Isolation, by Partial Pepsin Digestion, of the Three Collagen-Like Regions Present in Subcomponent C1q of the First Component of Human Complement

By KENNETH B. M. REID

*Medical Research Council Immunochemistry Unit, Department of Biochemistry,
University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.*

(Received 28 August 1975)

1. Digestion of human subcomponent C1q with pepsin at pH4.45 for 20h at 37°C fragmented most of the non-collagen-like amino acid sequences in the molecule to small peptides, whereas the entire regions of collagen-like sequence that comprised 38% by weight of the subcomponent C1q were left intact. 2. The collagen-like fraction of the digest was eluted in the void volume of a Sephadex G-200 column, and was shown to be composed of two major fragments when examined by electrophoresis on polyacrylamide gels run in buffers containing sodium dodecyl sulphate. These fragments were separated on CM-cellulose at pH4.9 in buffers containing 7.5M-urea. 3. Human subcomponent C1q on reduction and alkylation yields equimolar amounts of three chains, which have been designated A, B and C [Reid *et al.* (1972) *Biochem. J.* 130, 749-763]. One of the pepsin fragments was shown to be composed of the *N*-terminal 95 residues of the A chain linked, via residue A4, by a single disulphide bond to a residue in the sequence B2-B6 in the *N*-terminal 91 residues of the B chain. The second pepsin fragment was shown to be composed of a disulphide-linked dimer of the *N*-terminal 94 residues of the C chain, the only disulphide bond being located at residue C4. 4. The mol.wts. of the unoxidized and oxidized pepsin fragments were estimated from their amino acid compositions to be 20000 and 18200 for the A-B and C-C dimers and 11400, 8800 and 9600 for the collagen-like fragments of the A, B and C chains respectively. Estimation of the molecular weights of the peptic fragments by polyacrylamide-gel electrophoresis run in the presence of sodium dodecyl sulphate gave values that were approx. 50% higher than expected from the amino acid sequence data. This is probably due to the high collagen-like sequence content of these fragments.

Human subcomponent C1q is an unusual serum protein since approx. 40% of its amino acid sequence is composed of regions of the repeating triplet Gly-X-Y, where X is often proline and Y is often hydroxyproline or hydroxylysine (Reid *et al.* 1972; Reid, 1974). These regions of collagen-like amino sequence are readily digested by collagenase in the native subcomponent C1q molecule (Reid *et al.*, 1972; Lowe & Reid, 1974), and this suggests that they may contain triple-helix type structure, as is found in collagens (Traub & Piez, 1971). It is known that amino acid sequences in a collagen-type triple helix are resistant to pepsin digestion at temperatures in the range 4-15°C when dissolved in dilute acid (such as 500mM-acetic acid), whereas most of the non-triple-helical portions of collagen (Miller, 1972) or procollagen (Bellamy & Bornstein, 1971) are readily fragmented to small peptides under these conditions. Limited proteolysis of human subcomponent C1q by pepsin at 37°C and at pH4.45 yielded a fraction that contained the entire regions of collagen-like sequence present in subcomponent C1q. This fraction, in dissociating conditions, was shown

to be composed of two major fragments. Chemical studies showed that three distinct regions of collagen-like sequence were present in the two fragments, each sequence corresponding to the *N*-terminal 91-95 residues of one of the three chains of human subcomponent C1q. Two of these *N*-terminal collagen-like sequences were found to be linked to each other by a single disulphide bond, and the third *N*-terminal collagen-like sequence was shown to be present as a disulphide-linked dimer. The isolation and characterization of these distinctive fractions of subcomponent C1q are described in the present paper.

Materials

Carboxypeptidase A, carboxypeptidase B and collagenase (CLSPA, chromatographically purified) were obtained from Worthington Biochemical Corp., Freehold, NJ, U.S.A. Pepsin (twice crystallized) was obtained from Sigma (London) Chemical Co., London S.W.6, U.K.

Sephadex G-25 (fine grade) and Sephadex G-200

were obtained from Pharmacia Fine Chemicals, Uppsala, Sweden. CM-cellulose 32 was obtained from W. and R. Balston, Maidstone, Kent, U.K.

Sequencer chemicals were obtained from Beckman Instruments, Palo Alto, CA, U.S.A. Iodo[$^{14}\text{C}_2$]acetic acid (33mCi/mmol) was obtained from The Radiochemical Centre, Amersham, Bucks., U.K., and was diluted to 125 $\mu\text{Ci/mmol}$ before use with a solution of carrier iodoacetic acid that had been previously recrystallized from light petroleum (60–80°C). Urea (AnalaR grade) was obtained from BDH Chemicals, Poole, Dorset, U.K. Urea-containing buffers were made up from a 9M-urea solution that had been deionized on a column of Amberlite monobed resin type MB-3.

Methods

Partial pepsin digestion of human subcomponent C1q and fractionation of the digest

Human subcomponent C1q was isolated as described by Reid (1974), concentrated by ultrafiltration and dialysed into 100mM-sodium acetate buffer, pH4.45 (at 20°C). Subcomponent C1q (87mg, i.e. 19ml of a 4.6mg/ml solution) was incubated with pepsin (3.0mg) at 37°C for 20h. Immediately the pepsin was added the digest became turbid, but the turbidity cleared rapidly and gradually (3–20h) a fine precipitate settled out. A time-course study of the rate of progress of the digest, performed by running samples of the digest on 7% polyacrylamide gels in buffers containing sodium dodecyl sulphate, showed that after 6h approx. 70% of the subcomponent C1q had been fragmented and that 20h was required for near-complete digestion (i.e. when approx. 95% of the subcomponent C1q had been fragmented). After 20h the digest was spun at 30000g for 20min at 4°C and the precipitate washed with digestion buffer (5.0ml) and re-spun, as before. The two supernatants were pooled, adjusted to pH 8.0 with 4M-NaOH, kept at room temperature for 30min applied to a column (6cm $\times$ 85cm) of Sephadex G-200 that was equilibrated with 100mM-acetic acid and run at 4°C. Two major peaks were obtained: P1, which was eluted near the void volume of the column, and P2 (Fig. 1). The peak-P1 material was pooled and freeze-dried. The freeze-dried material (32mg), dissolved in 36mM-sodium acetate/7.5M-urea buffer, pH4.9 (3.0ml), was applied to a column (1.1cm $\times$ 14cm) of CM-cellulose equilibrated with the same buffer. The column was run at room temperature (22°C) at a flow rate of 26ml/h, and fractions (2.6ml) were collected. After 15 fractions had been collected, a linear gradient composed of 100ml of 36mM-sodium acetate/7.5M-urea buffer, pH4.9, and 100ml of 36mM-sodium acetate/7.5M-urea/100mM-NaCl buffer, pH4.9, was begun. The two major fractions

eluted with the gradient, P1-CMC1 and P1-CMC2 (Fig. 2a), were pooled, concentrated by ultrafiltration and desalted on a Sephadex G-25 column equilibrated with 100mM-acetic acid, and freeze-dried. The fractions P1-CMC1 (7.6mg) and P1-CMC2 (12.8mg) were both re-run on CM-cellulose under exactly the same conditions described above, and a single symmetrical peak was obtained in both runs (Fig. 2b shows the elution profile of the fraction P1-CMC2 re-run). Fractions P1-CMC1 and P1-CMC2 were oxidized with performic acid by the procedure of Hirs (1956), and then run on CM-cellulose at pH4.9 in sodium acetate/urea buffers under the conditions given above. Fraction P1-CMC1 yielded one polypeptide chain on oxidation and running on CM-cellulose. This chain, P1-CMC1-Ox1 (216nmol), was eluted immediately at the beginning of the gradient. Fraction P1-CMC2 yielded two major polypeptide chains on oxidation: P1-CMC2-Ox1 (229nmol), which was eluted with the starting buffer (Fig. 2c), and P1-CMC2-Ox2 (204nmol), which was eluted just over halfway through the gradient (Fig. 2c). The oxidized polypeptide-chain fractions were concentrated by ultrafiltration, desalted on a Sephadex G-25 column equilibrated with 100mM-acetic acid and freeze-dried.

Collagenase digestion of the oxidized B chain of subcomponent C1q and fractionation of the digest

Human subcomponent C1q was isolated as described by Reid (1974) and oxidized by the method of Hirs (1956). The isolation of the oxidized B chain of human subcomponent C1q was performed as described by Reid *et al.* (1972). Chromatographically purified collagenase was further purified on Sephadex G-200 (Peterkofsky & Diegelmann, 1971). This purification allowed the specificity of the collagenase used to be restricted to cleavage of susceptible Y-Gly bonds in the repeating sequence (Gly-X-Y) $_n$. The purified collagenase (400 μg in 750 μl of 50mM-Tris/HCl buffer, pH7.4, that was 5mM with respect to CaCl_2 and 0.25mM with respect to *N*-ethylmaleimide) was added to the oxidized B chain (10mg) and the digestion carried out at 37°C for 20h. The digest was spun at 30000g for 20min at 4°C. The precipitate was washed twice with one-half the original volume of digestion buffer and spun down as before. The three supernatants were pooled and the precipitate was suspended in 50mM-NH $_3$ (3.0ml).

Automated sequence determination of proteins and peptides

N-Terminal amino acid sequences were determined by automated Edman degradation in a Beckman 890C sequencer. Two programmes were used, a

DMAA (1.0M-dimethylallylamine trifluoroacetate, pH9.0)-fast peptide programme (described by Beckman Instruments for use in the 890c sequencer) and a 0.1M-Quadrol (0.1M-*N,N,N',N'*-tetrakis-(2-hydroxypropyl)ethylenediamine trifluoroacetate, pH9.5) programme, which employs a single-cleavage step, described by Brauer *et al.* (1975) for the micro-scale sequencing of proteins. The thiazolines released and collected in the sequencer were converted into the amino acid phenylthiohydantoin derivatives by heating at 80°C under N₂ for 10min in 1M-HCl (200μl) containing 0.1% (v/v) ethanethiol. After two extractions with ethyl acetate (700μl) 4–8% of the extracted amino acid phenylthiohydantoin derivatives was examined by g.l.c. Silylation of samples for g.l.c. was performed by mixing the sample (10μl) with *N,O*-bis(trimethylsilyl)acetamide (10μl) and heating at 50°C for 10min. The g.l.c. was performed on a 5830A Hewlett–Packard reporting gas chromatograph with a processor equipped with an integrator that measured the phenylthiohydantoin peak areas, which were then compared with the areas of known standards. Quantitative determination by g.l.c. was made only on the alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine and tyrosine peaks. The amino acid phenylthiohydantoin derivatives were further identified by two-dimensional t.l.c. (Summers *et al.*, 1973) on polyamide sheets and with the incorporation of fluorescent indicator in the first solvent, which allowed detection of 0.5nmol of amino acid phenylthiohydantoin derivative with ease. Some samples of amino acid phenylthiohydantoin derivatives were further identified by hydrolysis in HI in an autoclave at 145kPa (21lb/in²) pressure for 20h, as described by Smithies *et al.* (1971), which yielded amino acids that were identified by amino acid analysis.

Reduction and alkylation of fraction P1-CMC1

Fraction P1-CMC1 (50nmol) was dissolved in 200mM-Tris/HCl/8M-urea/2mM-EDTA, pH8.0 (1.0ml), and reduced by incubation at 37°C for 1h in the presence of 20mM-dithiothreitol. The solution was then made 40mM with respect to iodo[¹⁴C₂]-acetic acid plus carrier iodoacetic acid (to give a final specific radioactivity of 125μCi/mmol) and incubated at 37°C for a further 20min. The reduced and alkylated fraction P1-CMC1 was then separated from urea and salts by gel filtration on a column of Sephadex G-25 equilibrated with 50mM-NH₃.

Quantitative amino acid analysis, manual dansyl-Edman degradation, electrophoresis and chromatography and digestion with carboxypeptidases A and B

These were performed as described previously (Reid, 1974).

Polyacrylamide-gel electrophoresis in the presence of sodium dodecyl sulphate

Polyacrylamide-gel electrophoresis in the presence of sodium dodecyl sulphate was performed essentially as described by Fairbanks *et al.* (1971).

Results

Fractionation of the partial pepsin digest of subcomponent C1q by centrifugation and gel-filtration on Sephadex G-200

The amino acid composition of the precipitate obtained after partial digestion of subcomponent C1q with pepsin is shown in Table 1. The precipitate, which was approx. 9% of the total protein in the digest, had a non-collagen-like amino acid composition, since it contained no hydroxyproline, little hydroxylysine and had relatively low glycine content on comparison with the intact subcomponent C1q (Table 1). The precipitate was not fractionated further, but electrophoresis, performed in 7% polyacrylamide gels run in buffers containing sodium dodecyl sulphate, indicated that the apparent mol.wt. of the major portion of this precipitate was 14000 without reduction and alkylation of disulphide bonds. Reduction of disulphide bonds with dithiothreitol followed by alkylation with iodoacetamide did not change the apparent molecular weight of the major portion of the precipitate. It is possible that the precipitate is composed of the entire C-terminal, non-collagen-like, region of one of the A, B or C chains and which is not readily fragmented by pepsin.

The fraction P1, obtained by running the supernatant of the pepsin digest on Sephadex G-200 (Fig. 1), was enriched in the amino acids hydroxyproline, hydroxylysine and glycine relative to the intact subcomponent C1q (Table 1). The value of 27 residues of glycine/100 residues obtained for fraction P1 is consistent with 80% of this fraction being composed of Gly-X-Y collagen-like triplet sequence. Fraction P1 was 38% (by weight) of the intact subcomponent C1q present at the start of the digest and was, like the intact subcomponent C1q, eluted in the void volume of the Sephadex G-200 column, which suggests that it was behaving as an aggregate or had unusual gel-filtration properties. The fraction P2, which was eluted as a broad peak late in the Sephadex G-200 run, was found to be composed of mainly small-molecular-weight non-collagen-like peptides. Fraction P2 had a much lower glycine content (9.21%) than the intact subcomponent C1q (17.09%) (Table 1) and contained only trace amounts of hydroxyproline and hydroxylysine. These small non-collagen-like peptides in fraction P2 accounted for approx. 48% of the total protein in

Table 1. *Amino acid compositions of the precipitate and the two major peaks from Sephadex G-200 obtained after fractionation of a limited pepsin digest of human subcomponent C1q*

Samples were hydrolysed at 110°C for 24h. Experimental details are given in the text. No corrections were made for threonine and serine destruction, and tryptophan was not estimated. The results are expressed as amino acid residues/100 residues. The values for intact human subcomponent C1q are taken from Reid *et al.* (1972). Methionine and cystine were measured as methionine sulphone [Met(O₂)] and cysteic acid [Cys(O₃H)] throughout. N.D., Not determined.

Amino acid	Amino acid composition (residues/100 residues)			
	Pepsin precipitate	Sephadex G-200		Human subcomponent C1q
		fraction P1	fraction P2	
Cys(O ₃ H)	3.65	0.89	1.20	1.73
Hyp	—	10.02	0.22	4.51
Asp	9.43	5.71	10.80	8.19
Met(O ₂)	0.75	1.30	2.03	1.55
Thr	6.72	2.17	7.45	5.62
Ser	9.18	2.69	6.40	4.72
Glu	10.77	8.90	9.37	8.39
Pro	4.55	7.64	6.50	7.05
Gly	12.30	27.01	9.21	17.09
Ala	3.54	5.36	4.08	4.36
Val	8.60	2.50	8.89	5.95
Ile	4.58	5.09	3.68	4.26
Leu	8.00	4.26	7.85	6.20
Tyr	3.88	1.91	4.70	3.57
Phe	5.10	1.85	6.10	4.23
His	2.00	1.29	3.51	1.41
Hyl	0.42	3.85	0.20	2.09
Lys	3.13	3.88	3.32	4.04
Arg	3.37	3.74	4.56	4.55
Trp	N.D.	N.D.	N.D.	0.50
Proportion of digest	9%	38%	48%

Fig. 1. *Gel filtration on Sephadex G-200 of the supernatant from the partial pepsin digest of subcomponent C1q*

The supernatant (24ml) from the partial pepsin digest of human subcomponent C1q (87mg) was applied to a column (6cm×85cm) of Sephadex G-200, which was equilibrated with 100mM-acetic acid and run at 4°C. The column was run at 25ml/h, and 9.3ml fractions were collected. Subsequently the fractions within the horizontal bars were pooled and freeze-dried.

the digest, and therefore, along with the non-collagen-like material in the precipitate, gave a value of 57% by weight for the amount of non-collagen-like material in subcomponent C1q.

CM-cellulose chromatography of the collagen-like fraction P1

Fraction P1 yielded two major fractions, P1-CMC1 and P1-CMC2, when run on CM-cellulose at pH4.9 in buffers containing 7.5M-urea (Fig. 2a). These fractions were re-run on CM-cellulose, under the same conditions, to check their purity and homogeneity (Fig. 2b). The amino acid compositions of fractions P1-CMC1 and P1-CMC2 (Table 2) showed that both contained hydroxyproline and hydroxylysine and had a high (28%) glycine content. Assuming that there are two cysteic acid residues in the oxidized samples of fractions P1-CMC1 and P1-CMC2, both these fractions appeared to be composed of about 185 amino acid residues (Table 2). Their amino acid compositions were markedly different with respect to several

Fig. 2. Separation of the two major collagen-like fragments produced by pepsin digest of subcomponent C1q, and their oxidized chains, on CM-cellulose

All samples were applied to a column (1.1 cm $\times$ 14 cm) of CM-cellulose equilibrated with 36 mM-sodium acetate/7.5 M-urea buffer, pH 4.9, run at room temperature at a flow rate of 26 ml/h, and fractions (2.6 ml) were collected. In each run, after 15 fractions, a linear gradient composed of the buffer used for equilibration (100 ml) and the buffer used for equilibration made 100 mM with respect to NaCl (100 ml) was begun. The samples applied were: (a) fraction P1 from Sephadex G-200; (b) fraction P1-CMC2; (c) oxidized fraction P1-CMC2. Subsequently the fractions within the horizontal bars were pooled, desalted and freeze-dried.

phenylalanine, histidine and hydroxylysine values (Table 2). Since fractions P1-CMC1 and P1-CMC2-Ox1 appeared to display a monomer/dimer relationship, the integral values of amino acids for fraction P1-CMC1-Ox1 shown in Table 2 have been doubled for comparison.

Oxidation of fraction P1-CMC2 followed by fractionation of the oxidized material on CM-cellulose at pH 4.9 in buffers containing 7.5 M-urea yielded two polypeptide chains, P1-CMC2-Ox1 and P1-CMC2-Ox2 (Fig. 2c). These chains differed markedly in their amino acid compositions; for example, peptide P1-CMC2-Ox1 contained no arginine and one amino acid residue each of methionine and histidine, whereas peptide P1-CMC2-Ox2 contained eight arginine residues and no methionine or histidine (Table 2). Peptides P1-CMC2-Ox1 and P1-CMC2-Ox2 were estimated to be 91 and 94 amino acid residues long respectively, and the sum of the residues in both the chains was identical with that for the intact fraction P1-CMC2, except for minor differences in the serine and lysine values, which suggested that fraction P1-CMC2 was composed of equimolar amounts of peptides P1-CMC2-Ox1 and P1-CMC2-Ox2.

The overall yield of the three oxidized polypeptide chains isolated from fractions P1-CMC1 and P1-CMC2 was 16–18% in each case (Table 2).

N-Terminal and C-terminal sequence analysis of fractions P1-CMC1 and P1-CMC2 and of the polypeptide chains derived from them by performic oxidation

The *N*-terminal sequence of residues 1–10 of fraction P1-CMC1 is shown in Table 3. Alkylation with ^{14}C -labelled iodoacetic acid, after reduction of fraction P1-CMC1 allowed firm identification of an *S*-carboxymethylcysteine residue at position 4 (Table 3). The yields of amino acid phenylthiohydantoin derivatives released at each step were low, being only 16% for that of glycine at step 3 as measured by g.l.c. Since fraction P1-CMC1 was subjected to two purification steps and one labelling step in buffers that contained urea, it is possible that partial *N*-terminal blocking may have taken place. The *N*-terminal sequence of fraction P1-CMC1 was found to be identical with that of residues 1–10 of the intact C chain of subcomponent C1q (Mercer, 1975), except that residue 8 of fraction P1-CMC1 was not identified. The *N*-terminal amino acid of the polypeptide chain P1-CMC1-Ox1, which was derived from fraction P1-CMC1 by performic oxidation and fractionation on CM-cellulose, was shown to be Asx by the dansyl technique. The amino acids released from the C-terminal ends of both fraction P1-CMC1 and

amino acids, such as methionine, threonine, alanine, valine, tyrosine and histidine (Table 2).

Oxidation of fraction P1-CMC1 followed by fractionation of the oxidized material on CM-cellulose at pH 4.9 in buffers containing 7.5 M-urea yielded one polypeptide chain P1-CMC1-Ox1, which was estimated to be 94 amino acid residues long (Table 2). Peptide P1-CMC1-Ox1 had an amino acid composition that appeared to be identical with that obtained for the intact P1-CMC1 fraction except for small differences in the aspartic acid,

Table 2. Amino acid compositions of fractions P1-CMC1 and P1-CMC2 and the polypeptide chains derived from these fractions by performic acid oxidation

Samples were hydrolysed at 110°C for 24h. Experimental details are given in the text. No corrections were made for threonine and serine destruction, and tryptophan was not estimated. Fractions P1-CMC1 and P1-CMC2 were oxidized with performic acid before amino acid analysis. The results are expressed as the mol of amino acid residues/mol of fraction or chain, on the basis that there are 2 mol or 1 mol respectively of cystic acid/mol of fraction or chain. Samples were treated with carboxypeptidase A for 2h at 37°C before determination of liberated C-terminal amino acids by amino acid analysis.

Amino acid composition

Amino acid	Amino acid composition									
	Assuming Cys(O ₃ H) = 2.00 P1-CMC1 (determined)	(to nearest integer)	Assuming Cys(O ₃ H) = 1.00 P1-CMC1-Ox1 (determined)	[To nearest integer (values multiplied by two)]	Assuming Cys(O ₃ H) = 2.00 P1-CMC2 (determined)	(to nearest integer)	Assuming Cys(O ₃ H) = 1.00 P1-CMC2-Ox1 (determined)	Assuming Cys(O ₃ H) = 1.00 P1-CMC2-Ox2 (determined)	[to nearest integer (values for P1-CMC2-Ox1 plus values for P1-CMC2-Ox2)]	
Cys(O ₃ H)	2.0	2	1.0	2	2.0	2	1.0	1.0	2	
Hyp	28.4	28	14.1	28	18.1	18	10.5	7.3	18	
Asp	9.1	9	4.4	8	11.2	11	4.7	6.0	11	
Met(O ₂)	6.4	6	3.0	6	1.3	1	1.0	—	1	
Thr	2.4	2	1.3	2	5.2	5	3.2	1.9	5	
Ser	2.1	2	1.4	2	4.2	4	1.8	3.2	5	
Glu	15.7	16	8.2	16	16.0	16	7.2	9.2	16	
Pro	14.0	14	7.4	14	13.5	14	7.1	7.2	14	
Gly	52.0	52	26.0	52	52.0	52	26.0	26.0	52	
Ala	5.5	6	3.3	6	11.2	11	5.5	5.3	11	
Val	3.5	4	1.8	4	2.2	2	1.2	1.1	2	
Ile	10.2	10	5.1	10	11.0	11	5.7	4.8	11	
Leu	6.4	6	3.2	6	7.3	7	3.2	4.0	7	
Tyr	4.2	4	1.5	4	2.2	2	0.8	0.9	2	
Phe	2.6	3	1.5	4	3.0	3	1.7	1.0	3	
His	2.5	3	2.0	4	1.3	1	1.1	0.1	1	
Hyl	5.9	6	4.1	8	9.8	10	5.1	4.7	10	
Lys	7.5	8	3.7	8	6.7	7	3.1	3.4	6	
Arg	4.4	4	2.0	4	7.7	8	0.3	7.8	8	
Total residues		185		188		185			185	
Yield			17%				18%	16%		
N-terminal amino acid	Asp		Asx		Glu		None detected	Glx		
C-Terminal amino acids released by carboxypeptidase A as mol of residues/mol.	Phe (1.0) Ser* (0.7) Val (0.5)		Phe (1.0) Ser* (0.5) Val (0.5)		Phe (1.8) Ala (1.7) Ile (1.0) Ser* (0.8)		Phe (1.2) Ala (1.1) Ile (1.1)	Phe (1.0) Ala (0.8) Ser* (0.9)		

* Asparagine and glutamine are eluted along with serine under the conditions used for amino acid analysis.

Table 3. Sequence analysis of the *N*-terminal first ten residues of fractions P1-CMC1 and P1-CMC2

(a) Fraction P1-CMC1 (30nmol) and (b) fraction P1-CMC2 (30nmol) were each subjected to ten cycles of Edman degradation in the 890C Beckman protein sequencer with the 0.1 M-Quadrol protein programme. Fraction P1-CMC1 was reduced and then alkylated with iodo[¹⁴C₂]acetic acid. Full experimental details are given in the text. The yield of amino acid phenylthiohydantoin derivative released at each cycle was estimated by either the g.l.c. peak area or by amino acid analysis after hydrolysis in HI. The yields ranged from 15 to 17% at cycles 1–3 down to 3–4% at cycle 10 in both runs. The sequence of the first ten amino acid residues of the A chain of human subcomponent C1q is taken from Reid (1974, 1975). The sequence of the first ten amino acid residues of the C chain of human subcomponent C1q is taken from Mercer (1975). X represents an unidentified residue. Abbreviations: CmCys, *S*-carboxymethylcysteine; Aab, 2-aminobutyric acid.

Residue no.	Method of identification of amino acid phenylthiohydantoin derivative			Radioactivity released (c.p.m.)	Proposed sequence	Known sequence of:
	G.l.c.	T.l.c.	Amino acid analysis			
(a) P1-CMC1						
1	—	Asn	Asx	200	Asn	(a) C chain Asn
2	—	—	Aab*	200	Thr	Thr
3	—	Gly	Gly	780	Gly	Gly
4	CmCys	CmCys	—	9140	Cys	Cys
5	—	Tyr	Tyr	1040	Tyr	Tyr
6	—	Gly	Gly	360	Gly	Gly
7	Ile/Leu	Ile/Leu	Ile	260	Ile	Ile
8	—	—	—	200	X	Hyp
9	—	Gly	Gly	180	Gly	Gly
10	Met	Met	—	200	Met	Met
(b) P1-CMC2						
1	—	Glu	Glx		Glu	(b) A chain Glu
2	—	Asp	Asx		Asp	Asp
3	Ile/Leu	Ile/Leu	Leu		Leu	Leu
4	—	Cys	Ala†		Cys	Cys
5	—	—	Arg		Arg	Arg
6	Ala	Ala	Ala		Ala	Ala
7	—	Pro	Pro		Pro	Pro
8	—	Asp	Asx		Asp	Asp
9	Gly	Gly	Gly		Gly	Gly
10	—	Lys	Lys		Lys	Lys

* Threonine phenylthiohydantoin derivative yields Aab on hydrolysis in HI.

† Cysteine phenylthiohydantoin derivative yields Ala on hydrolysis in HI.

peptide P1-CMC1-Ox1 by carboxypeptidase A in 2h were phenylalanine, valine and 'serine' in approximately equimolar amounts (Table 2). A time-course study of the digestion of fractions P1-CMC1 and P1-CMC1-Ox1 by carboxypeptidase A showed that both had the C-terminal sequence -'Ser'-Val-Phe. Further experiments with carboxypeptidase B as well as carboxypeptidase A showed that the C-terminal sequence of peptide P1-CMC1-Ox1 was -Lys-'Ser'-Val-Phe. These results established that fractions P1-CMC1 and P1-CMC1-Ox1 had the same *N*-terminal and C-terminal amino acid sequences.

The *N*-terminal sequence of residues 1–10 in fraction P1-CMC2 is shown in Table 3. Only one sequence was found, and this was identical with that obtained for residues 1–10 of the oxidized A chain

of subcomponent C1q (Reid, 1974, 1975). The *N*-terminal amino acid of peptide P1-CMC2-Ox2, derived from fraction P1-CMC2 by performic acid oxidation and fractionation on CM-cellulose, was found to be Glx by the dansyl technique. The amino acid composition of peptide P1-CMC2-Ox2 (Table 2) was found to be identical with that of the *N*-terminal first 94 residues of the oxidized A chain of subcomponent C1q (Reid, 1974, 1975), except that it appeared to contain one more hydroxyproline residue. For technical reasons, it has been found that hydroxyproline is sometimes difficult to determine with great accuracy. Carboxypeptidase A digestion of fraction P1-CMC2 for 2h released phenylalanine (1.8), alanine (1.7), isoleucine (1.0) and 'serine' (0.8), whereas digestion of peptide P1-CMC2-Ox2 for the same time released only phenylalanine (1.0), alanine

Table 4. *Amino acid composition of the precipitate Col-B-P isolated after collagenase digestion of the oxidized B chain*

The sample was hydrolysed at 110°C for 24 h. No corrections were made for serine and threonine destruction, and tryptophan was not estimated. Full experimental details are given in the text. The results are expressed as residues/100 residues and as mol of amino acid residues/mol of oxidized B chain with integral values given in parentheses.

Amino acid	Amino acid composition	
	(residues/ 100 residues)	(mol of residue/mol)
Cys(O ₃ H)	2.28	2.7 (3)
Hyp	Nil	Nil (0)
Asp	11.16	13.1 (13)
Met(O ₂)	2.58	3.0 (3)
Thr	7.18	8.5 (9)
Ser	5.76	6.8 (7)
Glu	10.43	12.3 (12)
Pro	4.13	4.9 (5)
Gly	10.11	11.9 (12)
Ala	5.49	6.4 (6)
Val	5.77	6.8 (7)
Ile	3.67	4.3 (4)
Leu	7.15	8.4 (8)
Tyr	4.56	5.4 (5)
Phe	6.75	7.9 (8)
His	1.70	2.0 (2)
Hyl	0.77	0.9 (1)
Lys	4.39	5.2 (5)
Arg	6.12	7.2 (7)
Total residues		117

(0.8) and 'serine' (0.9) (Table 2). Residues 92–97 of the oxidized A chain are -Pro-Ala-Phe-Ser-Ala-Ile- (Reid, 1974, 1975), which suggested that cleavage by pepsin took place at the Ser-95 or Ala-96 residues. The amino acid composition of peptide P1-CMC2-Ox2 (Table 2), the *N*-terminal analyses of fractions P1-CMC2 and P1-CMC2-Ox2 (Tables 2 and 3) and *C*-terminal analysis of fractions P1-CMC2 and P1-CMC2-Ox2 (Table 2) were all consistent with the first 95 residues of the A chain of subcomponent C1q being disulphide-bonded to another polypeptide, approximately 91 residues long, to yield intact fraction P1-CMC2.

The second polypeptide chain, P1-CMC-Ox1, derived from fraction P1-CMC2 by performic acid oxidation and fractionation on CM-cellulose, had no detectable free *N*-terminal amino acid. It is known from previous work (Reid *et al.*, 1972) that the *N*-terminal amino acid of the oxidized B chain of subcomponent C1q is blocked, and this therefore suggested that peptide P1-CMC2-Ox1 might be

derived from the B chain. Digestion of peptide P1-CMC2-Ox1 with carboxypeptidase A for 2 h yielded phenylalanine (1.2), alanine (1.1) and isoleucine (1.1), which, on addition to the amino acids liberated from peptide P1-CMC2-Ox2 by carboxypeptidase A, gives the exact sum of amino acids released from the intact fraction P1-CMC2 by carboxypeptidase A. Further studies on the *C*-terminal sequence of peptide P1-CMC2-Ox1 were performed for time-intervals ranging from 1 min to 120 min with carboxypeptidase A alone and with carboxypeptidase A plus carboxypeptidase B. These studies showed that the *C*-terminal sequence of peptide P1-CMC2-Ox1 was consistent with the partial sequence of -Tyr-(Lys,Ala,Thr,'Ser')-Lys-Ile-Ala-Phe. The 120 min values expressed as mol of residues released/mol of peptide by carboxypeptidase A plus carboxypeptidase B were tyrosine (0.8), lysine (2.0), threonine (0.8), isoleucine (1.2), alanine (2.1), 'serine' (0.5) and phenylalanine (1.2). 'Serine' is probably a glutamine residue, as this is suggested from the sequence data given below, showing that peptide P1-CMC2-Ox1 is derived from the B chain of subcomponent C1q.

Sequence information obtained from a precipitate, 'Col-B-P', produced by collagenase digestion of the oxidized B chain of human subcomponent C1q

Prolonged collagenase digestion of the oxidized B chain of subcomponent C1q yielded a number of soluble peptides plus a precipitate. The precipitate was composed of a single polypeptide chain with an apparent mol.wt. of 13600, as estimated by polyacrylamide-gel electrophoresis run in buffers containing sodium dodecyl sulphate. The amino acid composition of this polypeptide chain, Col-B-P, shows that it has one hydroxylysine residue, no hydroxyproline and a relatively low glycine content (Table 4). The *N*-terminal amino acid of fraction Col-B-P was found to be glycine; and, since it is known that the *N*-terminal amino acid of the intact B chain is blocked, and that collagenase cleaves only susceptible Y-Gly bonds in the repeating (Gly-X-Y)_n sequence, it appeared likely that fraction Col-B-P represented the *C*-terminal 117 residues of the B chain. This positioning of fraction Col-B-P in the B chain is consistent with sequence data obtained from trypsin and chymotrypsin digests of the B chain (E. O. P. Thompson & K. B. M. Reid, unpublished work). Automated *N*-terminal sequence analysis of fraction Col-B-P showed that it had a short collagen-like sequence for the first nine *N*-terminal residues, this then being followed by a sequence that showed no collagen-like features (Table 5). At cycle 25 in the Edman degradation of fraction Col-B-P the yield of the amino acid phenylthiohydantoin derivative fell

Table 5. *Proposed sequence of the first 25 amino acid residues of the fragment Col-B-P produced by collagenase digestion of the oxidized B chain of human subcomponent C1q*

Precipitate Col-B-P (280 nmol) was subjected to 30 cycles of Edman degradation in the Beckman 890C protein sequencer with the DMAA-fast peptide programme. Full experimental details are given in the text. The amino acid phenylthiohydantoin derivatives were identified by g.l.c., t.l.c. and, in some cases, by back hydrolysis in HI followed by amino acid analysis. Yields of amino acid phenylthiohydantoin derivatives were calculated from the g.l.c. peak areas. The residues that can be accurately determined quantitatively by g.l.c. are alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine, tyrosine and tryptophan. A repetitive yield of 95% was calculated from the slope of a graph of a semi-logarithmic plot of the absolute yield in nmol against the residue number for the first 23 steps. N.D., Not done. X represents an unidentified residue which may be hydroxylysine (K. B. M. Reid, unpublished work). Abbreviation: Aab, 2-aminobutyric acid.

Residue no.	Method of identification of amino acid phenylthiohydantoin derivative			Proposed sequence	Yield of amino acid phenylthiohydantoin derivative (nmol)
	G.l.c.	T.l.c.	Amino acid analysis		
1	Gly	Gly	Gly	Gly	129
2	Pro/Thr	Pro	Pro	Pro	—
3	Pro/Thr	Pro	Pro	Pro	—
4	Gly	Gly	Gly	Gly	98
5	—	Glu	Glx	Glu	—
6	Ser	Glu	Ala†/Glx	X	—
7	Gly	Gly	Gly	Gly	90
8	—	Asp	Asx	Asp	—
9	Tyr*	Tyr	N.D.	Tyr	—
10	—	Lys	N.D.	Lys	—
11	Ala	Ala	Ala	Ala	78
12	Pro/Thr	Thr	Aab‡	Thr	—
13	—	Gln	Glx	Gln	—
14	—	Lys	Lys	Lys	—
15	Ile*	Ile/Leu	Ile	Ile	67
16	Ala	Ala	N.D.	Ala	64
17	Phe	Phe	Phe	Phe	59
18	Ser	—	Ala	Ser	—
19	Ala	Ala	Ala	Ala	50
20	Pro/Thr	Thr	Aab	Thr	—
21	—	—	Arg	Arg	—
22	Pro/Thr	Thr	N.D.	Thr	—
23	Ile*	Ile/Leu	N.D.	Ile	42
24	—	Asn	N.D.	Asn	—
25	Val	Val	N.D.	Val	19

* Identified after silylation of sample.

† Serine phenylthiohydantoin derivative yields Ala on hydrolysis in HI.

‡ Threonine phenylthiohydantoin derivative yields Aab on hydrolysis in HI.

sharply, and no further convincing sequence information was obtained beyond this cycle. The sequence of the residues at positions 8–17 in fraction Col-B-P was -Asp-Tyr-Lys-Ala-Thr-Gln-Lys-Ile-Ala-Phe-, and this is entirely consistent with the partial sequence of -X-Tyr-(Lys,Ala,Thr,'Ser')-Lys-Ile-Ala-Phe suggested for the C terminal end of peptide P1-CMC2-Ox1 by digestion with carboxypeptidase A and carboxypeptidase B. The Phe residue at position 17 in fraction Col-B-P (i.e. position 91 in the B chain) therefore appears to be the point at which the B chain is cleaved in the intact subcomponent C1q molecule (Fig. 4).

Polyacrylamide-gel electrophoresis, in buffers containing sodium dodecyl sulphate, of the fractions produced by pepsin digestion of subcomponent C1q

Unreduced intact subcomponent C1q was shown to be composed of two types of non-covalently linked subunit of apparent mol.wts. 67 000 and 54 000 and which were found in a ratio of 1.99:1.00 when the sample was loaded on the gel in a buffer of low ionic strength (Fig. 3a). The gels were scanned at 550 nm after staining with Coomassie Brilliant Blue. It was assumed that the two subunits, which have similar amino acid compositions, chain structure and which are close in their apparent molecular

Fig. 3. Scans of polyacrylamide-gel electrophoresis of intact subcomponent C1q and fractions derived from the pepsin digest of subcomponent C1q

All samples were run on 7% acrylamide gels in buffers containing sodium dodecyl sulphate. (a) Intact unreduced subcomponent C1q; the ionic strength of the applied sample was 0.05 mol/l; (b) unreduced supernatant of pepsin digest of subcomponent C1q; (c) unreduced fraction P1-CMC2; (d) unreduced fraction P1-CMC1.

weights, would stain with the same intensity with Coomassie Blue over the range 40–200 μ g of protein. The ratio of the two subunits found by this technique was found to be markedly affected by the ionic strength of the applied sample (Reid & Porter, 1976). However, it is thought that the ratio of 2:1 shown in Fig. 3(a) is the correct value to be used in any estimation of the subunit composition of subcomponent C1q (Reid & Porter, 1976).

Fig. 3(b) shows a gel scan of the band pattern obtained from the supernatant of the pepsin digest before its gel filtration on Sephadex G-200. It can be seen that there are two major bands of apparent mol.wts. 35800 and 28500, along with some material that may be residual intact subcomponent C1q. The molecular weights were calculated by using unreduced globular proteins as markers and, owing to the collagen-like nature of the major fractions in the pepsin digest, this has probably led to a gross overestimation of molecular weights, for the reasons outlined in the Discussion section. Fraction P1 eluted from Sephadex G-200 gave a gel band pattern identical with that obtained with the supernatant of the total digest (Fig. 3b) when run under the same conditions. This showed that the peptides in fraction P2 eluted from Sephadex G-200 are too small to be detected on 7% acrylamide gels run in buffers containing sodium dodecyl sulphate. When unreduced samples of fractions P1-CMC1 and P1-CMC2 were run on 7% acrylamide gels in buffers containing sodium dodecyl sulphate it was found that fraction P1-CMC1 corresponded to the 28500-apparent-mol.wt. fraction (Fig. 3d) and that fraction P1-CMC2 corresponded to the 35800-apparent-mol.wt. fraction (Fig. 3c). Reduction and alkylation of fractions P1-CMC1 and P1-CMC2 followed by electrophoresis on 7% acrylamide gels run in buffers containing sodium dodecyl sulphate showed that their apparent molecular weights were approximately halved. Values of 15000 and 16900 were obtained for reduced fraction P1-CMC1 and reduced fraction P1-CMC2 respectively. Reduced non-collagen-like proteins were used as markers, and this may have led to overestimation of the molecular weights, for the reasons outlined in the Discussion section.

Calculation of the molecular weights of fractions P1-CMC1, P1-CMC2, P1CMC1-Ox1, P1CMC2-Ox1 and P1-CMC2-Ox2 from their amino acid compositions

It has been shown above that peptide P1-CMC2-Ox2 is composed of the first 94 or 95 residues of the A chain of subcomponent C1q, and therefore an accurate estimate of the molecular weight of this fragment can be made from its amino acid composition. A value of 11400 was obtained for

the mol.wt. of peptide P1-CMC2-Ox2 when allowance was made for the fact that the five hydroxyl-lysine residues in this sequence probably have glucosylgalactose disaccharide units linked to them (Calcott & Müller-Eberhard, 1972; Reid, 1974). Similarly values of 8800 and 9600 were obtained for peptides P1-CMC2-Ox1 and P1-CMC1-Ox1 respectively, and values of 18700 and 20000 were obtained for fractions P1-CMC1 and P1-CMC2 respectively.

Discussion

It was found that the optimum conditions for digestion of subcomponent C1q with pepsin, such that there were no major cleavages in the collagen-like regions of the molecule, was by incubation at a substrate/enzyme ratio of 30:1 (w/w) for 20h at pH 4.4 at 37°C. The higher pH of 4.4, compared with that of approx. 2.6 used in pepsin digests of procollagen and collagen (Bellamy & Bornstein, 1971; Miller, 1972), was used in an attempt to enhance the specificity of the pepsin used in the digestion of subcomponent C1q. The use of a temperature of 37°C, rather than a temperature in the range 4°–15°C as has been used for digestion of procollagen and collagen non-triple-helical amino acid sequences, is consistent with the finding that some structure in subcomponent C1q, and in pepsin-treated subcomponent C1q, has a higher 'melting' temperature (47.7°C) than that of lathyrin rat skin collagen (37.8°C) as estimated by circular-dichroism studies (B. B. Doyle & K. B. M. Reid, unpublished work). It is noteworthy that rabbit subcomponent C1q was found to retain all its haemolytic activity (and presumably structure) after heating up to a temperature of 48°C, whereas at 50°C only 45% of the activity remained after cooling to 37°C, and at 52°C no activity was recovered on cooling (Lowe & Reid, 1974). Therefore these studies involving pepsin digestion, circular-dichroism measurements and heat inactivation all suggest that, if a collagen-like triple helix structure is present in subcomponent C1q, it may be more stable than that normally found in most collagens.

The collagen-like fraction P1, obtained by gel filtration of the pepsin digest of subcomponent C1q on Sephadex G-200 (Fig. 1), was isolated in a yield of 38% by weight of the amount of subcomponent C1q used for the digest, which corresponds to a yield of approx. 91% of that which would be expected if fraction P1 were composed of the entire sections of collagen-like amino acid sequence known to be present in subcomponent C1q (Reid *et al.*, 1972; Reid, 1974). Fraction P1 was eluted in the void volume of Sephadex G-200; this was unexpected, since it can be estimated that the maximum mol.wt. of fraction P1, if not aggregated, would have been

approx. 176000. However, it is well established that, in gel filtration, collagen peptides are eluted at positions that correspond to globular proteins of a much higher molecular weight (Piez, 1968). This may explain why fraction P1 is still eluted in the void volume of Sephadex G-200 despite having lost at least 58% of the subcomponent C1q non-collagen-like amino acid sequence (Table 1). Extensive aggregation of fraction P1 does not appear likely, since electron-microscopy studies indicated that it consisted almost entirely of single molecules of subcomponent C1q that had lost their six peripheral globular regions (K. Leonard, unpublished work). It was previously established by electron-microscopy studies (Shelton *et al.*, 1972; Knobel *et al.*, 1975) that the subcomponent C1q is composed of six peripheral globular regions [5.0nm×7.0nm (50Å×70Å) each] each joined by a connecting strand [1.5nm×11.5nm (15Å×115Å)] to a fibril-like central portion [4.5nm×11.2nm (45Å×112Å)]. Since fraction P1 is composed of the six connecting strands and fibril-like central portion, it appears likely that the six peripheral globular regions were fragmented by pepsin to yield the non-collagen-like peptides found in fraction P2 (Fig. 1 and Table 1).

The yields of fractions P1-CMC1 and P1-CMC2, after the first CM-cellulose run (Fig. 2a), were 50–60% of that expected if a mol.wt. of 410000 is taken for subcomponent C1q and it is accepted that intact subcomponent C1q is composed of 6mol of each of the A, B and C chains/mol (Reid *et al.*, 1972). The yields of fractions P1-CMC1-Ox1, P1-CMC2-Ox1 and P1-CMC2-Ox2, which correspond to the N-terminal 91–95 residues of the C, B and A chains respectively, were all of the same order (16–17% of that expected), which is additional evidence that there are equimolar amounts of the three chains if it is assumed that any losses during the chromatography and desalting steps would be the same for each fraction.

The mol.wts. of fractions P1-CMC1, P1-CMC2, P1-CMC1-Ox1, P1-CMC2-Ox1 and P1-CMC2-Ox2 were estimated to be 28500, 35800, 15000, 16900 and 16900 respectively on polyacrylamide gels run in the presence of sodium dodecyl sulphate with the use of non-collagen-like proteins as markers, whereas values of 18700, 20000, 9600, 8800 and 11400 were estimated from the amino acid compositions of the fractions. The results from the polyacrylamide gels run in the presence of sodium dodecyl sulphate show a dimer/monomer relationship between fractions P1-CMC1 and P1-CMC1-Ox1 and between fraction P1-CMC2 and fractions P1-CMC2-Ox1 plus P1-CMC2-Ox2; however, the large differences in the molecular-weight estimates are almost certainly due to the atypical behaviour of the collagen-like fractions on polyacrylamide gels run in the presence of sodium dodecyl sulphate. It has been shown that

Fig. 4. Probable sites of cleavage of subcomponent C1q by pepsin,

P, probable sites of pepsin cleavage of intact subcomponent C1q at pH4.4 which yield fractions P1-CMC1 and P1-CMC2; 1-98 of the A chain are taken from Reid (1974, 1975). ----, amino acid sequences for which only partial sequence data in the B chain and 94 in the C chain is only tentative because of the incomplete nature of the B chain and C chain amino

molecular weights 50% higher than expected are consistently obtained when collagen fragments are run on polyacrylamide gels in the presence of sodium dodecyl sulphate (Furthmayr & Timpl, 1971). It has been suggested that the principal reason for this unusual behaviour of the collagen fragments may be the existence of a certain rigidity of the polypeptide structure even in the denatured state.

The probable sites of pepsin cleavage of the B and C chains in the intact subcomponent C1q, at pH4.4, appear to be at the phenylalanine residues at positions B-91 and C-94 (Fig. 4). The site of pepsin cleavage in the A chain appears to be the serine or alanine residues at positions A-95 and A-96. Attempts to establish the exact C-terminal sequence of peptide P1-CMC2-Ox2 by digestion with carboxypeptidase A at times from 30s to 120min suggested that the sequence was -Phe-Ser-Ala, but that the alanine was present as a partial residue and that major splitting by pepsin took place at the serine at position A-95.

The location of a disulphide bond joining positions 4 of each chain of the C-chain-derived fraction P1-CMC1 (Fig. 4) was unexpected, since half-cystine has not been found in sequences of the Gly-X-Y repeating triplet type previously, although *Ascaris* collagen has been reported to have a high

half-cystine content, namely about three residues/100 residues (McBride & Harrington, 1967). The evidence showing that position 4 in the A chain is linked to a position in the section of residues 2-6 of the B chain (Fig. 4) comes from unpublished work (E. O. P. Thompson & K. B. M. Reid). The location of these two interchain disulphide bridges is consistent with the data present by Yonemasu & Stroud (1972) on the chain composition of the unreduced subunits, in which they showed that a 60000-apparent-mol.wt. subunit was a disulphide-linked dimer of the A and B chains and that 42000-apparent-mol.wt. subunit was a disulphide-linked dimer of the C chain.

The presence of a dimer composed of regions of the A and B chains and a dimer composed of regions of two C chains, as has been shown from the pepsin-digestion studies, is not consistent with the view that all three chains reside in the same type of non-covalently linked subunit, as was suggested previously (Reid *et al.*, 1972). Therefore the subunit structure of subcomponent C1q has been re-examined in a previous publication (Reid & Porter, 1975) and in the following paper (Reid & Porter, 1976).

I thank Professor R. R. Porter for his advice throughout this work and in the preparation of the manuscript, and also Mr. T. Gascoyne for performing the amino acid

⁹⁵
 Lys-Asp-Gln-Pro-Arg-Pro-Ala-Phe-Ser-Ala-Ile-Arg- (-----) C
 ↑ ↑
 P P

⁹¹
 Lys-Ala-Thr-Gln-Lys-Ile-Ala-Phe-Ser-Ala-Thr-Arg-Thr-Ile-Asn-Val- (-----) C
 ↑
 P

⁹⁴
 -----)-Lys'-Ser'-Val-Phe- (-----) C
 ↑
 P

⁹⁴
 -----)-Lys'-Ser'-Val-Phe- (-----) C
 ↑
 P

Col, site of collagenase cleavage of the oxidized B chain which yields fraction Col-B-P. The sequence of amino acid residues is available and which therefore has not been included in the figure. The exact numbering of the phenylalanine residues at 91 acid sequences. 'X', when not in Gly-X-Y triplet sequence, represents an unidentified residue.

Bellamy, P. & Bornstein, P. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1138-1142

Brauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14**, 3029-3035

Calcott, M. A. & Müller-Eberhard, H. J. (1972) *Biochemistry* **11**, 3443-3450

Fairbanks, G., Stech, T. L. & Wallach, D. F. H. (1971) *Biochemistry* **10**, 2606-2617

Furthmayr, H. & Timpl, R. (1971) *Anal. Biochem.* **41**, 510-516

Hirs, C. H. W. (1956) *J. Biol. Chem.* **219**, 611-621

Knobel, H. R., Villiger, W. & Isliker, H. (1975) *Eur. J. Immunol.* **5**, 78-82

Lowe, D. M. & Reid, K. B. M. (1974) *Biochem. J.* **143**, 265-272

McBride, O. W. & Harrington, W. F. (1967) *Biochemistry* **6**, 1484-1498

Mercer, C. P. (1975) M.Sc. Thesis, University of Oxford

Miller, E. J. (1972) *Biochemistry* **11**, 4903-4909

- Peterkofsky, B. & Diegelmann, R. (1971) *Biochemistry* **10**, 988-994
- Piez, K. A. (1968) *Anal. Biochem.* **26**, 305-312
- Reid, K. B. M. (1974) *Biochem. J.* **141**, 189-203
- Reid, K. B. M. (1975) in *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 2 (Dayhoff, M. O., ed.), National Biomedical Research Foundation, Silver Spring
- Reid, K. B. M. & Porter, R. R. (1975) in *Contemporary Topics in Molecular Immunology*, vol. 4 (Inman, F. P., ed.), pp. 1-22, Plenum Publishing Corp., New York
- Reid, K. B. M. & Porter, R. R. (1976) *Biochem. J.* **155**, 19-23
- Reid, K. B. M., Lowe, D. M. & Porter, R. R. (1972) *Biochem. J.* **130**, 749-763
- Shelton, E., Yonemasu, K. & Stroud, R. M. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 65-68
- Smithies, O., Gibson, D., Fanning, E. M., Goodfliesth, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10**, 4912-4921
- Summers, M. R., Smythers, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53**, 624-628
- Traub, W. & Piez, K. A. (1971) *Adv. Protein. Chem.* **25**, 243-352
- Yonemasu, K. & Stroud, R. M. (1972) *Immunochemistry* **9**, 545-554