in a like manner, and this possibility is presently being considered.

Registry No. Hm, 16009-13-5; lithium, 7439-93-2; sodium, 7440-23-5; nitrate, 14797-55-8; chloride, 16887-00-6; Im, 288-32-4.

References

Adams, P. A., & Berman, M. C. (1980) *Biochem. J. 191*, 95–102.

Bearden, A. J., Morgan, W. T., & Muller-Eberhard, U. (1974) *Biochem. Biophys. Res. Commun. 61*, 265–272.

Beaven, G. H., Chen, S., D'Albis, A., & Gratzer, W. B. (1974) *Eur. J. Biochem. 41*, 539–546.

Brown, S. B., Dean, T. C., & Jones, P. (1970) *Biochem. J. 117*, 733–779.

Brown, S. B., Shillcock, M., & Jones, P. (1976) *Biochem. J. 153*, 279–285.

Campbell, V. M. (1981) Ph.D. Thesis, University of Witwaterstrand.

Clark, P., Rachinsky, J. R., & Foster, J. F. (1962) *J. Biol. Chem. 237*, 2509–2513.

Collier, G. S., Pratt, J. M., De Wet, C. R., & Tshabalala, C. F. (1979) *Biochem. J. 179*, 281–289.

Conway, T. P., & Muller-Eberhard, U. (1976) *Arch. Biochem. Biophys. 172*, 558–564.

Davies, D. M., Smith, A., Muller-Eberhard, U., & Morgan, W. T. (1979) *Biochem. Biophys. Res. Commun. 91*, 1504–1511.

Gallagher, W., & Elliott, W. (1968) *Biochem. J. 108*, 131–136.

Gibbs, E., Skowronek, W. R., Jr., Morgan, W. T., Muller-Eberhard, U., & Pasternack, R. F. (1980) *J. Am. Chem. Soc. 102*, 3939–3944.

Hershko, C. (1975) *Br. J. Haematol. 29*, 199–204.

Hrkal, Z., & Muller-Eberhard, U. (1971) *Biochemistry 10*, 1746–1750.

Hrkal, Z., Vodrazka, Z., & Kalousek, I. (1974) *Eur. J. Biochem. 43*, 73–78.

Kodicek, M., Cvekl, A., & Vodrazka, Z. (1981) *Int. J. Biol. Macromol. 3*, 397–398.

Koshland, D. E., Jr., & Neet, K. E. (1968) *Annu. Rev. Biochem. 37*, 359–410.

Morgan, W. T. (1976) *Ann. Clin. Res. 8*, 223–232.

Morgan, W. T., & Vickery, L. E. (1978) *J. Biol. Chem. 253*, 2940–2945.

Morgan. W. T., Liem, H. H., Sutor, R. P., & Muller-Eberhard, U. (1976) *Biochim. Biophys. Acta 444*, 435–445.

Muller-Eberhard, U. (1970) *N. Engl. J. Med. 283*, 1090–1094.

Muller-Eberhard, U. (1978) in *Transport by Proteins* (Blauer, G., & Sund, H., Eds.) pp 295–310, Walter de Gruyter, Berlin and New York.

Muller-Eberhard, U., Bosman, C., & Liem, H. H. (1970) *J. Lab. Clin. Med. 76*, 426–431.

Parr, G. R., & Pasternack, R. F. (1977) *Bioinorg. Chem. 7*, 277–282.

Pasternack, R. F., Francesconi, L., Raff, D., & Spiro, E. (1973) *Inorg. Chem. 12*, 2602–2611.

Pasternack, R. F., Cobb, M. A., & Sutin, N. (1975) *Inorg. Chem. 14*, 866–873.

Seery, V. L., Hathaway, G., & Muller-Eberhard, U. (1972) *Arch. Biochem. Biophys. 150*, 269–272.

Smith, A., & Morgan, W. T. (1979) *Biochem. J. 182*, 47–54.

Tanford, C. (1968) *Adv. Protein Chem. 23*, 121–282.

Voet, J. G., Coe, J., Epstein, J., Matossian, V., & Shipley, T. (1981) *Biochemistry 20*, 7182–7185.

Complete Amino Acid Sequence of the Catalytic Chain of Human Complement Subcomponent $C\bar{1}r$[†]

Gérard J. Arlaud* and Jean Gagnon

ABSTRACT: The amino acid sequence of human $C\bar{1}r$ b chain has been determined, from sequence analysis performed on fragments obtained by CNBr cleavage, dilute acid hydrolysis, tryptic cleavage of the succinylated protein, and subcleavages by staphylococcal protease. The polypeptide chain contains 242 amino acids (M_r 27 096), and the sequence shows strong homology with other mammalian serine proteases. The histidine, aspartic acid, and serine residues of the active site (His-57, Asp-102, and Ser-195 in bovine chymotrypsinogen) are located at positions 39, 94, and 191, respectively. The chain, which lacks the "histidine-loop" disulfide bridge, contains five half-cystine residues, of which four (positions 157–176 and 187–217) are homologous to residues involved in disulfide bonds generally conserved in serine proteases, whereas the half-cystine residue at position 114 is likely to be involved in the single disulfide bridge connecting the catalytic b chain to the N-terminal a chain. Two carbohydrate moieties are attached to the polypeptide chain, both via asparagine residues at positions 51 and 118.

The first component of human complement classical pathway, C1, is a calcium-dependent complex consisting of glycoproteins C1q, C1r, and C1s.[1] C1q is considered as a recognition unit, whereas C1r and C1s, which are zymogens of proteases, are sequentially activated through limited proteolysis during C1 activation. The zymogen C1r is autocatalytically activated to its protease form $C\bar{1}r$ (EC 3.4.21.41), which in turn mediates C1s activation [see review by Sim (1981a)]. The active

[†] From the Département de Recherches Fondamentales (E.R.A. CNRS No. 695, Unité INSERM No. 238), Laboratoire de Biologie Moléculaire et Cellulaire, Centre d'Etudes Nucléaires de Grenoble 85X, 38041 Grenoble Cédex, France (G.J.A.), and the Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, U.K. (J.G.). *Received September 21, 1982.* G.J.A. held a NATO Research Fellowship. A preliminary report of this work was presented at the Federation of American Societies for Experimental Biology meeting in New Orleans, LA, April 1982.

[1] Abbreviations: iPr_2P-F, diisopropyl phosphorofluoridate; Tos-Phe-CH_2Cl, 1-chloro-4-phenyl-3-(L-tosylamido)butan-2-one; HPLC, high-pressure liquid chromatography; EDTA, ethylenediaminetetraacetic acid. The nomenclature of complement components is that recommended by the World Health Organization (1968); activated components are indicated by a superscript bar, e.g., $C\bar{1}r$.

0006-2960/83/0422-1758$01.50/0

protease C1s (EC 3.4.21.42) then cleaves and activates complement components C2 and C4 [see review by Reid & Porter (1981)].

Both C1r and C1r are isolated as dimers, composed of two identical, noncovalently linked monomers, which can be dissociated at acid pH (Arlaud et al., 1980) or in denaturing conditions. Each monomer, which, in proenzyme C1r, is a single-chain polypeptide of M_r 83 000–95 000, is cleaved upon activation, probably at a single site. C1r monomers therefore consist of two disulfide-linked polypeptide chains named a (M_r 56 000–60 000) and b (M_r 27 000–36 000) (de Bracco & Stroud, 1971; Valet & Cooper, 1974; Ziccardi & Cooper, 1976a,b; Sim et al., 1977). Both the proenzyme and activated forms of C1r are known to bind calcium (Villiers et al., 1980).

The N-terminal a chain, which is thought to mediate intermonomer and intersubcomponent interactions (Arlaud et al., 1980), contains 7.4% carbohydrate and has a blocked amino terminus (Sim et al., 1977). The iPr_2P-F reactive site of C1r is located in the C-terminal b-chain moiety (Takahashi et al., 1975; Sim & Porter, 1976), which contains 12.4% carbohydrate (Sim et al., 1977). The N-terminal sequence of 20 amino acid residues of C1r b chain, first determined by Sim et al. (1977), showed obvious homology with other serine proteases. More recently, N-terminal sequence analysis of the CNBr-cleavage peptides allowed the identification of about two-thirds of the amino acid residues of C1r b chain and revealed that it lacks the "histidine-loop" disulfide bridge (Arlaud et al., 1982), a characteristic that is shared by C1s b chain (Arlaud & Gagnon, 1981). This paper presents the detailed proof of the complete amino acid sequence of C1r b chain.

Experimental Procedures

Materials

Trypsin (treated with Tos-Phe-CH_2Cl) and carboxypeptidase Y were obtained from Worthington Biochemical Corp. V8 proteinase from *Staphylococcus aureus* was from Miles Laboratories.

Iodo[2-^{3}H]acetic acid (54 Ci/mol) was from the Radiochemical Centre, Amersham. Polybrene [1,5-diaza-1,5-dimethylundecamethylene poly(methobromide)] and CNBr were obtained from Aldrich Chemical Co. Quadrol [*N*,*N*,*N*′,*N*′-tetrakis(2-hydroxypropyl)ethylenediaminetrifluoroacetate] was from Fluka. Methanol, acetonitrile, and all other chemicals used for HPLC and automated sequencing were obtained from Rathburn Chemicals, Walkerburn, Peebleshire, Scotland. *p*-Toluenesulfonic acid was from Pierce Chemical Co. Spectrapor 6 dialysis tubing (M_r cutoff = 1000) was purchased from Spectrum Medical Industries Inc.

Methods

C1r was purified from human serum as described previously (Arlaud et al., 1979). Except for the protein that was subsequently used for CNBr cleavage of the b chain, isolated C1r was treated with iPr_2P-F, as described previously (Arlaud et al., 1982). Reduction of C1r, alkylation by iodo[2-^{3}H]acetic acid, and separation of C1r a and b chains by high-pressure gel-permeation chromatography were performed as described previously (Arlaud et al., 1982).

CNBr Cleavage. Reduced and S-[^{3}H]carboxymethylated C1r b chain (100 nmol) was cleaved by CNBr, and peptides CB1a, CB1b, CB2, and CB3 were purified by gel filtration on Sephadex G-50 and reversed-phase HPLC as published previously (Arlaud et al., 1982).

Dilute Acid Cleavage. Reduced and S-[^{3}H]carboxymethylated C1r b chain (140 nmol) was dissolved in 70% (v/v) formic acid (1.6 mL). The vial was flushed with nitrogen and sealed, and the mixture was incubated for 48 h at 40 °C with constant stirring. The solution was then made 6 M in urea, and the cleavage products were separated by high-pressure gel-permeation chromatography, by repetitive loading of 100-μL fractions on a 7.5 mm × 600 mm column of TSK-G 3000 SW (Toyo Soda Manufacturing, Tokyo, Japan) equilibrated in 6 M urea/0.2 M formic acid and pumped at 1 mL/min. Peptide AC3 was dialyzed against 1% (v/v) acetic acid in Spectrapor 6 dialysis tubing (M_r cutoff = 1000) and freeze-dried.

Peptide Separation by Reversed-Phase HPLC. All separations were performed at room temperature, with a μBondapak C_{18} column (3.9 mm × 300 mm) pumped at 1 mL/min by a Waters Associates system. Two solvent systems were used. In system 1, the column was equilibrated with a mixture of 0.1% NH_4HCO_3 and acetonitrile in the ratio 95:5 (v/v) and then eluted with a linear gradient to give a final ratio of 45:55 (v/v). In system 2, the column was equilibrated with a mixture of solutions A (0.1% trifluoroacetic acid) and B (acetonitrile/methanol/propan-2-ol, 1:1:1 v/v/v) in the ratio 95:5 (v/v) and then eluted with a linear gradient to give a final ratio of 45:55 (v/v). Peptides were detected from the absorbance at 206 nm.

Succinylation of C1r b Chain and Tryptic Cleavage. Reduced and S-[^{3}H]carboxymethylated C1r b chain (180 nmol) was dissolved in 6 M guanidine hydrochloride (2 mL), and the pH was adjusted to 8.0 with 4 M NaOH. A 10-fold excess of succinic anhydride with respect to protein (by weight) was added, and succinylation was carried out according to the procedure of Koide et al. (1978). Excess reagents were removed by dialysis against 0.1 M NH_4HCO_3. Reduced and S-[^{3}H]carboxymethylated succinylated C1r b chain was incubated in 0.1 M NH_4HCO_3 with Tos-Phe-CH_2Cl-treated trypsin (60 μg) for 2 h at 37 °C, and the incubation was repeated after addition of another 60 μg of enzyme for a final enzyme/protein ratio of 2% (w/w). After addition of 2.5 mM iPr_2P-F, the digest was applied to a column (2 cm × 95 cm) of Sephadex G-50 (superfine grade) equilibrated with 0.1 M NH_4HCO_3. Ten pools were collected and freeze-dried, and tryptic peptides were further analyzed by reversed-phase HPLC, in solvent system 1 described above.

Staphylococcal Protease Digestion. Reduced and S-[^{3}H]carboxymethylated C1r b chain (3.5 mg), CNBr-cleavage peptide CB3 (0.25 mg), and tryptic peptide T1 (0.7 mg) were dissolved at 0.5–1.5 mg/mL in 0.1 M NH_4HCO_3/2 mM EDTA (pH 7.8) and digested with staphylococcal protease (enzyme to substrate ratio 1/30 w/w) for 5–6 h at 37 °C. Subfragments from peptides T1 and CB3 were separated by reversed-phase HPLC in, respectively, solvent systems 1 and 2 described above.

Acetic acid was added to 10% (v/v) to the peptides obtained from C1r b chain. This treatment rendered insoluble part of the peptides, among which peptide SP8, which was purified by high-pressure gel-permeation chromatography on two 7.5 mm × 600 mm LKB Blue columns in series (TSK-G 3000 SW and TSK-G 2000 SW) equilibrated in 6 M urea/0.2 M formic acid at a flow rate of 1 mL/min.

Carboxypeptidase Y. Reduced and alkylated C1r b chain (1 mg) was incubated with carboxypeptidase Y in 50 mM pyridine–acetate (pH 5.5) for up to 1 h at 37 °C (enzyme to substrate ratio 1/15 w/w). Samples were analyzed on a Durrum D-500 amino acid analyzer.

Amino Acid and Carbohydrate Analyses. Peptides were hydrolyzed for 24 h under reduced pressure at 110 °C in

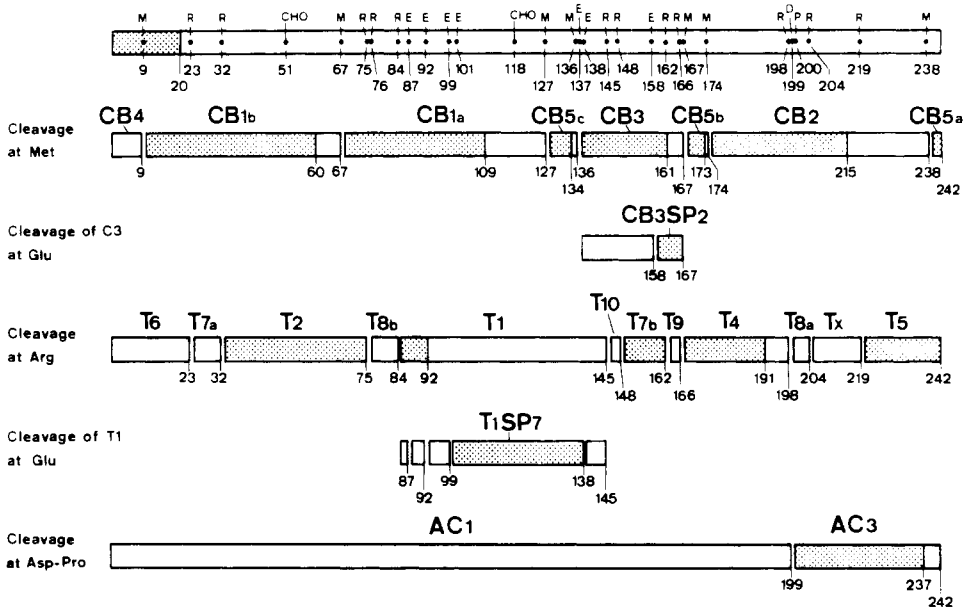

FIGURE 1: Summary of sequencing strategy. Standard one-letter abbreviations are used to designate residues defining the different fragments obtained from Clr b chain, and the two carbohydrate attachment sites are indicated by –CHO. Peptides generated by CNBr cleavage, tryptic cleavage at arginyl residues, dilute acid cleavage, and cleavage by staphylococcal protease are designated CB, T, AC, and SP, respectively. For subdigestion of peptides CB3 and T1 by staphylococcal protease, minor fragments originating from partial cleavages are not depicted. Stippling denotes those portions of Clr b chain that were sequenced.

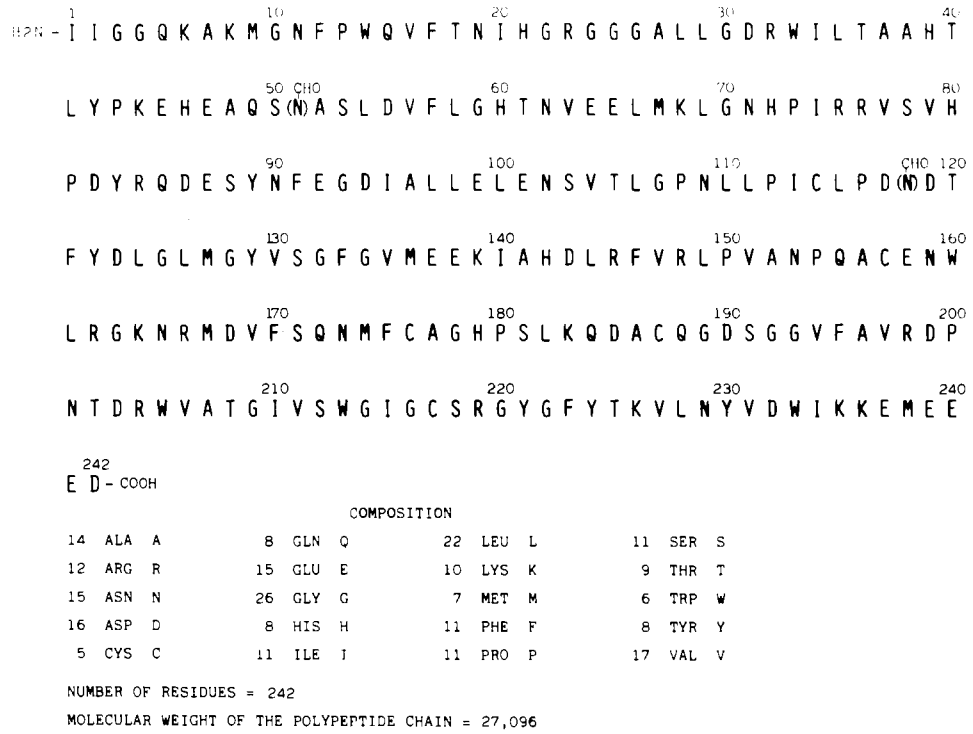

FIGURE 2: Amino acid sequence and composition of Clr b chain.

constant-boiling HCl containing 0.1% (v/v) 2-mercaptoethanol and 4 mM phenol. Half-cystine was determined as *S*-(carboxymethyl)cysteine. Tryptophan was not determined. Analyses were performed on a Durrum D-500 analyzer.

Hexosamines were determined after hydrolysis of samples in 3 M *p*-toluenesulfonic acid for 24 h at 100 °C, according to the procedure recommended by Allen & Neuberger (1975). Samples were analyzed on a Durrum D-500 amino acid analyzer.

Total carbohydrate analysis of the CNBr-cleavage glycopeptides CB1a and CB1b was performed by gas–liquid chromatography of *O*-(trimethylsilyl) ethers.

Automated Sequence Analysis. Automated Edman degradation was performed in a Beckman 890 C sequenator with the 0.3 M Quadrol program of Hunkapiller & Hood (1978). The procedure used for amino acid phenylthiohydantoin derivative identification by reversed-phase HPLC has been detailed by Christie & Gagnon (1982).

Results

Chemical cleavage of methionyl bonds and of the single Asp–Pro bond in Clr b chain, as well as enzymatic cleavage of arginyl and glutamyl bonds, was used to generate the necessary peptides for determining the amino acid sequence

Table I: Amino Acid Compositions[a] of C1r b Chain Selected Peptides and Details of Automated Sequence Analyses

	mol of amino acid/mol of peptide									
amino acid	CB3SP2	T2	T4	T5	T7b	T1	T1SP6	T1SP7	T1SP8	AC3
Asx	1.9 (2)	4.7 (4)	4.5 (4)	2.9 (3)	1.8 (2)	9.7 (10)	5.8 (6)	5.6 (6)	0.4	4.7 (5)
Thr[b]		2.7 (3)	0.5	1.1 (1)		2.3 (2)	1.7 (2)	1.9 (2)		2.4 (3)
Ser[b]		2.3 (2)	3.5 (3)			3.5 (3)	1.0 (1)	2.1 (2)	1.0 (1)	2.0 (2)
Hse[c]	0.7 (1)									
Glx		5.0 (5)	3.1 (3)	4.3 (4)	2.2 (2)	6.9 (7)	1.1 (1)	2.9 (3)	2.1 (2)	4.0 (4)
Pro		2.0 (2)	1.4 (1)		1.9 (2)	3.3 (3)	2.7 (3)	2.6 (3)		1.2 (1)
Gly	1.4 (1)	2.9 (2)	4.0 (4)	2.2 (2)	0.3	6.0 (6)	1.5 (1)	5.6 (5)	4.2 (4)	5.1 (5)
Ala		3.6 (4)	2.6 (3)		2.1 (2)	2.1 (2)	0.3			1.5 (1)
Val		2.5 (2)	2.7 (3)	2.1 (2)	1.0 (1)	3.2 (3)	1.2 (1)	3.2 (3)	2.4 (2)	3.8 (4)
Cys[d]			1.6 (2)		1.0 (1)	1.1 (1)	1.0 (1)	0.9 (1)		1.1 (1)
Met		0.9 (1)	1.3 (2)	0.9 (1)		1.6 (2)		1.8 (2)	1.6 (2)	1.0 (1)
Ile		1.9 (2)	0.3	1.0 (1)		2.8 (3)	1.1 (1)	1.0 (1)		2.3 (3)
Leu	1.0 (1)	6.3 (6)	1.2 (1)	1.1 (1)	2.0 (2)	9.7 (10)	4.6 (5)	6.7 (7)	1.9 (2)	1.8 (1)
Tyr		1.2 (1)	0.3	2.5 (3)		3.0 (3)	0.9 (1)	2.0 (2)	0.8 (1)	2.2 (3)
Phe		1.2 (1)	2.4 (3)	1.1 (1)		2.8 (3)	0.8 (1)	2.0 (2)	0.7 (1)	1.1 (1)
His		3.4 (4)	0.9 (1)			1.2 (1)				0.4
Lys	1.0 (1)	1.9 (2)	1.2 (1)	2.9 (3)		1.2 (1)	0.3	0.3		2.8 (3)
Arg	1.9 (2)	1.3 (1)	1.1 (1)		1.0 (1)	1.1 (1)				2.2 (2)
Trp	ND[e] (1)	ND (1)	ND	ND (1)	ND (1)	ND	ND	ND	ND	ND (3)
purificn yield (%)	35	47	36	27	28	74	25	38	8	35
amt used in automated sequence analysis (nmol)	21	72	64	40	51	21	24	36	7	46
initial yield (%)	83	67	39	59	41	38	42	50	53	33
repetitive yield (%)[f]	89	93	87	92	87	87	93	93	83	93

[a] Amino acid compositions were calculated from duplicate 24-h HCl hydrolysates. Values obtained from the sequence are indicated in parentheses. [b] Threonine and serine yields were corrected on the assumption of the same rate of destruction as observed for C1r b chain. [c] Quantified as the sum of homoserine and homoserine lactone values. [d] Estimated as *S*-(carboxymethyl)cysteine. [e] ND, not determined. [f] Calculated by regression analysis of all steps.

of C1r b chain. The sequencing strategy is summarized in Figure 1, and the amino acid sequence of the chain is shown in Figure 2.

Further Sequence Analysis of CNBr-Cleavage Peptides. N-Terminal sequence analysis of the CNBr peptides previously reported (Arlaud et al., 1982) had been performed on peptides prepared from [1,3-^{3}H]iPr_2P-labeled C1r b chain. In the case of CB2, the active serine containing peptide, pretreatment with iPr_2P-F enabled us to locate the [1,3-^{3}H]iPr_2P-labeled serine at position 191 (Figure 2). This, however, was likely the cause of a sudden drop of the stepwise yield after this position, which impaired subsequent identification of residues at positions 197, 198, and 202 (Arlaud et al., 1982). A new N-terminal sequence analysis performed on CB2 prepared from untreated C1r b chain allowed the positive identification at these positions of residues valine, arginine, and threonine, respectively, and extended the known sequence of CB2 to residue 210 (Figure 2).

Previous N-terminal sequence analysis of peptide CB3 allowed the identification of residues 137–161, with the exception of the residue at position 160 (Arlaud et al., 1982). $CB3SP_2$, the C-terminal peptide of CB3, was obtained after cleavage by staphylococcal protease of the Glu–Asn bond at position 158. N-Terminal sequence analysis of peptide $CB3SP_2$ provided the identification of residues 159–167, thus allowing the determination of the complete sequence of CB_3 (Figures 1 and 2).

Purification and Sequence Analysis of Tryptic Peptides from Succinylated C1r b Chain. The initial fractionation of the tryptic digest of reduced and S-[^{3}H]carboxymethylated, succinylated C1r b chain is illustrated in Figure 3. Ten pools were collected, freeze-dried, and further analyzed by reversed-phase HPLC, in the solvent system referred to under Methods as system 1.

Each of pools 1, 2, 5, 6, and 10 was found to contain a single major peptide, denoted T1, T2, T5, T6, and T10, respectively.

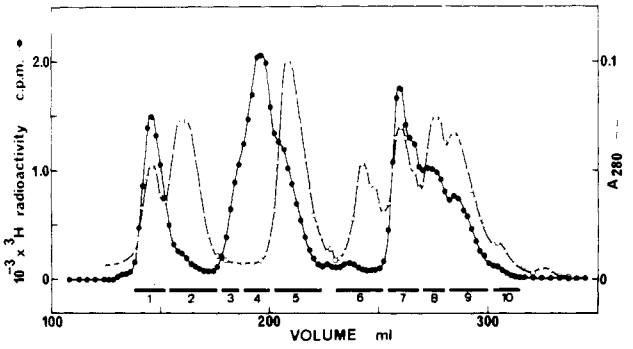

FIGURE 3: Fractionation of a tryptic digest from reduced and S-[^{3}H]carboxymethylated succinylated C1r b chain (180 nmol) on a column (2.0 × 95 cm) of Sephadex G-50 (superfine grade) equilibrated with 0.1 M NH_4HCO_3 (pH 7.8) and eluted at 15 mL/h. 2-mL fractions were collected, and radioactivity was determined on 30-μL aliquots. Pools were made as indicated by bars.

The N-terminal peptide T6 was also present in pool 7, which also contained two peptides, T7a and T7b, the latter being resolved into two peaks by reversed-phase HPLC (Figure 4). Pool 8 contained two main peptides, T8a and T8b, in addition to lower amounts of peptides T7a and T7b. T9 was the main peptide recovered from pool 9, which also contained peptides T8a and T8b.

No significant peptide was found in pool 3, whereas amino acid composition of pool 4 (Table I) indicated that it contained one peptide, denoted T4, which could be located within the chain by comparison with the amino acid compositions and the known part of the sequence of peptides CB5b and CB2 (Figure 1). As peptide T4 could not be recovered in good yield from HPLC, it was submitted to sequence analysis without further purification.

In summary, eleven peptides, T1, T2, T4, T5, T6, T7a, T7b, T8a, T8b, T9, and T10, resulting from the tryptic cleavage of succinylated C1r b chain were purified and could be located

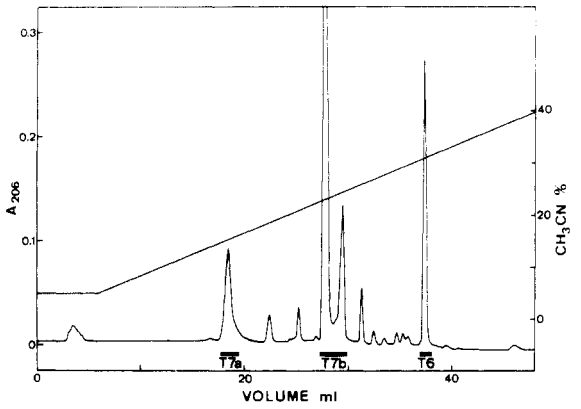

FIGURE 4: Separation by reversed-phase HPLC of tryptic peptides from pool 7 of Sephadex G-50. One-third of pool 7 from Sephadex G-50 (see Figure 3) was loaded on to a μBondapak C_{18} column (3.9 mm × 300 mm) equilibrated with 0.1% NH_4HCO_3 and acetonitrile in the ratio 95:5 (v/v). The column was eluted at 1 mL/min with a linear gradient to give a final ratio of 45:55 (v/v). Pools were made as indicated by bars.

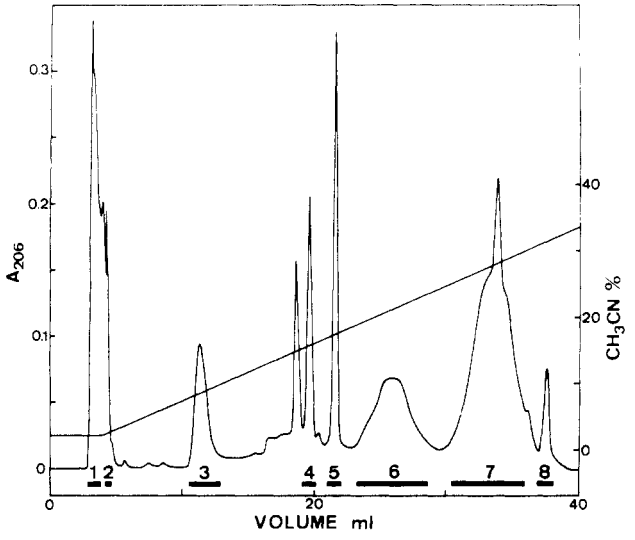

FIGURE 5: Separation by reversed-phase HPLC of peptides from staphylococcal protease subdigest of T1. Subdigestion of peptide T1 (0.7 mg) was carried out as described under Methods. One-third of the digest was loaded on to a μBondapak C_{18} column (3.9 mm × 300 mm), and separation was carried out as described in Figure 4. Pools were made as indicated by bars.

within the sequence of the polypeptide chain (Figure 1). The tryptic peptide corresponding to positions 205–219 (T*x*) (Figures 1 and 2) was not recovered. Considering its highly hydrophobic amino acid composition, it is likely that T*x* could not be desorbed from the C_{18} column and was lost during reversed-phase HPLC.

The complete sequence of peptide T2 was determined (positions 33–75) with the exception of the residue at position 51. This site is likely to represent an asparagine-linked carbohydrate moiety, as a serine residue was identified at position 53, which is consistent with an Asn-X-Ser sequence requirement for attachment of that type of carbohydrate (Neuberger et al., 1972). The sequence of T2 provided the overlap between peptides CB1b and CB1a (Figure 1), thus extending the N-terminal sequence of C$\bar{1}$r b chain to position 109.

Determination of the complete sequence of peptide T7b (positions 149–162) confirmed the identification already provided from the corresponding sequences of peptides CB3 and $CB3SP_2$. Sequence analysis of T7b, as well as A_{280nm} measurement (Edelhoch, 1967), confirmed the presence of a tryptophan residue at position 160.

The N-terminal sequence of T4 was determined for 25 cycles (residues 167–191). This clearly confirmed the overlap between peptides CB5b and CB2 (Arlaud et al., 1982) and provided a one-residue overlap between peptides $CB3SP_2$ and CB5b (Figures 1 and 2). Thus, the sequence of residues 137–215 of C$\bar{1}$r b chain was established (Figure 1).

T5 was the only tryptic peptide lacking arginine (Table I) and was therefore identified as the C-terminal peptide. Determination of the complete sequence of T5 (residues 220–242) confirmed that the C-terminal sequence of C$\bar{1}$r b chain is Met-Glu-Glu-Glu-Asp-COOH, as concluded previously from the amino acid composition and sequence of peptide CB5a (Arlaud et al., 1982).

From its amino acid composition and from a short N-terminal amino acid analysis (eight cycles), T1 was recognized as the peptide generated by cleavage of the Arg–Gln bond at position 84 and ending at position 145 (Figures 1 and 2). Comparison of the amino acid composition of T1 (Table I) with the portions of its structure already known from the sequences of CB1a, CB5c, and CB3 (Figure 1) indicated that no Glx residue was present in the unknown part of T1, which was therefore submitted to subdigestion by staphylococcal protease. From the digest, eight peptides, referred to as T1SP1 to T1SP8, were separated by reversed-phase HPLC (Figure 5). As summarized in Figure 6, these peptides resulted from expected cleavage of glutamyl bonds at positions 87, 92, 99, 101, and 138, as well as from the partial cleavage of the Asp–Leu bond at position 123.

The Glu–Glu bond at position 137 was not cleaved, as judged from amino acid composition of peptide T1SP8 (Table I). N-Terminal sequence analyses of T1SP6 and T1SP7 showed that each peptide contained two sequences, a major sequence commencing Leu-Glu-Asn-Ser-Val-... and a minor sequence commencing Asn-Ser-Val-..., a heterogeneity due to incomplete cleavage of the glutamyl bond at position 101. The complete amino acid sequence of peptide T1SP7 was determined (residues 100–138), with the exception of the residue at position 118. Again, the presence of a threonine residue at position 120 strongly favors the hypothesis of a carbohydrate

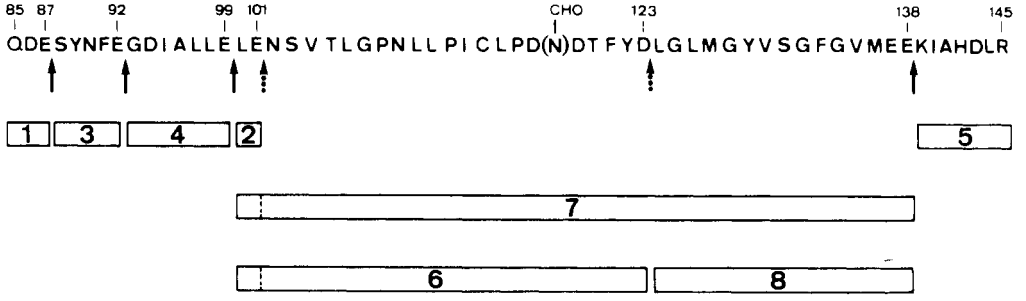

FIGURE 6: Diagrammatic representation of fragments generated by subdigestion of peptide T1 by staphylococcal protease. Dotted arrows indicate sites of partial cleavage. Fragments are numbered according to their elution order on HPLC (see Figure 5).

Table II: Carbohydrate Analyses of C1r b Chain and Selected Peptides

monosaccharide	mol/mol of protein or peptide				
	C1r b chain	CB1a	T1	CB1b	T2
mannose[a]	ND[c]	2.7	ND	3.5	ND
galactose[a]	ND	1.9	ND	2.1	ND
glucose[a]	ND	trace	ND	trace	ND
N-acetylglucosamine[b]	6.8	2.2	2.4	2.5	2.8
N-acetylgalactosamine[b]	0.3	0.1	0.1	0.2	0.1
sialic acid[a]	ND	1.2	ND	1.3	ND

[a] Analyzed by gas-liquid chromatography of *O*-(trimethylsilyl) ethers. [b] Hexosamines were determined after hydrolysis of samples in 3 M *p*-toluenesulfonic acid for 24 h at 100 °C. [c] ND, not determined.

moiety linked to an asparagine residue at position 118 (Neuberger et al., 1972). The presence of an asparagine residue at this position is consistent with the Asx content of peptides T1SP6 and T1SP7 (Table I, Figure 2), whereas the glycopeptidic nature of these two peptides partly explains that they are eluted as wide peaks on reversed-phase HPLC (Figure 5).

From the sequences of both peptides T1SP7 and T1SP8, the sequence of residues at positions 132–136, which had been tentatively proposed as Gly-Val-Gly-Phe-Met (Arlaud et al., 1982), was unambiguously established as Gly-Phe-Gly-Val-Met. The complete sequence of peptide T1SP7 provided the overlap between peptides CB1a, CB5c, and CB3, thus allowing the determination of a continuous sequence from residues 1 to 215 (Figure 1).

Dilute Acid Cleavage. N-Terminal sequence analysis of peptide CB2 allowed the identification of an Asp–Pro bond at position 199 (Arlaud et al., 1982). Dilute acid cleavage of C1r b chain generated the C-terminal fragment AC3, which was separated from the remaining N-terminal part of the C1r b chain (AC1) and from side-reaction products by gel-permeation HPLC. N-Terminal sequence analysis of peptide AC3 allowed the identification of residues at positions 200–237, thus providing the overlap between peptides CB2 and T5 (Figure 1) and completing the amino acid sequence of C1r b chain.

Carbohydrate Analyses. Amino sugar analyses of C1r b chain indicated the almost exclusive presence of *N*-acetylglucosamine, which was contained in peptides CB1a, CB1b, T1, and T2 (Table II), and the absence of significant amounts of *N*-acetylgalactosamine ruled out the possibility of O-glycosidic linkage of carbohydrate (Kornfeld & Kornfeld, 1976). As peptides CB1b and T2 on one hand, CB1a and T1 on the other hand, partially overlap (Figure 1), the presence of two asparagine-linked oligosaccharides, located respectively between positions 33 and 67 and between positions 85 and 127, can therefore be postulated. Carbohydrate analyses of peptides CB1a and CB1b showed that, in addition to *N*-acetylglucosamine, both oligosaccharides contain mannose, galactose, and sialic acid.

Discussion

N-Terminal sequence analyses of 14 selected peptides obtained from cleavage of methionyl bonds, tryptic cleavage at arginyl residues, dilute acid hydrolysis, and subcleavages by staphylococcal protease provided sufficient information to establish the complete amino acid sequence of human C1r b chain. The validity of the sequence Arg-Met-Asp at positions 166–168, which was based on a one-residue overlap between peptides $CB3SP_2$ and T4 (Figure 1), was confirmed from N-terminal analysis of peptide SP8, obtained by direct cleavage of C1r b chain by staphylococcal protease, originating from cleavage of the glutamyl bond at position 158, which gave the expected sequence Asn-Trp-Leu-Arg-Gly-Lys-Asn-Arg-Met-Asp-Val-Phe-....

The single polypeptide core of human C1r b chain comprises 242 amino acid residues, with an M_r of 27 096. The residues involved in the active site of serine proteases (His-57, Asp-102, and Ser-195 in bovine chymotrypsinogen A) are located at positions 39, 94, and 191 in C1r b chain, respectively. Since C1r b chain contains an amino-terminal isoleucine and an aspartic acid residue at position 190 (Asp-194 in bovine chymotrypsinogen A), it is likely that, in activated C1r, these residues form an ion pair as in chymotrypsin. The presence of an aspartic acid residue at position 185 likely indicates that C1r is a "trypsin-like" serine protease, which is in agreement with the esterolytic and proteolytic specificity of human C1r (Naff & Ratnoff, 1968; Volanakis et al., 1977; Sim et al., 1977; Andrews & Baillie, 1979; Sim, 1981b).

Five half-cystine residues have been located in C1r b chain. Although the position of the disulfide bonds remains to be clearly demonstrated, it can be noticed that residues at positions 157 and 176 are homologous to those forming the "methionine loop" in other serine proteases, whereas residues at positions 187 and 217 are equivalent to those involved in another conserved disulfide bridge linking the primary and secondary binding sites of serine proteases (Young et al., 1978). In the same way, the half-cystine residue at position 114 is homologous to that implicated in the interchain disulfide bond in chymotrypsinogen, plasminogen, factor X, and prothrombin (Young et al., 1978) and is likely involved in the single disulfide bridge connecting the C-terminal b chain to the N-terminal a chain in human C1r.

C1r b chain lacks the "histidine loop, a disulfide bond that is present in all other known mammalian serine proteases but C1s (Arlaud & Gagnon, 1981). It must be emphasized that C1r b chain is poorly disulfide bonded, as it contains only two disulfide bridges, whereas the number of these bonds in the catalytic chain of known mammalian serine proteases ranges from three in thrombin (Young et al., 1978; Elion et al., 1977) and group-specific protease (Woodbury et al., 1978) up to six in trypsin (Young et al., 1978).

Four out of the five C-terminal residues of C1r b chain are acidic amino acids, which is not an usual feature of the C-terminal end of serine proteases (de Haën et al., 1975; Young et al., 1978). C1r b chain was found resistant to carboxypeptidase Y digestion at pH 5.5, and this could be explained by its polyacidic C-terminal structure.

Overall comparison of C1r b chain sequence with the corresponding sequences of bovine chymotrypsin A, porcine elastase, human plasmin, porcine kallikrein, bovine trypsin, bovine factor X (Young et al., 1978), and human thrombin (Elion et al., 1977) indicates 25%, 25%, 26%, 26%, 30%, 33%, and 36% homology, respectively. This confirms the trypsin-like nature of human C1r and its closer relationship to blood clotting factors, which was already apparent from partial C1r b chain sequence (Arlaud et al., 1982). Out of 113 residues that are generally conserved in the catalytic chain of serine proteases (Young et al., 1978), 68 are present in C1r b chain. In addition to the half-cystine residues of the His-loop disulfide bridge, another two residues invariant in known serine proteases (Young et al., 1978) are missing: valine at position 34 and proline at position 194 are replaced by isoleucine and valine, respectively.

Earlier work has suggested the presence of asparagine-linked oligosaccharides in C1r b chain (Sim et al., 1977). From

amino acid and sequence data and from carbohydrate analyses, it can be concluded that two carbohydrate moieties are attached to the polypeptide core of $C\bar{1}r$ b chain. Both are linked via asparagine residues at positions 51 and 118, in sequences Asn-Ala-Ser and Asn-Asp-Thr, respectively. The carbohydrate moiety at position 51 is found at an equivalent position in thrombin (Elion et al., 1977; Young et al., 1978), whereas the second carbohydrate moiety might play an important role, due to its location in the neighborhood of half-cystine-114, which is supposed to be involved in the disulfide bridge connecting the two chains of $C\bar{1}r$.

The crystallization of $C\bar{1}r$ (Haupt & Baudner, 1981), together with the extended knowledge of the primary structure reported here, should lead to a clearer understanding of the mechanism of the autocatalytic activation of C1r, which initiates the activation of the classical pathway of complement.

Acknowledgments

We are indebted to Dr. Glynn Wilson, from Unilever Research Colworth Laboratory, who kindly performed the carbohydrate analyses. We are grateful to Professor R. R. Porter and Professor M. G. Colomb for their continuous help and encouragement. We thank A. Gascoyne for performing the amino acid analyses, A. C. Willis for sequence analysis, and M. Lacroix for excellent technical assistance.

Registry No. Human C1r b chain, 84693-76-5; serine proteinase, 37259-58-8.

References

Allen, A. K., & Neuberger, A. (1975) *FEBS Lett. 60*, 76–80.

Andrews, J. M., & Baillie, R. D. (1979) *J. Immunol. 123*, 1403–1408.

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

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) *Biochim. Biophys. Acta 616*, 105–115.

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

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

de Bracco, M. M. E., & Stroud, R. M. (1971) *J. Clin. Invest. 50*, 838–848.

de Haën, C., Neurath, H., & Teller, D. C. (1975) *J. Mol. Biol. 92*, 225–259.

Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.

Elion, J., Downing, M. R., Butkowski, R. J., & Mann, K. G. (1977) in *Chemistry and Biology of Thrombin* (Lundblad, R. L., Fenton, J. W., & Mann, K. G., Eds.) pp 97–111, Ann Arbor Science Publishers, Ann Arbor, MI.

Haupt, H., & Baudner, S. (1981) *Hoppe-Seyler's Z. Physiol. Chem. 362*, 1147–1150.

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

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

Kornfeld, R., & Kornfeld, S. (1976) *Annu. Rev. Biochem. 45*, 217–237.

Naff, G. B., & Ratnoff, O. D. (1968) *J. Exp. Med. 128*, 571–593.

Neuberger, A., Gottschalk, A., Marshall, R. D., & Spiro, R. G. (1972) in *Glycoproteins* (Gottschalk, A., Ed.) 2nd ed., 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. (1981a) *Methods Enzymol. 80*, 6–16.

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

Sim, R. B., & Porter, R. R. (1976) *Biochem. Soc. Trans. 4*, 127–129.

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

Takahashi, K., Nagasawa, S., & Koyama, J. (1975) *FEBS Lett. 55*, 156–160.

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

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

Volanakis, J. E., Schrohenloher, R. E., & Stroud, R. H. (1977) *J. Immunol. 119*, 337–342.

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

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

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

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

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