Biochem. J. (1982) **203**, 559–569
Printed in Great Britain

Completion of the amino acid sequences of the A and B chains of subcomponent C1q of the first component of human complement

Kenneth B. M. REID, Jean GAGNON and Jon FRAMPTON
Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

(*Received 30 October 1981/Accepted 17 February 1982*)

The sequences of amino acid residues 109–224 of the A chain, and residues 109–226 of the B chain, of human subcomponent C1q are given. These results, along with previously published sequence data on the *N*-terminal, collagen-like, regions of the A and B chains [Reid (1979) *Biochem. J.* **179**, 367–371] yield the complete amino acid sequences of the A and B chains of subcomponent C1q. The asparagine residue at position A-124 has been identified as the major site of asparagine-linked carbohydrate in subcomponent C1q. When the sequences of the *C*-terminal, 135-residue-long, 'globular' regions of A and B chains are compared they show 40% homology. The degree of homology over certain stretches of 15–20 residues, within the *C*-terminal regions, rises up to values of 73%, indicating the presence of strongly conserved structures. Structure prediction studies indicate that both the A and B chain *C*-terminal regions may adopt a predominantly β-type structure with apparently little α-helical structure.

Human subcomponent C1q is composed of 18 polypeptide chains (six A, six B and six C) each of which consists of a short, non-collagen-like, *N*-terminal region of two to seven residues followed by a collagen-like sequence of approx. 80 residues and terminating in a, non-collagen-like, *C*-terminal region of approx. 130 residues (Reid & Porter, 1976; Reid, 1979). The amino acid sequences of the *N*-terminal 100–110 amino acid residues of the A, B and C chains have been published (Reid, 1979). The sequences of amino acid residues 109–224 of the A chain and 109–226 of the B chain are given in this paper, thus allowing comparison of the complete amino acid sequences of these chains.

When viewed in the electron microscope subcomponent C1q can be seen to be composed of six peripheral globular regions which are each joined by a collagen-like connecting strand to a fibril-like central portion (Shelton *et al.*, 1972; Knobel *et al.*, 1975; Brodsky-Doyle *et al.*, 1976). It is possible that these collagen-like structures (connecting-strands + central portion) in the molecule may be involved in: the binding and activation of subcomponents C1r and C1s (Reid *et al.*, 1977; Siegel & Schumaker, 1981); the binding of subcomponent C1q to receptors on lymphoid cells (Gabay *et al.*, 1979; Tenner & Cooper, 1980) or platelets (Wautier *et al.*, 1980); the binding to serum fibronectin (Menzel *et al.*, 1981). The six globular 'head' regions are considered to contain the binding sites in subcomponent C1q which interact with aggregated immunoglobulin G (Knobel *et al.*, 1974; Hughes-Jones & Gardner, 1979; Pâques *et al.*, 1979). The *C*-terminal 130 amino acid residues of one A, one B and one C chain form one globular 'head' region; therefore examination of the amino acid sequences of these regions will aid studies aimed at locating the precise sites in subcomponent C1q which are involved in the binding to immunoglobulin G.

Subcomponent C1q contains approx. 8.0% carbohydrate (Reid *et al.*, 1972; Calcott & Muller-Eberhard, 1972; Yonemasu *et al.*, 1971); 69% of this carbohydrate is present as glucosylgalactosyl disaccharide units linked to certain hydroxylysine residues in the collagen-like regions (Shinkai & Yonemasu, 1979; Reid, 1979). The remaining 31% of the carbohydrate is composed of six asparagine-linked sugar chains, which are located in the *C*-terminal globular regions of C1q (Mizuochi *et al.*, 1979). The structural studies given in this paper indicate that there is one asparagine-linked sugar chain located at position 124 in each A chain.

Abbreviation used: SDS, sodium dodecyl sulphate.

0306-3275/82/060559-11$01.50/1

Materials and Methods

Automated sequence determination, quantitative amino acid analysis, manual dansyl-Edman degradation, electrophoresis and chromatography, complete reduction and alkylation, succinylation

The procedures were performed as described previously (Reid, 1974, 1976; Reid & Thompson, 1978; Johnson *et al.*, 1980). The sources of all the enzymes and gel-filtration and ion-exchange media used in this study are also given in these references. For tryptophan estimation samples were hydrolysed in 3 M-mercaptoethane-sulphonic acid for 24, 48 and 72 h, at 110°C, under vacuum (Penke *et al.*, 1974). Tryptophan was estimated by extrapolation to zero time.

Purification of the reduced and alkylated A and B chains of subcomponent C1q

The reduced and alkylated A and B chains of subcomponent C1q were prepared by ion-exchange chromatography on DEAE-cellulose in a similar manner to that used for the preparation of the oxidized chains (Reid *et al.*, 1972; Lowe & Reid, 1974). Completely reduced and alkylated subcomponent C1q (30 mg in 3 mM-sodium phosphate/9 M urea buffer, pH 7.8) was fractionated on a column (1 cm × 80 cm) of DEAE-cellulose (DE-32) equilibrated with 3 mM-sodium phosphate/9 M-urea buffer, pH 7.8 and eluted with a linear gradient (chamber 1: 100 ml of 3 mM-sodium phosphate/9 M-urea, pH 7.8; chamber 2: 100 ml of 50 mM-NaCl/3 mM-sodium phosphate/9 M-urea, pH 7.8). The reduced and alkylated A chain was eluted at approx. 9–15 mM-NaCl and the B chain was eluted at approx. 20–30 mM-NaCl. The A chain was generally pure enough for structural studies (as judged by SDS/polyacrylamide-gel electrophoresis) but the B chain pool always required refractionation on the DEAE-cellulose, under the same conditions, to remove traces of the A and C chains. The purified chains were removed from the urea/phosphate buffer by dialysis against 0.4 M-acetic acid followed by freeze-drying.

Preparation of the C-terminal, non-collagen-like, regions (A-col-P and B-col-P) of the reduced and alkylated A and B chains

The *C*-terminal, non-collagen-like regions, of approximately 140 amino acid residues, of the reduced and alkylated A and B chains were prepared by collagenase digestion exactly as has been described previously for the oxidized A chain (Reid, 1974) and oxidized B chain (Reid & Thompson, 1978). The resulting A-col-P and B-col-P fractions, which had apparent M_r values on SDS/polyacrylamide-gel electrophoresis of 19600 and 16600 respectively, were succinylated with the procedure employed for the oxidized A chain (Reid, 1974).

Isolation of the trypsin-derived peptide from the A chain of subcomponent C1q

The succinylated A-col-P fraction was digested with trypsin under conditions previously described for the intact chain (Reid, 1974). Peptide T-2 was isolated, in approx. 70% yield, as a precipitate which formed during the trypic digest. Peptides T-Succ-4, T-1 and T-3 were initially fractionated from the supernatant of the tryptic digest by gel-filtration on a column (1 cm × 80 cm) of Sephadex G-75 (superfine grade), equilibrated with 0.1 M-NH_4HCO_3, on which they eluted at a V_e/V_0 of 1.07, 1.25 and 1.95 respectively. Peptide T-Succ-4 was further purified on a column (1 cm × 8.5 cm) of DEAE-Sephacel equilibrated with 10 mM-NH_4HCO_3, pH 8.0, and eluted with a linear gradient (chamber 1: 200 ml of 10 mM-NH_4HCO_3, pH 8.0; chamber 2: 200 ml of 750 mM-NH_4HCO_3). Peptide T-Succ-4 was obtained, in 50% yield, in fractions eluted with 400–500 mM-NH_4HCO_3. Peptide T-1 was further purified by gel filtration on Sephadex G-50 (superfine grade) and Sephadex G-25, equilibrated with 0.1 M-NH_3, where it eluted at V_e/V_0 of 1.16 and 1.00, respectively, and was obtained in 40% yield. Peptide T-3, which was obtained in 48% yield, was further purified by gel filtration on Sephadex G-25 in 0.1 M-NH_3, where it had a V_e/V_0 of 1.64 and by paper chromatography in butan-1-ol/acetic acid/water/pyridine (45:9:36:30, by vol.) where it had an R_F of 0.16, relative to Phenol Red.

Purification of the 'V-8 proteinase', thermolysin and chymotrypsin derived peptides used in the sequencing of residues 100–224 of the A chain of subcomponent C1q

Peptide V8-T1-2. Peptide T1 (100 nmol in 0.5 ml of 0.1 M-NH_4HCO_3 pH 8.0) was digested with 'V-8 proteinase' (25 μg), for 16 h at 37°C, and the digest fractionated on a column (1 cm × 80 cm) of Sephadex G-75 (Superfine grade) equilibrated with 0.1 M-NH_4HCO_3, pH 8.0. Peptide V8-T1-2 was eluted at a V_e/V_0 of 1.23 in 60% yield.

Peptides TH-3-6, TH-5-4 and TH-4-8. These peptides were isolated from a thermolysin (crystalline, 6000 PU/mg, Merck) digest of fragment A-col-P. Thermolysin (100 μg in 100 μl of 0.2 M-NH_4HCO_3, pH 7.9) was added to the reduced and alkylated A-col-P fragment (10 mg) which was suspended in 2.0 ml of 0.2 M-NH_4HCO_3, pH 7.9, and digestion was carried out for 3 h at 37°C. The digest was centrifuged at 20000 *g* for 20 min at 4°C and the peptides in the supernatant were separated on a column (2.5 cm × 80 cm) of Sephadex G-25 (Fine grade) that was equilibrated with 50 mM-NH_3, pH 10.5, and then further purified by paper chromatography in butan-1-ol/acetic acid/water/pyridine (45:9:36:30, by vol.). Peptides TH-3-6, TH-4-8 and TH-5-4 were eluted at V_e/V_0 of 1.53,

1.68 and 1.61 respectively and had R_F values, relative to Phenol Red, on paper chromatography of 0.21, 0.50 and 0.15 respectively.

Peptide CH-IV-3. This peptide was isolated from a chymotryptic digest of the oxidized A chain (Reid, 1974) by gel filtration and paper chromatography in the same manner as the thermolysin derived peptides. Peptide CH-IV-3 was eluted at a V_e/V_0 of 1.63 and had an R_F value, on paper chromatography, of 0.40.

Purification of the CNBr-derived peptides from the C-terminal regions of the A and B chains of subcomponent C1q

CNBr cleavage of the intact, reduced and alkylated A chain (10mg), or fragment B-col-P (7.5mg) was performed by dissolving the chain, or fragment, in 1.5ml of 70% (v/v) formic acid containing 28mg of CNBr and leaving in the dark, at room temperature, for 16h. The samples were then freeze-dried after the addition of distilled water (10ml).

Peptide CNBr-3 from the A chain. This peptide was isolated by gel filtration of the CNBr digest of the reduced and alkylated A chain, first on a column (1cm × 80cm) of Sephadex G-25 (Superfine grade) in 10% (v/v) formic acid, and then on a column (1cm × 80cm) of Sephadex G-50 (Superfine grade) in 5% (v/v) formic acid. Peptide CNBr-3 was eluted at a V_e/V_0 of 1.95 on Sephadex G-75, and at a V_e/V_0 of 1.58 on Sephadex G-50, in approx. 40% yield.

CNBr-derived peptides from the B chain. These peptides were initially fractionated by gel filtration of a CNBr digest of the reduced and alkylated B-col-P fraction on a column (1cm × 80cm) of Sephadex G-50 (Superfine grade) in 5% (v/v) formic acid. Peptides CNBr-1, CNBr-2, CNBr-3, CNBr-4, CNBr-5 and CNBr-6 were eluted at V_e/V_0 of 1.48, 1.75, 1.75, 1.81, 1.95 and 2.20 respectively. Peptides CNBr-5 and CNBr-6 were pure enough for structural studies but the other four required further purification by high pressure liquid chromatography. A μ-Bondapak C-18 column was used in conjunction with a Waters Associates pumping system (model 6000A) and a solvent system composed of 0.1% (v/v) trifluoroacetic acid (solution A) and methanol/propan-1-ol/CH_3CN (1:1:1, by vol.; solution B). A freeze-dried mixture of peptides CNBr-2, CNBr-3 and CNBr-4 (approx. 60nmol of each peptide) was dissolved in 400μl of solution A and loaded on to the μ-Bondapak C-18 column which was pumped at 1ml/min and equilibrated with a 19:1 (v/v) mixture of solution A and solution B. After the breakthrough peak had eluted, a 1h linear gradient was begun to give a final ratio of 9:11 (v/v) of solution A to solution B. The effluent was monitored at 206nm and for radioactivity associated with *S*-[^{14}C]carboxymethylcysteine. Peptides CNBr-2, CNBr-3 and CNBr-4 were eluted at 47–49ml, 49–50.5ml and 50.5–52.5ml of the gradient, respectively, which corresponded to concentrations of solvent B of 41–43%, 43–45% and 45–48% (v/v) respectively. Peptide CNBr-1 was purified in a similar manner and was eluted at 42.5–43.3ml, which corresponded to a concentration of solvent B of 36.8–37.5% (v/v).

Purification of the pepsin, chymotrypsin- and trypsin-derived peptides used in the sequencing of residues 81–226 of the B chain of subcomponent C1q

Pepsin-derived peptides. The B-col-P fragment (400nmol) was suspended in 0.6ml of 0.6M-acetic acid and pepsin (100μg in 100μl of 0.6M-acetic acid) was added. After 16h at 37°C the digest was fractionated on a column (1cm × 80cm) of Sephadex G-50 (Superfine grade) equilibrated with 5% (v/v) formic acid and then on a column (1cm × 80cm) of Sephadex G-25 (Superfine grade) equilibrated with 0.1M-NH_4HCO_3. Peptides P-6-II-k-4, P-6-II-i-3 and P-6-II-i-6 were eluted together at a V_e/V_0 of 1.95 on the Sephadex G-50 and approx. 1.42 on the Sephadex G-25. Further purification was performed by gradient elution off a column of DEAE-Sephacel by using the conditions described above. Peptides P-6-II-i-3 and P-6-II-i-6 were eluted with 10mM-NH_4HCO_3 while peptide P-6-II-k-4 was eluted with 80–90mM-NH_4HCO_3. Peptides P-6-II-i-3 and P-6-II-i-6 were further purified by high pressure liquid chromatography on a μ-Bondapak C-18 column using the conditions described above. Peptides P-6-II-i-3 and P-6-II-i-6 were eluted at 17–19ml and 31–33ml of the gradient respectively, which corresponded to concentrations of solvent B of 21–23% and 34–35.5% respectively.

Peptide C-5-II was isolated from a chymotryptic digest of the oxidized B chain (Reid & Thompson, 1978). It was purified by gel filtration on Sephadex G-25 (fine grade) in 50mM-NH_3 where it was eluted with a V_e/V_0 of 1.70 and by paper chromatography in butanol/acetic acid/pyridine/water (45:9:30:36, by vol.) where it had an R_F, relative to Phenol Red, of 0.31.

Trypsin-derived peptides. The succinylated, reduced and alkylated, fragment B-col-P (400nmol) was suspended in 1.95ml of 0.2M-NH_4HCO_3, pH8.0, and trypsin (100μg in 100μl of 1mM-HCl) was added. After 4h at 37°C 2μl of 2.5M-diisopropyl phosphorofluoridate was added, the digest was centrifuged at 20000*g* for 20min and the supernatant was applied to a column (1cm × 80cm) of Sephadex G-50 (Superfine grade) run in 0.1M-NH_4HCO_3. Peptides T-Succ-1-a and T-5-I were eluted at a V_e/V_0 of 1.05 and 1.86 respectively. Peptide T-Succ-1-a was further purified by gradient elution off a column of DEAE-Sephacel,

using the conditions described above, and it was eluted, in 38% yield, with 450–550mM-NH_4HCO_3. Peptide T-5-I was further purified, in 37% yield, by paper electrophoresis at pH 3.5.

Isolation of peptide containing asparagine-linked carbohydrate from the A chain of subcomponent C1q

Peptide V8-T1-2 was isolated by trypsin and 'V-8 proteinase' digestion as described above. Peptide V8-T1-2 (50nmol) in 500μl of 0.2M-Tris/HCl/10mM-$CaCl_2$ was digested with Pronase (30μg), for 16h at 37°C, and then the digest was applied to a column (1.5cm × 80cm) of Sephadex G-50 (Superfine grade) equilibrated with 0.2M-NH_4HCO_3. A peptide which accounted for all the glucosamine present in the A chain of C1q was eluted between 85–95ml. Hexosamine analysis was performed by the method of Allen & Neuberger (1975) in which the sample is heated in 3M-*p*-toluenesulphonic acid at 100°C, under vacuum, for 24h. Recoveries of glucosamine and galactosamine were 86% and 73% respectively.

Results

Amino acid sequence of residues 100–224 of the A chain of subcomponent C1q

The amino acid compositions, and details concerning automated sequencing of the large tryptic peptides, peptide CNBr-3 and peptide TH-3-6, are given in Table 1. The amino acid compositions of the small peptides shown in Fig. 1 are: peptide CH-IV-3, $Ser_{0.7}$ $Gly_{1.2}$ $Phe_{0.8}$ $Arg_{1.0}$; peptide T-3, $Glx_{1.0}$, $Gly_{1.1}$, $Val_{0.8}$, $Arg_{1.0}$; peptide TH-5-4, $Ser_{0.8}$, $Val_{0.9}$, $Arg_{2.0}$; peptide TH-4-8, $Ser_{0.9}$, $Ala_{1.0}$, $Phe_{1.2}$. In nearly every case the sequence postulated in Fig. 1 agrees closely with the amino acid composition. The exception is peptide T-2, since alanine (0.6 of a residue) and lysine (0.5 of a residue) were seen in the composition but were not accounted for in the sequence analysis. This difference was probably due

Table 1. *Amino acid compositions of the large peptides used in the determination of the amino acid sequence of residues 100–224 of the A-chain of subcomponent C1q*

Compositions are given as mol of residue/mol of peptide and values less than 0.1 are omitted. The compositions of the small (three to four residue) peptides shown in Fig. 1 are given in the text. Peptide isolation is described in the text and samples were hydrolysed in 5.7M-HCl at 110°C for 24h. No corrections were made for serine and threonine destruction and tryptophan was only determined by sequence analysis. Cysteine was determined as *S*-carboxymethylcysteine or cysteic acid. The values given in parentheses are those obtained by sequence analysis. N.E., not estimated.

	Composition (mol/mol) of peptide					
Amino acid	T-1	T-2	T-Succ-4	V8-T1-2	CNBr-3	TH-3-6
Asx	4.8 (5)	1.2 (1)	5.4 (5)	2.1 (2)	3.2 (3)	—
Thr	1.8 (2)	1.9 (2)	2.1 (2)	1.5 (2)	0.3 (0)	—
Ser	1.1 (1)	4.0 (4)	5.4 (6)	0.8 (1)	3.5 (4)	2.1 (2)
Glx	3.9 (4)	3.6 (4)	8.1 (8)	3.5 (4)	6.8 (7)	1.0 (1)
Pro	3.0 (3)	1.3 (1)	2.0 (2)	1.1 (1)	2.0 (2)	—
Gly	3.2 (3)	2.4 (1)	8.7 (9)	1.4 (1)	4.5 (5)	1.2 (1)
Ala	0.3 (0)	0.6 (0)	2.4 (2)	0.3 (0)	2.2 (2)	—
Cys	—	1.0 (1)	0.8 (1)	—	—	—
Val	2.5 (3)	3.9 (4)	5.5 (6)	1.0 (1)	3.9 (4)	0.9 (1)
Met	0.8 (1)	—	0.7 (1)	—	—	—
Ile	1.7 (2)	1.4 (1)	1.8 (2)	1.1 (1)	1.9 (2)	0.7 (1)
Leu	0.2 (0)	2.3 (2)	4.5 (5)	0.2 (0)	3.0 (3)	—
Tyr	1.3 (1)	2.5 (3)	0.6 (1)	0.6 (1)	1.0 (1)	—
Phe	1.0 (1)	2.6 (3)	4.5 (5)	0.1 (0)	3.2 (3)	—
His	1.1 (1)	0.2 (0)	0.8 (1)	0.8 (1)	1.0 (1)	—
Lys	—	0.5 (0)	3.5 (4)	0.2 (0)	3.0 (3)	—
Arg	1.8 (2)	1.0 (1)	0.1 (0)	0.9 (1)	0.2 (0)	1.0 (1)
Trp	N.E. (0)	N.E. (2)	N.E. (1)	N.E. (0)	N.E. (1)	N.E. (1)
Total residues	29	30	61	16	41	8
Amount used in sequencer (nmol)	55	110	72	50	62	32
Recovery at first step (nmol)	21	45	40	30	41	20
Repetitive yield (%)	89	95	95	95	95	93

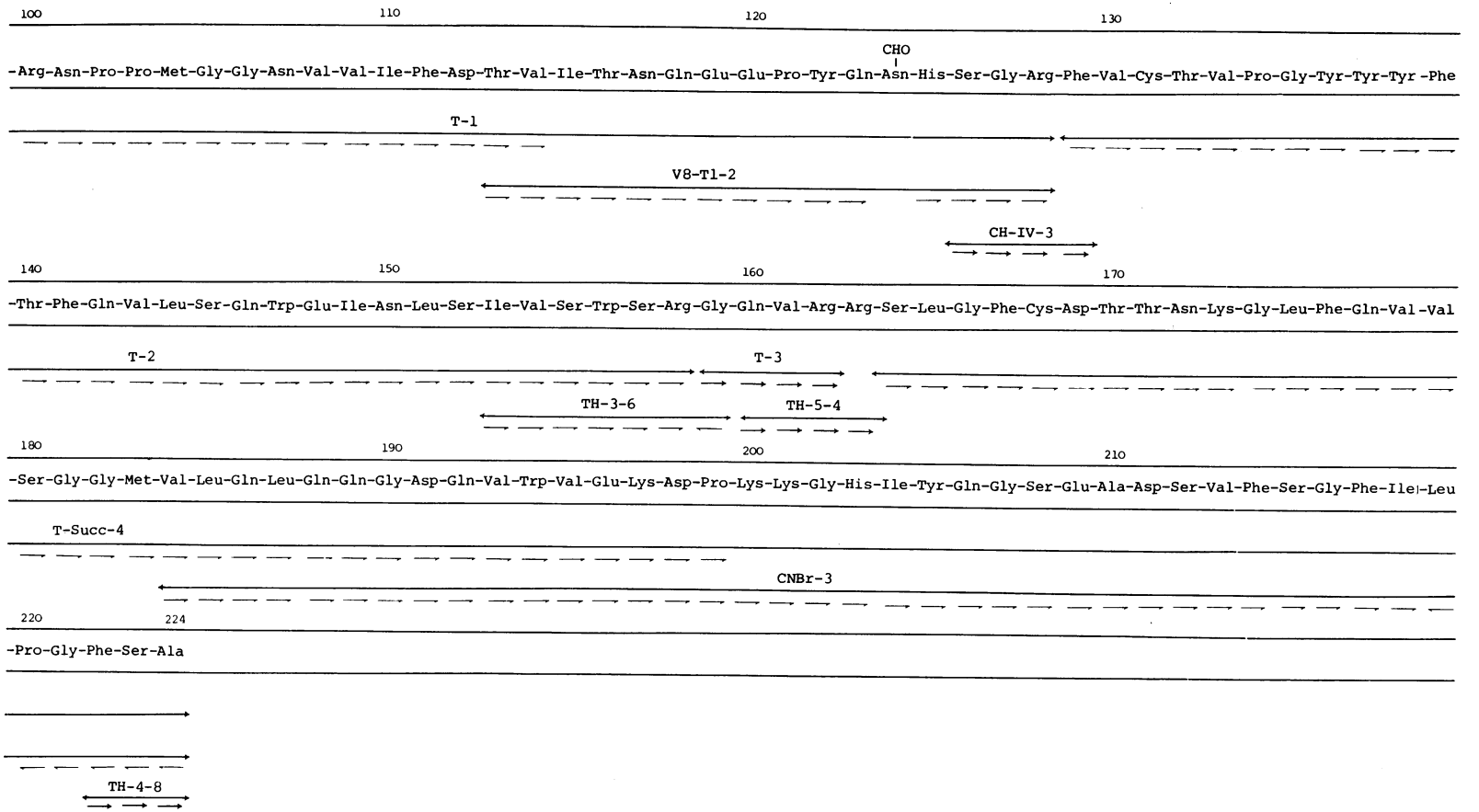

Fig. 1. *Amino acid sequence of residues 100–224 of the A-chain of subcomponent C1q*

Peptides prefixed by T-, TH-, V8-, CH- and CNBr- were derived from trypsin, thermolysin, V8-proteinase, chymotrypsin and CNBr digests respectively. The sequence of residues 100–108 has been published previously (Reid. 1979). Sequence determination from the *N*-terminus by use of the Beckman 890 C protein sequencer or by the dansyl procedure is shown by ⇀ and → respectively. Sequence determination from the *C*-terminus by use of carboxypeptidases A and Y is shown by ↽. CHO denotes the major, asparagine-linked, carbohydrate attachment site.

to the presence of a contaminating peptide, since peptide T-2 was unusually difficult to fractionate because of its very hydrophobic nature (Fig. 1) Also, the 'overlapping' sequence data indicates that peptide T-2 cannot be longer than is shown in Fig. 1. Only single-residue 'overlaps' were obtained at positions A-129 and A-164; however, the limited number of arginine residues makes any alternative structure very unlikely.

In Table 2 the amino acid composition of the entire A chain as determined by amino acid analysis is compared with the composition calculated from the sequence data given in this paper and in a previous publication (Reid, 1979). There is good agreement between the two sets of values, with only the aspartic acid and lysine contents differing by approx. 1 residue.

Table 2. *Amino acid compositions of the A and B chains of subcomponent C1q as determined by total amino acid analysis and by amino acid sequence analysis*

Samples of reduced and alkylated A or B chain were hydrolysed at 110°C for 24, 48 and 72h, under vacuum, in 5.7M-HCl. Half-cystine was estimated as *S*-carboxymethylcysteine and tryptophan was estimated as described in the text. The results are expressed as mol of amino acid/mol of polypeptide chain.

	Residues/mol in:			
	A chain		B chain	
Amino acid	Amino acid analysis	Amino acid sequence	Amino acid analysis	Amino acid sequence
Hyp	7.7	8	10.3	11
Asx*	15.8	17	20.7	21
Thr	8.0	8	14.5	15
Ser	14.5	15	9.4	9
Glx†	24.5	25	19.3	18
Pro	12.7	13	13.7	14
Gly	39.3	40	40.1	37
Ala	7.9	8	13.0	13
Cys	3.1	3	3.8	4
Val	14.7	15	11.2	10
Met	2.1	2	5.2	6
Ile	12.4	12	10.4	10
Leu	10.4	11	14.1	14
Tyr	5.5	6	6.6	7
Phe	9.8	10	10.7	11
His	2.1	2	3.1	3
Hyl	5.4	5	5.5	6
Lys	7.1	7	8.2	8
Arg	14.4	14	9.3	9
Trp	2.2	3	0.2	0
		224		226

* A chain, 9 Asp + 8 Asn; B chain, 10 Asp + 11 Asn.
† A chain, 9 Glu + 16 Gln; B chain, 11 Glu + 7 Gln.

Isolation of a peptide containing asparagine-linked carbohydrate from the A-chain of subcomponent C1q

Hexosamine analysis of peptides T-1, T-2 and T-Succ-4 showed that peptide T-1 contained approx. 3.6 residues of glucosamine/mol, while peptides T-2 and T-Succ-4 contained less than 0.1 of a residue of glucosamine/mol. All the glucosamine values were corrected assuming an 86% recovery. Peptide V8-T1-2 was isolated from peptide T-1 (as described above) and then was further digested with Pronase to give a peptide with the composition: $Asx_{1.2}$, $Ser_{1.0}$, $Glx_{1.2}$, $Gly_{1.0}$, $His_{0.7}$, $GlcN_{3.6}$. This peptide had an *N*-terminal amino acid sequence of Glx-Asx-His-, as judged by quantitative subtractive Edman degradation, and therefore must be derived from residues 123–127 of the A chain. The presence of glucosamine in this peptide accounts for the non-identification of the asparagine residue at position 124 in the A chain by automated sequencing (Fig. 1).

Amino acid sequence of residues 81–226 of the B chain of subcomponent C1q

The amino acid compositions, and details concerning automated sequencing, of the peptides derived by CNBr digestion of the B chain are given in Table 3. There is only one discrepancy between the compositions determined by amino acid analysis on comparison with those determined from the postulated sequence, and this is a peptide CNBr-2. By amino acid analysis peptide CNBr-2 appears to have two residues of both serine and glutamic acid; however, the automated sequence analysis indicates there are three residues of serine and one of glutamic acid (Fig. 2). The serine residues at positions 130, 148 and 149 were readily identified in the automated sequence run and there was no evidence for glutamine or glutamic acid at these positions. However partial, manual, sequence data, from a chymotryptic peptide isolated from the B-col-P fragment, yielded a sequence of X-Ala-Ser-Glx-Arg-, indicating that it is possible that there may be heterogeneity at position 149 with both serine and glutamine (or glutamic acid) being present.

Table 4 shows the compositions of the peptides used to establish the sequential order of the CNBr peptides of B-col-P. In each case the composition determined by amino acid analysis agrees with the composition calculated from the postulated amino acid sequence (Fig. 2). It was also found that there is also good agreement between the amino acid composition of the entire B chain, as determined by amino acid analysis, compared with the composition taken from the sequence data given in this paper and in a previous publication (Reid, 1979) (Table 2).

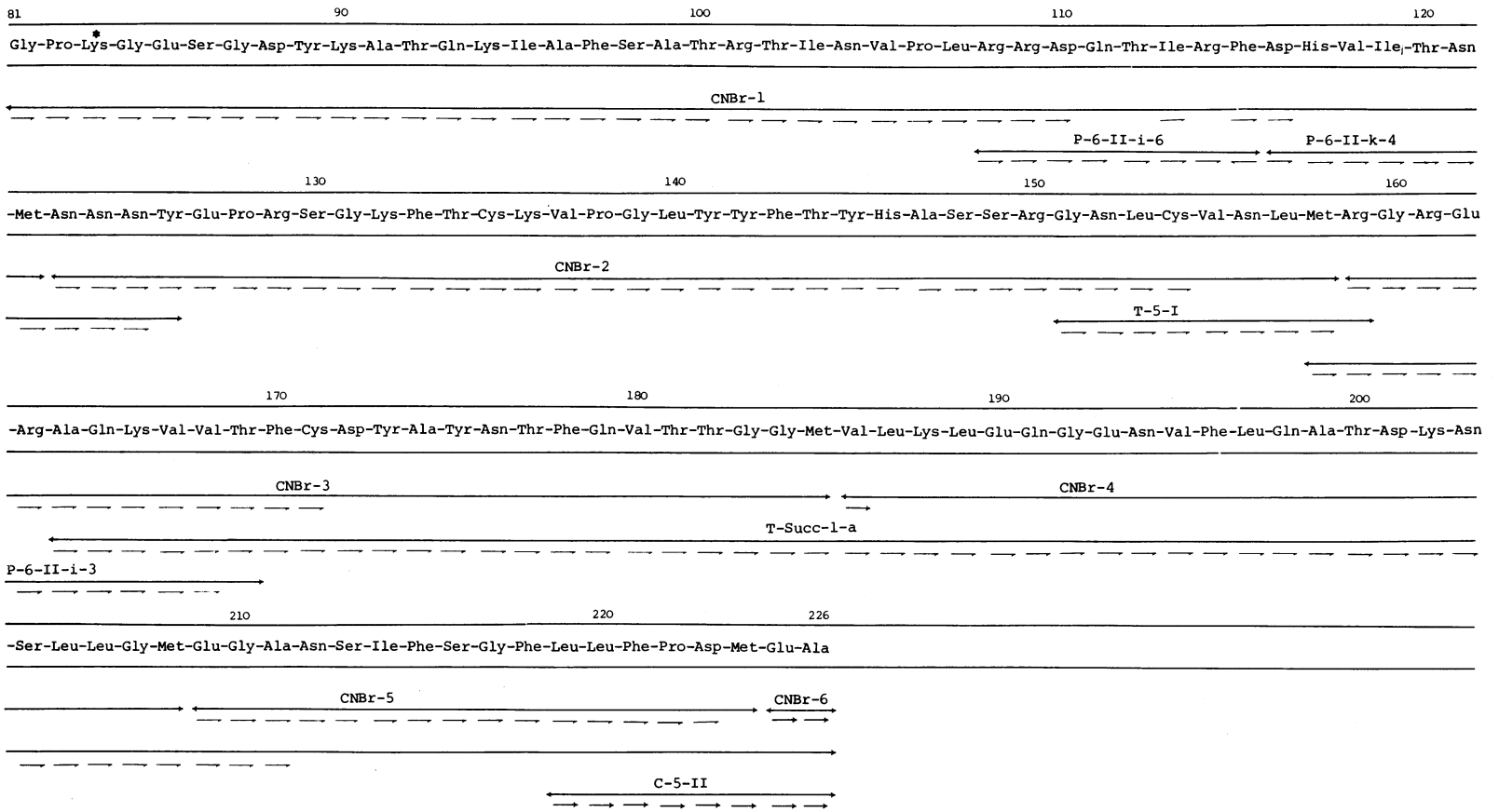

Fig. 2. *Amino acid sequence of residues 81–226 of the B-chain of subcomponent C1q*

Peptides prefixed by CNBr-, P-, T- and C- were derived from CNBr, pepsin, trypsin and chymotrypsin digests respectively. The sequence of residues 81–108 have been published previously (Reid & Thompson, 1978). Sequence determination from the *N*-terminus by use of the Beckman 890 C protein sequencer or by the dansyl procedure is shown by → and → respectively. * Denotes a hydroxylated residue.

Table 3. *Amino acid compositions of the CNBr peptides used in the determination of the sequence of residues 81–226 of the B-chain of subcomponent C1q*

Compositions are given as mol of residue/mol of peptide and values less than 0.1 are omitted. The isolation of the peptides is described in the text; samples were hydrolysed in 5.7M-HCl at 110°C for 24h. No corrections were made for serine and threonine destruction and tryptophan was not determined. Cysteine was determined as *S*-carboxymethylcysteine. The values given in parentheses are those obtained by amino acid sequence analysis.

	Composition (mol/mol) of the CNBr-peptides:					
Amino acid	CNBr-1	CNBr-2	CNBr-3	CNBr-4	CNBr-5	CNBr-6
Asx	5.0 (5)	4.6 (5)	2.3 (2)	2.7 (3)	2.1 (2)	—
Thr	4.5 (5)	1.7 (2)	3.5 (4)	0.8 (1)	—	—
Ser	2.0 (2)	2.1 (3)	0.4 (0)	0.9 (1)	1.7 (2)	—
Hse	0.9 (1)	1.0 (1)	1.0 (1)	0.9 (1)	1.0 (1)	—
Glx	3.0 (3)	2.1 (1)	3.2 (3)	3.5 (4)	1.0 (1)	1.0 (1)
Pro	1.8 (2)	1.5 (2)	—	—	0.6 (1)	—
Gly	3.4 (3)	3.3 (3)	3.3 (3)	2.0 (2)	2.1 (2)	—
Ala	3.1 (3)	1.2 (1)	2.1 (2)	1.0 (1)	0.9 (1)	1.0 (1)
Cys	—	1.5 (2)	1.0 (1)	—	—	—
Val	2.2 (2)	2.4 (2)	3.0 (3)	2.0 (2)	—	—
Ile	3.7 (4)	—	0.2 (0)	—	1.0 (1)	—
Leu	1.1 (1)	3.4 (3)	0.3 (0)	4.5 (5)	2.0 (2)	—
Tyr	1.0 (1)	3.5 (4)	2.0 (2)	0.4 (0)	—	—
Phe	2.0 (2)	2.2 (2)	2.1 (2)	1.4 (1)	3.1 (3)	—
His	1.1 (1)	0.8 (1)	0.1 (0)	—	—	—
Hyl	1.0 (1)	—	—	—	—	—
Lys	2.0 (2)	2.2 (2)	1.1 (1)	1.9 (2)	—	—
Arg	3.9 (4)	1.8 (2)	2.8 (3)	—	—	—
Total residues	42	36	27	23	16	2
Amount used in sequencer (nmol)	240	33	22	—	70	—
Recovery at first step (nmol)	89	19	15	—	40	—
Repetitive yield (%)	94	92	94	—	93	—

Discussion

The completion of the amino acid sequence of the A and B chains of subcomponent C1q confirms, and emphasizes, the expected marked difference in primary structure between the, collagen-like, *N*-terminal half and the non-collagen-like, *C*-terminal half of each chain (Fig. 3). Using two methods for predicting secondary structure (Chou & Fasman, 1974: Lim, 1974), it was estimated that the *C*-terminal regions (residues 88–226), of both A and B chains, may be composed of predominantly β-sheet-type structures involving approx. 35% of the A chain residues and 34% of the B chain residues (Fig. 3). There was a relatively low amount of α-helical structure predicted, since it was estimated that only approx. 5% of the A chain, and approx. 10% of the B chain, *C*-terminal 140 residues would be involved in that type of structure.

There is 40% homology between the *C*-terminal amino acid sequences of the *C*-terminal regions of A and B chains when alignment of the chains is made starting at the lysine residue immediately after the last collagen-like triplet in each chain (residues A-88 and B-90) and with the insertion of 'gaps' to maximize the homology (Fig. 3). Over certain sections in the chains, such as B-130–B144, and B-170–B-192, the homology between the A and B chains rises to 73% and 58% respectively. These sections include some of the longer stretches of predicted β-structure found in the chains (Fig. 3), and their high degree of conservation implies that they may be of importance in maintaining the critical structure features in subcomponent C1q which are involved in the binding of aggregated immunoglobulin G.

There are two potential sites in the A chain of subcomponent C1q to which carbohydrate could become linked via an asparagine residue, i.e. the two sites, of the form Asn-Xaa-Ser/Thr-, which are located at A-124 and A-150. No asparagine was observed, during automated sequencing, at position A-124 and an unexpectedly low yield of asparagine was obtained at position A-150. However, hexos-

Table 4. *Amino acid compositions of the peptides used to establish the sequential order of the CNBr peptides derived from residues 81–226 of the B chain of subcomponent C1q*

Compositions are given as mol of residue/mol of peptide and values less than 0.1 are omitted. The isolation of the peptides is described in the text; samples were hydrolysed at 110°C for 24h. No corrections were made for serine and threonine destruction and tryptophan was not measured. Cysteine was determined as cysteic acid and methionine as methionine sulphone. The values given in parentheses are those obtained by sequence analysis.

	Composition (mol/mol) of peptide:					
Amino acid	P-6-II-i-3	P-6-II-i-6	P-6-II-k-4	T-5-I	C-5-II	T-Succ-1-a
Asx	—	1.0 (1)	4.5 (5)	1.9 (2)	1.0 (1)	7.4 (7)
Thr	1.0 (1)	0.9 (1)	0.8 (1)	1.1 (1)	—	4.5 (5)
Ser	—	—	—	—	—	3.2 (3)
Glx	1.9 (2)	1.1 (1)	—	—	1.1 (1)	8.0 (8)
Pro	—	—	—	—	1.2 (1)	1.2 (1)
Gly	1.2 (1)	—	—	1.3 (1)	—	6.8 (6)
Ala	1.1 (1)	—	—	—	—	4.5 (5)
Cys	—	—	—	1.0 (1)	—	0.7 (1)
Val	2.1 (2)	—	0.8 (1)	1.1 (1)	—	5.4 (5)
Met	1.0 (1)	—	0.9 (1)	—	1.1 (1)	2.7 (3)
Ile	—	0.9 (1)	0.9 (1)	—	—	1.3 (1)
Leu	—	—	—	1.9 (2)	1.9 (2)	7.1 (7)
Tyr	—	0.8 (1)	0.8 (1)	—	—	1.5 (2)
Phe	—	—	—	—	1.1 (1)	5.5 (6)
His	—	—	1.0 (1)	—	—	0.4 (0)
Lys	0.9 (1)	—	—	—	—	3.4 (3)
Arg	2.9 (3)	3.0 (3)	—	1.0 (1)	—	0.2 (0)
Total residues	12	8	11	9	7	63
Amount used in sequencer (nmol)	48	25	45	—	—	80
Recovery at first step (nmol)	10	8	10	—	—	53
Repetitive yield (%)	90	89	90	—	—	95

amine analysis indicated clearly that the asparagine residue at position A-124 is the major site of glycosylation, since 3.6mol of glucosamine were associated with one mol of this residue and less than 0.1mol of hexosamine was estimated to be associated with one mol of residue A-150. The presence of 3.6mol of glucosamine at position A-124 on each A chain would account for at least 90% of the glucosamine estimated to be present in the globular head regions of subcomponent C1q (Mizuochi *et al.*, 1979). The asparagine residue at position A-124 is probably situated at the beginning of a β-turn, as judged by structure-prediction studies, which would be consistent with the findings on other proteins which show that asparagine-linked carbohydrate is most commonly found at β-turns (Beeley, 1977; Aubert *et al.*, 1976). On the other hand, suitable asparagine residues, such as the asparagine at position A-150 (Fig. 1), which are surrounded by hydrophobic residues, are considered less likely to be glycosylated (Bause *et al.*, 1979). The location of the major asparagine-linked carbohydrate site to the A chain, rather than the B or C chains, of subcomponent C1q is consistent with the finding by Tenner *et al.* (1981) that tritiation of sialic acid residues with NaB^3H_4 after mild periodate treatment, selectively labelled the globular head region of the A chain of subcomponent C1q.

Although the complete amino acid sequences of the A and B chains are now known, the positions of the intrachain disulphide bonds, and free sulphydryl groups, present in the *C*-terminal globular regions of the chains of subcomponent C1q have not been finally elucidated. However, the presence of an *S*-carboxymethylcysteine residue at position B-154 which has no counterpart in the A chain (Fig. 3) is suggestive that this residue may be present as a free thiol group in the native protein and that an intra-chain disulphide bond will be present between residues A-131 and A-168 and possibly between residues B-135 and B-171.

No striking similarity could be found between the amino acid sequences of the *C*-terminal regions of the A and B chains of subcomponent C1q and the published sequences of a variety of other proteins, such as complement proteins, those of the blood

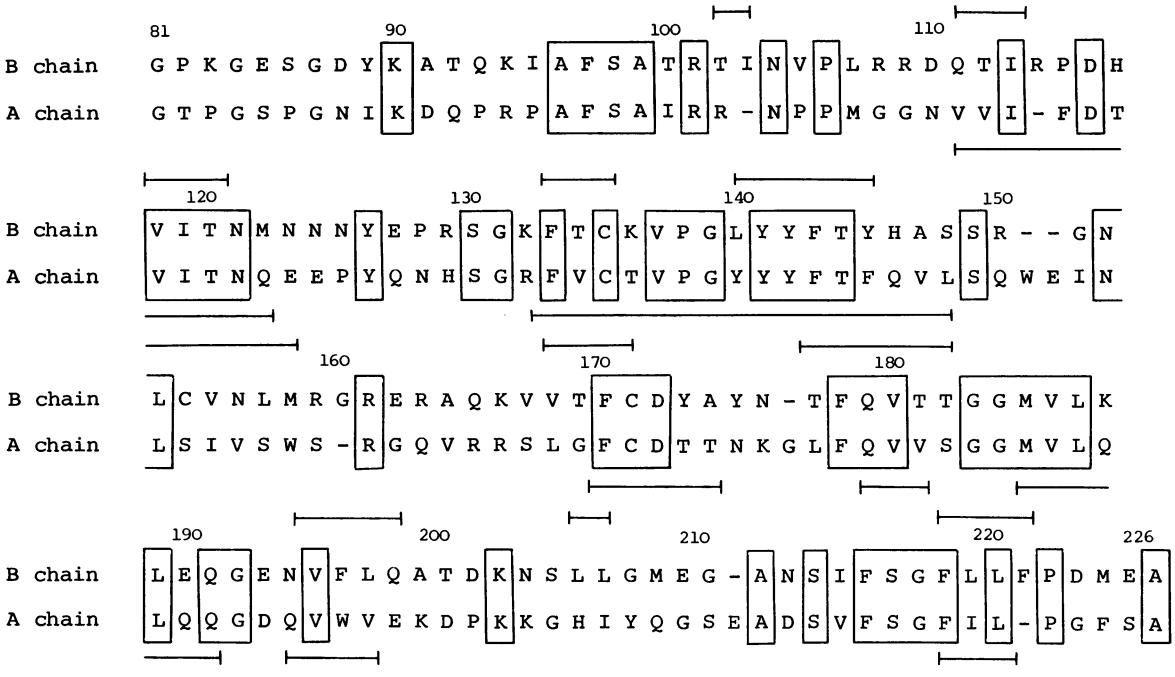

Fig. 3. *Alignment of the amino acid sequences of the non-collagen-like regions of the A and B chains of subcomponent C1q*
The alignment, starting at the lysine residues at positions A-88 and B-90, shows that there is approx. 40% homology between the chains over their *C*-terminal regions, as indicated by the boxed residues. The numbering is based on the B chain sequence; – denotes that a gap was left to maximize homology. Areas of the chains which are predicted, by the method of Chou & Fasman (1974), to be composed of β structure are indicated by ——, above the B chain sequence or below the A chain sequence. The sequence data are taken from this paper and Reid (1979). The single-letter code for amino acids is used: A, Ala; 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; Y, Tyr.

clotting system, immunoglobulins, and the non-collagenous regions of procollagen chains. However, one short region of α_2-macroglobulin has some homology with portions of both the A and B chains of subcomponent C1q. Fibrin, α_2-macroglobulin and cold-insoluble globulin all appear to be major substrates for plasma transglutaminase as shown by the incorporation of labelled putrescine, or cadaverine, into these proteins when plasma is clotted in the presence of the labelled diamines (Mosher, 1976; Mortensen *et al.*, 1981). The putrescine appears to be specifically incorporated into one glutamine residue in the case of α_2-macroglobulin, as judged by amino acid sequence studies on the labelled protein (Mortensen *et al.*, 1981). The sequence around this particular glutamine residue shows five identities over nine residues, when compared with the A and B chain sequences starting at the methionines at positions A-183 and B-185, (Fig. 3). The relevant amino acid sequence from α_2-macroglobulin is: -Met-Cys-Pro-Gln-Leu-Gln-Gln*-Tyr-Glu-, where the asterisk denotes the major site of incorporation of labelled putrescine. It can be seen that there is an equivalent glutamine residue at this position in both the A and B chain sequences when this alignment is made. The functional significance of the postulated crosslinking site in α_2-macroglobulin is not clear, but it has been suggested that cellular transaminases are involved in receptor-mediated endocytosis of α_2-macroglobulin–proteinase complexes (Davies *et al.*, 1980). It is not known if putrescine will bind covalently to subcomponent C1q under the same conditions but it has been reported that putrescine is a strong inhibitor of the interaction between subcomponent C1q and immunoglobulin (Sledge & Bing, 1973).

References

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

Aubert, J. P., Biserte, G. & Loucheux-Lefevre, M. H. (1976) *Arch. Biochem. Biophys.* **175**, 410–418

Bause, E. & Lehle, L. (1979) *Eur. J. Biochem.* **101**, 531–540
Beeley, J. G. (1977) *Biochem. Biophys. Res. Commun.* **76**, 1051–1055.
Brodsky-Doyle, B., Leonard, K. R. & Reid, K. B. M. (1976) *Biochem. J.* **159**, 279–286
Calcott, M. A. & Muller-Eberhard, H. J. (1972) *Biochemistry* **11**, 3443–3450
Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
Davies, P. J. A., Davies, D. R., Levitzki, A., Maxfield, F. R., Milhaud, P., Willingham, M. C. & Pastan, I. H. (1980) *Nature (London)* **283**, 162–167
Gabay, Y., Perlmann, H., Perlmann, P. & Sobel, A. T. (1979) *Eur. J. Immunol.* **9**, 797–801
Hughes-Jones, N. C. & Gardner, B. (1979) *Mol. Immunol.* **16**, 697–701
Johnson, D. M. A., Gagnon, J. & Reid, K. B. M. (1980) *Biochem. J.* **187**, 863–874
Knobel, H. R., Heusser, C., Rodrick, M. L. & Isliker, H. (1974) *J. Immunol.* **112**, 2094–2101
Knobel, H. R., Villiger, W. & Isliker, H. (1975) *Eur. J. Immunol.* **5**, 78–82
Lim, V. I. (1974) *J. Mol. Biol.* **88**, 873–894
Lowe, D. M. & Reid, K. B. M. (1974) *Biochem. J.* **143**, 265–272
Menzel, E. J., Smolen, J. S., Liotta, L. & Reid, K. B. M. (1981) *FEBS Lett.* **129**, 188–192
Mizuochi, T., Yonemasu, K., Yamashita, K. & Kobata, A. (1979) *J. Biol. Chem.* **253**, 7404–7409
Mortensen, S. B., Sottrup-Jensen, L., Hansen, H. F., Rider, D., Petersen, T. E. & Magnusson, S. (1981) *FEBS Lett.* **129**, 314–317
Mosher, D. F. (1976) *J. Biol. Chem.* **251**, 1639–1645
Pâques, E. P., Huber, R. & Preiss, H. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 177–183
Penke, B., Ferenczi, R. & Kovaks, K. (1974) *Anal. Biochem.* **60**, 45–50
Reid, K. B. M. (1974) *Biochem. J.* **141**, 189–203
Reid, K. B. M. (1976) *Biochem. J.* **155**, 5–17
Reid, K. B. M. (1979) *Biochem. J.* **179**, 367–371
Reid, K. B. M. & Porter, R. R. (1976) *Biochem. J.* **155**, 19–23
Reid, K. B. M. & Thompson, E. O. P. (1978) *Biochem. J.* **173**, 863–868
Reid, K. B. M., Lowe, D. M. & Porter, R. R. (1972) *Biochem. J.* **130**, 749–763
Reid, K. B. M., Sim, R. B. & Faiers, A. P. (1977) *Biochem. J.* **161**, 239–245
Shelton, E., Yonemasu, K. & Stroud, R. M. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 65–68
Shinkai, H. & Yonemasu, K. (1979) *Biochem. J.* **177**, 847–852
Siegel, R. C. & Schumaker, V. N. (1981) *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 24
Sledge, C. R. & Bing, D. H. (1973) *J. Biol. Chem.* **248**, 2818–2823
Tenner, A. J. & Cooper, N. R. (1980) *J. Immunol.* **125**, 1658–1664
Tenner, A. J., Lesavre, P. H. & Cooper, N. R. (1981) *J. Immunol.* **127**, 648–653
Wautier, J. L., Reid, K. B. M., Legrand, Y. & Caen, J. P. (1980) *Mol. Immunol.* **17**, 1399–1405
Yonemasu, K., Stroud, R. M., Niedermeier, W. & Butler, W. T. (1971) *Biochem. Biophys. Res. Commun.* **43**, 1388–1394