Complete Amino Acid Sequences 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 3 October 1978)

The sequences of amino acid residues 38–51 of the A-chain, and residues 42–90 of the C-chain, of human subcomponent C1q are given. These results, along with previously published sequence data [Reid (1974) *Biochem. J.* **141**, 189–203; Reid (1977) *Biochem. J.* **161**, 247–251; Reid & Thompson (1978) *Biochem. J.* **173**, 863–868] allow the presentation, and comparison with each other, of the complete amino acid sequences of the collagen-like regions found in the A-, B- and C-chains of human subcomponent C1q. Each chain has the continuity of its collagen-like Gly-X-Y repeating triplet amino acid sequence broken. The B- and C-chains have alanine residues at positions B-9 and C-36 where glycine might be expected. The A-chain has a threonine residue at position A-39, which is located between two Gly-X-Y triplets.

Human subcomponent C1q is composed of 18 polypeptide chains (6A, 6B and 6C), each of which is approx. 200 amino acid residues long (Reid *et al.*, 1972; Reid & Porter, 1976). Previous amino acid-sequence studies have established that there is a collagen-like amino acid sequence of approx. 80 residues in each of the three types of chain present in human subcomponent C1q (Reid, 1974, 1976, 1977; Reid & Thompson, 1978). The present paper completes the sequences of the collagen-like regions of the A- and C-chains, thus allowing the comparison of these sequences with that of the B-chain (Reid & Thompson, 1978).

Materials and Methods

Automated sequence determinations, quantitative amino acid analysis, manual dansyl-Edman degradation, electrophoresis and chromatography

The procedures were performed as described previously (Reid, 1974, 1976; Reid & Thompson, 1978).

Isolation of peptides from the collagen-like region of A-chain of subcomponent C1q

Peptide T-B3b (which corresponds to residues A-39 to A-59) was isolated from a tryptic digest of the oxidized A-chain as described previously (Reid, 1974).

Peptide TH 4-4 (which corresponds to residues A-37 to A-40) was isolated from a thermolysin (crystalline, 6000 [PU]_{μmol of Tyr}^{Casein, 35°C, 280nm/mg}; Merck, Darmstadt, Germany) digest of the oxidized A-chain. Thermolysin (200 μg in 100 μl of 0.2M-NH₄HCO₃,

pH 7.9) was added to the oxidized A-chain (20mg), which was suspended in 2.0ml of 0.2M-NH₄HCO₃, pH 7.9. Digestion was carried out at 37°C for 3 h. The digest was centrifuged at 20000g for 20 min at 4°C. The supernatant was applied to a column (2.5 cm × 80 cm) of Sephadex G-25 (fine grade) that was equilibrated with 50 mM-NH₃, pH 10.5. Peptide TH 4-4 was eluted at $V/V_0 = 1.62$ and was further purified by paper chromatography in butan-1-ol/acetic acid/water/pyridine (45:9:36:30, by vol.), where it had an R_F of 0.28 relative to Phenol Red.

Isolation of peptides from the collagen-like region of the C-chain of subcomponent C1q

The collagen-like region of the C-chain corresponding to residues C-1 to C-90 was isolated by limited digestion of intact subcomponent C1q with pepsin, followed by gel filtration on Sephadex G-200 and ion-exchange chromatography on CM-cellulose as previously described (Reid, 1976).

The peptide corresponding to residues C-1 to C-90 (700 nmol) was dissolved in aq. 70% (v/v) formic acid (3.0 ml) containing CNBr (100 mg), and the mixture incubated, under N₂, for 16 h at 25°C; water (100 ml) was then added and the digest freeze-dried. The freeze-dried material was dissolved in 0.1M-acetic acid (5 ml) and applied to a column (2.5 cm × 80 cm) of Sephadex G-150 that had been equilibrated with 0.1M-acetic acid. The peptides CNBr-2 and CNBr-4 were eluted together between 386 ml and 426 ml. Peptide CNBr-3 was eluted, along with at least two other peptides, between 427 ml and 490 ml. Peptides CNBr-2, CNBr-3 and CNBr-4 were further purified by ion-exchange chromatography on CM-cellulose as

follows. The freeze-dried peptide fraction was dissolved in 10mM-sodium citrate/citric acid buffer, pH3.6, and applied to a CM-cellulose column (2cm×15cm) equilibrated with the same buffer. A linear gradient of NaCl (0–0.2M) in the starting buffer and in a total volume of 400ml was applied to the column. Peptide CNBr-3 was eluted from the CM-cellulose column before the gradient was started, whereas peptides CNBr-4 and CNBr-2 were eluted after 100ml and 150ml (respectively) of the gradient had been run. Peptides CNBr-3 and CNBr-4 were eluted at V/V_0 of 1.90 and 1.60 respectively on a column (2.5cm×80cm) of Sephadex G-50 equilibrated with 50mM-NH₃, pH10.5. The amino acid compositions of CNBr-3 and CNBr-4 are given in Table 1. The amino acid composition of CNBr-2 was reported previously (Reid, 1977).

Peptides T-3, T-4, T-5 and T-6 were purified from a tryptic digest of the pepsin-derived peptide corresponding to residues C-1 to C-90. Peptide C-1 to C-90 (300nmol) was dissolved in 200mM-NH₄HCO₃ buffer (200μl), pH8.1, and digested with trypsin (70μg) for 2.5h at 37°C. The digest was fractionated by gel

filtration on columns (2.5cm×80cm) of Sephadex G-25 and Sephadex G-50, which were equilibrated with 50mM-NH₃, pH10.5, and finally by electrophoresis at pH3.5. Peptides T-3, T-4, T-5 and T-6 were eluted at V/V_0 of 1.20, 1.02, 2.00, 1.90 respectively, on Sephadex G-25. Peptides T-3 and T-4 were eluted at V/V_0 of 1.60 and 1.18 respectively, on Sephadex G-50. Peptides T-3, T-5 and T-6 had mobilities of +0.73, +0.64 and +0.80 respectively at pH3.5 (relative to lysine = +1.0; no marker was used to define zero mobility). The amino acid compositions of peptides T-3, T-4, T-5 and T-6 are given in Table 1.

Results

Amino acid sequence of the collagen-like region of the A-chain

The tryptic peptide T-B3b (residues A-39 to A-59) had the following amino acid composition: Hyp_{1.7} Asp_{1.9} Thr_{0.9} Ser_{1.1} Glu_{3.2} Pro_{1.3} Gly_{6.7} Ile_{0.9} Leu_{1.0} Hyl_{0.9} Lys_{1.0}. This composition is identical with that published previously (Reid, 1974), except that the

Table 1. *Amino acid compositions of the tryptic and CNBr peptides used in the determination of the sequence of residues 41–94 of the C-chain of subcomponent C1q*

Compositions are given as mol of residue/mol of peptide and values less than 0.1 are omitted. All the peptides in Fig. 1 are described here with the exception of CNBr-2, which was described previously (Reid, 1977). The samples were hydrolysed at 110°C for 24h. No corrections were made for serine and threonine destruction.

Amino acid	Composition (mol/mol)					
	Tryptic peptides from residues C-42 to C-89				CNBr peptides from residues C-68 to C-90	
	T-3	T-4	T-5	T-6	CNBr-3	CNBr-4
Hyp	1.9 (2)	3.5 (4)	—	—	1.8 (2)	0.8 (1)
Asp	—	1.5 (2)	—	—	—	—
Met(O ₂)	—	2.4 (2)	—	—	—	—
Thr	—	—	—	—	—	—
Ser	—	0.3 (0)	—	—	—	—
Hse	—	—	—	—	0.5 (1)	—
Glu	2.0 (2)	3.4 (3)	—	0.9 (1)	—	3.7 (4)
Pro	1.5 (2)	3.7 (4)	—	—	0.8 (1)	0.9 (1)
Gly	6.0 (6)	8.5 (9)	—	—	2.3 (2)	3.8 (4)
Ala	—	—	—	—	—	—
Val	—	1.1 (1)	—	—	1.0 (1)	—
Ile	—	1.0 (1)	—	—	—	1.0 (1)
Leu	1.0 (1)	—	—	—	—	—
Tyr	—	—	0.8 (1)	—	—	0.9 (1)
Phe	—	—	—	—	—	0.7 (1)
His	1.3 (1)	—	—	—	—	—
Hyl	1.6 (2)	—	—	—	—	—
Lys	0.8 (1)	—	1.0 (1)	1.0 (1)	—	1.9 (2)
Arg	—	0.8 (1)	—	—	—	0.9 (1)
Total residues	17	27	2	2	7	16
Yield (%)	55	40	40	43	35	39
N-terminal residue	Gly	Asp	Tyr	Gln	Hyp	Gly
Position in the chain	43–58	59–85	86–87	88–89	68–74	75–90

hydroxyproline and proline values were previously found to be 0.3 and 2.0 respectively. Therefore the degree of hydroxylation of the proline appeared to be greater in the T-B3b fraction used in the present study than in the previous study. The entire amino acid sequence (A-39-A-59, Fig. 2) of peptide T-B3b (150nmol) was obtained, with the aid of Polybrene (2.0mg), by automated sequential Edman degradation using the 0.1M-Quadrol program as described previously (Reid & Thompson, 1978). By using this procedure, even the C-terminal lysine residue was obtained, as its bisphenylthiohydantoin derivative, in good yield at step 21. The phenylthiohydantoin derivatives were identified by t.l.c. and by amino acid analysis (Reid, 1977). T.l.c. readily differentiates between the phenylthiohydantoin derivatives of aspartic acid or asparagine and glutamic acid or glutamine, thus the identification of the charges on the Asx and Glx residues at positions A-46, A-47 and A-49 were made.

The amino acid composition of peptide Therm 4-4 was Thr_{1.0} Gly_{1.2} Ile_{0.9} Arg_{1.0} and its amino acid sequence was found, by the subtractive Edman procedure, to be Ile-Arg-Thr-Gly, which established the overlap of residues 37 to 40 of the A-chain (Fig. 2).

The results above give the complete sequence of the collagen-like region of the A-chain of subcomponent C1q when taken in conjunction with previous work (Reid, 1974, 1977).

Amino acid sequence of the collagen-like region of the C-chain

Peptides CNBr-3, CNBr-4, T-5 and T-6 were sequenced completely by either using the Beckman 890c protein sequencer or by the dansyl-Edman procedure (Fig. 1). When peptide T-3 was sequenced by using the sequencer, no amino acid residue could be found at steps 6 or 15, and it was suspected that these positions were occupied by glycosylated hydroxylysine residues. The presence of hydroxylysine at position 6 in peptide T-3 was shown by the dansyl technique. The first 15 residues of peptide T-4 were identified by automated sequence determination, which established the overlap between peptides CNBr-2 and CNBr-3.

The results shown in Fig. 1, along with previously published results (Reid, 1976, 1977), give the complete sequence of the collagen-like region of the C-chain of subcomponent C1q (Fig. 2). In a previous study (Reid, 1977) a low yield of aspartic acid was found at position 31 in a sequencer run of peptide CNBr-2. Position 31 in this peptide corresponds to position 41 in the intact C-chain (Fig. 1). It is now clear from a second sequencer run on peptide CNBr-2 and by tryptic digestion of peptide CNBr-2 that arginine is the major amino acid at position 31 in the peptide

and consequently at position 41 in the intact chain (K. B. M. Reid, unpublished work).

Discussion

The N-terminal amino acid sequence of each of the three types of chain found in human C1q is shown in Fig. 2. These sequences represent approximately half the entire amino acid sequence of subcomponent C1q, since each chain is approx. 200 residues long. The numbering of the amino acid residues in Fig. 2 is based on the B-chain sequence; and, for ease of comparison of the three chains, gaps have been left in the B- and C-chain sequences between positions 38 and 39, and in the A- and C-chain sequences between positions 41 and 45. The insertion of these gaps yields the maximum homology between the three sequences, i.e. in terms of aligning the repeating glycine residues and several 'key' residues (such as the lysine residues at position 90 and the phenylalanine residues at position 97) in each chain. The lysine residue at position 90 indicates the transition from collagen-like sequence to non-collagen-like sequence in each chain, and the bond C-terminal to the phenylalanine at position 97 is the major point of cleavage of each chain by pepsin during limited proteolysis of subcomponent C1q at pH4.5 (Reid, 1976).

In contrast with the collagens, which have been found to have strictly continuous stretches of Gly-X-Y repeating triplet sequence, the B- and C-chains have alanine residues at positions in their collagen-like sequences where glycine might be expected (positions B-9 and C-36; Fig. 2). The A-chain also has the continuity of its collagen-like sequence broken by the insertion of a threonine residue between positions 38 and 39. These breaks in the continuity of the Gly-X-Y repeating triplets of the A- and C-chains would be expected to be located $(30-33) \times 0.29$ nm i.e. 8.7nm-9.6nm (30-33 being the number of residues proposed to be in the triple helix before the break positions, and 0.29 nm being the length of the cross-section of a one-residue-long stretch of a collagen-like triple helix) from the short N-terminal non-collagen-like sequences (Fig. 2), if it is assumed that the initial glycine residue in each chain is involved in triple-helix formation. The initial glycine residues in the collagen-like regions of the A-, B- and C-chains are at positions 9, 6 and 3 respectively, and this fact, along with the location of the only interchain disulphide bonds between A-4 and B-4 and two C-4 positions (Reid & Thompson, 1978), may be of importance in the alignment of the chains to form the triple-helical structure known to be present in subcomponent C1q (Brodsky-Doyle *et al.*, 1976; Reid & Porter, 1976). There is an absolute requirement for a glycine residue in every third position in order to form a continuous collagen-type molecule

Fig. 1. Amino acid sequence of residues 39-94 of the C-chain of human subcomponent C1q

Peptides prefixed by T- and CNBr- were derived from tryptic and CNBr digests respectively. The sequence of residues C-1-C-40 and C-91-C-94 have been published previously (Reid, 1976, 1977). Fragment CNBr-4 does not end in homoserine, since it was derived from the CNBr digest of residues C-1-C-90. Sequence determination from the N-terminus by use of the Beckman 890c protein sequencer or by the dansyl procedure is shown by $\rightarrow$ and $\leftarrow$ respectively. * Denotes a hydroxylated residue. The hydroxyllysine residues at C-47 and C-56 may be glycosylated. The one-letter notation for amino acids is as follows: A, alanine; B, aspartic acid or asparagine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; X, unknown or 'other'; Y, tyrosine; Z, glutamic acid or glutamine.

	1	10	20	30	40	50	60
A-chain	E D L C R A P D G K K G E A G R P* G R R G R P* G L K G E Q G E P* G A P* G I R T G I Q - - - G L K G D Q G E P* G P S G N P G						
B-chain	E L S C T G P P A I P* G I P G I P G T P G P D G Q P* G T P* G I K G E K G L P* - G L A G D H G E F G E K G D P G I P* G D P G						
C-chain	N T G C Y G I P* G M P* G L P* G A P* G K D G Y D G L P* G P* G E P G I P* A I K - G I R - - - G P P* G Q K G E P G L P* G H K G						

	61	70	80	90	100	110
A-chain	K V G Y P* G P S G P L G A R G I K* G I K* G T P* G S P G N I K D Q P R P A P S A I R R N P P M G G N V					
B-chain	K V G P K* G P M G P K* G P P* G A P* G A P* G P K G E S G D Y K A T Q K I A F S A T R T I N V P L R					
C-chain	K D G P N G P P* G M P* G V P* G P M G I P* G E P G E E G R Y K Q K F Q S V F					

Fig. 2. N-Terminal amino acid sequences of the A-, B- and C-chains of human subcomponent C1q

The numbering of the amino acid residues in all three chains is based on the B-chain sequence. The optimal alignment for showing the maximum homology between the three chains is obtained if a gap is left between positions 38 and 39 in the B- and C-chain sequences to allow for the 'extra' threonine residue in the A-chain sequence; a triplet gap is left between residues 41-45 in the A- and C-chain sequences to allow for the 'extra' triplet in the B-chain sequence. The sequences shown are given in Reid (1974, 1976, 1977), Reid & Thompson (1978) and the present paper. * Denotes a hydroxylated residue. All except three (at positions B-50, B-65 and C-38) of the fourteen hydroxyllysine residues found in the collagen-like sequences appear to be glycosylated. The key to the one-letter notation for amino acids is given in the legend to Fig. 1.

in which glycine residues are packed, at the centre of the cylindrical triple-helical structure, in such a way that there is no room for an L-amino acid side chain (Traub & Piez, 1971). Therefore the insertion of a threonine residue between two Gly-X-Y triplets in the A-chain and the presence of an alanine at position C-36 would be expected to cause a distortion in the proposed triple helix at approx. 8.7 nm–9.6 nm from the short non-collagen-like, *N*-terminal regions. The B-chain collagen-like region has an 'extra' Gly-X-Y triplet, when compared with the collagen-like regions of the A- and C-chains, and by alignment of the chains, to give maximum homology, it appears that this 'extra' triplet may be located in the proposed distortion region, i.e. in positions B-43 to B-45 (Fig. 2). Electron-microscopic studies of subcomponent C1q have shown that the collagen-like region of the molecule is composed of six connecting strands (1.5 nm × 11.5 nm, each) which are each set at an angle to a fibril-like central portion (45 nm × 11.2 nm) (Shelton *et al.*, 1972; Brodsky-Doyle *et al.*, 1976). This apparent bending, or distortion, of the triple-helical sections of the molecule takes place 11.2 nm from the foot of the fibril-like central portion (known to be the location of the *N*-terminal ends of the chains; see Brodsky-Doyle *et al.*, 1976; Reid, 1976) which is similar to the minimum value of 8.7 nm–9.6 nm predicted from the sequence studies, to be the distance from the *N*-terminal ends of the molecule to the points of distortion in the triple-helical structures (8.7 nm–9.6 nm is a minimum value, since it accounts only for the 30–33 residues of collagen-like sequence

and not the short *N*-terminal stretch of non-collagen-like sequence). Non-collagen-like sequences of eight, five and two amino acid residues are found at the *N*-terminal ends of the A-, B- and C-chains respectively (Fig. 2).

Of the fourteen hydroxylysine residues found in the collagen-like regions of the A-, B- and C-chains, only three (at positions B-50, B-65 and C-38) were assumed not to be glycosylated, since cleavage by trypsin was obtained at these positions. The other eleven hydroxylysine residues were assumed to be glycosylated (by glucosyl–galactosyl disaccharide units) based on their resistance to digestion by trypsin and oxidation by periodic acid, along with the non-appearance of an expected *N*-terminal hydroxylysine residue during the automated sequencing of peptides.

References

- Brodsky-Doyle, B., Leonard, K. R. & Reid, K. B. M. (1976) *Biochem. J.* **159**, 279–286
- Reid, K. B. M. (1974) *Biochem. J.* **141**, 189–203
- Reid, K. B. M. (1976) *Biochem. J.* **155**, 5–17
- Reid, K. B. M. (1977) *Biochem. J.* **161**, 247–251
- 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
- Shelton, E., Yonemasu, K. & Stroud, R. M. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 65–68
- Traub, W. & Piez, K. A. (1971) *Adv. Protein Chem.* **25**, 243–352