequence of peptides CN3, CN4b and CN2b (Fig. 2). Superimposition of autolytic fragments and CNBr-cleavage peptides (Fig. 5) clearly shows, therefore, that fragment β cannot be *C*-terminal, which leads to the conclusion that autolytic fragments are aligned in the order: α–β–γ.

Discussion

N-Terminal sequence analysis of autolytic fragments, CNBr-cleavage peptides and methionine-containing tryptic peptides allowed the identification of about 60% of the sequence of C$\bar{1}$r A chain, which contains 440–460 residues, with a minimum M_r of 51 700, as estimated from the average amino acid composition derived from the whole A chain, the autolytic fragments and the CNBr-cleavage peptides. At least two carbohydrate moieties are attached to the polypeptide core of C$\bar{1}$r A chain: one site has been located in peptide CN2a, whereas peptide CN1a contains one, or possibly two, carbohydrate-binding site(s). In agreement with earlier results (Sim *et al.*, 1977), these carbohydrates all contain *N*-acetylglucosamine, indicating a linkage to the peptide chain via asparagine residues (Neuberger *et al.*, 1972). This is established for the site located in peptide CN2a, which occurs in the typical Asn-Gly-Thr sequence.

In agreement with previous findings (Arlaud *et al.*, 1980*b*), autolytic cleavage of activated C$\bar{1}$r occurs in the *N*-terminal A chain, which undergoes two major cleavages yielding three fragments, α, β and γ, located in that order within the A chain. These major fragments (approx. 210, 66 and 176 residues long respectively) were characterized and they can account for the whole C$\bar{1}$r A chain, although the possibility of other small peptides cannot be excluded. With regard to the specificity of the major autolytic cleavages, it is very likely that both occur at arginine–glycine bonds. This is supported by the following evidence: (i) both β and γ fragments have *N*-terminal glycine; (ii) tryptic cleavage of the succinylated C$\bar{1}$r A chain generates a peptide with the same *N*-terminal sequence as fragment β; (iii) the arginine–tyrosine bond at position 5 of fragment γ is only partially cleaved during the autolytic process.

The last piece of evidence also indicates that the autolytic cleavage process is not strictly specific for Arg–Gly bonds, but rather for arginyl bonds, the cleavage sites being selected primarily from their accessibility, as expected for limited proteolytic cleavages. The strict arginyl specificity of the enzyme responsible for the autolytic process, i.e. C$\bar{1}$r, is also supported by the absence of cleavage of the two lysyl bonds at positions 7 and 3 of fragments β and γ respectively.

C$\bar{1}$r A chain contains six methionine residues, and CNBr cleavage yielded seven major peptides, CN1b, CN4a, CN2a, CN1a, CN3, CN4b and CN2b, located in that order within the chain. From the sequence of peptides CN3, CN4b and CN2b, the continuous sequence of the 95 *C*-terminal residues of C$\bar{1}$r A chain was determined. The *C*-terminal end of this region shows homology with the corresponding part of plasminogen and chymotrypsinogen A (Fig. 6). In this respect, it is noteworthy that the half-cystine residue located at position 38 of peptide CN2b (Figs. 2 and 6) is homologous with those participating in interchain disulphide bridges in these serine proteinases. On this basis, we suggest that this residue is involved in a disulphide bridge linking the A and B chains, which, in C$\bar{1}$r, are very likely to be connected by a single disulphide bridge, considering that fragment γ (Table 1) and C$\bar{1}$r B chain (Arlaud & Gagnon, 1983), which remain disulphide-bonded after the autolytic cleavage, both contain odd numbers of half-cystine residues (i.e. nine and five respectively).

$C\bar{1}r$ A chain peptide CN2b	C	G	K	P	-	V	N	P	V	E	Q	R	Q	R	
Plasminogen	C	G	K	P	-	-	Q	-	V	E	P	K	K	C	P	G	R
Haptoglobin	C	G	K	P	-	K	N	P	A	N	P	V	Q				
Chymotrypsinogen A	C	G	V	P	A	I	Q	P	V	L	S	G	L	S	R		
	1				5					10					15		

Fig. 6. *Structure of the C-terminal end of $C\bar{1}r$ A chain: homology with other serine proteinases and with the α1 chain of haptoglobin 1-1*

The sequence of the *C*-terminal 13 residues of peptide CN2b is compared with the homologous parts of human plasminogen (Young *et al.*, 1978), α1 chain of human haptoglobin 1-1 (Kurosky *et al.*, 1980) and bovine chymotrypsinogen A (Young *et al.*, 1978). - denotes that a gap was left to give maximum homology on alignment of amino acid residues. The residue numbering indicated is that of chymotrypsinogen.

As the entire sequence of peptide CN2b has been unambiguously determined, it can be postulated that, if a single bond is cleaved during C1r activation, this bond links, in proenzyme C1r, the *C*-terminal arginine residue of CN2b and the *N*-terminal isoleucine residue of the B chain moiety (Arlaud & Gagnon, 1983). However, a secondary cleavage cannot be excluded, so long as the peptide overlapping the A and B chain moieties has not been isolated and sequenced. The *C*-terminal region of $C\bar{1}r$ A chain also shows a striking homology with the *C*-terminal end of the α1 chain of human haptoglobin 1-1, a haemoglobin-binding protein that is a serine proteinase homologue (Kurosky *et al.*, 1980). This homology is not restricted to the extreme *C*-terminal end of $C\bar{1}r$ A chain, shown in Fig. 6, but is also found in the rest of the sequence of peptide CN2b, in which it amounts to 43% for the last 44 residues. From the point of view of evolution, this high degree of homology is interesting, as the haptoglobin gene is thought to result from a duplication of a serine proteinase precursor gene that subsequently diverged (Kurosky *et al.*, 1980). Within this context, it must be noticed that the other known parts of $C\bar{1}r$ A chain do not show obvious homology with the Kringle regions of plasminogen or prothrombin (Young *et al.*, 1978).

We are grateful to Professor R. R. Porter and Professor M. G. Colomb for their continuous help and encouragement. We thank Mr. A. C. Willis for amino acid sequence determination, Mr. A. Gascoyne for performing the amino acid analysis and Mrs. L. A. Weaver for technical assistance.

References

Arlaud, G. J. & Gagnon, J. (1983) *Biochemistry* **22**, 1758–1764

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. (1980*a*) *Biochim. Biophys. Acta* **616**, 105–115

Arlaud, G. J., Villiers, C. L., Chesne, S. & Colomb, M. G. (1980*b*) *Biochim. Biophys. Acta* **616**, 116–129

Arlaud, G. J., Gagnon, J. & Porter, R. R. (1982) *Biochem. J.* **201**, 49–59

Assimeh, S. N., Chapuis, R. M. & Isliker, H. (1978) *Immunochemistry* **15**, 13–17

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

Dodds, A. W., Sim, R. B., Porter, R. R. & Kerr, M. A. (1978) *Biochem. J.* **175**, 383–390

Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry* **17**, 2124–2133

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

Koide, A., Titani, K., Ericson, L. H., Kumar, S., Neurath, H. & Walsh, K. A. (1978) *Biochemistry* **17**, 5657–5672

Kurosky, A., Barnett, D. R., Lee, T. H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H. & Fitch, W. M. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3388–3392

Neuberger, A., Gottschalk, A., Marshall, R. D. & Spiro, R. G. (1972) in *Glycoproteins* (Gottschalk, A., ed.), 2nd edn., vol. 5, part A, pp. 450–490, Elsevier, Amsterdam

Okamura. K. & Fujii, S. (1978) *Biochim. Biophys. Acta* **534**, 258–266

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

Sim, R. B. (1981) *Methods Enzymol.* **80**, 6–16

Sim, R. B., Porter, R. R., Reid, K. B. M. & Gigli, I. (1977) *Biochem. J.* **163**, 219–227

Valet, G. & Cooper, N. R. (1974) *J. Immunol.* **112**, 1667–1673

Villiers, C. L., Arlaud, G. J., Painter, R. H. & Colomb, M. G. (1980) *FEBS Lett.* **117**, 289–294

World Health Organisation (1968) *Bull. W. H. O.* **39**, 935–938

World Health Organisation (1981) *Bull. W. H. O.* **59**, 489–490

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. (1982) *J. Immunol.* **128**, 2500–2504

Ziccardi, R. J. & Cooper, N. R. (1976*a*) *J. Immunol.* **116**, 496–503

Ziccardi, R. J. & Cooper, N. R. (1976*b*) *J. Immunol.* **116**, 504–509