*Biochem. J.* (1974) **141**, 189–203
*Printed in Great Britain*



# A Collagen-Like Amino Acid Sequence in a Polypeptide Chain of Human C1q (a Subcomponent of the First Component of 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* 18 *January* 1974)

1. A partial amino acid sequence of 95 residues of the 191 residues in the oxidized A chain of human subcomponent C1q was determined. The partial nature of the sequence is because one overlapping peptide is missing in the proposed sequence, also the proof of some of the overlapping peptides depends partly on their amino acid composition and not on their complete sequence. 2. This region of the A chain contained a repeating sequence of glycine-X-Y (where X is often proline and Y is often hydroxyproline) for 78 residues. 3. The five hydroxylysine residues and the five hydroxyproline residues present in the oxidized A chain were all in these 78 residues and only in the Y position of the repeating sequence. 4. Prolonged collagenase digestion of the oxidized A chain yielded a large, apparently *C*-terminal, peptide which contained most of the non-collagenous sequences present in the chain. 5. It is concluded that there is a collagen-like region in the A chain of subcomponent C1q which constitutes most of the *N*-terminal half of the chain and that similar collagen-like regions will be found in the B and C chains.

C1q, a subcomponent of the first component of complement, has a mol.wt. of 410000 and is probably composed of six non-covalently linked subunits of approx. 65000 mol.wt. (Calcott & Müller-Eberhard, 1972; Reid *et al.*, 1972). Each of these non-covalently linked subunits contains three covalently linked polypeptide chains which differ in amino acid sequence as judged by their amino acid compositions, *N*-terminal amino acids and peptide 'mapping' (Reid *et al.*, 1972). Each of the three polypeptide chains has a high glycine content (approx. 18%) and contains the amino acids hydroxyproline and hydroxylysine, which are characteristic of collagen and basement-membrane proteins (Traub & Piez, 1971; Kefalides, 1973) and which have not been found in globular proteins. The unusual amino acid composition of subcomponent C1q and its high degree of susceptibility to collagenase (Reid *et al.*, 1972) was strong indirect evidence that there are collagen-like sequences in the polypeptide chains of subcomponent C1q. A partial amino acid sequence of 95 of the 191 residues in the oxidized A chain of human subcomponent C1q is presented in this paper. Over a region of 78 residues there is a repeating sequence of Gly-X-Y (where X is often proline, five residues out of 26; and where Y is often hydroxyproline, five residues out of 26), and all the hydroxylysine in the oxidized A chain is in this collagen-like region and is restricted to the Y position of the repeating triplet sequence.

## Experimental

### *Materials*

Phenyl isothiocyanate, *N*-ethylmorpholine, trifluoroacetic acid, amino acids and dansyl chloride were obtained from BDH Chemicals Ltd., Poole, Dorset, U.K. The chemicals used in the Edman degradation procedures were purified and redistilled before use.

Sephadex G-25, G-50 and G-75 (all fine grade) and DEAE-Sephadex A-25 (fine particle size) were obtained from Pharmacia Fine Chemicals, Uppsala, Sweden.

*Enzymes.* Carboxypeptidase A, carboxypeptidase B, α-chymotrypsin, trypsin and collagenase (CLSPA, chromatographically purified) were obtained from Worthington Biochemical Corp., Freehold, N.J., U.S.A. Leucine aminopeptidase M was obtained from Röhm G.m.b.H., D-6100 Darmstadt, West Germany.

### *Methods*

*Isolation of the oxidized A chain of human subcomponent C1q.* The oxidation of human subcomponent C1q and the isolation of the oxidized A chain of human subcomponent C1q were carried out as described by Reid *et al.* (1972). The preparation of human subcomponent C1q used for oxidation was also isolated as described by Reid *et al.* (1972) except that 55mM-sodium phosphate–1 mM-EDTA

buffer, pH7.4, was used in the DEAE-cellulose step and 230mM-sodium acetate buffer, pH5.2, was used as the equilibrating and starting buffer in the CM-cellulose chromatography.

*Succinylation (2-carboxypropionylation) of the oxidized A chain of human subcomponent C1q.* The oxidized A chain (15mg) was dissolved in 1.0M-Tris–HCl buffer, pH10.5 (3.5ml), and left at 37°C for 30min before cooling to 4°C. Succinic anhydride (220mg) was added slowly, with stirring, over a period of 1h. The pH was kept at 8.1 by the addition of 4M-NaOH. After all the succinic anhydride had been added the mixture was left at 4°C for 1h then dialysed extensively against ice-cold 5mM-$NH_3$.

*Enzymic digestion of the oxidized A chain of subcomponent C1q.* Trypsin. Trypsin was treated with L-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone to remove chymotryptic activity as described by Kostka & Carpenter (1964). The treated trypsin (400$\mu$g in 500$\mu$l of 1mM-HCl) was added to the oxidized A chain or the succinylated oxidized A chain (24mg) which was suspended in 1.95ml of 200mM-$NH_4HCO_3$ buffer, pH8.1. Digestion was carried out at 37°C for 4h. The digest was then fractionated as described below.

Collagenase. Chromatographically purified collagenase (Worthington Biochemical Corp.) was further purified by gel filtration on Sephadex G-200 as described by Peterkofsky & Diegelmann (1971). This purification, along with the inclusion of *N*-ethylmaleimide in the digestion buffer, allowed the specificity of the collagenase used to be restricted to cleavage of susceptible Y–Gly bonds in the repeating sequence $(Gly\text{-}X\text{-}Y)_n$. The purified collagenase (400$\mu$g in 750$\mu$l of 50mM-Tris–HCl buffer, pH7.4, which was 5mM with respect to $CaCl_2$ and 0.25mM with respect to *N*-ethylmaleimide) was added to the oxidized A chain (10mg) and the digestion was carried out at 37°C for 8h, to give limited digestion conditions, or carried out at 37°C for 16h to give prolonged digestion conditions. Each digest was fractionated as described below.

Chymotrypsin. Bovine $\alpha$-chymotrypsin (200$\mu$g in 250$\mu$l of 1mM-HCl) was added to the oxidized A chain (16mg) which was suspended in 1.25ml of 200mM-$NH_4HCO_3$ buffer, pH8.1. Digestion was carried out at 37°C for 4h. The digest was fractionated as described below.

*Fractionation of the enzyme digests of the oxidized A chain of subcomponent C1q.* All enzyme digests of the oxidized A chain or succinylated oxidized A chain were fractionated in the same manner. The digest was spun at 30000*g* for 20min at 4°C. Any precipitate was washed twice with one-half the original volume of digestion buffer and spun down, as before, and the three supernatants pooled. The supernatants were first subjected to gel filtration on a column (2.5cm×80cm) of Sephadex G-25 (fine grade) that was equilibrated with 50mM-$NH_3$, pH10.5. Peptides eluted in fractions which were eluted between $V/V_0 = 1.00$ and $V/V_0 = 1.40$ ($V_0$, void volume) were further fractionated on a column (2.5cm×80cm) of Sephadex G-50 (fine grade) that was equilibrated with 50mM-$NH_3$, pH10.5. Most peptide fractions were then further fractionated on a column (1.6cm×25cm) of DEAE-Sephadex A-25 equilibrated with 5mM-$NH_4HCO_3$ buffer, pH8.5, and eluted with a linear gradient (chamber 1, 250ml of 5mM-$NH_4HCO_3$, pH8.5; chamber 2, 250ml of 500mM-$NH_4HCO_3$, pH8.2) and finally with 1.0M-$NH_4HCO_3$ buffer. Further purification of peptides, where necessary, was achieved by paper chromatography and by paper electrophoresis.

*Enzymic digestion of peptides derived from the oxidized A chain of subcomponent C1q.* Trypsin and chymotrypsin. The enzyme was dissolved in 1mM-HCl as a 0.16% (w/v) solution and 2$\mu$l of trypsin or 4$\mu$l of chymotrypsin was added to 0.1$\mu$mol of peptide in 25$\mu$l of 200mM-$NH_4HCO_3$ buffer, pH8.1. Digestion was carried out at 37°C for 2h. The reaction was stopped by acidification with acetic acid and freeze-drying.

Collagenase. Purified collagenase (100$\mu$g in 200$\mu$l of 200mM-$NH_4HCO_3$ buffer, pH7.8, which was 1mM with respect to $CaCl_2$ and 0.25mM with respect to *N*-ethylmaleimide) was added to the peptide (0.2$\mu$mol) and the digestion was carried out at 37°C for 16h, after which the digest was freeze-dried before fractionation.

Carboxypeptidases A and B and leucine aminopeptidase M. Peptide (0.05$\mu$mol) was dissolved in 100mM-$NaH_2PO_4$ buffer (45$\mu$l; pH7.4) containing 0.1% (w/v) sodium dodecyl sulphate. Enzyme (25$\mu$g in 5$\mu$l of buffer for carboxypeptidases A and B and 40$\mu$g, i.e. 500munits, in 50$\mu$l of 5mM-$NaH_2PO_4$, pH7.4, for leucine aminopeptidase M) was added and the mixture incubated at 37°C for periods from 2 to 16h. The reaction was stopped by acidification with HCl followed, immediately, by freeze-drying, then the sample was analysed on the amino acid analyser.

*Electrophoresis and chromatography.* Descending paper chromatography was performed on Whatman 3MM paper in butan-1-ol–acetic acid–water–pyridine (45:9:36:30, by vol.) for 16h. In this time Phenol Red ran 34–36cm and the $R_F$ values of the peptides were measured relative to this marker. High-voltage paper electrophoresis at pH3.5 and pH6.5 was carried out as described by Crumpton & Wilkinson (1965).

Polyacrylamide-gel electrophoresis in sodium dodecyl sulphate was performed as described by Reid *et al.* (1972). Dansyl-amino acids were resolved by t.l.c. on 5cm×5cm polyamide sheets (Woods & Wang, 1967), 0.3–1.0nmol of the sample being applied to each side.

*Edman degradation.* Qualitative identification of *N*-terminal sequences was performed by the dansyl–Edman procedure as described by Gray (1967) with the modifications suggested by Hartley (1970).

Quantitative identification of *N*-terminal sequences was carried out by the subtractive method by analysing the residual peptide after release of the *N*-terminal residue as described by Dopheide & Jones (1968).

*Amino acid analysis.* Peptides were hydrolysed in twice-glass-distilled 5.7M-HCl at 110°±1°C for 22h, unless indicated otherwise, in sealed evacuated tubes as described by Moore & Stein (1963). When a peptide was hydrolysed for 24, 48 and 72h the serine, threonine and tyrosine values were extrapolated back to zero time of hydrolysis, the valine and isoleucine values were taken as found in the 72h hydrolysate and the values for the other amino acids were averaged over the three hydrolysis times. The dried hydrolysate was examined on a Locarte analyser or a modified Beckman analyser with a 0.9cm×25cm column in each case. The minimum amount of peptide applied was 5nmol and 10nmol respectively. Each run was performed by stepwise elution with sodium citrate buffers of pH3.25 and 4.25 and a sodium borate buffer of pH9.40. Under these conditions, and with a 10min period at 30°C at the beginning of each run before reverting to 48°C for the remainder of the run, all the basic, neutral and acidic amino acids encountered were eluted in the order given in the tables.

## Results

### *Amino acid composition of the oxidized A chain of subcomponent C1q*

The amino acid composition of the oxidized A chain of subcomponent C1q is shown in Table 1. No estimate of tryptophan was possible since this amino acid is destroyed by performate oxidation. A previous estimate of the tryptophan content of unoxidized human subcomponent C1q showed that, on average, there could not be more than one residue of tryptophan/23500-mol.-wt. chain (Reid *et al.*, 1972).

### *Trypsin digestion of the oxidized A chain of subcomponent C1q*

The peptides isolated from the trypsin digest of the oxidized A chain were divided into two types, collagenous and non-collagenous (Table 2). Those peptides which contained 30% of glycine and one, or both, of the amino acids hydroxyproline and hydroxylysine were designated 'collagenous'.

The yields of all the tryptic peptides isolated, except T-D3c, were good, in the range 22–48%. The low yield (6%) of peptide T-D3c is suggestive that it may have arisen from a non-trypsin-like cleavage in the A chain or from a contaminant present in the A-chain preparation. However, since the *N*-terminal sequence of peptide T-D3c (Fig. 1) and the *N*-terminal sequence of the intact oxidized A chain (Reid *et al.*, 1972) were both Glx-Asx-Leu-, it is possible that peptide T-D3c is the *N*-terminal tryptic peptide of the A chain and therefore it has been included in Table 2. The total number of amino acid residues in the peptides isolated from the tryptic digest of the oxidized A chain is very close to the theoretical number predicted from acid hydrolysis of the oxidized A chain (Table 2), the only major discrepancy being that the number of methionine sulphone residues found in the tryptic peptides is one when two were expected from the analysis of the whole chain (Table 1 and Table 2). Methionine sulphone is probably more easily missed in the amino acid-analysis procedure used than any other amino acid, since small changes in pH or temperature can cause it to run with aspartic acid. Re-analysis of all the large peptides containing

Table 1. *Amino acid composition of the oxidized A chain of human subcomponent C1q*

Samples of oxidized A chain (0.170mg) were hydrolysed for 24, 48 and 72h. Experimental details are given in the text. The results are finally expressed as the number of amino acid residues/mol of oxidized A chain on the basis that there are two residues of histidine/mol of peptide. Similar results are obtained if it is assumed that there are two residues of methionine sulphone/mol of peptide. The following abbreviations are used throughout the tables and figures: Met($O_2$), methionine sulphone; Cys($O_3$H), cysteic acid.

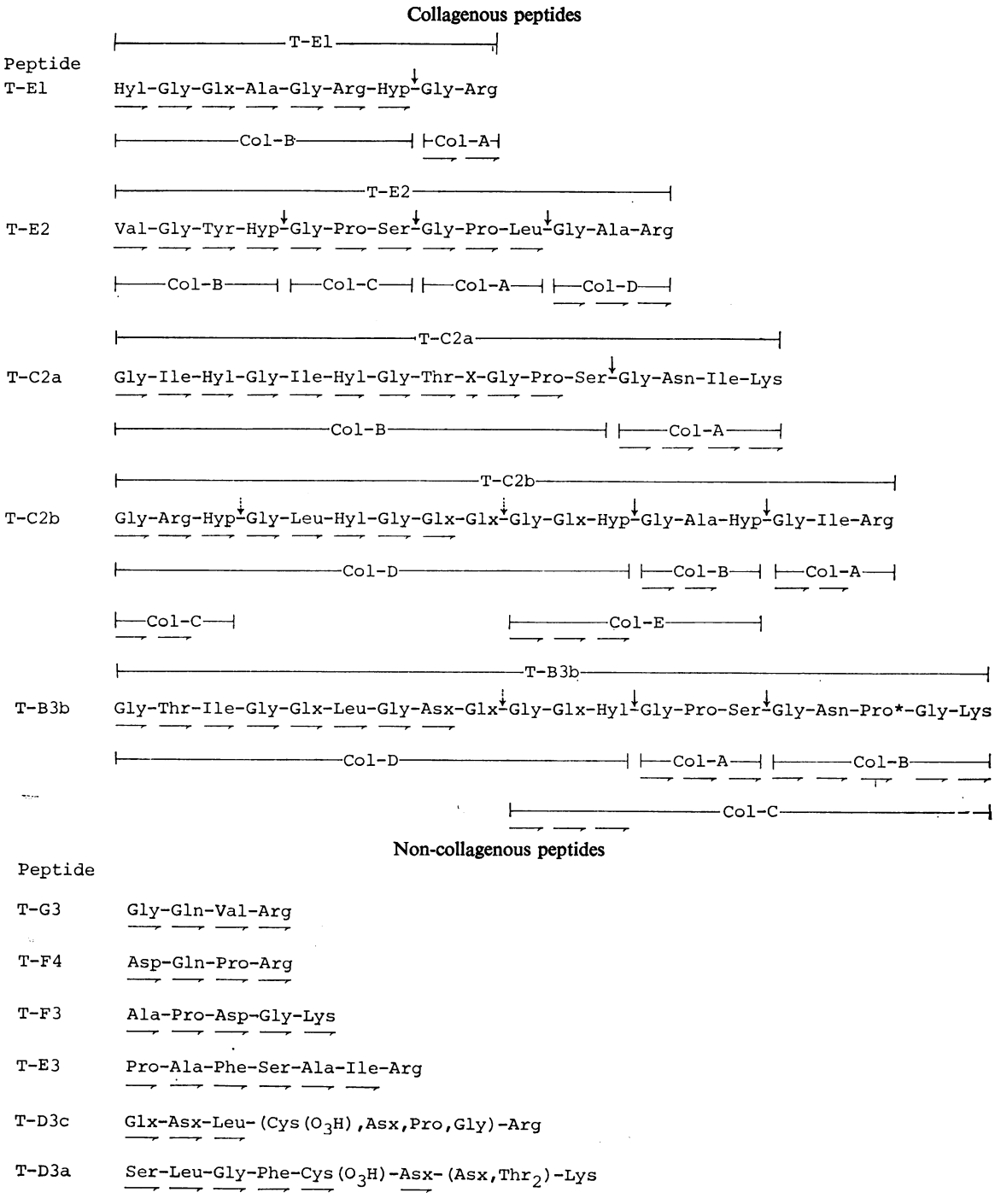
| Amino acid | Corrected amino acid composition (nmol) | Amino acid composition assuming His = 2.00 (mol/mol) | |
|---|---|---|---|
| | | Corrected | To nearest integer |
| Cys($O_3$H) | 24.3 | 3.17 | 3 |
| Hyp | 38.5 | 5.02 | 5 |
| Asp | 110.0 | 14.50 | 15 |
| Met($O_2$) | 16.0 | 2.09 | 2 |
| Thr | 59.6 | 7.63 | 8 |
| Ser | 93.4 | 12.21 | 12 |
| Glu | 162.5 | 21.35 | 21 |
| Pro | 96.2 | 12.59 | 13 |
| Gly | 262.7 | 34.50 | 35 |
| Ala | 60.0 | 7.83 | 8 |
| Val | 85.2 | 11.29 | 11 |
| Ile | 75.3 | 9.70 | 10 |
| Leu | 74.3 | 9.68 | 10 |
| Tyr | 38.4 | 5.01 | 5 |
| Phe | 61.2 | 8.00 | 8 |
| His | 15.6 | 2.00 | 2 |
| Hyl | 33.8 | 4.52 | 5 |
| Lys | 50.8 | 6.41 | 6 |
| Arg | 91.4 | 11.73 | 12 |
| Total residues | | | 191 |

Table 2. *Amino acid compositions of the peptides isolated from a trypsin digest of the oxidized A chain of human subcomponent C1q*

Full experimental details are given in the text. The results are expressed as the number of amino acid residues/mol of peptide on the basis that there is either one residue of lysine or one residue (in certain peptides two residues) of arginine/mol of peptide. Values less than 0.1 of a residue are omitted. Peptides which contained 30% of glycine and one, or both, of the amino acids hydroxyproline and hydroxylysine were designated collagenous peptides. No corrections have been made in the calculation of the percentage yields of the peptides. The yields have been calculated assuming a mol. wt. of 23500 for the intact oxidized A chain. The theoretical total number of amino acid residues in the A chain is taken from Table 1. Mobilities at pH3.5 and 6.5, with respect to migration to the cathode and anode and in the neutral zone, are compared relative to lysine = +1, aspartic acid = −1 and Dnp-lysine = 0. The $R_F$ values in paper chromatography were measured relative to Phenol Red. Elution from DEAE-Sephadex A-25 is given as the concentration (mM) of $NH_4HCO_3$ in the peak tube of the peptide fraction.

| | Collagenous peptides | | | | | Non-collagenous peptides | | | | | | | | | | Total amino acid residues | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | Combined | |
| Amino acid | T-E1 | T-E2 | T-C2a | T-C2b | T-B3b | T-H2 | T-G3 | T-F4 | T-E3 | T-F3 | T-D3c | T-D3a | T-A3a | T-C2c | T-P | Collagenous | Non-collagenous | Found | Theoretical |
| Cys($O_3$H) | — | — | — | — | — | — | — | — | — | — | 1.05 | 0.84 | — | — | 1.02 | — | 3 | 3 | 3 |
| Hyp | 1.30 | 1.16 | — | 3.10 | 0.30 | — | — | — | — | — | — | — | — | — | — | 5 | — | 5 | 5 |
| Asp | — | — | 0.91 | — | 1.77 | — | — | 1.26 | 0.96 | — | 1.75 | 1.94 | 3.04 | 2.35 | 1.18 | 3 | 12 | 15 | 15 |
| Met($O_2$) | — | — | — | — | — | — | — | — | — | — | — | — | 0.68 | — | — | — | 1 | 1 | 2 |
| Thr | — | — | 0.94 | — | 0.79 | — | — | — | — | — | — | 2.10 | 1.19 | 0.91 | 1.90 | 2 | 6 | 8 | 8 |
| Ser | — | 0.92 | 0.96 | — | 0.95 | — | — | — | — | 1.09 | — | 1.23 | 0.91 | 2.02 | 4.03 | 3 | 9 | 12 | 12 |
| Glu | 1.12 | — | — | 2.86 | 2.72 | — | 0.95 | 1.14 | — | — | 1.20 | — | 2.51 | 2.95 | 3.64 | 7 | 13 | 20 | 21 |
| Pro | — | 2.42 | 1.02 | — | 1.95 | — | — | 1.41 | 0.96 | 1.06 | 0.92 | — | 2.14 | 1.14 | 1.32 | 5 | 8 | 13 | 13 |
| Gly | 3.13 | 4.25 | 5.18 | 6.00 | 7.15 | — | 1.05 | — | 1.22 | — | 0.59 | 1.12 | 2.25 | 3.92 | 2.76 | 25 | 13 | 38 | 35 |
| Ala | 1.11 | 1.06 | — | 1.00 | — | — | — | — | 0.86 | 1.91 | — | — | — | 1.73 | 1.21 | 3 | 6 | 9 | 8 |
| Val | — | 1.03 | — | — | — | — | 0.80 | — | — | — | — | — | 1.54 | 1.20 | 3.88 | 1 | 8 | 9 | 11 |
| Ile | — | — | 3.10 | 1.02 | 0.82 | — | — | — | — | 1.00 | — | — | 0.68 | 1.28 | 1.59 | 5 | 5 | 10 | 10 |
| Leu | — | 1.06 | — | 0.96 | 0.82 | — | — | — | — | — | 1.07 | 0.97 | — | 1.39 | 2.30 | 3 | 5 | 8 | 10 |
| Tyr | — | 0.98 | — | — | — | — | — | — | — | — | — | — | 1.05 | 0.61 | 2.40 | 1 | 4 | 5 | 5 |
| Phe | — | — | — | — | — | — | — | — | — | 0.95 | — | 0.93 | 0.63 | 1.44 | 2.55 | — | 7 | 7 | 8 |
| His | — | — | — | — | — | — | — | — | — | — | — | — | 0.71 | 0.52 | — | — | 2 | 2 | 2 |
| Hyl | 1.09 | — | 2.35 | 0.99 | 1.20 | — | — | — | — | — | — | — | — | — | — | 5 | — | 5 | 5 |
| Lys | — | — | 1.00 | — | 1.00 | — | — | — | 1.00 | — | — | 1.00 | — | 1.00 | 0.66 | 2 | 4 | 6 | 6 |
| Arg | 2.00 | 1.00 | — | 2.00 | — | 2.00* | 1.00 | 1.00 | — | 1.00 | 1.00 | — | 1.00 | — | 1.00 | 5 | 8 | 13 | 12 |
| Total residues | 9 | 13 | 15 | 18 | 20 | 2 | 4 | 4 | 5 | 7 | 8 | 10 | 19 | 22 | 32 | 75 | 114 | 189 | 191 |
| % yield | 31 | 42 | 40 | 35 | 45 | 30 | 48 | 40 | 46 | 32 | 6 | 25 | 40 | 22 | 90 | | | | |
| Gel filtration, $V/V_0$ | | | | | | | | | | | | | | | | | | | |
| Sephadex G-25 | 1.41 | 1.41 | 1.20 | 1.20 | 1.09 | 2.00 | 1.64 | 1.52 | 1.41 | 1.52 | 1.33 | 1.33 | 1.00 | 1.20 | — | | | | |
| Sephadex G-50 | — | — | 1.57 | 1.57 | 1.48 | — | — | — | — | — | — | — | 1.16 | 1.57 | — | | | | |
| DEAE-Sephadex A-25 | 5 | 25 | 15 | 70 | 113 | — | — | 68 | 113 | 38 | 500 | 500 | 80 | 500 | — | | | | |
| Mobility pH3.5 | — | — | +0.68 | — | — | — | — | — | — | +0.52 | +0.04 | +0.14 | — | — | — | | | | |
| pH6.5 | — | — | — | −0.09 | — | +0.91 | +0.50 | +0.08 | +0.07 | — | — | — | — | — | — | | | | |
| $R_F$ value in paper chromatography | — | — | — | — | — | 0.16 | 0.16 | — | 0.12 | — | — | 0.13 | — | — | — | | | | |

* Sequence analysis shows two Arg-Arg sequences.




*There may be partial hydroxylation of this proline residue.

Fig. 1. *Amino acid sequences determined in the tryptic peptides of the oxidized A chain*

Sequence determination from the *N*-terminal end was by the dansyl–Edman or subtractive Edman procedure, shown by—→. ↓, indicates extensive splitting by collagenase, ↓, indicates limited splitting by collagenase. X represents an unidentified residue.

Table 3. *Amino acid compositions of peptides produced by collagenase digestion of the A chain tryptic peptides which were designated collagenous in Table 2*

Full experimental details are given in the text. Values are residues/mol of peptide; values less than 0.1 are omitted. No corrections were made in estimating the % yield of each peptide. Mobilities at pH 3.5 and 6.5, with respect to migration to the cathode and anode and in the neutral zone, are compared relative to lysine = +1, aspartic acid = −1 and Dnp-lysine = 0. The $R_F$ values in paper chromatography were measured relative to Phenol Red. No cysteic acid, methionine sulphone, phenylalanine or histidine were found in any of these collagenous peptides.

| | T-E1 | | T-E2 | | | | T-C2a | | T-C2b | | | | | T-B3b | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amino acid | Col-A | Col-B | Col-A | Col-B | Col-C | Col-D | Col-A | Col-B | Col-A | Col-B | Col-C | Col-D | Col-E | Col-A | Col-B | Col-C | Col-D |
| Hyp | — | 0.83 | — | 1.20 | — | — | — | — | — | 0.60 | 0.72 | 2.20 | 1.60 | — | 0.87 | — | — |
| Asp | — | — | — | — | — | — | 0.98 | — | — | — | — | — | — | — | 1.00 | 0.78 | 1.02 |
| Thr | — | — | — | — | — | — | — | 1.06 | — | — | — | — | — | — | — | — | 0.74 |
| Ser | — | — | — | — | 1.00 | — | — | 1.18 | — | — | — | — | — | 0.71 | — | 0.86 | — |
| Glu | — | 0.77 | — | — | — | — | — | — | — | — | — | 2.82 | 1.12 | — | — | 0.64 | 2.81 |
| Pro | — | — | 1.15 | — | 0.99 | — | — | 0.58 | — | — | — | — | — | 0.71 | — | 1.83 | — |
| Gly | 1.08 | 1.88 | 1.00 | 1.00 | 1.16 | 1.03 | 1.16 | 4.31 | 1.00 | 1.00 | 1.18 | 4.19 | 2.20 | 1.00 | 2.24 | 4.30 | 3.90 |
| Ala | — | 0.76 | — | — | — | 1.02 | — | — | — | 1.02 | — | — | 1.00 | — | — | — | — |
| Val | — | — | — | 0.93 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| Ile | — | — | — | — | — | — | 1.00 | 2.01 | 1.04 | — | — | — | — | — | — | — | 0.93 |
| Leu | — | — | 0.87 | — | — | — | — | — | — | — | — | 1.00 | — | — | — | — | 1.00 |
| Tyr | — | — | — | 0.83 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| Hyl | — | 0.66 | — | — | — | — | — | 1.88 | — | — | — | 1.11 | — | — | — | 0.91 | 0.91 |
| Lys | — | — | — | — | — | — | 0.97 | — | — | — | — | — | — | — | — | 0.81 | — |
| Arg | 1.00 | 1.00 | — | — | — | 1.00 | — | — | 1.00 | — | 1.00 | 0.91 | — | — | — | — | — |
| Total residues | 2 | 7 | 3 | 4 | 3 | 3 | 4 | 11 | 3 | 3 | 3 | 12 | 6 | 3 | 5 | 11 | 12 |
| % yield | 27 | 63 | 91 | 67 | 66 | 79 | 26 | 18 | 83 | 14 | 8 | 78 | 9 | 50 | 25 | 6 | 25 |
| Mobility | | | | | | | | | | | | | | | | | |
| pH 3.5 | — | — | — | — | — | — | — | — | — | +0.31 | — | — | — | — | — | — | — |
| pH 6.5 | — | — | +0.03 | — | +0.03 | — | — | — | — | — | +0.40 | +0.07 | — | — | +0.56 | +0.40 | −0.05 |
| $R_F$ value in paper chromatography | 0.23 | 0.09 | 0.74 | 0.66 | 0.30 | 0.25 | 0.31 | 0.12 | 0.45 | 0.24 | 0.07 | 0.03 | 0.19 | 0.29 | 0.09 | 0.09 | 0.09 |

aspartic acid yielded compositions similar to those shown in Table 2. The good agreement between the found and theoretical values for the other amino acids (Table 2) suggests that probably no trypic peptide, of any length, was missed in the isolation procedures.

All the tryptic peptides, except T-D3c and T-P, behaved satisfactorily in the purification procedures used. Peptide T-P was isolated in the very high yield of 90% and was present as a precipitate after trypsin digestion of the oxidized A chain, since it was insoluble in the digestion buffer, 200mM-$NH_4HCO_3$, pH8.1. It was soluble in 50mM-$NH_3$, pH10.5, and when applied to a column of Sephadex G-75, equilibrated with 50mM-$NH_3$, it was eluted in the void volume of the column. Polyacrylamide-gel electrophoresis of peptide T-P in buffers containing sodium dodecyl sulphate showed that the approximate molecular weight of peptide T-P was 3800, indicating that it was behaving as an aggregate on Sephadex G-75. This insolubility of peptide T-P at pH8.1 (and also at acid pH) and tendency to aggregate at pH10.5 made it difficult to work with despite the ease with which it was initially separated from the other tryptic peptides.

When sequencing the soluble peptides produced by the trypsin, and other enzyme, digests of the oxidized A chain most attention was concentrated on the collagenous peptides since these peptides were readily characterized, by nature of their unusual amino acid composition, thus allowing overlapping peptides to be quickly isolated and identified.

*Amino acid sequences in the collagenous peptides isolated from a tryptic digest of the oxidized A chain*

*Peptide T-E1.* The *N*-terminal sequence of the first seven residues of this peptide was determined by both the subtractive Edman procedure and the dansyl–Edman procedure (Fig. 1). Collagenase digestion of peptide T-E1, followed by paper chromatography, gave two peptides, T-E1-Col-A and T-E1-Col-B (Table 3 and Fig. 1). The sequence of peptide T-E1-Col-A was found by the dansyl procedure to be Gly-Arg, thus giving the complete sequence of peptide T-E1 as shown in Fig. 1.

*Peptide T-E2.* The *N*-terminal sequence of the first ten residues of this peptide was determined by the subtractive Edman procedure (Fig. 1). Digestion of peptide T-E2 with collagenase, followed by paper chromatography, gave four peptides, T-E2-Col-A, T-E2-Col-B, T-E2-Col-C and T-E2-Col-D (Fig. 1 and Table 3). The compositions of the first three peptides agree with the sequence determined for the *N*-terminal ten residues of peptide T-E2. The sequence of peptide T-E2-Col-D was found by the dansyl–Edman procedure to be Gly-Ala-Arg, thus giving the complete sequence of peptide T-E2 as shown in Fig. 1.

*Peptide T-C2a.* The *N*-terminal sequence of the first eleven residues of this peptide was determined by the subtractive Edman procedure (Fig. 1). The Edman degradation procedure went through the ninth step without any apparent loss of amino acid as judged by amino acid analysis. It is possible that this position in the sequence is occupied by an amino acid residue, or modified residue, which is either destroyed on acid hydrolysis or not fractionated from $NH_3$ during amino acid analysis. Collagenase digestion of peptide T-C2a, followed by paper chromatography, gave two peptides, T-C2a-Col-A and T-C2a-Col-B (Fig. 1 and Table 3). The sequence of peptide T-C2a-Col-A was found by the dansyl–Edman procedure to be Gly-Asx-Ile-Lys; this peptide was taken as the C-terminal sequence of peptide T-C2a. Electrophoresis at pH6.5 indicated that the Asx residue in peptide T-C2a-Col-A was asparagine, thus giving the sequence of peptide T-C2a shown in Fig. 1.

*Peptide T-C2b.* The *N*-terminal sequence of the first eight residues of this peptide was determined by the subtractive Edman procedure (Fig. 1). Digestion of peptide T-C2b with collagenase, followed by paper chromatography and electrophoresis, gave two peptides in high yield and three peptides in low yield (Fig. 1 and Table 3). The sequences of peptide T-C2b-Col-A, one of the peptides obtained in high yield, was found to be Gly-Ile-Arg by the dansyl–Edman procedure, This sequence was taken to be the *C*-terminal tripeptide of peptide T-C2b. The other peptide isolated in high yield, peptide T-C2b-Col-D, was taken to be the *N*-terminal twelve amino acids of peptide T-C2b. The three peptides, isolated in low yield, T-C2b-Col-B, T-C2b-Col-C and T-C2b-Col-E, had *N*-terminal sequences of Gly-Ala-, Gly-Arg- and Gly-Glx-Hyp- respectively, as determined by the dansyl–Edman procedure (Fig. 1). The above sequence information gives the proposed sequence of peptide T-C2b shown in Fig. 1.

*Peptide T-B3b.* The *N*-terminal sequence of the first eight residues of this peptide was determined by both the subtractive Edman procedure and the dansyl–Edman procedure (Fig. 1). Digestion of peptide T-B3b with collagenase, followed by paper chromatography and electrophoresis, gave three peptides in high yield, T-B3b-Col-A, T-B3b-Col-B and T-B3b-Col-D (Fig. 1 and Table 3). The sequence of peptide T-B3b-Col-A was found to be Gly-Pro-Ser by the dansyl–Edman procedure. The sequence of peptide T-B3b-Col-B was found to be Gly-Asx-Hyp-Gly-Lys by the dansyl–Edman procedure and this was taken as being the *C*-terminal sequence of peptide T-B3b. Electrophoresis at pH6.5 of peptide T-B3b-Col-B indicated that the Asx residue in this peptide was asparagine. Peptide T-B3b-Col-D had the known *N*-terminal sequence of peptide T-B3b of Gly-Thr-Ile-Gly-Glx-Leu-Gly-Asx-, along with the residues

$Glx_2$, Gly and Hyl in its amino acid composition. The peptide T-B3b-Col-C, which was isolated in low yield (Fig. 1 and Table 3), contained the peptide T-B3b-Col-B, which has been placed at the *C*-terminal end of peptide T-B3b. Peptide T-B3b-Col-C was found to have an *N*-terminal sequence of Gly-Glx-Hyl- by the dansyl–Edman procedure. All the sequence information obtained from the peptides produced by collagenase digestion taken with the known *N*-terminal sequence gives the proposed sequence for peptide T-B3b shown in Fig. 1.

*Amino acid sequences found in the non-collagenous peptides isolated from a tryptic digest of the oxidized A chain*

*Peptide T-G3*. The sequence of peptide T-G3 was determined by the dansyl–Edman procedure (Fig. 1). Electrophoresis at pH6.5 indicated that the Glx residue in this peptide was glutamine.

*Peptide T-F4*. The sequence of this peptide was determined by the dansyl–Edman procedure (Fig. 1). Complete digestion of this peptide with leucine aminopeptidase M showed that it contained one residue of aspartic acid and one residue of glutamine.

*Peptide T-E3*. The sequence of this peptide was determined by the subtractive Edman procedure (Fig. 1). Electrophoresis at pH6.5 indicated that this peptide contained one aspartic acid residue.

*Peptides T-F3, T-D3c and T-D3a*. The partial sequences of these peptides were determined by the subtractive Edman procedure (Fig. 1). Arginine or lysine was taken as the *C*-terminal residue in each case since the peptides were produced by trypsin digestion of the oxidized A chain.

*Peptides T-A3a, T-C2c and T-P*. So far little sequence information is available concerning these large non-collagenous peptides; however, the isolation of peptides with similar properties and amino acid compositions as peptides T-A3a, T-C2c and T-P from a tryptic digest of the precipitate produced by collagenase digestion of the A chain, would appear to confine them to the *C*-terminal half of the A chain for reasons that are outlined below.

*Amino acid sequence studies of the overlapping peptides isolated from collagenase, or chymotrypsin, digests of the oxidized A chain or from a trypsin digest of the succinylated oxidized A chain*

The amino acid compositions of the overlapping peptides are given in Table 4 and are compared with the expected number of residues in the proposed overlapping sequence found in the sequence studies carried out on the tryptic peptides of the oxidized A chain (Fig. 1). The yields of these overlapping peptides ranged from 20 to 59% (Table 4) indicating that they all represent major sequences in the A-chain preparation.

*Peptide T-S-4c (overlap of peptides T-E3 and T-E1)*. This peptide was obtained from a trypsin digest of the succinylated oxidized A chain and its amino acid composition equals, exactly, the sum of compositions of peptides T-E3 and T-E1 (Table 2 and Table 4). The *N*-terminal sequence of peptide T-S-4c was found to be Ala-Pro-Asx-Gly-Lys-Hyl-Gly- by the subtractive Edman procedure. This sequence information shows that the order of peptides T-E3 and T-E1 is as shown in Fig. 2.

*Peptide Col-C-7 (overlap of peptides T-E1 and T-H2)*. This peptide was isolated from a collagenase digest of the oxidized A chain and its sequence was found by the dansyl–Edman procedure to be Gly-Arg-Arg. From consideration of the amino acid sequences of the collagenous tryptic peptides isolated from the oxidized A chain (Fig. 1), the presence of free arginine in the trypsin digest of the oxidized A chain (Table 2) and the known specificity of collagenase (i.e. cleavage *N*-terminal to glycine in susceptible regions of the repeating sequence Gly-X-Y), it can be concluded that peptide Col-C-7 could come from the *C*-terminal end of peptide T-E1 thus linking peptide T-E1 to T-H2 (which is free arginine) as shown in Fig. 2. It is also possible that peptide Col-C-7 could be derived from the *C*-terminal end of peptide T-D3c if peptide T-D3c were to be followed by an arginine residue.

*Overlap of peptides T-E1 and T-H2 to peptide T-C2b*. No overlapping peptide from this region of the A chain has been isolated, but the alignment of the other collagenous peptides (Fig. 2) makes the tentative order peptides T-E1, T-H2 and T-C2b, shown in Fig. 2, a likely possibility. Also the absence of peptides T-C2b and T-E1 from a tryptic digest of a peptide, isolated from a collagenase digest of the A chain, that is 100 amino acid residues long [and which appears to be the *C*-terminal portion of the A chain (see below)], supports the view that peptides T-E1, T-H2 and T-C2b are in the same region of the oxidized A chain. It should also be noted that peptide T-C2b can be isolated, in good yield, from a tryptic digest of the succinylated oxidized A chain, thus establishing that peptide T-C2b must be preceded by an arginine residue in the A chain sequence.

*Peptide Col-C-2d (overlap of peptides T-C2b and T-B3b)*. This peptide was isolated from a collagenase digest of the oxidized A chain, and its amino acid composition (Table 4) establishes that the *C*-terminal end of peptide T-C2b is probably linked to the *N*-terminal end of peptide T-B3b (Fig. 2). This was confirmed by digesting peptide Col-C-2d with trypsin to produce two peptides, Col-C-2d-T-1 and Col-C-2d-T-2 (Table 5). The *N*-terminal sequence of peptide Col-C-2d-T-1 was found by the dansyl–Edman procedure to be Gly-Ala-Hyp-, and the *N*-terminal sequence of the first six amino acids of Col-C-2d-T-2 was found by the subtractive Edman procedure to

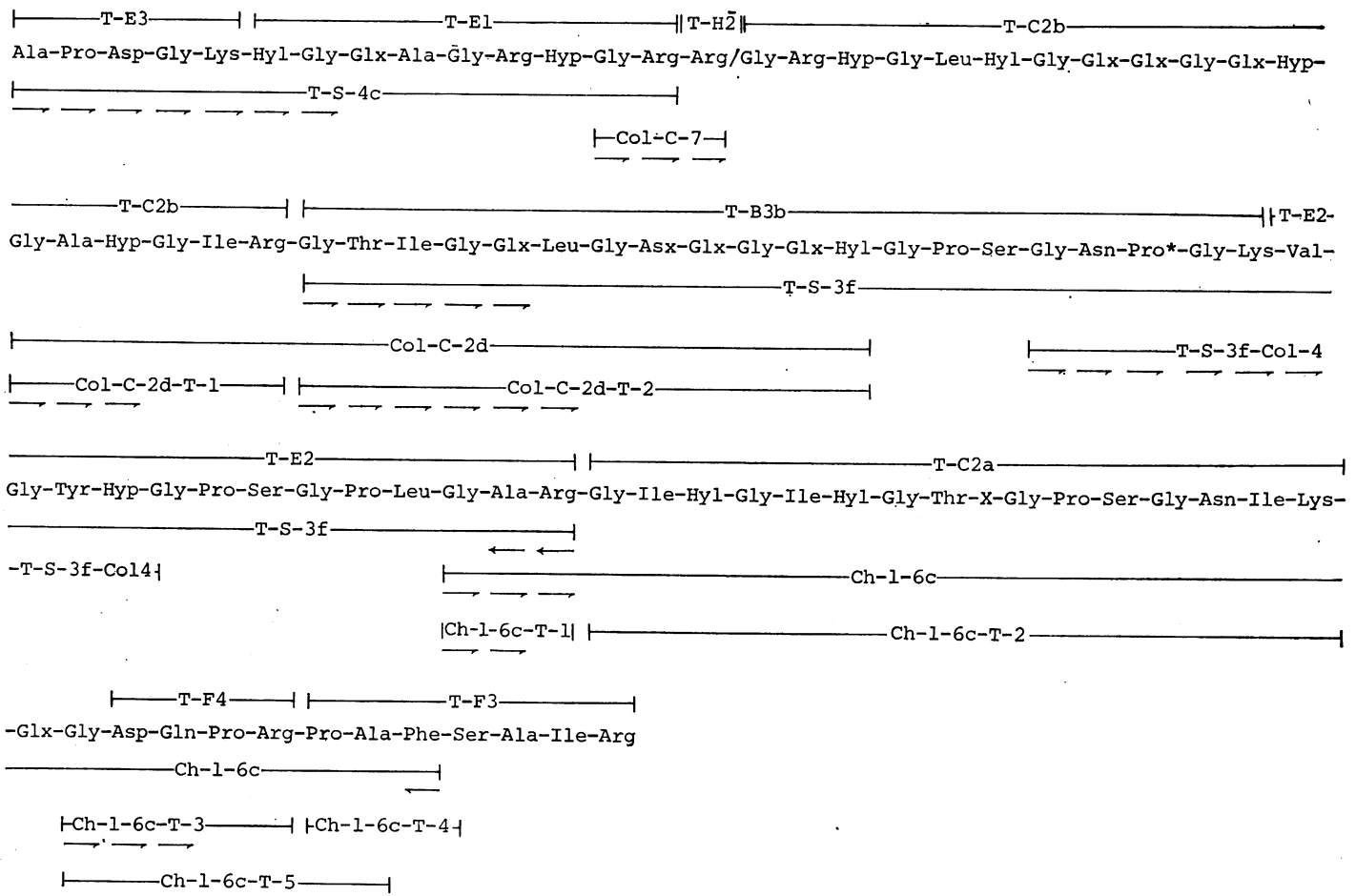
Table 4. *Amino acid compositions of overlapping peptides obtained from collagenase and chymotrypsin digests of the oxidized A chain and from a trypsin digest of the succinylated oxidized A chain*

Full experimental details are given in the text. Values are residues/mol of peptide; values less than 0.1 are omitted. The amino acid compositions of the overlapping peptides are compared with the expected number of residues in the proposed overlapping sequence found from the sequence studies performed on the tryptic peptides of the oxidized A chain. Mobilities at pH3.5 and 6.5, with respect to migration to the cathode and anode and in the neutral zone, are compared relative to lysine = +1, aspartic acid = −1 and Dnp-lysine = 0. The $R_F$ values in paper chromatography were measured relative to Phenol Red. Elution from DEAE-Sephadex A-25 is given as the concentration (mM) of $NH_4HCO_3$ in the peak tube of the peptide fraction.

| Overlapping peptide | ... T-S-4c | | Col-C-7 | | Col-C-2d | | T-S-3f | | Ch-1-6c | |
|---|---|---|---|---|---|---|---|---|---|---|
| Peptides overlapped | ... T-E3–T-E1 | | T-E1–T-H2 | | T-C2b–T-B3b | | T-B3b–T-E2 | | T-E2–T-C2a–T-F4–T-F3 | |
| Amino acid | Found | Expected | Found | Expected | Found | Expected | Found | Expected | Found | Expected |
| Cys($O_3$H) | — | — | — | — | — | — | — | — | — | — |
| Hyp | 0.97 | 1 | — | — | 0.60 | 1 | 1.60 | 1 | — | — |
| Asp | 1.04 | 1 | — | — | 1.10 | 1 | 2.19 | 2 | 2.38 | 2 |
| Met($O_2$) | — | — | — | — | — | — | — | — | — | — |
| Thr | — | — | — | — | 0.88 | 1 | 1.00 | 1 | 1.12 | 1 |
| Ser | — | — | — | — | — | — | 1.56 | 2 | 1.18 | 1 |
| Glu | 1.16 | 1 | — | — | 2.90 | 3 | 2.76 | 3 | 1.87 | 2 |
| Pro | 1.14 | 1 | — | — | — | — | 3.04 | 4 | 3.05 | 3 |
| Gly | 3.86 | 4 | 1.30 | 1 | 5.56 | 6 | 10.90 | 11 | 7.21 | 7 |
| Ala | 1.94 | 2 | — | — | 0.92 | 1 | 0.93 | 1 | 1.45 | 2 |
| Val | — | — | — | — | — | — | 1.02 | 1 | — | — |
| Ile | — | — | — | — | 1.72 | 2 | 1.45 | 1 | 2.88 | 3 |
| Leu | — | — | — | — | 1.07 | 1 | 1.55 | 2 | — | — |
| Tyr | — | — | — | — | — | — | 0.83 | 1 | — | — |
| Phe | — | — | — | — | — | — | — | — | 0.83 | 1 |
| His | — | — | — | — | — | — | — | — | — | — |
| Hyl | 0.86 | 1 | — | — | 0.91 | 1 | 1.10 | 1 | 2.15 | 2 |
| Lys | 0.93 | 1 | — | — | — | — | 0.98 | 1 | 1.21 | 1 |
| Arg | 2.00 | 2 | 2.00 | 2 | 1.00 | 1 | 1.00 | 1 | 1.87 | 2 |
| Total residues | 14 | 14 | 3 | 3 | 18 | 18 | 33 | 33 | 27 | 27 |
| % yield | 59 | | 25 | | 20 | | 35 | | 40 | |
| Gel filtration, $V/V_0$ | | | | | | | | | | |
| Sephadex G-25 | 1.11 | | — | | 1.09 | | 1.00 | | 1.06 | |
| Sephadex G-50 | 1.70 | | — | | 1.68 | | 1.59 | | 1.73 | |
| DEAE-Sephadex A-25 | — | | — | | — | | 328 | | 60 | |
| Mobility | | | | | | | | | | |
| pH3.5 | — | | — | | — | | — | | — | |
| pH6.5 | −0.28 | | — | | — | | — | | — | |
| $R_F$ value in paper chromatography | — | | 0.16 | | — | | — | | — | |

be identical with the *N*-terminal sequence of peptide T-B3b (Fig. 2), thus conclusively establishing the overlapping sequence between peptides T-C2b and T-B3b shown in Fig. 2.

*Peptide T-S-3f (overlap of peptides T-B3b and T-E2).* This peptide was obtained from a trypsin digest of the succinylated oxidized A chain and its amino acid composition was very close to the sum of the amino acid compositions of tryptic peptides T-B3b and T-E2 (Table 4). The proline value for peptide T-S-3f is low compared with that expected for proline from peptide T-B3b plus peptide T-E2, i.e. a value of 3.04 was observed when a value of 4.00 was expected. On the other hand, the hydroxyproline value for T-S-3f is higher (1.60) than expected (1.00) (see Table 4). This may indicate that there is some heterogeneity in the hydroxylation of the proline residues in this region of the A chain, although in the tryptic peptides T-E2 and T-B3b, isolated from another oxidized A chain preparation, integral values for proline and hydroxyproline were found (Table 2). The other values in peptide T-S-3f, which did not correlate exactly with what was expected from the sum of peptides T-E2 and T-B3b, were the isoleucine and leucine values which were 1.45 and 1.55 respectively in peptide T-S-3f, when values of 1.00 and 2.00 were expected (Table 4). These findings suggest that there may be some heterogeneity in amino acid




* There may be partial hydroxylation of this proline residue.

Fig. 2. *Partial amino acid sequence of the collagenous portion of the A chain of human subcomponent C1q*

Peptides prefixed by T-, Col-, Ch- were derived from trypsin, collagenase and chymotrypsin digests of the oxidized A chain respectively. Peptides prefixed by T-S- were derived from a trypsin digest of the succinylated oxidized A chain. /, indicates a position in the partial sequence where no overlapping peptide has been obtained. (In this case the reasons for the positioning of the peptides is given in the text.) Sequence determination from the *N*-terminal end was by the dansyl–Edman or subtractive Edman procedure and is shown by —→. Sequence determination from the *C*-terminal end by use of carboxypeptidase A and carboxypeptidase B is shown by ⟵ and ← respectively. X represents an unidentified residue.

Table 5. *Peptides derived by trypsin or collagenase treatment of the overlapping peptides*

Full experimental details are given in the text. Values are residues/mol of peptide; values less than 0.1 are omitted. Mobilities at pH3.5 and 6.5, with respect to migration to the cathode and anode and in the neutral zone, are compared relative to lysine = +1, aspartic acid = −1 and Dnp-lysine = 0. The $R_F$ values in paper chromatography were measured relative to Phenol Red.

| Peptide ... Amino acid | Col-C-2d-T-1 | Col-C-2d-T-2 | T-S-3f-Col-4 | Ch-1-6c-T-1 | Ch-1-6c-T-2 | Ch-1-6c-T-3 | Ch-1-6c-T-4 | Ch-1-6c-T-5 |
|---|---|---|---|---|---|---|---|---|
| Cys($O_3$H) | — | — | — | — | — | — | — | — |
| Hyp | 0.62 | — | 1.52 | — | — | — | — | — |
| Asp | — | 0.99 | 0.67 | — | 1.13 | 1.16 | — | 1.17 |
| Thr | — | 0.75 | — | — | 0.99 | — | — | — |
| Ser | — | — | — | — | 1.22 | — | — | — |
| Glu | — | 3.01 | — | — | — | 1.08 | — | 1.20 |
| Pro | — | — | 0.32 | — | 0.70 | 1.12 | 1.12 | 1.77 |
| Gly | 1.97 | 3.99 | 3.20 | 1.34 | 5.00 | 0.72 | — | 0.84 |
| Ala | 1.00 | — | — | 0.69 | — | — | 1.16 | 0.78 |
| Val | — | — | 1.09 | — | — | — | — | — |
| Ile | 0.95 | 0.84 | — | — | 2.67 | — | — | — |
| Leu | — | 1.00 | — | — | — | — | — | — |
| Tyr | — | — | 0.62 | — | — | — | — | — |
| Phe | — | — | — | — | — | — | 1.00 | — |
| His | — | — | — | — | — | — | — | — |
| Hyl | — | 1.04 | — | — | 1.74 | — | — | — |
| Lys | — | — | 1.04 | — | 1.00 | — | — | — |
| Arg | 1.00 | — | — | 1.00 | — | 1.00 | — | 1.00 |
| Total residues | 6 | 12 | 9 | 3 | 15 | 5 | 3 | 7 |
| % yield | 91 | 93 | 20 | 11 | 34 | 24 | 12 | 4 |
| Mobility pH3.5 | — | — | +0.13 | +0.90 | +0.68 | +0.54 | +0.10 | +0.38 |
| pH6.5 | — | — | — | — | — | — | — | — |
| $R_F$ value in paper chromatography | 0.28 | 0.06 | — | — | — | — | — | — |

sequence in the oxidized-A-chain preparations. This would be possible since pooled human plasma from a large number of donors was used as the starting material of each preparation subcomponent C1q. The *N*-terminal sequence of the first five amino acids of peptide T-S-3f was found by the subtractive Edman procedure to be identical with the *N*-terminal sequence of peptide T-B3b (Fig. 2). Carboxypeptidase B treatment of peptide T-S-3f gave a 20% yield of arginine plus a 7% yield of alanine, suggesting that the C-terminal sequence of peptide T-S-3f was the same as that of peptide T-E2 (Fig. 2). Collagenase digestion of peptide T-S-3f yielded a variety of peptides one of which, T-S-3f-Col-4 (Table 5), had an *N*-terminal sequence of Gly-Asx-Hyp-Gly-Lys-Val- as determined by the dansyl–Edman procedure. This sequence information is further evidence of the overlap between peptides T-B3b and T-E2 and illustrates that there is probably partial hydroxylation of the proline residue in this region of the A chain.

*Peptide Ch-1-6c (overlap of peptides T-E2, T-C2a, T-F4 and T-F3).* This peptide was isolated from a chymotryptic digest of the oxidized A chain. Digestion of peptide Ch-1-6c with trypsin yielded four peptides in high yield along with one in low yield (Table 5 and Fig. 2). Peptide Ch-1-6c-T-2 was identical with tryptic peptide T-C2a in amino acid composition (Table 2 and Table 5) and electrophoretic behaviour. The sequence of peptide Ch-1-6c-T-1 and the *N*-terminal sequence of peptide Ch-1-6c were both found by the dansyl–Edman procedure to be Gly-Ala-Arg. Peptide Ch-1-6c-T-3 appeared similar to peptide T-F4 except that it had, as judged by the dansyl procedure, an *N*-terminal glycine residue. There is one glutamic acid (or glutamine) residue not accounted for in the peptides isolated from the trypsin digest of peptide Ch-1-6c and this residue, along with the glycine residue in peptide Ch-1-6c-T-3, has been tentatively placed between tryptic peptides T-C2a and T-F4 (Fig. 2). Treatment of peptide Ch-1-6c with carboxypeptidase A gave only phenylalanine, in a yield of 60%, thus placing peptide Ch-1-6c-T-4 at the *C*-terminal end of peptide Ch-1-6c (Fig. 2). The amino acid composition of peptide Ch-1-6c-T-5, isolated in low yield (Table 5), appears to support the alignment shown in Fig. 2, i.e. it must be assumed that

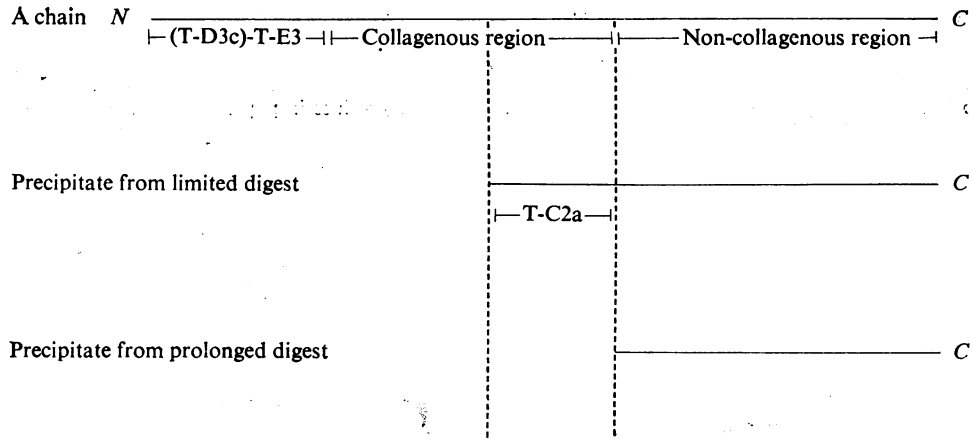



Fig. 3. *Possible alignment of the precipitates isolated from limited and prolonged collagenase digestion of the oxidized A chain*

Full experimental details are given in the text.

Table 6. *Amino acid compositions of the precipitates isolated after limited and prolonged collagenase digestion of the oxidized A chain*

The amino acid compositions of the precipitates isolated after limited and prolonged collagenase digestion of the oxidized A chain are compared with the number of residues predicted to be present in the *C*-terminal region of the oxidized A chain as estimated from the sequence studies. Full experimental details are given in the text. Arginine was taken as six residues in each case and values less than 0.1 of a residue are omitted.

| | Amino acid composition | | | Amino acid composition | |
|---|---|---|---|---|---|
| | Precipitate from limited digest | | Sum of non-collagenous residues plus peptide T-C2a minus peptides T-D3c and T-E3 | Precipitate from prolonged digest | Sum of non-collagenous residues minus peptides T-D3c and T-E3 |
| Amino acid | Digest 1 | Digest 2 | | | |
| Cys($O_3H$) | 1.90 | 2.12 | 2 | 1.82 | 2 |
| Hyp | — | — | — | — | — |
| Asp | 9.50 | 10.00 | 10 | 9.00 | 9 |
| Met($O_2$) | 1.38 | 1.97 | 1 | 2.02 | 1 |
| Thr | 5.91 | 6.30 | 7 | 4.44 | 6 |
| Ser | 9.60 | 7.49 | 10 | 4.90 | 9 |
| Glu | 13.10 | 13.45 | 12 | 11.45 | 12 |
| Pro | 7.21 | 7.12 | 7 | 6.39 | 6 |
| Gly | 16.02 | 13.95 | 16 | 10.30 | 11 |
| Ala | 3.92 | 3.74 | 5 | 3.25 | 5 |
| Val | 8.90 | 9.61 | 8 | 8.70 | 8 |
| Ile | 6.79 | 6.32 | 8 | 4.99 | 5 |
| Leu | 5.86 | 5.78 | 4 | 4.95 | 4 |
| Tyr | 3.83 | 4.98 | 4 | 3.94 | 4 |
| Phe | 7.71 | 6.98 | 7 | 6.62 | 7 |
| His | 1.68 | 1.66 | 2 | 2.02 | 2 |
| Hyl | 1.41 | 1.48 | 2 | — | — |
| Lys | 4.23 | 4.08 | 4 | 3.45 | 3 |
| Arg | 6.00 | 6.00 | 6 | 6.00 | 6 |

the *C*-terminal phenylalanine has been lost and there has been incomplete cleavage of the arginine–proline bond.

The positioning of peptide T-F3 after peptide T-F4 (Fig. 2) is based on the fact that the *N*-terminal sequence of peptide T-F3 (Fig. 1) is identical with the three *C*-terminal amino acids of peptide Ch-1-6c.

*Sequence information obtained by limited and prolonged collagenase digestion of the oxidized A chain*

The sequence *C*-terminal to peptide T-C2a appears to mark the end of the collagenous region in the A chain (Fig. 2). The remaining tryptic peptides (Table 2), except for T-D3c, which are not accounted for in the proposed sequence of residues given in Fig. 2 can all be isolated from a large peptide produced by collagenase digestion of the oxidized A chain. Prolonged collagenase digestion of the oxidized A chain yielded a number of soluble peptides plus a precipitate, which was insoluble in neutral buffers and dilute $NH_3$ but soluble, although aggregated, in dilute acetic acid. This precipitate was composed mainly of a single chain and had an apparent mol.wt. of about 13000 as estimated by polyacrylamide-gel electrophoresis run in buffers containing sodium dodecyl sulphate. Trypsin digestion of this large polypeptide yielded all the non-collagenous peptides (Table 2), except T-D3c and T-E3, i.e. the peptides thought to constitute the *N*-terminal portion of the A chain. Strong evidence that peptide T-E3 is located near the *N*-terminal end of the A chain was obtained when the oxidized A chain was examined in the protein sequenator (L. E. Mole, unpublished work). A sequence of Ala-Pro-Asp-Gly-Lys-, which is identical with the sequence of peptide T-E3, was established for positions 6–10 of the chain.

Limited collagenase digestion of the oxidized A chain gave a precipitate that appeared larger and more heterogeneous (as judged by polyacrylamide-gel electrophoresis run in buffers containing sodium dodecyl sulphate) than that obtained in the prolonged digestion. Trypsin digestion of this precipitate allowed isolation of the collagenous peptide T-C2a (but no other collagenous peptides) as well as all the non-collagenase peptides except T-D3c and T-E3 (Fig. 3). These observations point to the precipitates produced during collagenase digestion being derived from the *C*-terminal end of the A chain.

The amino acid compositions of these proposed *C*-terminal fragments isolated after prolonged and limited collagenase digestion of the oxidized A chain are shown in Table 6. The composition of the precipitate produced by (*a*) prolonged digestion and by (*b*) limited digestion are compared (Table 6) with (i) the sum of the residues in the non-collagenous peptides (Table 2) minus the residues derived from peptides T-D3c and T-E3 (Table 2) and with (ii) the sum of the residues in the non-collagenous peptides plus residues derived from peptide T-C2a (Table 2) but minus residues derived from peptides T-D3c and T-E3 (see also Fig. 3). All the values agree reasonably well except for the serine value in the analysis of the prolonged collagenase digest precipitate which was much lower than predicted. Other minor differences between the obtained and predicted values may reflect some heterogeneity in the collagenase precipitate fractions that were not subjected to rigorous purification procedures.

Further evidence that the precipitate produced by collagenase digestion of the oxidized A chain contains the *C*-terminal portion of the intact chain was obtained by digesting both the precipitate and the intact oxidized A chain with carboxypeptidase A [in a buffer containing 0.1% (w/v) sodium dodecyl sulphate to solubilize the precipitate]. The results in Table 7 show that alanine was released in a yield of 60% from both the precipitate and the intact A chain and that valine, leucine and phenylalanine were also released in comparable yields and all in the range 21–35%. The amino acids present in the serine peak

Table 7. *Amino acids released by carboxypeptidase A treatment of the oxidized A chain and of the precipitate produced by limited collagenase digestion of the oxidized A chain*

Oxidized A chain (37.4nmol) and precipitate (50nmol), produced by limited collagenase digestion of the oxidized A chain, were treated with carboxypeptidase A for 15h. The results are corrected for carboxypeptidase A incubated on its own. Full experimental details are given in the text.

| | Amount released (nmol) | | Amino acids released (%of material digested) | |
|---|---|---|---|---|
| Amino acid | A chain (corrected) | Precipitate (corrected) | A chain | Precipitate |
| $Cys(O_3H)$ | — | — | — | — |
| Hyp | — | — | — | — |
| Asp | 2.10 | 2.00 | 5.5 | 4.0 |
| $Met(O_2)$ | — | — | — | — |
| Thr | 5.99 | 8.20 | 16.6 | 16.4 |
| Ser* | 13.90 | 17.00 | 36.5 | 28.0 |
| Glu | 3.03 | 3.00 | 7.9 | 6.0 |
| Pro | — | — | — | — |
| Gly | 7.00 | 10.00 | 18.3 | 20.0 |
| Ala | 23.46 | 29.00 | 61.2 | 58.0 |
| Val | 11.20 | 10.60 | 29.4 | 21.2 |
| Ile | 4.91 | 4.70 | 12.8 | 9.4 |
| Leu | 8.60 | 13.70 | 22.5 | 27.2 |
| Tyr | 4.14 | 5.60 | 10.8 | 11.2 |
| Phe | 13.44 | 12.40 | 35.5 | 24.8 |
| His | 6.46 | 10.50 | 16.9 | 21.0 |
| Hyl | — | — | — | — |
| Lys | — | — | — | — |
| Arg | — | — | — | — |

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

(which could include glutamine and asparagine as well as serine) were also present in similar amounts in both samples. The distribution of yields, below 21% for the other amino acids, was also similar, all of which suggests that the intact oxidized A chain and the precipitate produced by collagenase digestion have a similar *C*-terminal sequence.

## Discussion

Previous chemical studies of human and rabbit subcomponent C1q (Yonemasu *et al.*, 1971; Calcott & Müller-Eberhard, 1972; Reid *et al.*, 1972) provided indirect evidence that there might be collagen-like amino acid sequences in subcomponent C1q, since it has an unusually high glycine content, contains hydroxylysine and hydroxyproline residues and disaccharide units of glucosylgalactose linked to hydroxylysine and has a great susceptibility to collagenase. It should be emphasized that this is a partial sequence in which one overlapping peptide is missing, i.e. to join peptide T-E1, T-H2 (free arginine) and peptide T-C2b (Fig. 2). Some of the overlapping peptides described depend partly on amino acid composition, and not on their complete sequence, which could possibly give rise to misleading results when dealing with a repeating sequence of the type $(Gly\text{-}X\text{-}Y\text{-})_n$.

Nevertheless there is little reason to doubt that the sequence of Fig. 2 is correct and shows that the A chain of subcomponent C1q has a repeating sequence of Gly-X-Y for 78 of its 191 residues, and that this collagen-like region is located near the *N*-terminal end of the chain. In addition to having glycine as every third residue this region of the A chain contains all the hydroxylysine and hydroxyproline in the A chain (Table 2). These hydroxylated amino acid residues are restricted to position Y of the repeating sequence Gly-X-Y as is found in vertebrate collagens (Dayhoff, 1972) and for glycosylated hydroxylysine in invertebrate collagens (Isemura *et al.*, 1973). There appears to be no simple rule for predicting when a lysine or proline residue will be hydroxylated in subcomponent C1q. However, on examination of the tripeptide sequences containing hydroxylated amino acids (Fig. 2) it can be seen that besides the hydroxylysine, or hydroxyproline, residue being in the Y position the X position is occupied by a non-acidic amino acid in nine positions out of eleven, the remaining two X positions being occupied by Glx residues that have not yet had their charge determined.

The specificity of the collagenase used in the sequence studies and the true collagen-like nature of the collagenous peptides isolated from the tryptic digest of the A chain are illustrated in Fig. 1. Cleavage was restricted entirely to bonds *N*-terminal to glycine in the repeating sequence Gly-X-Y, although some of these bonds, especially those adjacent to sequences which contained hydroxylysine residues (which may be glycosylated), were resistant to collagenase. The resistance of Y–Gly bonds joining triplets containing an acidic amino acid has been pointed out by Balian *et al.* (1971). The presence of carbohydrate on the hydroxylysine residues present in the A chain may explain the lack of digestion of Hyl-Gly bonds by collagenase. The apparent inability of trypsin to cleave Hyl–Gly bonds (Fig. 2) could also be due to the presence of carbohydrate on the hydroxylysine residues.

Previous results (Reid *et al.*, 1972) indicated that the B and C chains, the other two chains of subcomponent C1q, may also contain collagen-like regions of the same length as has been found in the A chain. Peptide 'mapping' of the B and C chains (Reid *et al.*, 1972) and initial sequence studies of these chains (K. B. M. Reid & C. P. Mercer, unpublished work) have shown that their collagen-like regions will probably differ in amino acid sequence from each other and from that of the A chain.

In electron-microscopy studies Shelton *et al.* (1972) found that subcomponent C1q appeared to be composed of six 'terminal' globular units each connected by rod-like 'connecting strands' to a large 'central subunit'. Preparations of subcomponent C1q, used in the present paper to prepare the A chain, had a similar structure in the electron microscope (E. A. Munn & A. Feinstein, unpublished work) to that described by Shelton *et al.* (1972). The 'connecting strands' were reported to have dimensions of 1.5nm × 10–13nm (15Å × 100–130Å) and therefore are strikingly similar in shape and dimensions to a stretch of triple-stranded collagen (Traub & Piez, 1971). Since the A, B and C chains are covalently linked, and therefore present in the same subunit (Reid *et al.*, 1972), it is possible that each 'connecting strand' may represent a region of triple helix comprising portions of the collagen-like sequences known to be present in each chain. The presence of a collagen-like triple helix in subcomponent C1q appears likely, as digestion of native subcomponent C1q with purified collagenase at neutral pH yields small collagenous peptides and a high-molecular-weight non-collagenous precipitate (Reid *et al.*, 1972).

The existence of collagenous and non-collagenous sequences in the same polypeptide chain is now well documented for collagen, which is synthesized in the form of procollagen (Bellamy & Bornstein, 1971; Layman *et al.*, 1971; Bornstein *et al.*, 1972). Basement-membrane proteins may also contain collagenous and non-collagenous sequences in the same polypeptide chain (Grant *et al.*, 1973; Kefalides, 1973). It has been established that procollagen has non-collagenous *N*-terminal and *C*-terminal regions. Much of the non-collagenous *N*-terminal region is

removed by proteolytic digestion after the three chains have coiled together in a triple helix (Pontz *et al.*, 1973), which leads to the speculation that the function of the non-collagenous *N*-terminal region may be to align the chains correctly and initiate helix formation. The sequence studies of the A chain of subcomponent C1q (presented here) suggest that it has a relatively short non-collagenous *N*-terminal amino acid sequence and a much longer non-collagenous *C*-terminal sequence. Although the A-chain collagen-like sequence contains the prerequisite for triple helix formation of having glycine as every third residue, it has a relatively low imino acid content (13%; Fig. 2) compared with that of vertebrate collagen partial sequences (18–27%; Dayhoff, 1972). The imino acid content of collagen is thought to play a role in stabilizing the triple helix (Traub & Piez, 1971), therefore any triple helix involving a portion of the collagen-like sequence from the A chain might be expected to be relatively unstable, unless it were stabilized by some factor other than a high imino acid content.

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 analyses and Mr. A. W. Dodds for excellent technical assistance.

## References

Balian, G., Click, E. M. & Bornstein, P. (1971) *Biochemistry* **10**, 4470–4478

Bellamy, G. & Bornstein, P. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 1138–1142

Bornstein, P., von der Mark, K., Wyke, A. W., Erhlich, H. P. & Monson, J. M. (1972) *J. Biol. Chem.* **247**, 2808–2813

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

Crumpton, M. J. & Wilkinson, J. M. (1965) *Biochem. J.* **94**, 545–556

Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, vol. 5, National Biomedical Research Foundation, Silver Spring, Md.

Dopheide, T. A. A. & Jones, W. M. (1968) *J. Biol. Chem.* **243**, 3906–3911

Grant, M. E., Schofield, J. D., Kefalides, N. A. & Prockop, D. J. (1973) *J. Biol. Chem.* **248**, 7432–7437

Gray, W. R. (1967) *Methods Enzymol.* **11**, 469–475

Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822

Isemura, M., Zahn, R. K. & Schmid, K. (1973) *Biochem. J.* **131**, 509–521

Kefalides, N. A. (1973) *Int. Rev. Connect. Tissue Res.* **6**, 63–107

Kostka, V. & Carpenter, F. H. (1964) *J. Biol. Chem.* **239**, 1799–1803

Layman, D. L., McGoodwin, E. B. & Martin, G. R. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 454–458

Moore, S. & Stein, W. H. (1963) *Methods Enzymol.* **6**, 819–831

Peterkofsky, B. & Diegelmann, R. (1971) *Biochemistry* **10**, 988–994

Pontz, B. F., Müller, P. K. & Meigel, W. N. (1973) *J. Biol. Chem.* **248**, 7558–7564

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. Nat. Acad. Sci. U.S.* **69**, 65–68

Traub, W. & Piez, K. A. (1971) *Advan. Protein Chem.* **25**, 243–352

Woods, K. R. & Wang, K.-T. (1967) *Biochim. Biophys. Acta* **133**, 369–370

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