Biochem. J. (1987) 241, 711–720 (Printed in Great Britain)

Complete amino acid sequence of the A chain of human complement-classical-pathway enzyme C$\bar{1}$r*

Gérard J. ARLAUD,†§ Anthony C. WILLIS‡ and Jean GAGNON‡||

†D.R.F. Groupe Immunochimie (Unité I.N.S.E.R.M. 238), Centre d'Etudes Nucléaires de Grenoble, B.P. 85 X, 38041 Grenoble Cédex, France, and ‡M.R.C. Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

The amino acid sequence of human C$\bar{1}$r A chain was determined, from sequence analysis performed on fragments obtained from C$\bar{1}$r autolytic cleavage, cleavage of methionyl bonds, tryptic cleavages at arginine and lysine residues, and cleavages by staphylococcal proteinase. The polypeptide chain has an *N*-terminal serine residue and contains 446 amino acid residues (M_r 51200). The sequence data allow chemical characterization of fragments α (positions 1–211), β (positions 212–279) and γ (positions 280–446) yielded from C$\bar{1}$r autolytic cleavage, and identification of the two major cleavage sites generating these fragments. Position 150 of C$\bar{1}$r A chain is occupied by a modified amino acid residue that, upon acid hydrolysis, yields *erythro*-β-hydroxyaspartic acid, and that is located in a sequence homologous to the β-hydroxyaspartic acid-containing regions of Factor IX, Factor X, protein C and protein Z. Sequence comparison reveals internal homology between two segments (positions 10–78 and 186–257). Two carbohydrate moieties are attached to the polypeptide chain, both via asparagine residues at positions 108 and 204. Combined with the previously determined sequence of C$\bar{1}$r B chain [Arlaud & Gagnon (1983) Biochemistry 22, 1758–1764], these data give the complete sequence of human C1r.

INTRODUCTION

The first component of the classical pathway of complement, C1, is a complex comprising two loosely interacting entities: C1q, the recognition unit, and the catalytic unit, which is a Ca^{2+}-dependent tetrameric association of two serine proteinases, C1r and C1s, sequentially activated through limited proteolysis during C1 activation (Reid & Porter, 1981; Cooper, 1985). Further studies (Villiers *et al.*, 1985) have established the domain structure and associated functions of C1r and C1s. A model has been proposed for the isolated form of the C1s–C1r–C1r–C1s catalytic unit, in which the $(C1r)_2$ dimer forms a core and the C1s molecules are located at each extremity of the complex. Upon interaction with C1q, the C1s–C1r–C1r–C1s complex is thought to adopt a figure-8-shaped conformation in which each C1s molecule comes into contact with the central $(C1r)_2$ dimer (Colomb *et al.*, 1984).

The first event of C1 activation is the activation of C1r by itself. The proenzyme form of human C1r is a dimer comprising two identical single-chain polypeptides of M_r approx. 90000. Each monomer is activated through proteolysis of a single Arg–Ile bond (Arlaud & Gagnon, 1985), which produces two-chain molecules consisting of the disulphide-linked heavy (A) and light (B) chains, of M_r approx. 58000 and 34000 respectively (see review by Sim, 1981). Prolonged incubation of isolated activated C$\bar{1}$r at 37 °C produces further limited cleavages, occurring in the *N*-terminal A chain, which splits into three fragments, α, β and γ, located in that order within the chain (Gagnon & Arlaud, 1985). Fragments α and β, which are removed from each monomer, are believed to correspond to an interaction domain and a connecting strand respectively (Villiers *et al.*, 1985). The resulting molecule, C$\bar{1}$r II, is still a dimer, each monomer consisting of fragment γ disulphide-linked to the *C*-terminal B chain and corresponding to a catalytic domain that retains the functional properties of the serine active site (Arlaud *et al.*, 1980; Villiers *et al.*, 1985).

The complete sequence of the catalytic B chain of human C$\bar{1}$r has been determined (Arlaud & Gagnon, 1983) and a partial sequence of the A chain has been obtained, from *N*-terminal sequences of fragments generated from autolytic and CNBr cleavage (Gagnon & Arlaud, 1985). The present paper provides the detailed proof of the entire amino acid sequence of human C$\bar{1}$r A chain, and thus completes the elucidation of the sequence of human C1r.

MATERIALS AND METHODS

Materials

Trypsin [treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one ('TPCK')] was obtained from Worthington Biochemical Corp. V8 proteinase from *Staphylococcus aureus* was from Miles Laboratories. Iodo[2-^{3}H]acetic acid (106 mCi/mmol) was obtained from Amersham International. Spectrapor 6 dialysis tubing (M_r cut-off 1000) was purchased from Spectrum Medical Industries. Reagents used for h.p.l.c. were purchased from Merck, and those used for automated

* This paper is dedicated to the memory of Professor R. R. Porter.

§ To whom correspondence should be addressed.

|| Present address: Conseil National de Recherches du Canada, Institut de Recherches en Biotechnologie, Hôpital Royal Victoria, 687 Avenue des Pins Ouest, Pavillon Hersey, Montréal, Qué. H3A 1A1, Canada

amino acid sequence analysis were obtained as described by Christie & Gagnon (1982).

Methods

C$\bar{1}$r was purified from human serum as described previously (Arlaud *et al.*, 1979). C$\bar{1}$r autolytic cleavage and purification of the reduced and *S*-[^{3}H]carboxymethylated α, β and γ fragments were performed as described by Gagnon & Arlaud (1985). 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 were performed as described previously (Arlaud *et al.*, 1982).

CNBr cleavage of C$\bar{1}$r A chain. Reduced and *S*-[^{3}H]carboxymethylated C$\bar{1}$r A chain (300 nmol) was dissolved in 70% (v/v) formic acid (3.5 ml). CNBr (5 mmol) was added, and the mixture was kept in the dark for 24 h at 4 °C. The digest was fractionated by gel filtration on a Bio-Gel P-100 column (2.5 cm × 110 cm) equilibrated in 9% (v/v) formic acid, and six pools were collected. This method allowed direct separation of CNBr-cleavage peptides CN1a and CN1b, which were obtained from the second and the third pool respectively and used for further proteolytic cleavages.

Peptide separation by reversed-phase h.p.l.c. Separations were performed on a μBondapak C_{18} column (Waters Associates) with two solvent systems as previously described (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 215 nm.

Staphylococcal-proteinase digestion. Reduced and *S*-[^{3}H]carboxymethylated succinylated (3-carboxypropionylated) fragment γ (2.0 mg) and CNBr-cleavage peptide CN1b (1.0 mg), *S*-[^{3}H]carboxymethylated fragment β (1.1 mg) and CNBr-cleavage peptide CN1a (4.2 mg) were dissolved at 0.5–2.0 mg/ml in 0.1 M-NH_4HCO_3/2 mM-EDTA, pH 7.8, and digested with staphylococcal proteinase (enzyme/substrate ratio 1:30, w/w) for 5–6 h at 37 °C. The digest of fragment γ was fractionated by gel filtration on a Sephadex G-50 (fine grade) column (2.5 cm × 110 cm) equilibrated in 0.1 M-NH_4HCO_3, pH 7.8. Three pools were collected, and peptide γSP1 was obtained from pool 1. The digests of fragment β and of CNBr-cleavage peptides CN1a and CN1b were fractionated by reversed-phase h.p.l.c. with solvent system 1. This allowed purification of peptides βSP1, βSP4, CN1b SP8, CN1b SP12, CN1a SP1, CN1a SP2, CN1a SP3 and CN1a SP9, whereas peptides CN1a SP4, CN1a SP7, CN1a SP8a and CN1a SP8b required further purification by reversed-phase h.p.l.c. with solvent system 2.

Tryptic cleavages. Reduced and *S*-[^{3}H]carboxymethylated succinylated fragment α (170 nmol), reduced and *S*-[^{3}H]carboxymethylated fragment β (150 nmol) and CNBr-cleavage peptide CN1b (70 nmol) were cleaved by trypsin, as described previously for C$\bar{1}$r B chain (Arlaud & Gagnon, 1983). The tryptic digest of fragment α was fractionated by gel filtration on a Sephadex G-50 (fine grade) column (2.5 cm × 100 cm) equilibrated in 0.1 M-NH_4HCO_3. This allowed purification of peptides αST1 (Gagnon & Arlaud, 1985) and αST4, whereas peptides αST3a, αST3b, αST3c, αST7a and αST7b were further purified by reversed-phase h.p.l.c. with solvent system 1. The tryptic digest of peptide CN1b was fractionated by reversed-phase h.p.l.c. with solvent system 2, which allowed purification of peptides CN1b T5 and CN1bT8. Isolation of peptide βT4c was obtained by successive fractionations of the tryptic digest of fragment β by reversed-phase h.p.l.c. with solvent systems 1 and 2.

Amino acid analysis. Peptides were hydrolysed for 24 h under reduced pressure at 110 °C in constant-boiling HCl containing 0.1% (v/v) 2-mercaptoethanol and 4 mM-phenol. Cysteine was determined as *S*-carboxymethylcysteine. Positive identification and quantification of hexosamines was done in the same hydrolysates as amino acids, although no correction was made for destruction during acid hydrolysis. Analyses were performed on an LKB 4400 analyser with the standard LKB sodium citrate eluting buffers. For analysis of DL-*erythro*-β-hydroxyaspartic acid, the pH of the starting buffer was lowered from 3.25 to 3.0, in order to increase resolution of acidic amino acids. β-Hydroxyaspartic acid was synthesized and purified in accordance with Benoiton *et al.* (1959) and Kornguth & Sallach (1960). DL-*threo*- and DL-*erythro*-β-Hydroxyaspartic acid standards were kindly provided by Dr. J. G. Morris, University College of Wales, Aberystwyth, Dyfed, U.K.

Automated sequence analysis. Automated Edman degradation was performed in a Beckman 890C Sequenator with the 0.3 M-Quadrol programme of Hunkapiller & Hood (1978), or in an Applied Biosystems model 470 A gas-phase protein sequencer.

Identification of amino acid phenylthiohydantoin derivatives by reversed-phase h.p.l.c. was performed as described previously (Arlaud & Gagnon, 1985). The protocol recommended by Applied Biosystems was also used (Hunkapiller, 1985). The phenylthiohydantoin derivative of β-hydroxyaspartic acid was obtained by performing, in the sequencer, one cycle of Edman degradation on synthetic DL-*erythro*-β-hydroxyaspartic acid.

Peptide nomenclature. Autolytic fragments are designated α, β and γ. Peptides obtained from CNBr cleavage, staphylococcal-proteinase digestion, full tryptic cleavage and tryptic cleavage after succinylation are designated CN, SP, T and ST respectively. The next number indicates the elution position of the peptide in the initial separation used, and it is followed by a letter indicating its elution position in the second separation.

RESULTS

Completion of the sequence of the *C*-terminal γ fragment

Previous studies (Gagnon & Arlaud, 1985) had established the sequence of the 62 *N*-terminal residues of fragment γ of C$\bar{1}$r A chain, and the sequence of its *C*-terminal 95 residues. Cleavage of fragment γ by staphylococcal proteinase yielded a single large peptide, γSP1, which was isolated by gel filtration on Sephadex G-50. As judged from its *N*-terminal sequence and its amino acid composition (Table 1), peptide γSP1 resulted from cleavage of the glutamyl bonds at positions 329 and 406 of C$\bar{1}$r A chain (Figs. 1 and 3). Its *N*-terminal

Table 1. Amino acid compositions of selected peptides from the β and γ regions of C$\bar{1}$r A chain

Amino acid compositions were calculated from duplicate 24 h HCl hydrolysates. Values less than 0.2 mol/mol are omitted, and values obtained from the sequence are indicated in parentheses. Tryptophan was not determined.

Amino acid	Amino acid composition (mol of residue/mol)					
Peptide... Position in sequence...	γSP1 330–406	ST8i 364–371	βSP1 275–279	βSP4 250–274	βT4c 254–279	CN1aSP8a 275–288
Asx	8.0 (8)	3.0 (3)	0.9 (1)	4.8 (5)	5.8 (6)	1.0 (1)
Thr	5.9 (7)			1.9 (2)	2.1 (2)	1.9 (2)
Ser	2.5 (3)		1.6 (2)	1.8 (2)	3.5 (4)	1.8 (2)
Glx	8.4 (8)			1.8 (2)	1.9 (2)	1.0 (1)
Pro	3.8 (4)	1.1 (1)		2.0 (2)	2.1 (2)	
Gly	6.4 (6)	0.9 (1)	1.1 (1)	1.0 (1)	1.1 (1)	2.0 (2)
Ala	4.5 (4)			0.9 (1)	1.2 (1)	
Val	3.5 (3)			1.0 (1)	1.1 (1)	
Cys	3.8 (4)			1.0 (1)		
Met	2.0 (3)					
Ile	2.3 (2)					
Leu	2.5 (2)	1.0 (1)		2.9 (3)	2.9 (3)	1.0 (1)
Tyr	6.1 (6)					1.0 (1)
Phe	2.1 (2)	1.0 (1)		2.8 (3)	1.9 (2)	
His	2.7 (3)					
Lys	3.6 (4)			1.0 (1)		1.3 (1)
Arg	6.8 (7)	1.0 (1)	1.0 (1)	1.0 (1)	2.1 (2)	2.0 (2)
Trp	(1)					(1)

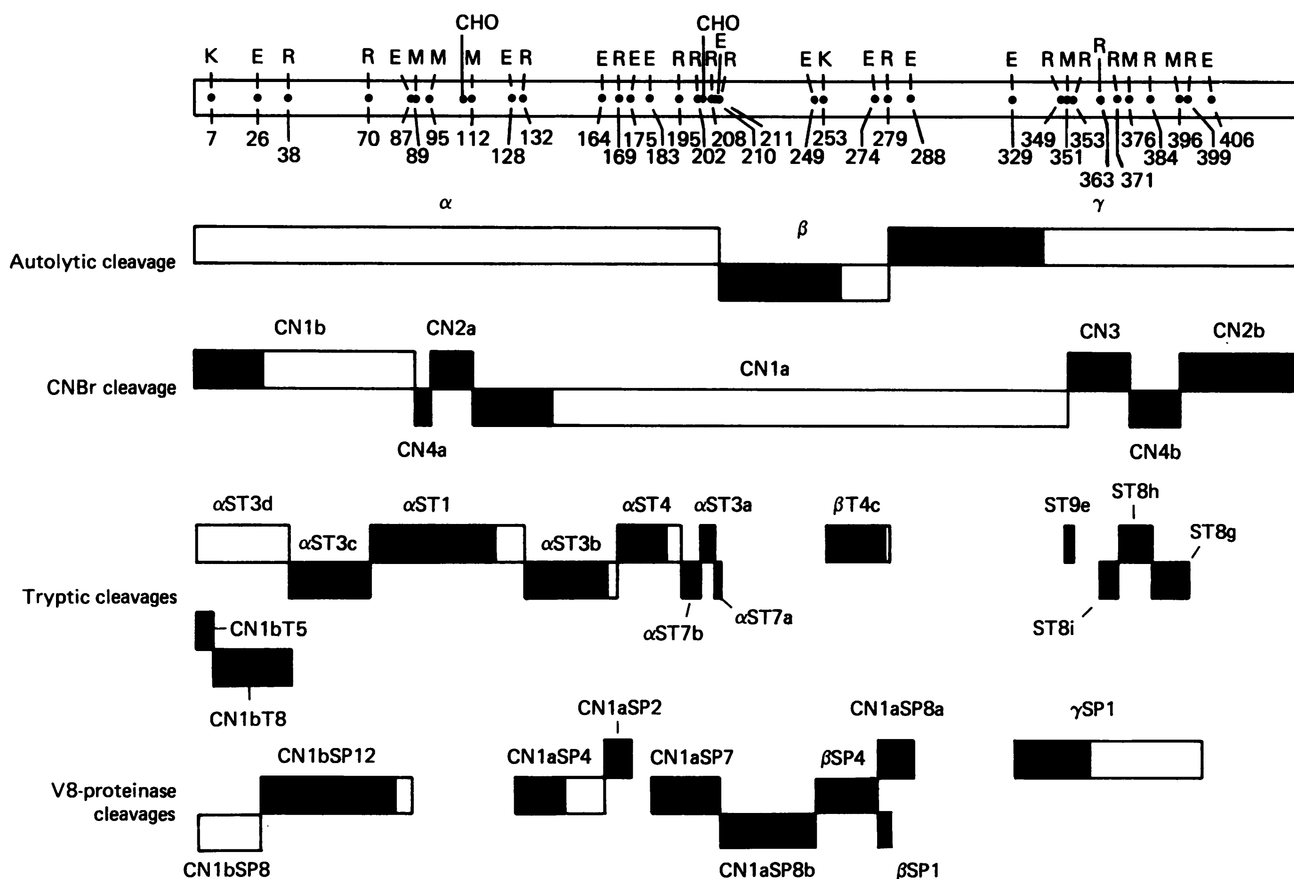

Fig. 1. Summary of the sequencing strategy

Standard one-letter abbreviations are used to designate residues defining the different fragments obtained from C$\bar{1}$r A chain, and the two carbohydrate attachment sites are indicated by –CHO. Black areas correspond to the portions of C$\bar{1}$r A chain that have been sequenced.

```
Peptide

                             10                  20                  30
γSP1       G N Q V L H S F T A V C Q D D G T W H R A M P R C K I K D C

ST8i       N L P N G D F R

βSP1       S G D S R

                             10                  20
βSP4       F C G K Q R P P D L D T S S N A V D L L F F T D E

                             10                  20
βT4c       Q R P P D L D T S S N A V D L L F F T D E X G D

                             10
CN1aSP8a   S G D S R G W K L R Y T T E

αST3a      C _ Y S I R

                               10                  20                  30
αST3b      S K S/L G E E D P Q P Q C Q H L C H X Y V G G Y F C S C R P G Y E L Q E D R

                             10                  20                  30
αST3c      V K L V F Q Q F D L E P S E G C F Y D Y V K I S A D K K S L G R

                             10
αST4       H S C Q A E C S S E L Y T E A S G Y I

αST7a      V E R

αST7b      S Y P P D L R

                             10
CN1aSP2    L Q E D R H S C Q A E

                               10                  20
CN1aSP4    C A S R S K S/L G E E D P Q P Q C Q H L C

                             10                  20
CN1aSP7    A S G Y I S S L E Y P R S Y P P D L R C _ Y S I R V E

                             10                  20                  30
CN1aSP8b   R G L T L H L K F L E P F D I D D H Q Q V H C P Y D Q L Q I Y A N G K N I G E

                             10                  20                  30                  40                  50
CN1bSP12   T T T V I T V P T G Y R V K L V F Q Q F D L E P S E G C F Y D Y V K I S A D K K S L G R F C G Q L G S P L G

CN1bT5     S I P I P Q K

                             10                  20                  30
CN1bT8     L F G E V T S P L F P K P Y P N N F E T T T V I T V P T G Y R

                             10                  20
CN1b       S I P I P Q K L F G E V T S P L F P K P Y P N N F E
```

Fig. 2. *N*-Terminal sequences of selected peptides from human C$\bar{1}$r A chain

The single-letter abbreviations are used. _ indicates the presence of an asparagine-linked carbohydrate moiety.

```
1                 10                  20                  30                  40                  50
S I P I P Q K L F G E V T S P L F P K P Y P N N F E T T T V I T V P T G Y R V K L V F Q Q F D L E P
                  60                  70                  80                  90                 100
S E G C F Y D Y V K I S A D K K S L G R F C G Q L G S P L G N P P G K K E F M S Q G N K M L L T F H
              CHO 110                 120                 130       S         140                150
T D F S N E E N G T I M F Y K G F L A Y Y Q A V D L D E C A S R S K S/L G E E D P Q P Q C Q H L C H β
                  160                 170                 180                 190                200
Y V G G Y F C S C R P G Y E L Q E D R H S C Q A E C S S E L Y T E A S G Y I S S L E Y P R S Y P P D
      CHO         210                 220                 230                 240                250
L R C N Y S I R V E R G L T L H L K F L E P F D I D D H Q Q V H C P Y D Q L Q I Y A N G K N I G E F
                  260                 270                 280                 290                300
C G K Q R P P D L D T S S N A V D L L F F T D E S G D S R G W K L R Y T T E I I K C P Q P K T L D E
                  310                 320                 330                 340                350
F T I I Q N L Q P Q Y Q F R D Y F I A T C K Q G Y Q L I E G N Q V L H S F T A V C Q D D G T W H R A
                  360                 370                 380                 390                400
M P R C K I K D C G Q P R N L P N G D F R Y T T T M G V N T Y K A R I Q Y Y C H E P Y Y K M Q T R A
                  410                 420                 430                 440         446
G S R E S E Q G V Y T C T A Q G I W K N E Q K G E K I P R C L P V C G K P V N P V E Q R Q R
```

14	Ala	A	21	Cys	C	20	Ile	I	32	Pro	P	18	Val	V
23	Arg	R	33	Gln	Q	31	Leu	L	27	Ser	S			
17	Asn	N	29	Glu	E	26	Lys	K	27	Thr	T			
24	Asp	D	33	Gly	G	6	Met	M	3	Trp	W			
1	β-OH-Asp		10	His	H	23	Phe	F	28	Tyr	Y			

No. of residues: 446

M_r of the polypeptide chain: 51 208

Fig. 3. Amino acid sequence and composition of human C$\overline{1}$r A chain

The amino acid composition indicated is that of a chain bearing a serine residue at position 135.

sequence was determined for 30 cycles (Fig. 2), providing the overlap between the known *N*-terminal sequences of fragment γ and of CNBr-cleavage peptide CN3 (Fig. 1), and confirming the location of peptide ST9e at the junction between peptides CN1a and CN3, as proposed previously (Gagnon & Arlaud, 1985).

The residue at position 367 of C$\overline{1}$r A chain (Fig. 3) had been provisionally assigned as aspartic acid/asparagine, on the basis of the sequence of peptide CN3, which yielded equivalent amounts of aspartic acid and asparagine at this position (Gagnon & Arlaud, 1985). Peptide ST8i, originating from cleavage of the arginyl bonds at positions 363 and 371 of C$\overline{1}$r A chain (Figs. 1 and 3), was obtained from tryptic cleavage of reduced and alkylated succinylated C$\overline{1}$r A chain and purified as described for the methionine-containing tryptic peptides (Gagnon & Arlaud, 1985). The sequence of peptide ST8i (Fig. 2) provided unambiguous identification of the residue at position 367 as asparagine, suggesting that the unusually low relative yield of asparagine previously observed at this position was due to partial resistance of the Asn–Gly bond to Edman degradation (Allen, 1981).

The above results completed the elucidation of the sequence of fragment γ, extending from the glycine residue at position 280 of C$\overline{1}$r A chain to the *C*-terminus of this chain (position 446).

Completion of the sequence of fragment β and overlap β–γ

From previous studies (Gagnon & Arlaud, 1985), the sequence of the *N*-terminal 50 residues of fragment β (positions 212–261 of C$\overline{1}$r A chain) had been established. Digestion of fragment β with staphylococcal proteinase yielded peptide βSP4, originating from cleavage of the glutamyl bonds at positions 249 and 274 of C$\overline{1}$r A chain, and peptide βSP1, the *C*-terminal peptide of fragment β (Figs. 1 and 3), and the complete sequences of these peptides were determined (Fig. 2). Full tryptic cleavage of fragment β yielded peptide βT4c, generated by cleavage of the lysyl bond at position 253 of C$\overline{1}$r A chain (Figs. 1 and 3). A partial sequence of peptide βT4c was obtained (Fig. 2), providing the overlap between peptides βSP4 and βSP1 and thereby establishing the complete sequence of fragment β (positions 212–279 of C$\overline{1}$r A chain). The overlap between fragments β and γ was provided by peptide CN1aSP8a, obtained from digestion of peptide CN1a with staphylococcal proteinase. Peptide CN1aSP8a originated from cleavage of the glutamyl bonds at positions 274 and 288 of C$\overline{1}$r A chain (Figs. 1 and 3) and its complete sequence was established (Fig. 2). These results completed the elucidation of the sequence of the β–γ region of C$\overline{1}$r A chain (positions 212–446) and established that the major peptide bond cleaved between fragments β and γ is the Arg–Gly bond at position 279, as suggested previously (Gagnon & Arlaud, 1985).

Completion of the sequence of fragment α and overlap α–β

From previous studies based on the sequences of CNBr-cleavage peptides CN4a, CN2a and CN1a, and of αST1, the methionine-containing peptide obtained from tryptic cleavage of succinylated fragment α, a continuous sequence of 74 residues (positions 71–144 of C$\overline{1}$r A chain) had been established, representing about one-third of the estimated length of fragment α (Gagnon & Arlaud, 1985). Tryptic cleavage of succinylated fragment α, staphylococcal-proteinase digestion of CNBr-cleavage peptides

Table 2. Amino acid compositions of tryptic peptides from succinylated fragment *α*

Amino acid compositions were calculated from duplicate 24 h HCl hydrolysates. Values less than 0.2 mol/mol are omitted, and values obtained from the sequence are indicated in parentheses.

Amino acid		Amino acid composition (mol of residue/mol)							
	Peptide...	αST1	αST3a	αST3b	αST3c	αST3d	αST4	αST7a	αST7b
	Position in sequence...	71–132	203–208	133–169	39–70	1–38	170–195	209–211	196–202
β-OH-Asx				0.7 (1)					
Asx		6.4 (7)	1.1 (1)	2.2 (2)	3.1 (3)	2.0 (2)			1.1 (1)
Thr		2.6 (3)				5.5 (6)	1.1 (1)		
Ser		3.5 (4)	1.1 (1)	2.7 (3–2)	2.7 (3)	1.5 (2)	5.7 (6)		1.0 (1)
Glx		7.4 (7)		7.5 (8)	4.0 (4)	3.2 (3)	4.5 (5)	1.0 (1)	
Pro		2.6 (3)		3.0 (3)	1.2 (1)	6.3 (7)	0.9 (1)		1.6 (2)
Gly		7.1 (7)		3.9 (4)	2.2 (2)	1.9 (2)	1.3 (1)		
Ala		2.3 (3)			1.1 (1)		1.9 (2)		
Val		1.0 (1)		1.1 (1)	2.8 (3)	2.8 (3)		1.0 (1)	
Cys		1.9 (2)	0.9 (1)	3.5 (4)	1.0 (1)		1.9 (2)		
Met		1.9 (3)							
Ile		1.0 (1)	1.0 (1)		0.9 (1)	2.6 (3)	1.1 (1)		
Leu		5.6 (6)		2.3 (2–3)	3.1 (3)	2.0 (2)	2.1 (2)		1.1 (1)
Tyr		3.5 (3)	0.7 (1)	3.0 (3)	2.0 (2)	1.9 (2)	2.6 (3)		0.8 (1)
Phe		5.9 (6)		1.0 (1)	2.9 (3)	2.9 (3)			
His		1.1 (1)		1.8 (2)			1.1 (1)		
Lys		3.6 (4)		1.2 (1)	3.5 (4)	1.8 (2)			
Arg		1.0 (1)	1.1 (1)	2.1 (1)	1.2 (1)	1.3 (1)	1.0 (1)	1.0 (1)	1.3 (1)

CN1a and CN1b and full tryptic cleavage of peptide CN1b were used to generate the necessary peptides for completing the establishment of the sequence of fragment *α*. Tryptic cleavage of succinylated fragment *α* yielded eight peptides, purified by gel filtration on Sephadex G-50 and reversed-phase h.p.l.c. These peptides, denoted αST1, αST3a, αST3b, αST3c, αST3d, αST4, αST7a and αST7b, originated from cleavage of arginyl bonds at positions 38, 70, 132, 169, 195, 202 and 208 of C$\bar{1}$r A chain, the only arginyl bond of fragment *α* resistant to enzymic cleavage being the Arg–Pro bond at position 160 (Figs. 1 and 3). The amino acid compositions of all the tryptic peptides from succinylated fragment *α* are shown in Table 2, and their *N*-terminal amino acid sequences are indicated in Fig. 2.

The *N*-terminal peptide, αST3d (positions 1–38), had a blocked *N*-terminus, whereas the complete sequence of peptide αST3c was obtained (positions 39–70). The complete sequence of peptide αST3b was determined with the exception of the residue at position 150. Amino acid analysis of this peptide with a modified elution programme (see the Materials and methods section) indicated the presence of *erythro-β*-hydroxyaspartic acid, estimated as 0.7 mol/mol of peptide. Analysis of the phenylthiohydantoin derivative of residue 150 also gave positive identification of *erythro-β*-hydroxyaspartic acid, although two unidentified peaks were also detected. The sequence of peptide αST3b also indicated that position 135 of C$\bar{1}$r A chain contained two different residues, serine and leucine (Figs. 2 and 3), in good agreement with the amino acid composition and full sequence of peptide αST3b (Fig. 3 and Table 2). The *N*-terminal sequence of peptide αST4 (positions 170–195) was determined for 19 cycles (positions 170–188), and peptides αST7a (positions 209–211) and αST7b (positions 196–202) were completely sequenced. The complete sequence of peptide αST3a (positions 203–208) was also obtained, with the exception of the residue at position 204. This site probably represents an asparagine-linked carbohydrate moiety, in agreement with the amino acid composition of peptide αST3a, which indicates the presence of aspartic acid/asparagine (1.1 mol/mol of peptide) (Table 2), and with identification of *N*-acetylglucosamine (1.7 mol/mol of peptide). This is consistent with the characteristic Asn-Tyr-Ser sequence, compatible with attachment of that type of carbohydrate (Neuberger *et al.*, 1972) and with the previous detection of *N*-acetylglucosamine in CNBr-cleavage peptide CN1a (Gagnon & Arlaud, 1985).

Reduced and *S*-[^{3}H]carboxymethylated CNBr-cleavage peptide CN1a was digested with staphylococcal proteinase, and 12 major peptides, accounting for the whole sequence of peptide CN1a, were purified by reversed-phase h.p.l.c. The amino acid compositions of those peptides that originate from the *α* region are shown in Table 3, and *N*-terminal amino acid sequences of some of these peptides are presented in Fig. 2. Peptide CN1aSP2 resulted from cleavage of the glutamyl bonds at positions 164 and 175 and contained an uncleaved Glu–Asp bond (Figs. 1 and 3). The complete sequence of peptide CN1aSP2 was determined (Fig. 2), providing the overlap between peptides αST3b and αST4 (Figs. 1 and 3). Peptide CN1aSP7 (positions 184–210) contained a Glu–Tyr bond (position 192) resistant to enzymic cleavage. With the exception of position 204, corresponding to the carbohydrate-attachment site already identified in peptide αST3a, the complete sequence of peptide CN1aSP7 was determined (Fig. 2), overlapping peptides αST4, αST7b, αST3a and αST7a (Figs. 1 and 3). The complete sequence of peptide CN1aSP8b (positions 211–249) (Fig. 2) provided a one-residue overlap

Table 3. Amino acid compositions of the peptides from staphylococcal-proteinase digestion of peptide CN1a originating from the α region

Amino acid compositions were calculated from duplicate 24 h HCl hydrolysates. Values less than 0.2 mol/mol are omitted, and values obtained from the sequence are indicated in parentheses.

	Amino acid composition (mol of residue/mol)						
Amino acid — Peptide... / Position in sequence...	CN1aSP1 176–179	CN1aSP2 165–175	CN1aSP3 180–183	CN1aSP4 129–164	CN1aSP7 184–210	CN1aSP8b 211–249	CN1aSP9 113–128
β-OH-Asx				0.6 (1)			
Asx		1.1 (1)		1.2 (1)	2.0 (2)	5.6 (6)	2.2 (2)
Thr			0.9 (1)			1.0 (1)	
Ser	1.6 (2)	0.9 (1)		3.4 (4–3)	4.2 (5)		
Glx	1.1 (1)	3.7 (4)	1.0 (1)	5.3 (6)	2.0 (2)	5.9 (6)	2.2 (2)
Pro				2.8 (3)	2.9 (3)	2.1 (2)	
Gly				3.6 (4)	1.3 (1)	3.3 (3)	1.2 (1)
Ala		1.1 (1)		1.1 (1)	1.3 (1)	1.0 (1)	1.7 (2)
Val				1.0 (1)	1.2 (1)	1.2 (1)	1.0 (1)
Cys	0.9 (1)	1.1 (1)		4.7 (5)	1.0 (1)	0.9 (1)	
Met							
Ile					1.9 (2)	2.8 (3)	
Leu		1.1 (1)	1.2 (1)	1.5 (1–2)	2.0 (2)	4.9 (5)	2.1 (2)
Tyr			1.0 (1)	2.8 (3)	3.8 (4)	1.9 (2)	2.6 (3)
Phe				0.9 (1)		1.9 (2)	1.7 (2)
His		0.9 (1)		1.9 (2)		2.7 (3)	
Lys				1.1 (1)		2.0 (2)	1.1 (1)
Arg		1.1 (1)		1.9 (2)	3.1 (3)	1.0 (1)	

between fragments α and β, and established that the peptide bond cleaved between these fragments is the Arg–Gly bond at position 211, as suggested by previous studies (Gagnon & Arlaud, 1985). The above results thus established the continuous sequence of residues 71–446 of C$\bar{1}$r A chain. The amino acid composition of peptide CN1aSP4 (positions 129–164) again indicated the presence of *erythro*-β-hydroxy aspartic acid (Table 3), estimated as 0.6 mol/mol of peptide, and thus confirmed the analysis of peptide αST3b. The presence of serine and leucine at position 135 was also confirmed by the *N*-terminal sequence of CN1aSP4, which was determined for 20 cycles (Fig. 2). Again, this result was in agreement with the amino acid composition of peptide CN1aSP4 (Table 3).

The *N*-terminal CNBr-cleavage peptide of C$\bar{1}$r A chain, CN1b, was succinylated and digested with staphylococcal proteinase. Complete cleavage of the glutamyl bonds at positions 26 and 87 and partial cleavage of the Glu–Val bond at position 11 were obtained, whereas the Glu–Pro and Glu–Gly bonds at positions 49 and 52 were resistant to enzymic cleavage (Figs. 1 and 3). The *N*-terminal peptide CN1bSP8 (positions 1–26) was blocked with a succinyl group, whereas the *N*-terminal sequence of peptide CN1bSP12 (positions 27–87) was determined for 54 cycles (Fig. 2), providing the overlap between peptides αST3c and αST1 and establishing the continuous sequence of residues 27–446 (Figs. 1 and 3).

In order to establish the structure of the unknown *N*-terminal part of C$\bar{1}$r A chain (residues 1–26), full tryptic cleavage of peptide CN1b was performed. This led to cleavage of the arginyl bonds at positions 38 and 70 and of the lysyl bonds at positions 7, 40, 60, 66 and 86, whereas the Lys–Pro bond at position 19 was resistant to enzymic cleavage (Fig. 3). The complete sequence of peptide CN1bT8 was determined (Fig. 2), allowing its location at positions 8–38 (Figs. 1 and 3).

Table 4. Amino acid compositions of selected peptides from cleavage of peptide CN1b by staphylococcal proteinase and trypsin

Amino acid compositions were calculated from duplicate 24 h HCl hydrolysates. Values less than 0.2 mol/mol are omitted, and values obtained from the sequence are indicated in parentheses.

	Amino acid composition (mol of residue/mol)			
Amino acid — Peptide... / Position in sequence...	CN1bSP8 1–26	CN1bSP12 27–87	CN1bT5 1–7	CN1bT8 8–38
Asx	1.8 (2)	4.1 (4)		1.9 (2)
Thr	0.8 (1)	4.4 (5)		4.8 (6)
Ser	1.5 (2)	3.4 (4)	1.0 (1)	1.0 (1)
Glx	3.1 (3)	5.9 (6)	1.1 (1)	2.2 (2)
Pro	5.6 (6)	4.7 (5)	1.8 (2)	4.7 (5)
Gly	1.2 (1)	6.5 (7)		2.1 (2)
Ala		1.1 (1)		
Val	1.0 (1)	5.1 (5)		2.8 (3)
Cys		2.0 (2)		
Met				
Ile	2.0 (2)	1.9 (2)	1.7 (2)	1.0 (1)
Leu	2.0 (2)	5.1 (5)		2.1 (2)
Tyr	1.0 (1)	3.1 (3)		2.1 (2)
Phe	2.6 (3)	4.0 (4)		2.8 (3)
His				
Lys	2.2 (2)	5.9 (6)	1.1 (1)	1.1 (1)
Arg		1.7 (2)		0.9 (1)

Full tryptic cleavage of peptide CN1b also yielded peptide CN1bT5, a heptapeptide that was submitted to Edman degradation and sequenced completely (Fig. 2). *N*-Terminal sequence analysis of peptide CN1b was performed on the gas-phase sequencer and showed that this peptide had the same *N*-terminal sequence as peptide CN1bT5, giving a sequence of 26 residues that clearly established the overlap between peptides CN1bT5 and CN1bT8, thus completing the amino acid sequence of C$\bar{1}$r A chain.

DISCUSSION

N-Terminal sequence analysis of 31 selected peptides obtained from C$\bar{1}$r autolytic cleavage, cleavage of methionyl bonds, tryptic cleavage at arginine and lysine residues and cleavages by staphylococcal proteinase provided sufficient information to establish the complete amino acid sequence of human C$\bar{1}$r A chain. The polypeptide core of human C$\bar{1}$r A chain comprises 446 amino acid residues, with an M_r of 51200. The validity of the sequence Val-Glu-Arg-Gly at positions 209–212, based on a one-residue overlap between peptides αST7a and CN1aSP8b (Figs. 1 and 3), has been confirmed by DNA sequence analysis (A. Journet, M. Tosi, T. Meo & M. G. Colomb, unpublished work).

Previous *N*-terminal sequence analyses performed on the whole A chain (Sim *et al.*, 1977) and on the α fragment, the *N*-terminal fragment of this chain obtained from C$\bar{1}$r autolytic cleavage (Gagnon & Arlaud, 1985), proved unsuccessful, and led to the conclusion that the A chain had a blocked *N*-terminus. The same conclusion was drawn from the first *N*-terminal sequence analysis of peptide CN1b, the *N*-terminal CNBr-cleavage peptide (Gagnon & Arlaud, 1985), although this analysis, performed in the liquid-phase Sequenator, gave a faint sequence, identified as Ser-Xaa-Pro-Ile-Pro-Gln-Lys-Xaa-Phe-Gly, and which, because of its very low recovery yield (< 5%), was attributed to a contaminant. The present study shows that this sequence represented the real *N*-terminal sequence of C$\bar{1}$r A chain. This sequence has been unambiguously established from a new sequence analysis of peptide CN1b, and from the complete sequence of peptide CN1bT5, the *N*-terminal heptapeptide of the chain, both performed in the gas-phase Sequenator. It should be mentioned that the resistance of peptides αST3d and CN1bSP8 to Edman degradation is explained by the succinylation of their *N*-terminal serine residue, as both were obtained from succinylated material. However, no satisfactory explanation has been found of the apparent resistance of the A chain and its α fragment to Edman degradation, although it is noteworthy that both are large polypeptides and were analysed only by the liquid-phase technique. Another partial explanation could arise from the particular structure of the *N*-terminal end of C$\bar{1}$r A chain (Fig. 3), which shows an unusually high percentage of proline residues.

Position 150 of human C$\bar{1}$r A chain is occupied by a residue that, upon acid hydrolysis, yields *erythro-β*-hydroxyaspartic acid. This conclusion is based on amino acid analysis of peptides αST3b (positions 133–169) and CN1aSP4 (positions 129–164), which were found to contain 0.6–0.7 mol of *erythro-β*-hydroxyaspartic acid/mol of peptide. Under standard analysis conditions, *erythro-β*-hydroxyaspartic acid was not resolved from *S*-carboxymethylcysteine, whereas decreasing the pH of the starting elution buffer from 3.25 to 3.0 improved the resolution and allowed clear identification of *erythro-β*-hydroxyaspartic acid, eluted between cysteic acid and *S*-carboxymethylcysteine. *Erythro-β*-Hydroxyaspartic acid has already been identified in several proteins, namely the bovine and/or human species of Factor VII (McMullen *et al.*, 1983*b*), Factor IX (Fernlund & Stenflo, 1983; Sugo *et al.*, 1984*a*), Factor X (McMullen *et al.*, 1983*a*,*b*; Sugo *et al.*, 1984*a*), protein C (Drakenberg *et al.*, 1983), protein S and protein Z (Fernlund & Stenflo, 1983). In human C$\bar{1}$r A chain, the residue at position 150 is located in a sequence that shows pronounced homology with the *β*-hydroxyaspartic acid-containing regions of protein C, Factor IX, Factor X and protein Z (Fig. 4), and that, like these regions, contains six half-cystine residues distributed in the characteristic pattern of epidermal-growth-factor-like domains (Doolittle, 1985). In bovine and human protein C (Long *et al.*, 1984; Foster *et al.*, 1985) and in human Factor IX (Anson *et al.*, 1984) it is known that *erythro-β*-hydroxyaspartic acid is the result of a post-translational hydroxylation of an aspartic acid residue, and its involvement in the binding of Ca^{2+} has been proposed in the case of bovine protein C (Esmon *et al.*, 1983), Factor IX (Morita *et al.*, 1984) and Factor X (Sugo *et al.*, 1984*b*). Although definitive identification of the residue at position 150 of C$\bar{1}$r A chain will require further analysis at the residue level, its identity with *erythro-β*-hydroxyaspartic acid appears likely. This residue could be involved in the binding of Ca^{2+} by human C1r, in agreement with previous studies suggesting that the Ca^{2+}-binding site of this protein is located in its *N*-terminal α interaction domain (Villiers *et al.*, 1980;

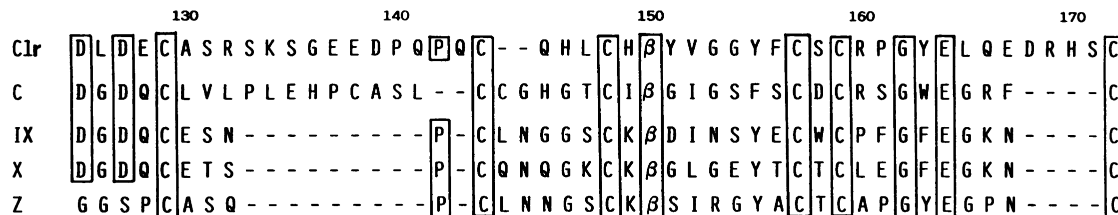

Fig. 4. Sequence homology between the *β*-hydroxyaspartic acid-containing regions of human protein C, Factor IX, Factor X, bovine protein Z and the region comprised between residues 125 and 172 of human C$\bar{1}$r A chain

Sequence data for human protein C, Factor IX, Factor X and bovine protein Z are taken from Foster *et al.* (1985), Anson *et al.* (1984), McMullen *et al.* (1983*b*) and Hojrup *et al.* (1985) respectively. – denotes that a gap was left to give maximum homology. The residue numbering indicated is that of C1r. *β*, *β*-Hydroxyaspartic acid.

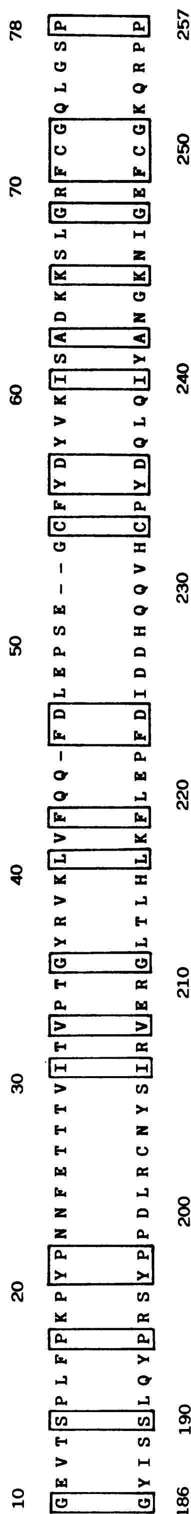

Fig. 5. Internal homology between segments 10–78 and 186–257 of human C$\overline{1}$r A chain

Arlaud *et al.*, 1985). In this hypothesis, it should be stressed that C1r would represent a new class of β-hydroxyaspartic acid-containing proteins, as in contrast with Factors VII, IX and X and proteins C, S and Z, this protein is not vitamin K-dependent.

The complete sequence of human C$\overline{1}$r A chain allows precise chemical definition of the three major fragments α (211 residues), β (68 residues) and γ (167 residues) yielded from C$\overline{1}$r autolytic cleavage. As indicated previously (Gagnon & Arlaud, 1985), fragment β is released from C$\overline{1}$r A chain without reduction, which, given that no free thiol group is detectable in C$\overline{1}$r (G. J. Arlaud & J. Gagnon, unpublished work), implies that the two half-cystine residues located in this fragment form a disulphide bridge in the native protein. For the same reasons it can be concluded that the ten half-cystine residues of fragment α are associated in five internal disulphide bridges. Fragment γ contains nine half-cystine residues, of which one (position 434) is probably involved in the single disulphide bridge connecting the A and B chains (Gagnon & Arlaud, 1985), the other eight residues forming four internal disulphide bridges.

The sequence data also show that the autolytic fragments originate from the cleavage of two Arg–Gly bonds, located at positions 211 and 279 of C$\overline{1}$r A chain, as suggested previously (Gagnon & Arlaud, 1985). Taking into account that C1r autoactivation (Arlaud & Gagnon, 1985) and C1s activation by C$\overline{1}$r (Spycher *et al.*, 1986) both involve cleavage of an Arg–Ile bond, our data strongly suggest that human C1r is a 'trypsin-like' serine proteinase with a restricted specificity for arginyl bonds.

Inspection of the amino acid sequence of human C$\overline{1}$r A chain reveals the presence of internal homology between two segments (Fig. 5), located in the *N*-terminal part of the chain (positions 10–78) and at the junction between the α and β regions (positions 186–257). These segments share 30% of identical residues, among them two half-cystine residues (positions 54 and 72 and positions 233 and 251) that in the native protein form two homologous disulphide bridges (G. J. Arlaud & J. Gagnon, unpublished work). This intrachain homology probably corresponds to that detected from partial sequence of human C$\overline{1}$s A chain (Spycher *et al.*, 1986), and its biological significance is unknown.

The data presented in this paper and previous studies (Gagnon & Arlaud, 1985) indicate that two carbohydrate moieties are attached to the polypeptide core of human C$\overline{1}$r A chain. These are linked to asparagine residues at positions 108 and 204, in sequences Asn-Gly-Thr and Asn-Tyr-Ser respectively. It is noteworthy that both sites are located in the *N*-terminal α fragment, whereas, in contrast with human C$\overline{1}$s A chain (Spycher *et al.*, 1986), the *C*-terminal region of human C$\overline{1}$r A chain contains no carbohydrate-attachment site.

N-terminal sequence analysis and amino acid composition of peptides αST3b and CN1aSP4 clearly establish that two different amino acid residues, serine and leucine, occur at position 135 of C$\overline{1}$r A chain, in relative proportions 2.5/1.0, as estimated from amino acid analysis. The exact origin of this heterogeneity is presently unknown, although it should be stressed that C$\overline{1}$r used for our sequence studies was always obtained from pools of human plasma originating from several donors. As this heterogeneity occurs in the vicinity of the putative β-hydroxyaspartic acid residue, it could have

important functional consequences, which justify further studies at the molecular level.

We are grateful to Professor M. G. Coulomb for his continuous help and encouragement. We thank Mrs. M. Lacroix for expert technical assistance and Ms. C. Chabot for skilful secretarial services.

REFERENCES

Allen, G. (1981) Lab. Tech. Biochem. Mol. Biol. **9**, 1–327

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A. & Brownlee, G. G. (1984) EMBO J. **3**, 1053–1060

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

Arlaud, G. J. & Gagnon, J. (1985) FEBS Lett. **117**, 289–294

Arlaud, G. J., Sim, R. B., Duplaa, A.-M. & Colomb, M. G. (1979) Mol. Immunol. **16**, 445–450

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

Arlaud, G. J., Gagnon, J. & Porter, R. R. (1982) Biochem. J. **202**, 49–59

Arlaud, G. J., Colomb, M. G. & Villiers, C. L. (1985) Biosci. Rep. **5**, 831–837

Benoiton, L., Winitz, M., Coleman, R. F., Birnbaum, S. M. & Greenstein, J. P. (1959) J. Am. Chem. Soc. **81**, 1726–1729

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

Colomb, M. G., Arlaud, G. J. & Villiers, C. L. (1984) Philos. Trans. R. Soc. London Ser. B **306**, 283–292

Cooper, N. R. (1985) Adv. Immunol. **37**, 151–216

Doolittle, R. F. (1985) Trends Biochem. Sci. **10**, 233–237

Drakenberg, T., Fernlund, P., Roepstorff, P. & Sternflo, J. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 1802–1806

Esmon, N. L., De Bault, L. E. & Esmon, C. T. (1983) J. Biol. Chem. **259**, 5698–5704

Fernlund, P. & Stenflo, J. (1983) J. Biol. Chem. **258**, 12509–12512

Foster, D. C., Yoshitake, S. & Davie, E. W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4673–4677

Gagnon, J. & Arlaud, G. J. (1985) Biochem. J. **225**, 135–142

Hojrup, P., Jensen, M. S. & Petersen, T. E. (1985) FEBS Lett. **184**, 333–338

Hunkapiller, M. W. (1985) User Bulletin no. 14, Applied Biosystems, Fostir City

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

Kornguth, M. L. & Sallach, H. J. (1960) Arch. Biochem. Biophys. **91**, 39–42

Long, G. L., Belagaje, R. M. & McGillivray, R. T. A. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 5653–5656

McMullen, B. A., Fujikawa, K. & Kisel, W. (1983*a*) Biochem. Biophys. Res. Commun. **115**, 8–14

McMullen, B. A., Fujikawa, K., Kisiel, W., Sasagawa, T., Howald, W. N., Kwa, E. Y. & Weinstein, B. (1983*b*) Biochemistry **22**, 2875–2884

Morita, T., Isaacs, B. S., Esmon, C. T. & Johnson, A. E. (1984) J. Biol. Chem. **259**, 5698–5704

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

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

Sim, R. B. (1981) Methods Enzymol. **80**, 26–42

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

Spycher, S. E., Nick, H. & Rickli, E. E. (1986) Eur. J. Biochem. **156**, 49–57

Sugo, T., Fernlund, P. & Stenflo, J. (1984*a*) FEBS Lett. **165**, 102–106

Sugo, T., Björk, I., Holmgren, A. & Stenflo, J. (1984*b*) J. Biol. Chem. **259**, 5705–5710

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

Villiers, C. L., Arlaud, G. J. & Colomb, M. G. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4477–4481

Received 3 July 1986/29 August 1986; accepted 6 October 1986