Biochem. J. (1978) **173**, 863–868
Printed in Great Britain

Amino Acid Sequence of the *N*-Terminal 108 Amino Acid Residues of the B chain of Subcomponent C1q of the First Component of Human Complement

By KENNETH B. M. REID and EDWARD O. P. THOMPSON*
Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

(*Received* 5 *January* 1978)

The amino acid sequence of the *N*-terminal 108 residues of the B chain of subcomponent C1q of the first component of human complement was determined. The B chain has a blocked *N*-terminal amino acid, which was judged to be 5-oxopyrrolidine-2-carboxylic acid. A collagen-like region of 84 residues was found, which started at position B-6, and all of the six hydroxylysine residues and 12 hydroxyproline residues present in the chain were found in this region. Four of the six hydroxylysine residues may be glycosylated. The repeating nature of the collagen-like region is broken at position B-9, where alanine is found in a position where glycine would be expected. The exact position of the inter-chain disulphide bond joining the A and B chains of human subcomponent C1q was shown to be between residues A4 and B4.

Human subcomponent C1q is composed of six A, six B and six C polypeptide chains each approx. 200 residues long and of mol.wt. 23000–24000 (Reid *et al.*, 1972). The A and B chains are linked by a single disulphide bond to yield six A–B dimer subunits, and pairs of C chains are also linked by a single disulphide bond to yield three C–C dimer subunits (Yonemasu & Stroud, 1972; Reid, 1976; Reid & Porter, 1976). There is a region of collagen-like amino acid sequence which accounts for approx. 80 residues out of the first 90 *N*-terminal residues of each of the three chains of human subcomponent C1q. The *N*-terminal amino acid sequences of residues 1–98 of the A chain and residues 1–45 of the C chain have been published (Reid, 1974, 1976, 1977). The *N*-terminal amino acid sequence of residues 1–108 of the B chain, and the exact positioning of the A–B interchain disulphide bond, are given in the present paper.

Materials and Methods

Materials

Polybrene [poly(*NNNN'*-tetramethyl-*N*-trimethylenehexamethylenediammonium bromide, hexadimethrine bromide] was from Aldrich Chemical Co., R. N. Emanuel, Wembley, Middlesex, U.K.

Trypsin (treated with 7-amino-1-chloro-3-L-tosylamidoheptan-2-one), α-chymotrypsin and collagenase (catalogue no. CLSPA) were from Worthington Biochemical Corp., Freehold, NJ, U.S.A. Pepsin (twice crystallized) was from Sigma (London) Chemical Co., London S.W.6, U.K. Pyroglutamate aminopeptidase (EC 3.4.11.8) was from Boehringer Corp., Lewes, Sussex, U.K. Pronase was from Kaken Chemical Co., Tokyo, Japan. 'V-8 protease' from *Staphylococcus aureus* V8 was from Miles Laboratories, P.O. Box 37, Stoke Poges, Slough, Bucks., U.K. Iodo[$^{14}C_2$]acetic acid (33mCi/mmol) was from The Radiochemical Centre, Amersham, Bucks., U.K.

Isolation of the oxidized B chain of human subcomponent C1q

Human subcomponent C1q was isolated as described by Reid (1974). Oxidation of subcomponent C1q and isolation of the oxidized B chain of subcomponent C1q was performed as described by Reid *et al.* (1972).

Partial pepsin digestion of human subcomponent C1q

Partial pepsin digestion of human subcomponent C1q and fractionation of the digest, to allow the isolation of peptide P1-CMC2-OX1 (residues B1–97), was performed as described by Reid (1976).

Enzymic digestion of the oxidized B chain of subcomponent C1q

Chymotrypsin. The oxidized B chain (25mg) was suspended in 1% (w/v) NH_4HCO_3 buffer, pH7.8 (2.5ml), then bovine α-chymotrypsin (250μg) was added and digestion carried out at 37°C for 4.5h. The digest was centrifuged at 30000*g* for 20min at 4°C. Any precipitate was washed twice with half the original volume of digestion buffer and spun down as before, and the three supernatants were pooled. The peptides in the supernatants were purified by gel filtration and paper electrophoresis.

* Present address: School of Biochemistry, University of New South Wales, Kensington, N.S.W. 2033, Australia.

Collagenase. Collagenase digestion of the oxidized B chain of subcomponent C1q and fractionation of the digest was performed as described previously (Reid, 1976).

Trypsin. Digestion of the oxidized B chain of subcomponent C1q and fractionation of the digest was performed in exactly the same manner as described for the digestion of the oxidized A chain of subcomponent C1q (Reid, 1974).

'*V-8 protease*'. 'V-8 protease' (0.9mg) was added to the oxidized B chain (15.4mg) which was suspended in 50mM-NH_4HCO_3 buffer, pH7.8 (2.2ml), and digestion carried out at 37°C for 18h. The digest was fractionated in exactly the same manner as the chymotryptic digest of the B chain.

Enzymic digestion of peptides derived from the oxidized B chain of subcomponent C1q

Peptides derived from tryptic, chymotryptic, 'V-8 protease', collagenase and Pronase are prefixed by T-, CHY-, V8-, Col- and PRON- respectively.

Collagenase. Purified collagenase [400μg in 1ml of 50mM-Tris/HCl buffer (pH7.4)/5mM-$CaCl_2$/0.25mM-*N*-ethylmaleimide] was added to peptide CHY-1 (220nmol) or peptide V8-1 (300nmol) and digestion carried out at 37°C for 20h. The digests were fractionated by gel filtration and electrophoresis.

Pronase. Pronase [25μg in 400μl of 1% (w/v) NH_4HCO_3, pH7.8] was added to peptide Col 1-a (100nmol) and digestion carried out at 37°C for 3h. The digest was fractionated by electrophoresis at pH3.5.

Pyroglutamate aminopeptidase. Pyroglutamate aminopeptidase (1mg in 1ml of 65mM-sodium phosphate buffer, pH7.4) was added to 200nmol of peptide P1-CMC2-OX1 (which is composed of the *N*-terminal 97 residues of the oxidized B chain) and digestion was carried out at 37°C for 6h. The digest was then applied to a column (2.5cm×80cm) of Sephadex G-75 that was equilibrated with 50mM-NH_3, pH10.5. The pyroglutamate aminopeptidase was eluted near the void volume of the column, well ahead of the unblocked peptide.

Quantitative amino acid analysis, manual dansyl-Edman degradation, electrophoresis, gel filtration and ion-exchange chromatography

These procedures were performed as described previously (Reid, 1974, 1976).

Automated sequence determination

N-Terminal amino acid sequences were determined by automated Edman degradation in a Beckman 890C sequencer by using a 0.1M-Quadrol [*NNN'N'*-tetrakis-(2-hydroxypropyl)ethenediamine trifluoroacetate, pH9.5] programme of Brauer *et al.* (1975) as described previously (Reid, 1976). As a routine 2mg of Polybrene was added to each sample, before its application to the sequencer cup, to prevent loss of the sample during solvent washes in the sequencer programme (Klapper *et al.*, 1978).

Results

Digestion of peptide P1-CMC-Ox1 with pyroglutamate aminopeptidase

Digestion of peptide P1-CMC-Ox1 (which is composed of the *N*-terminal 97 residues of the oxidized B chain) with pyroglutamate aminopeptidase removed a 'blocked' *N*-terminal amino acid and exposed leucine as the new *N*-terminal amino acid (Fig. 1). The unblocked peptide was well separated from the enzyme and low-molecular-weight materials by gel filtration on Sephadex G-75. Approx. 150nmol of the unblocked peptide P1-CMC2-Ox1 was applied to the sequencer cup along with Polybrene (2mg). Automated sequence analysis was performed for 70 cycles. Positive identifications of the amino acid phenylthiohydantoins released were made at every cycle up to 62 (i.e. residue B-63), except for the phenylthiohydantoins of the amino acids at positions B-4, B-32, B-35, B-44, B-49, B-50, B-54, B-56 and B-59 (Fig. 1). The residues in these positions were identified as described below.

Digestion of the oxidized B chain with chymotrypsin and 'V-8 protease'

Digestion with chymotrypsin produced two major fragments, CHY-1 and CHY-2, which were derived from residues B1-89 and corresponded to residues B1-47 and B48-89 respectively (Table 1, Fig. 1). Digestion with 'V-8 protease' also produced two major fragments V8-1 and V8-2, which were derived from residues B1-85 and which corresponded to residues B1-46 and B47-85 respectively (Table 1, Fig. 1). Peptides CHY-1 and V8-1 were subjected to collagenase and Pronase digestion, as described below, to allow the isolation of peptides from the region composed of residues B1-8.

Only limited cleavage *C*-terminal to the glutamic residue at B-34 by 'V-8 protease' was observed, and the glutamic residue at B-49 appeared to be completely resistant to digestion by the enzyme.

Digestion of the oxidized B chain with trypsin

The tryptic peptides T-1–T-7 accounted for residues B1–B108 of the B chain (Table 1, Fig. 1). Hydroxylysine was taken as the *C*-terminal residue of peptide T-1, thus indicating that residue B-50 was hydroxylysine (Fig. 1). Manual sequencing of peptides T-2 and T-3 allowed positive identification of residues B-56 and B-65 respectively. Edman degradation of peptide T-4 followed by amino acid analysis at each step allowed identification of residue B-71 as hydroxylysine. Manual sequencing of peptides T-5, T-6 and T-7 confirmed and extended the sequence data obtained from fragment Col-B-P (Fig. 1).

1 10 20 30
<Glu-Leu-Ser-Cys-Thr-Gly-Pro-Hyp-Ala-Ile-Hyp-Gly-Ile-Hyp-Gly-Ile-Hyp-Gly-Thr-Pro-Gly-Pro-Asp-Gly-Gln-Hyp-Gly-Thr-Hyp-Gly-Ile-Hyl-

T-1
CHY-1
V8-1
Col 1-a
PRON-1 PRON-2

40 50 60
-Gly-Glu-Hyl-Gly-Leu-Hyp-Gly-Leu-Ala-Gly-Asp-His-Gly-Glu-Phe-Gly-Glu-Hyl-Gly-Asp-Pro-Gly-Ile-Hyp-Gly-Asn-Hyp-Gly-Lys-Val-Gly-Pro-

T-1 T-2 T-3
CHY-1 CHY-2
V8-1 V8-2
Col 1-d

70 80 90
-Hyl-Gly-Pro-Met-Gly-Pro-Hyl-Gly-Pro-Hyp-Gly-Ala-Hyp-Gly-Ala-Hyp-Gly-Pro-Hyl-Gly-Glu-Ser-Gly-Asp-Tyr-Lys-Ala-Thr-Gln-Lys-Ile-Ala-

T-3 T-4 T-5 T-6
CHY-2
V8-2
Col-B-P

100 108
-Phe-Ser-Ala-Thr-Arg-Thr-Ile-Asn-Val-Pro-Leu-Arg-

T-6 T-7
Col-B-P to C-terminal of B chain

Fig. 1. *Amino acid sequence of the N-terminal* 108 *amino acid residues of the B chain of human subcomponent C1q*

Peptides prefixed by T-, CHY-, V8-, Col- and PRON- were derived from tryptic, chymotryptic, 'V8-protease', collagenase and Pronase digests respectively. <Glu denotes a 'blocked' glutamic acid residue. The *N*-terminal amino acid of the B chain was judged to be 5-oxopyrrolidine-2-carboxylic acid, since it was readily removed by pyroglutamate aminopeptidase before sequence analysis of a peptide containing residues 1–97 of the B chain. Sequence determination from the *N*-terminus by use of the Beckman 890c protein sequencer or by the dansyl procedure is shown by ⇀ and → respectively. Residues 59 and 74 appear to be only partially hydroxylated. Residue 73 showed an unusually high glycine 'carry-over' and a low yield of proline by automated sequencer analysis, therefore the possible presence of glycine in this position cannot be completely excluded. The hydroxylysine residues 32, 35, 71 and 83 may be glycosylated.

Table 1. *Isolation and amino acid compositions of the tryptic, chymotryptic, 'V-8 protease', collagenase and Pronase peptides used in the determination of the sequence of the N-terminal* 108 *amino acid residues 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. All the peptides in Fig. 1 are described here, except peptides P1-CMC2-OX1 and Col-B-P, which have been described previously (Reid, 1976). Mobilities at pH3.5 and pH6.5, with respect to the cathode and anode and in the neutral zone, are compared relative to lysine=+1, aspartic acid=−1 and 2,4-dinitrophenyl-lysine=0. Abbreviations: Cys(O_3H), cysteic acid; Met(O_2), methionine sulphone.

Amino acid	Tryptic peptides from residues B1–B108							Chymotryptic peptides from residues B1–B89		'V-8 protease' peptides from residues B1–B85		Collagenase peptides		Pronase peptides derived from peptide Col 1-a	
	T-1	T-2	T-3	T-4	T-5	T-6	T-7	CHY-1†	CHY-2†	V8-1	V8-2	Col 1-d	Col 1-a	PRON-1	PRON-2
Cys(O_3H)	1.0	—	—	—	—	—	—	1.0	—	1.0	—	—	0.9‡	—	1.0‡
Hyp	6.5	1.6	—	1.8*	—	—	—	7.2	4.8	6.8	4.5	1.0	0.6	—	0.7
Asp	2.4	1.7	—	1.2	—	—	1.1	2.4	3.3	2.1	2.0	2.1	—	—	—
Met(O_2)	—	—	—	1.0	—	—	—	—	1.1	—	1.1	—	—	—	—
Thr	2.8	—	—	—	1.0	0.9	0.7	2.7	—	2.6	—	—	1.0	—	0.9
Ser	1.3	—	—	0.9	—	0.8	—	1.1	1.0	1.1	—	—	1.0	—	1.2
Glu	4.5	—	—	1.2	1.1	—	—	4.3	2.4	4.1	2.1	2.2	1.1	1.0	—
Pro	2.9	0.8	0.9	3.8	—	—	0.9	2.8	5.6	2.7	5.6	1.1	1.0	—	0.8
Gly	13.6	3.7	1.0	8.3	—	—	—	12.6	13.9	13.1	13.0	5.6	1.3	—	1.3
Ala	2.2	—	—	2.1	1.1	1.8	—	2.2	2.3	2.0	2.0	1.2	—	—	—
Val	—	—	0.9	—	—	—	0.9	—	1.1	—	0.8	—	—	—	—
Ile	3.8	0.8	—	—	—	1.0	0.9	3.7	1.0	3.8	1.0	1.0	—	—	—
Leu	2.5	—	—	—	—	—	0.9	3.0	—	2.9	—	1.1	1.1	1.0	—
Tyr	—	—	—	0.9	—	—	—	—	1.0	—	—	—	—	—	—
Phe	0.9	—	—	—	—	0.8	—	0.9	—	—	0.7	1.0	—	—	—
His	1.0	—	—	—	—	—	—	1.0	—	0.9	—	0.9	—	—	—
Hyl	2.7	—	0.6	2.0	—	—	—	2.1	4.0	1.9	3.5	0.9	—	—	—
Lys	—	0.9	—	1.0	1.0	—	—	—	0.9	—	0.9	—	—	—	—
Arg	—	—	—	—	—	1.0	1.0	—	—	—	—	—	—	—	—
Total residues	50	11	4	24*	4	7	7	47	42	46	39	18	8	2	6
Yield (%)	70	30	42	38	60	33	37	24	26	19	46	80	34	80	30
N-Terminal	Blocked	Asp	Val	Gly	Ala	Ile	Thr	Blocked	Gly	Blocked	Phe	Gly	Blocked	Blocked	Ser
Gel filtration in 50mM-NH_3, V_e/V_0															
Sephadex G-25	1.00	1.40	1.72	1.06	1.76	1.69	1.60	1.00	1.00	1.00	1.00	1.25	1.40	—	—
Sephadex G-50	1.09	—	—	1.48	—	—	—	1.15	1.24	1.16	1.25	1.62	1.85	—	—
Mobility															
pH3.5	—	+0.30	+0.74	+0.51	+0.74	+0.54	+0.50	—	—	+0.03	+0.30	+0.36	−0.90	−1.0	+0.01
pH6.5	—	—	+0.54	—	+0.54	+0.33	—	−0.26	−0.03	—	—	—	—	—	—
Position in the chain	1–50	51–61	62–65	66–90	91–94	95–101	102–108	1–47	48–89	1–46	47–85	39–56	1–8	1–2	3–8

* By sequence analysis three hydroxyproline residues were found in this peptide.

† Purified further by electrophoresis at pH1.9.

‡ Half-cystine was estimated as *S*-carboxymethylcysteine in these peptides.

No cleavage, by trypsin, of the bonds *C*-terminal to the hydroxylysine residues at positions B-32, B-35, B-71 and B-83 was observed.

Collagenase digestion of the oxidized B chain

Collagenase digestion of the oxidized B chain allowed the isolation of fragment Col-B-P (which corresponded to residues B81–B207) as described previously by Reid (1976) and peptide Col-1-d (Table 1, Fig. 1). Automated sequencing of fragment Col-B-P has been described (Reid, 1976). Manual sequencing of peptide Col-1-d allowed the identification of residue B-44 as histidine.

Collagenase digestion of peptides CHY-1 and V8-1

Peptide Col-1-a (Table 1, Fig. 1) was isolated from collagenase digests of either peptide CHY-1 or peptide V8-1. Pronase digestion of peptide Col-1-a yielded peptides PRON-1 and PRON-2 (Table 1). Peptide PRON-1 was found to have a blocked *N*-terminal residue and clearly was derived from residues B1 and B2. Manual sequencing of peptide PRON-2, when isolated from peptide V8-1 (in which the one half-cystine residue present was labelled with iodo-$[^{14}C_2]$acetic acid) allowed the identification of residue B-4 as *S*-carboxymethylcysteine since all the radioactivity in the peptide was found in that position.

Discussion

The region of collagen-like sequence found in the B chain starts at the glycine residue in position B-6 and finishes at the tyrosine residue at position B-89 (Fig. 1). There is one break in the repeating nature of this collagen-like sequence, at position B-9, where alanine is found where glycine might be expected. Residue B-9 is considered to be in the collagen-like sequence for two reasons: one, because there is a hydroxyproline residue at position B-8 (Fig. 1); two, because bacterial collagenase efficiently cleaves the hydroxyproline–alanine bond between residues B-8 and B-9 (Fig. 1, Table 1). The second observation was unexpected, since it indicates that the hydroxyproline–alanine bond at B-8 to B-9 is as susceptible to collagenase as the Y–glycine bonds normally split by collagenases in X-Y-glycine repeating triplet sequences. However, if the residues B-6 to B-12 are involved in the formation of a collagen-like triple helix, then the helix would have to be distorted in some way to allow for the presence of an alanine residue at position B-9.

Since it is considered that the presence of hydroxyproline may be important in stabilizing triple-helix formation (Rosebloom *et al.*, 1973; Uitto & Prockop, 1974) it is of interest that there are hydroxyproline residues in the first four 'Y' positions and three out of the last six 'Y' positions in the B-chain X-Y-glycine repeating triplet sequence. Of the remaining 18 'Y' positions, only five are hydroxyproline residues. Thus both ends of the B chain collagen-like sequence have a high content of hydroxyproline. Another interesting feature of the sequence given, which concerns the hydroxylated residues, is that of the six Y-glycine bonds, in which hydroxylysine is found in the 'Y' position, trypsin is only able to cleave two, i.e. at residues B-50 and B-65 (Fig. 1). It is known that the B chain is glycosylated, containing glucose–galactose disaccharide units linked to hydroxylysine residues (Yonemasu *et al.*, 1971; Calcott & Muller-Eberhard, 1972; Reid *et al.*, 1972); therefore it is likely that the bonds *C*-terminal to the hydroxylysine residues at positions B-32, B-35, B-71 and B-83 are resistant to trypsin because of the presence of carbohydrate. The stability of at least three of these residues to periodate treatment appears to confirm the possibility that there is carbohydrate on them (K. B. M. Reid, unpublished work).

Residues 1–95 of the A chain and 1–97 of the B chain are linked by a single disulphide bond, and residues 1–94 of the C chain form dimers which also contain a single disulphide bond (Reid, 1976). Collagenase digestion of intact subcomponent C1q yields *C*-terminal fragments from the A, B and C chains in the form of single-chain polypeptides composed of 110–120 amino acid residues (Reid *et al.*, 1972; Lowe & Reid, 1974; K. B. M. Reid, unpublished work). Each *C*-terminal fragment of 110–120 amino acid residues contains a single intrachain disulphide bond, therefore the only interchain disulphide bonds in human subcomponent C1q are located between the residues A-4 and B-4 in the A–B

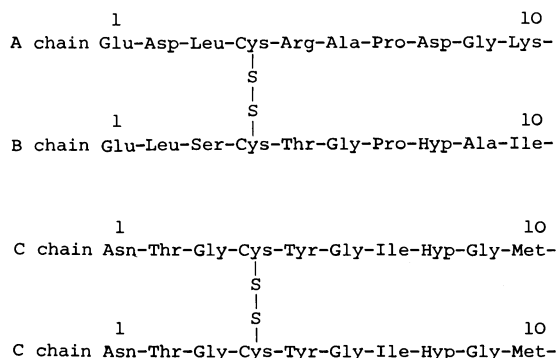

Fig. 2. *Positions of the interchain disulphide bonds in human subcomponent C1q*

The *N*-terminal amino acid sequences of the A and C chains are taken from Reid (1976). For the reasons given in the text it can be concluded that there are only two types of interchain disulphide bonds in human subcomponent C1q, i.e. an A–B type linking residues A-4 and B-4 and a C–C type linking two C-4 positions.

dimers and residues C-4 and C-4 in the C–C dimers (Fig. 2). The initial glycine residues in the collagen-like sequences of the A, B and C chains are at positions A-9, B-6 and C-3 respectively (Fig. 2), but it is not clear how the three chains might be aligned to form the triple-helical-type structure known to be present in subcomponent C1q (Brodsky-Doyle *et al.*, 1976). However, the presence of the interchain disulphide bonds between residues A-4 and B-4 and between two C-4 residues must set some restriction on the number of ways in which the alignment can be made.

The sequencing of the *N*-terminal region of the B chain was facilitated by two factors: the use of calf liver pyroglutamate aminopeptidase on the fragment composed of residues B-1–B-97; the use of Polybrene in the automated sequencing. The B chain had a blocked *N*-terminal amino acid, which was readily removed by pyroglutamate aminopeptidase in 60–70% yield. The good yield was unexpected considering the relatively large size of fragment used in the digest. There is only one arginine residue in the first 108 residues of the B chain (at B-108, Fig. 1), therefore most of the peptides prepared for automated sequencing contained no arginine. These peptides, unless modified in some way, would be rapidly washed out of the sequencer cup when the 0.1 M-Quadrol programme was used. It was found that the addition of Polybrene (as recommended by Klapper *et al.*, 1978) to peptides that contained no arginine, e.g. the fragment corresponding to residues B-2–B-97, allowed automated sequencing to be performed efficiently with the 01.M-Quadrol programme, which involves extensive solvent washes (Brauer *et al.*, 1975).

References

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

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

Klapper, D. G., Wilde, C. E. & Capra, J. D. (1978) *Anal. Biochem.* **85**, 126–131

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

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., Lowe, D. M. & Porter, R. R. (1972) *Biochem. J.* **130**, 749–763

Rosebloom, J., Harsch, M. & Jimenez, S. (1973) *Arch. Biochem. Biophys.* **158**, 478–484

Uitto, J. & Prockop, D. J. (1974) *Biochemistry* **13**, 4586–4591

Yonemasu, K. & Stroud, R. M. (1972) *Immunochemistry* **9**, 545–554

Yonemasu, K., Stroud, R. M., Niedermeier, W. & Butler, W. T. (1971) *Biochem. Biophys. Res. Commun.* **43**, 1388–1394