Biochem. J. (1983) **215**, 565–571
Printed in Great Britain

The serine proteinase chain of human complement component C$\overline{1s}$

Cyanogen bromide cleavage and *N*-terminal sequences of the fragments

Philip E. CARTER, Bryan DUNBAR and John E. FOTHERGILL
Department of Biochemistry, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, Scotland, U.K.

(*Received 26 April 1983/Accepted 5 August 1983*)

Human complement component C1s was purified from fresh blood by conventional methods of precipitation and chromatography. The single-chain zymogen form was activated by treatment with C$\overline{1r}$. Reduction and carboxymethylation then allowed the light chain and heavy chain to be separated on DEAE-Sepharose CL-6B in 8M-urea. Liquid-phase sequencing of the light chain determined 50 residues from the *N*-terminus. CNBr-cleavage fragments of the light chain were separated by high-pressure liquid chromatography on gel-permeation and reverse-phase columns. *N*-Terminal sequencing of these fragments determined the order of a further 138 residues, giving a total of 188 residues or about 75% of the light chain. Seven of these eight sequences could be readily aligned with the amino acid sequences of other serine proteinases. The typical serine proteinase active-site residues are clearly conserved in C$\overline{1s}$, and the specificity-related side chain of the substrate-binding pocket is aspartic acid, as in trypsin, consistent with the proteolytic action of C$\overline{1s}$ on C4 at an arginine residue. Somewhat surprisingly, when the C$\overline{1s}$ sequence is compared with that of complement subcomponent C$\overline{1r}$, the percentage difference (59%) is approximately the same as that found between the other mammalian serine proteinases (56–71%).

The complement system consists of more than a dozen plasma proteins than can be activated by either the classical pathway or the alternative pathway to produce cell lysis as well as inflammatory and other reactions important in the body's defence mechanisms (for review see Fothergill & Anderson, 1978; Reid & Porter, 1981). Activation of the classical pathway is initiated when antigen–antibody complexes or antibodies bound to cell-surface antigens interact with the first component of complement, C1. The first component consists of three subcomponents C1q, C1r and C1s that in plasma exist as a Ca^{2+}-dependent complex (Lepow *et al.*, 1963). Binding takes place between the globular heads of the six-headed C1q subcomponent and the C_H2 domain of the immunoglobulin. C1q contains more than 30% collagen-like sequence arranged in a 'bunch of tulips' structure (Reid & Porter, 1976). Binding of C1q to immunoglobulin leads to the conversion from zymogen into active enzyme of the two other subcomponents C1r and C1s, that are thought to be located near the collagen-fibril-like or 'flower pot' end of the C1q structure. The proteinase activity of C$\overline{1s}$ then leads to proteolytic activation of C4 and C2, giving rise to the C3 convertase of the classical pathway.

Activation of the C1 complex is not well understood. Both C1r and C1s must be present with C1q for activation to occur, although C1r is activated even if the C1s component is in an inactivated form (Dodds *et al.*, 1978). Both C1r and C1s are single-chain zymogen molecules (M_r approx. 85000) that are split on activation to disulphide-linked heavy (M_r approx. 58000) and light (M_r approx. 27000) chains. The active serine proteinase site that can be inhibited by iPr_2P-F (Becker, 1956) is located on the light chain (Barkas *et al.*, 1973; Sim & Porter, 1976).

The *N*-terminal sequences of the CNBr-cleavage fragments of the light chain of human C$\overline{1r}$ have been reported by Arlaud *et al.* (1982), and show that C$\overline{1r}$ contains a classical serine proteinase of the trypsin type with an active-site sequence similar to that of C$\overline{1s}$ (Carter *et al.*, 1982). We now report here the *N*-terminal sequences of the light chain of human C$\overline{1s}$ and its CNBr-cleavage fragments.

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

Materials and methods

Materials

The sources of materials have been described previously (Scott & Fothergill, 1982; Smith *et al.*, 1982; Campbell *et al.*, 1979*a*,*b*; Booth *et al.*, 1979).

Purification of C1s

A combination of euglobulin precipitation and chromatography on DEAE-cellulose, Ultrogel AcA-34 and DEAE-Sephadex was used to purify C1s (Barkas *et al.*, 1973; Campbell *et al.*, 1979*b*; Ziccardi & Cooper, 1976). The purification schedule is shown in Scheme 1. Euglobulin precipitation by dialysis of fresh serum (3 litres) at pH 6.1, was followed by gradient elution (2 × 4 litres, 40–200 mM-NaCl) from a column (5 cm × 40 cm) of DEAE-cellulose in 50 mM-sodium phosphate buffer (pH 7.4)/1 mM-EDTA/40 mM-NaCl, conductivity 9.3 mS/cm. Fractions containing C1s, located by immunodiffusion, were pooled and concentrated before application to an Ultrogel AcA-34 column (2.5 cm × 140 cm) in 10 mM-sodium phosphate buffer (pH 6.6)/1 mM-EDTA, conductivity 11 mS/cm. After location by immunodiffusion, C1s-containing fractions were applied to a DEAE-Sephadex A-50 column (2.5 cm × 30 cm) in 10 mM-sodium phosphate buffer (pH 6.6)/1 mM-EDTA/100 mM-NaCl, conductivity 11 mS/cm. A linear gradient (2 × 0.5 litres, 100–500 mM-NaCl) in the same buffer was used to elute the C1s.

Purification of C1r

C1r was purified by the method of Ziccardi & Cooper (1976), except that Ultrogel AcA-34 was used as above for the gel-filtration step.

Activation of C1s

C1s (4.5 mg/ml) was activated by incubation with C1r (which is autoactivated to C$\overline{1\text{r}}$ under the incubation conditions used) at 37°C for 3 h in 10 mM-sodium phosphate buffer, pH 6.6. The C1r was added in four equal amounts at intervals of 30 min to give a final C1r:C1s ratio of 1:25 (w/w). The degree of activation was checked by SDS/polyacrylamide-gel electrophoresis under reducing conditions (Campbell *et al.*, 1979*b*).

Treatment of C$\overline{1\text{s}}$ with iPr$_2$P-F

A 1 mCi portion of [1,3-^{3}H]iPr$_2$*P*-F (3 Ci/mmol) was incubated with 78 mg of C$\overline{1\text{s}}$ in 10 mM-sodium phosphate buffer, pH 6.6, at 37°C for 1 h. A 5-fold molar excess of iPr$_2$*P*-F was then added, and the reaction mixture was incubated for a further 2 h. The labelled C$\overline{1\text{s}}$ was then exhaustively dialysed against 1% NaCl followed by 1% NH_4HCO_3, and subsequently freeze-dried.

Reduction and carboxymethylation

This was carried out with the use of iodo[2-^{14}C]acetic acid as previously described (Campbell *et al.*, 1979*b*).

Separation of heavy chain and light chain of reduced carboxymethylated C$\overline{1\text{s}}$

This was carried out by ion-exchange chromatography in 8 M-urea on DEAE-Sepharose CL-6B (Campbell *et al.*, 1979*b*), and checked by SDS/polyacrylamide-gel electrophoresis.

Digestion with CNBr

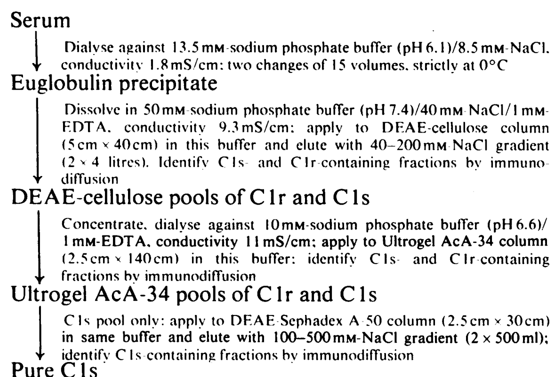
Reduced carboxymethylated C$\overline{1\text{s}}$ light chain (160 nmol) was dissolved in 70% (v/v) formic acid containing 1 mmol of CNBr and kept in the dark for 24 h at 4°C. The solution was diluted 1:10 with water and freeze-dried.

Separation of CNBr-cleavage fragments

This was carried out initially by h.p.l.c. on TSK SW-3000 and SW-2000 columns in 0.1% (v/v) trifluoroacetic acid as previously described (Smith *et al.*, 1982). Separation of fragments was monitored by SDS/polyacrylamide-(15%)-gel electrophoresis. Unresolved mixtures of fragments were subjected to reverse-phase h.p.l.c. on a Waters C-18 μBondapak column (0.39 cm × 30 cm) in 0.1% (v/v) trifluoroacetic acid with a gradient (0–60%) of propan-2-ol (Mahoney & Hermodson, 1980) or a gradient (0–30%) of acetonitrile/methanol/propan-2-ol (1:1:1, by vol.) (Arlaud *et al.*, 1982). Eluted peptides were detected by A_{214} and checked for purity by SDS/polyacrylamide-gel electrophoresis as above.

Amino acid analysis

This was done as previously described (Campbell *et al.*, 1979*a*), with a ninhydrin detection system, or by h.p.l.c. elution followed by hypochlorite treatment and detection by fluorescence after reaction with



Scheme 1. *Schedule for the purification of C1s*

o-phthalaldehyde with the use of the Waters amino acid analysis accessory.

N-Terminal sequencing

This was carried out in a Beckman 890C liquid-phase sequencer fitted with the Beckman cold-trap accessory, by using a 0.25M-Quadrol programme with Polybrene; identification of phenylthiohydantoin derivatives was done by reverse-phase h.p.l.c. on a C-18 μBondapak column with a pH 3.4 acetate/methanol buffer system (Smith *et al.*, 1982), or on a Waters 5μ Spherical C18 Resolve column (0.39 cm × 15 cm) with a pH 5.0 acetate/acetonitrile buffer system.

Results

Purification of C1s

The purification schedule outlined in Scheme 1 gave about 50 mg of pure C1s from 3 litres of fresh serum. The purified material showed only a single band of M_r approx. 85000 in SDS/polyacrylamide-gel electrophoresis with or without reduction, indicating that it was in its single-chain zymogen form.

Activation of C1s

Incubation of C1s with C$\overline{1r}$ converted the zymogen C1s into the active form, which showed two bands (M_r approx. 57000 and approx. 28000) in SDS/polyacrylamide-gel electrophoresis run under reducing conditions.

Treatment of C$\overline{1s}$ with iPr$_2$P-F

C$\overline{1s}$ incubated with [1,3-^{3}H]iPr$_2$*P*-F was found to contain tritium, which was subsequently located in the light chain, as shown previously (Barkas *et al.*, 1973).

Separation of heavy chain and light chain of C$\overline{1s}$

The reduced carboxymethylated chains of iPr$_2$*P*-F-treated C$\overline{1s}$ were separated on DEAE-Sepharose CL-6B. The light chain was eluted with the starting buffer, and the heavy chain was eluted about half way up the gradient, in a manner very similar to that observed with bovine C$\overline{1s}$ (Campbell *et al.*, 1979*b*). Both heavy chain and light chain were found to be labelled with ^{14}C, but only the light chain contained ^{3}H. Amino acid analysis of the light chain is shown in Table 1.

Separation of CNBr-cleavage fragments of C$\overline{1s}$ light chain

Gel-permeation h.p.l.c. separation of CNBr-cleavage fragments of C$\overline{1s}$ light chain is shown in Fig. 1. Fractions were pooled as indicated. Fractions CN-1, CN-2 and CN3 were found to be homogeneous by SDS/polyacrylamide-gel electrophoresis and were sequenced without further purification. Fractions CN-4 and CN-5 required further purification and were concentrated by freeze-drying

Table 1. *Amino acid composition of C$\overline{1s}$ light chain and its CNBr-cleavage fragments*

The results shown are those for 24 h hydrolysis time and have not been corrected. Cysteine was measured as *S*-carboxymethylcysteine. + represents detection of an amino acid corresponding to approximately one residue that could not be accurately measured.

	Amino acid composition (residues/molecule)									
	CN-1	CN-2	CN-3	CN-4a	CN-4b	CN-5a	CN-5b	CN-6	Sum CN-1–CN-6	C$\overline{1s}$ light chain
Cys	—	1.2	—	—	0.9	—	1.0	1.2	4.3	6.5
Asx	6.8	5.3	5.4	—	5.4	1.6	0.8	2.3	27.6	27.5
Thr	2.1	2.5	2.2	1.8	2.3	0.9	—	2.1	13.9	12.3
Ser	2.0	1.1	0.7	3.8	3.1	0.9	—	2.5	14.1	12.5
Glx	4.3	4.1	3.3	1.1	5.7	3.0	1.3	1.1	23.9	24.3
Pro	3.3	4.2	3.8	—	2.6	1.0	—	2.5	17.4	15.5
Gly	4.9	4.5	2.5	1.2	8.6	0.8	2.7	2.1	27.3	25.2
Ala	4.1	6.5	1.4	1.1	3.3	—	1.2	—	17.6	16.4
Val	3.5	4.3	3.2	1.7	2.2	—	—	1.2	16.1	14.5
Met	—	—	—	—	—	—	—	—	—	6.2
Ile	2.2	0.9	1.6	—	1.1	—	0.8	0.9	7.5	7.9
Leu	2.4	4.3	4.1	1.0	2.1	—	—	1.5	15.4	12.9
Tyr	1.7	0.9	—	1.0	1.3	—	—	0.8	5.7	7.8
Phe	2.4	1.0	1.6	—	1.3	—	—	—	6.3	8.1
His	2.6	—	1.6	—	—	—	—	—	4.2	4.1
Lys	1.9	5.0	2.7	1.5	2.1	—	—	—	13.2	15.4
Arg	1.4	5.6	1.9	+	1.3	+	—	—	12.2	12.2
Hse	+	+	+	+	+	—	+	+	7	
Total	46.6	52.4	37.0	16.2	44.3	9.2	8.8	19.2	233.7	229.3

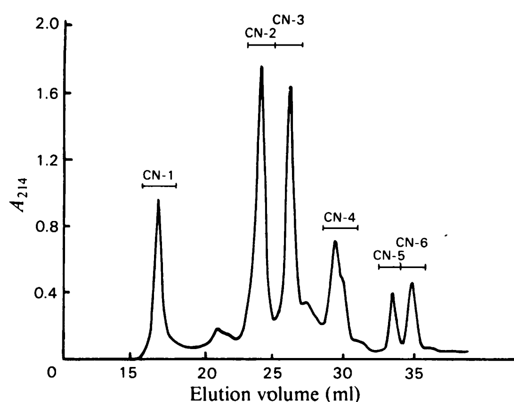


Fig. 1. *Separation of the CNBr-cleavage fragments of the light chain of reduced carboxymethylated human $C\overline{1s}$ by gel-permeation h.p.l.c.*

A CNBr digest of $C\overline{1s}$ light chain (160nmol) was dissolved in 3.0ml of 0.1% trifluoroacetic acid. Samples (150μl) were applied to columns (0.37cm x 60cm) of TSK SW-3000 and SW-2000 used together in order of decreasing pore size. Flow rate was 1.0ml/min from a Waters 6000A pump. Detection was at 214nm, with a Waters 441 monitor fitted with a cadmium lamp source. Six pools (CN-1 to CN-6) were collected as indicated.

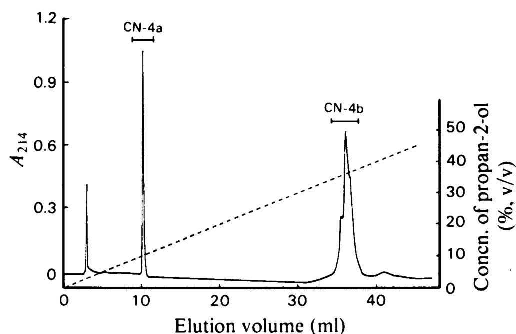


Fig. 2. *Separation of fragments of CN-4a and CN-4b by reverse-phase h.p.l.c.*

The CN-4 pool from the gel-permeation separation was freeze-dried and dissolved in 0.1% trifluoroacetic acid. A sample (13nmol in 100μl) was applied to a Waters C-18μBondapak column (0.39cm x 30cm) equilibrated in 0.1% trifluoroacetic acid. A gradient of 0–60% propan-2-ol was applied over 60min at 1.0ml/min. Fragments CN-4a and CN-4b were pooled as indicated. ——, A_{214}; -----, concn. of propan-2-ol.

before being subjected to reverse-phase h.p.l.c. Two fragments CN-4a and CN-4b were obtained from fraction CN-4 (Fig. 2); fragments CN-5a and CN-5b were obtained from fraction CN-5. Sequencing of fraction CN-6 gave a single sequence so it was not purified further.

Amino acid analyses of CNBr-cleavage fragments

These are shown in Table 1 together with the analysis of whole $C\overline{1s}$ light chain. The sum of analyses for the CNBr-cleavage fragments is quite similar to the analysis of the whole $C\overline{1s}$ light chain, suggesting that all the CNBr-cleavage fragments have been found, or that any remaining fragment is very small. The number of fragments is also consistent with the methionine content of the whole chain.

N-Terminal sequencing of CNBr-cleavage fragments

The results of sequencer experiments on the purified CNBr-cleavage fragments are shown in Fig. 3, together with the results of sequencing the whole $C\overline{1s}$ light chain. The amino acid sequences of seven of the eight fragments are sufficiently homologous with the sequences of other serine proteinases to allow alignment of the fragments in an obvious order within the molecule. The remaining fragment, CN-5a, gave the sequence Gln-Glu-Asn-Ser-Thr-Pro-Arg-Glu-Asp-, which is not obviously homologous with other serine proteinases but probably fits in the *C*-terminal region, where variability among the homologous sequences is quite large. The absence of homoserine from the analysis of fragment CN-5a further supports the *C*-terminal location of the fragment.

Discussion

We have determined the amino acid sequence of 188 residues of the $C\overline{1s}$ light chain, representing about three-quarters of the complete chain. Nearly all the CNBr-cleavage fragments were sequenced to within a few residues of their *C*-termini. Most of the residues not so far sequenced are in the *C*-terminal part of the largest fragment CN-4b located between the sequence containing the active-site serine residue and the *C*-terminal fragment CN-5a.

The active-site residue was identified by the detection of radioactivity in the seventh cycle of Edman degradation of fragment CN-4b obtained from $C\overline{1s}$ treated with [^{3}H]iPr$_2$*P*-F. This residue was characterized as a serine residue by sequencing an unlabelled sample of fragment CN-4b,

Our *N*-terminal sequence for $C\overline{1s}$ is identical with results obtained by sequencer runs on the intact light chain (Sim *et al.*, 1977; Arlaud & Gagnon, 1981).

The obvious homology of the sequences of the $C\overline{1s}$ CNBr-cleavage fragments with the sequences of the established serine proteinases allows the alignment of sequences shown in Fig. 3. The conservation of the active-site residues His-57, Asp-102

and Ser-195, as well as the *N*-terminal hydrophobic residue Ile-16 generated during activation from the single-chain zymogen, is obvious (chymotrypsinogen numbering system of Hartley, 1964). There is also clear conservation of a large number of other amino acid residues, that can be interpreted more easily in the context of the tertiary structure of the molecule. We have built a model of the C$\overline{1s}$ light-chain sequence by using the co-ordinates established for chymotrypsin by Birktoft & Blow (1972). Identical residues in the homologous sequences have been assumed to occupy identical co-ordinates, and rotational angles of the polypeptide backbone have been retained as closely as possible. In almost all cases the C$\overline{1s}$ sequence fitted the chymotrypsin structure quite satisfactorily, the major discrepancies being the insertion or deletion of external loops of residues, or in several cases the replacement of proline. The final structure shows retention of many of the internal residues, or in some cases conservative replacements. The disulphide bridge at residues 168–182 (chymotrypsin numbering) is conserved, and probably the bridge at 191–220. The cysteine residue at position 122, which is connected to the A-chain in chymotrypsin, probably forms the disulphide link to the heavy chain in C$\overline{1s}$, as its analogue in prothrombin and plasminogen does. The disulphide bridge at residues

```
                         |-- C1s L -------------------------------------------------------------------------
                         |-- CN-1 -->
                         1                   11                  21                  31                  41
C1s light chain (human)  I I G G S D A D I K N F P W Q V F F D N P - - - - - - - - - W A G G A L I N E Y W V L T A A H V V E
C1r light chain (human)  I I G G Q K A K M G N F P W Q V F T N I H G R - - - - - - - - G G G A L L G D R W I L T A A H T L Y
t-PA (human)             I K G G L F A D I A S H P W Q A A I F A K H R R S P G E R F L C G G I L I S S C W I L S A A H C F Q
Plasmin (human)          V V G G C V A H P H S W P W Q V S L R T R F G - - - - - M H F C G G T L I S P E W V L T A A H C L E
Factor X (bovine)        I V G G R D C A E G E C P W Q A L L V N E E N - - - - - E G F C G G T I L N E F Y V L T A A H C L H
Thrombin (bovine)        I V E G Q D A E V G L S P W Q V M L F R K S P Q - - - - E L L C G A S L I S D R W V L T A A H C L L
Kallikrein (pig)         I I G G R E C E K N S H P W Q V A I Y H Y S - - - - - - S F Q C G G V L V N P K W V L T A A H C K N
Elastase (pig)           V V G G T E A Q R N S W P S Q I S L Q Y R S G S S - - W A H T C G G T L I R Q N W V M T A A H C V D
Trypsin (bovine)         I V G G Y T C G A N T V P Y Q V S L N - - S G - - - - - Y H F C G G S L I N S Q W V V S A A H C Y K
Chymotrypsin A (bovine)  I V N G E E A V P G S W P W Q V S L Q D K T G - - - - - F H F C G G S L I N E N W V V T A A H C G V
                        (16)                          (31)                          (41)                (51)           ★

                         ------------------------------>
                                                         |-------- CN-4a ------->     |------------ CN-3 ------------
                         51                  61                  71                  81                  91
C1s light chain (human)  G N R E P X - - - - - - - - - - M Y V G S T S V Q T S R L A K S     L T - P E H V F I H P G W K L L
C1r light chain (human)  P K E H E A Q S N - - - A S L D V F L G H T N V - - - - - - - - - - E E - L M K L G N H P I R R V S
t-PA (human)             E R F P P H H - - L T - - - - - V I L G R T Y R V V P G E E E Q - K F E - V E K Y I V H K E F - - -
Plasmin (human)          K S P R P S S - - Y K - - - - - V I L G A H Q E V N L E P H V Q E I E V - S R L F L E P T R K - - -
Factor X (bovine)        Q A K R - - - - - F T - - - - - V R V G - - D R N T Q E G D E E M A H E - V E M T V K H S R F - - -
Thrombin (bovine)        - - Y P P W B K N F T V D D L L V R I G K H S R T R Y E R K V E K I S M - L D K I Y I H P R Y - - -
Kallikrein (pig)         D N Y E - - - - - - - - - - - - - - V G W L R H N L F E N E N T A Q F F G V T A D F P H P G F - - -
Elastase (pig)           R E L T - - - - - F R - - - - - V V V G E H N L N Q N N G T E Q Y V G - - V Q K I V V H P Y W - - -
Trypsin (bovine)         S G I Q - - - - - - - - - - - - V R L G Q D N I N V V E G N Q Q F I S - - A S K S I V H P S Y - - -
Chymotrypsin A (bovine)  - - - - - - - - - - T T S D - V V V A G E F D Q G S S S E K I Q K L K - - I A K V F K N S K Y - - -
                                            (61)                    (71)                (81)                  (91)
```

Fig. 3. *Alignment of the amino acid sequences of the CNBr-cleavage fragments of human C$\overline{1s}$ light chain with sequences of other serine proteinases*

The initial and repetitive yields for the sequencer runs were: C1s light chain, 40% (20nmol), 95% (Ala-7 to Ala-36); fragment CN-1, 7% (1nmol), 99% (linear-regression analysis); fragment CN-2, 25% (2.5nmol), 95% (Leu-4 to Leu-27); fragment CN-3, 28% (5.5nmol), 96% (Leu-1 to Leu-14); fragment CN-4a, 17% (4.3nmol), 97% (Val-2 to Val-7); fragment CN-4b, 13% (12nmol), 97% (Gly-8 to Gly-31); fragment CN-5a, 33% (3.0nmol), 78% (Glu-2 to Glu-8); fragment CN-5b, 23% (13nmol), 99% (linear-regression analysis); fragment CN-6, 36% (14nmol), 96% (linear-regression analysis). The human C$\overline{1r}$ light-chain sequence is from Arlaud *et al.* (1982) and Arlaud & Gagnon (1983), and the human tissue plasminogen activator (t-PA) sequence is from Pennica *et al.* (1983). All other sequences are from Young *et al.* (1978). Gaps introduced to increase homology are represented by –. Active-site residues His-57, Asp-102 and Ser-195, and side-chain-specificity-determining residue 189, are marked by ★. The chymotrypsinogen numbering system (Hartley, 1964) is shown in parentheses. 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. 3—*continued*

```
                             ———— CN-3 ————————→                     |———— CN-6 ——————————→
                            101        111         121           131          141
C1s light chain (human)     E V P E G R T N F D - - - - N D I A L V R           G - - - - P T V S P I C L P G T S S D Y N
C1r light chain (human)     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 L L P I C L P - - - - - D N D
t-PA (human)                - - - - - - - - D D D T Y D N D I A L L Q L K S D S S R C A Q E S S V V R T V C L P P A D L - - - Q
Plasmin (human)             - - - - - - - - - - - - - - - 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 - - - V
Factor X (bovine)           - - - - - - - - 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 A E T
Thrombin (bovine)           - - - - - - - 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 A K L
Kallikrein (pig)            - - - - N L S A D G K D Y S H D L M L L R L Q S P A K I - - - - T D A V K V L E L P T Q E P - - E L
Elastase (pig)              - - - - - - N T D D V A A 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 - - - I
Trypsin (bovine)            - - - - - - - - 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 - - - -
Chymotrypsin A (bovine)     - - - - - - - - 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 - - - D
                                           (95)      ★             (111)                 (121)

                              |——————————— CN-2 ——————————————————————————————————————————
                            151          161            171            181           191
C1s light chain (human)       D G D L G L - - - I S G W G R - - - - - - - A E K R D R A V R L K A A R L P V A P L R K C K E V
C1r light chain (human)     T F Y - D L G L M G Y V S G F G V - - - - - - - M E E K - I A H D L R F V R L P V A N P Q A C E N W
t-PA (human)                L P D W T E C E L - - - S G Y G K - - - - - H E A L S P F Y S E R L K E A H V R L Y P S S R C T S Q
Plasmin (human)             V A D R T E C F I - - - T G W G E - - - - - T Q - - G T F G A G L L K E A Q L P V I E N K V C N R Y
Factor X (bovine)           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 C K L S
Thrombin (bovine)           L H A G F K - - - G R V T G W G N 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 C K A S
Kallikrein (pig)            - - - G S T C Q A - - - S G W G
Elastase (pig)              L A N N S P C Y I - - - T G W G L - - - - - T R - T N G Q L A Q T L Q Q A Y L P T V D Y A I C S S S
Trypsin (bovine)            - 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 C K S A
Chymotrypsin A (bovine)     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 L S N T N C K K -
                            (131)          (141)                  (151)          (161)

                            ——→                    |—— CN-5b ——→          |———— CN-4b ————————————
                            201                         221                231           241
C1s light chain (human)     K V E                   I C A G - - - G E K            D S C K G D S G G A F A V Q D P N D K G
C1r light chain (human)     L R G K N R 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 V R D P N T D R
t-PA (human)                - - - H L L N R T V T D N M L C A G D T R S G G P Q A N L H D A C Q G D S G G P L V - - C L N D - G
Plasmin (human)             - - - 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 - D
Factor X (bovine)           - - - - 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 - D
Thrombin (bovine)           - - - - 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 - N
Kallikrein (pig)            - - - - - - - - K V T E S M L C A G Y L P G G K - - - - - - D T C M G D S G G P L I - - C N G M - -
Elastase (pig)              - - - 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 - G
Trypsin (bovine)            - - - - 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 - -
Chymotrypsin A (bovine)     - - - - 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 - G
                                (171)          (181)                    ★(191)    ★                 (201)

                            ————————————→
                            251         261          271          281          291
C1s light chain (human)     R F Y A A G L X S W G
C1r light chain (human)     - W V A T G I V S W G - - I G C S R - - G Y G F Y T K V L N Y V D W I K K E M E E E D
t-PA (human)                R M T L V G I I S W G - - L G C G Q K D V P G V Y T K V T N Y L D W I R D N M R P
Plasmin (human)             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 (bovine)           T Y F V T G I V S W G - - E G C A R K G K F G V Y T K V S N F L K W I D K I M K A R A G
Thrombin (bovine)           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
Kallikrein (pig)            - - - W Q G I T S W G - H T P C G S A N K P S I Y T K L I F Y L D W I B B T I T E N P
Elastase (pig)              Q Y A V H G V T S F V S R L G C N V T R K P T V F T R V S A Y I S W I N N V I A S N
Trypsin (bovine)            - - - 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
Chymotrypsin A (bovine)     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
                                 (211)          (221)           (231)          (241)
```

136–201 of chymotrypsin is not found in C1s, C1r or enzymes like plasmin and thrombin. More significantly, the disulphide bridge at residues 42–58, which stabilizes the position of the active-site histidine 57, is found in all the serine proteinases except C1s and C1r (Arlaud & Gagnon, 1981, 1983).

When the aligned sequence of C1s and C1r are

compared, the differences are found to be 59%, which is within the range (56–71%) found when comparing the other mammalian serine proteinases (Young *et al.*, 1978). This suggests that C1s and C1r are more different than might have been expected in view of their closely related functions. It is noticeable from the alignment that C1s shows strong similarities over short sequences with many of the other serine proteinases, but overall only the 'core' of conserved residues is retained in all sequences. Both C1s and C1r lack the proline residue at 198 that is found in all the other serine proteinases. The residue at position 189 is thought to contribute to the specificity of side-chain binding in the substrate pocket. In trypsin and a number of other serine proteinases this residue is aspartic acid, giving these enzymes their specificity for the positively charged lysine or arginine residues. In $C\overline{1s}$ this residue is also aspartic acid, consistent with the action of $C\overline{1s}$ in splitting C4 on the carboxy side of the arginine residue found at the *C*-terminus of C4a (Moon *et al.*, 1981; Smith *et al.*, 1982). Why $C\overline{1s}$ cleaves C4 (and C2), but not C3 and C5, which also have arginine residues in this position (Hugli & Müller-Eberhard, 1978), is less obvious. When the amino acid sequences of C3a, C4a and C5a are aligned (Smith *et al.*, 1982), the conserved residues in the *C*-terminal regions are found to show a different distribution of positively charged side chains. It is noteworthy that $C\overline{1s}$ has a number of carboxy groups, residues Glu-170, Glu-174 and Glu-186, near the active site in a position that could interact with the positive charges of the substrate, helping to distinguish between C4, C3 and C5.

We thank colleagues who donated blood and the North East of Scotland Blood Transfusion Service who collected it, Dr. Linda Fothergill for amino acid analyses, Mrs. Jean Bathgate and Mr. Ian Davidson for technical help, and the Medical Research Council and the Science and Engineering Research Council for financial support.

References

Arlaud, G. J. & Gagnon, J. (1981) *Biosci. Rep.* **1**, 779–784

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

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

Barkas, T., Scott, G. K. & Fothergill, J. E. (1973) *Biochem. Soc. Trans.* **1**, 1219–1220

Becker, E. L. (1956) *J. Immunol.* **77**, 462–468

Birktoft, J. J. & Blow, D. M (1972) *J. Mol. Biol.* **68**, 187–240

Booth, N. A., Campbell, R. D. & Fothergill, J. E. (1979) *Biochem. J.* **177**, 959–965

Campbell, R. D., Booth, N. A. & Fothergill, J. E. (1979*a*) *Biochem. J.* **177**, 531–540

Campbell, R. D., Booth, N. A. & Fothergill, J. E. (1979*b*) *Biochem. J* **183**, 579–588

Carter, P. E., Dunbar, B. & Fothergill, J. E. (1982) *Biochem. Soc. Trans.* **10**, 441–442

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

Fothergill, J. E. & Anderson, W. H. K. (1978) *Curr. Top. Cell. Regul.* **13**, 259–311

Hartley, B. S. (1964) *Nature (London)* **201**, 1284–1287

Hugli, T. E. & Müller-Eberhard, H. J. (1978) *Adv. Immunol.* **26**, 1–53

Lepow, I. H., Naff, G. B., Todd, E. W., Pensky, J. & Hinz, C. F. (1963) *J. Exp. Med.* **117**, 938–1008

Mahoney, W. C. & Hermodson, M. A. (1980) *J. Biol. Chem.* **255**, 11199–11203

Moon, K. E., Gorski, J. P. & Hugli, T. E. (1981) *J. Biol. Chem.* **256**, 8685–8692

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V. & Collen, D. (1983) *Nature (London)* **301**, 214–221

Reid, K. B. M. & Porter, R. R. (1976) *Biochem. J.* **155**, 19–23

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

Scott, J. D. & Fothergill, J. E. (1982) *Biochem. J.* **205**, 575–580

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

Smith, M. A., Gerrie, L. M., Dunbar, B. & Fothergill, J. E. (1982) *Biochem. J.* **207**, 253–260

World Health Organization (1968) *W. H. O. Bull.* **39**, 935–938 [or (1970) *Immunochemistry* **7**, 137–142]

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. (1976) *J. Immunol.* **116**, 496–503