*Biochem. J.* (1978) **173**, 373–385
*Printed in Great Britain*



# Amino Acid Sequences of α-Helical Segments from *S*-Carboxymethylkerateine-A

## COMPLETE SEQUENCE OF A TYPE-I SEGMENT

By KEITH H. GOUGH, ADAM S. INGLIS and W. GORDON CREWTHER
*Division of Protein Chemistry, CSIRO, Parkville (Melbourne), Vic. 3052, Australia*

(*Received* 2 *November* 1977)

The amino acid sequence of a type-I helical segment from the low-sulphur protein (*S*-carboxymethylkerateine-A) of wool was determined by combining automatic and manual-sequencing data. Whereas in the type-II helical segment most of the cationic groups occur in pairs, 11 of the 22 anionic residues in the sequence of the type-I segment were situated next to a second anionic residue. This suggests possible interactions between type-I and type-II helical segments in α-keratin. As observed with the sequence of a type-II helical segment a model constructed on 3.6 residues per turn of helix shows a line of hydrophobic residues along the helix, thereby supporting the physicochemical evidence that the molecule is predominantly helical and forms part of a coiled-coil structure. Examination of the sequence data by predictive methods indicates the possibility of extensive sections of α-helix interspersed with discontinuities. The molecule contains a number of regions with peptide sequences identical with those found by other workers after enzymic digestion of fractions from oxidized wool.

The two preceding papers (Hogg *et al.*, 1978; Crewther *et al.*, 1978*b*) concern the determination of the amino acid sequence of one of the segments (type II) constituting the highly helical fragment obtained from a partial chymotryptic digest of the low-sulphur fraction of *S*-carboxymethylkerateine from wool.

In the present study a segment of a different type (type I) has been isolated from the various type-I and type-II segments released by the disintegration of the helical fragments in 8M-urea (Crewther & Dowling, 1971). Its complete amino acid sequence has been determined.

## Materials and Methods

### *Enzymes*

Trypsin (1-chloro-4-phenyl-3-L-toluene-*p*-sulphonamido-2-one-treated; lot 33J790; Worthington Biochemical Corp., Freehold, NJ, U.S.A.), trypsin inhibitor (lima-bean; lot 5553t, batch 45950; Koch–Light, Colnbrook, Bucks., U.K.), thermolysin (B-grade, crystalline; lot 73326; Calbiochem, Los Angeles, CA, U.S.A.), pepsin (lot PMOEA; Worthington), carboxypeptidase A (di-isopropyl phosphorofluoridate-treated; Sigma Chemical Co., St. Louis, MO, U.S.A.) and carboxypeptidase B (di-isopropyl phosphorofluoridate-treated; Worthington) were used. Carboxypeptidase A was treated by the method of Ambler (1967), and carboxypeptidase B was re-treated with di-isopropyl phosphorofluoridate (Fraenkel-Conrat *et al.*, 1955) and dialysed against de-ionized water because the enzyme preparation was found to contain free amino acids.

Abbreviation used: dansyl, 5-dimethylaminonaphthalene-1-sulphonyl.

### *Reagents*

All reagents were of analytical grade and were used without further purification unless stated. The reagents for use with the sequenator were purified by the procedures of Edman & Begg (1967).

### *O.r.d.*

The helix content of protein in aqueous solutions was determined from o.r.d. measurements and estimation of the parameter ($b_0$) in the Moffitt–Yang (1956) equation as described by Crewther & Harrap (1967).

### *Polyacrylamide-gel electrophoresis*

Vertical slab polyacrylamide-gel electrophoresis was performed in 8M-urea buffer at pH 8.9 essentially as described by Sparrow & Crewther (1972).

### *Type-I helical segment*

Partial chymotryptic digestion of *S*-carboxymethylkerateine-A, subsequent gel filtration and ion-exchange chromatography in 8M-urea were essenti-

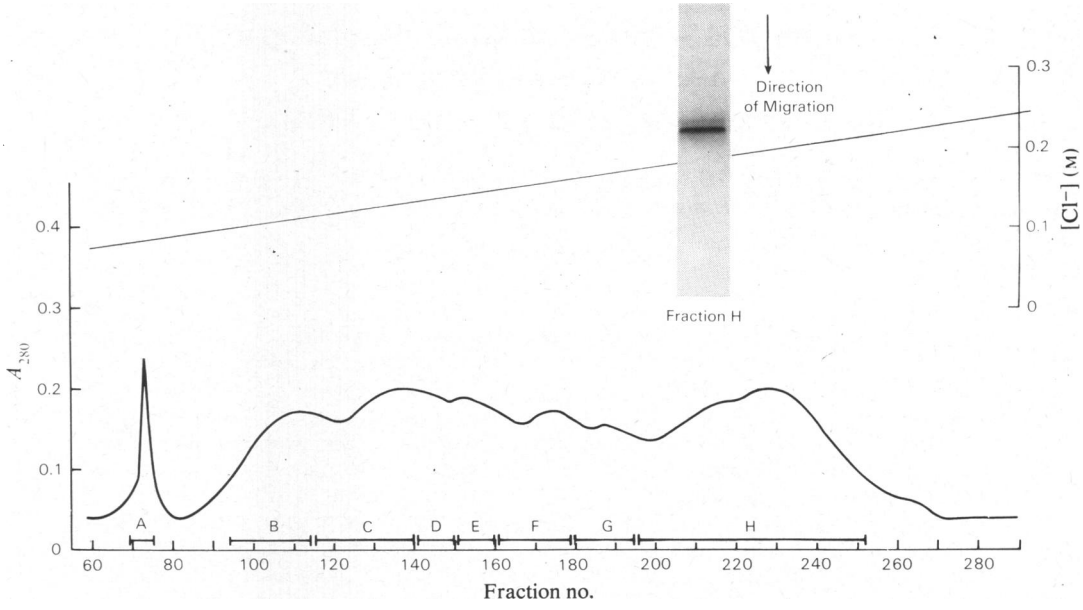


Fig. 1. *Purification of the type-I helical segment*

Fraction ChC'2 (from Crewther *et al.*, 1978) was fractionated on a column (3.0cm×115cm) of SP-Sephadex equilibrated and eluted at 30ml/h with 3.5 litres of buffer containing 8M-urea/0.01M-citric acid/1.0mM-EDTA/0.12–0.30M-KCl, pH3.7. The $A_{280}$ of fractions of volume 20ml was monitored. Fractions were pooled as shown and fraction H was selected for sequence determination on the basis of its gel-electrophoresis pattern (inset).

ally as described by Crewther & Dowling (1971). Considerable care was taken to prevent carbamoylation of amino groups by fractionation in the cold and the addition of ethanolamine to scavenge cyanate ions (Crewther *et al.*, 1978*b*; L. G. Sparrow, personal communication).

A fraction (fraction ChC'2; Crewther *et al.*, 1978) rich in one of the type-I helical segments was further purified by ion-exchange chromatography on SP (sulphopropyl)-Sephadex in 8M-urea/0.01M-citric acid/1mM-EDTA buffer at pH3.7 with a linear (0.12–0.30M) KCl gradient. Most of the protein was eluted in the first 3 litres of eluate and eight fractions (A–H) were obtained (Fig. 1). The fractionation was monitored by polyacrylamide-gel electrophoresis and by $A_{280}$. Fraction H, which gave a single main band on gel electrophoresis (Fig. 1), was selected for sequence determination. It was exhaustively dialysed against deionized water at 2°C to remove urea and salt, freeze-dried and stored at 2°C.

*Amino acid analysis*

Samples of the complete helical segment (approx. 1mg) and of peptides derived from it (10–200nmol) were hydrolysed *in vacuo* at 106°C for 24h in 0.5ml of constant-boiling HCl containing 50$\mu$l of 0.1M-phenol. The hydrolysates were dried *in vacuo* and analysed on a Beckman model 120C automatic amino acid analyser equipped with a single column. No corrections have been made for the destruction of labile amino acids during hydrolysis.

*Proteolysis of the helical segment*

Tryptic peptides were obtained by dissolving 50mg (approx. 4$\mu$mol) of the helical segment in 5ml of 0.5M-$NH_4HCO_3$, adding 0.5mg of treated trypsin and incubating at 37°C for 3h. Digestion was stopped by addition of 0.5mg of trypsin inhibitor and incubating for a further 30min. The buffer was removed by rotary evaporation with the addition of a few drops of octan-1-ol to prevent frothing. The resultant peptides were dissolved in 3ml of 5mM-$NH_4HCO_3$ at pH8.0 and separated by ion-exchange chromatography on a column (0.8cm×20cm) of DEAE-cellulose (see Fig. 2).

Thermolytic peptides were prepared by digesting 50mg of the helical segment with 0.5mg of thermolysin in 10ml of 0.1M-*N*-ethylmorpholine, adjusted to pH8.0 with acetic acid, for 2.75h at 40°C. The digest was freeze-dried and dissolved in 5ml of 5mM-$NH_4HCO_3$ for application to a column (0.75cm×17cm) of DEAE-cellulose (Fig. 3).

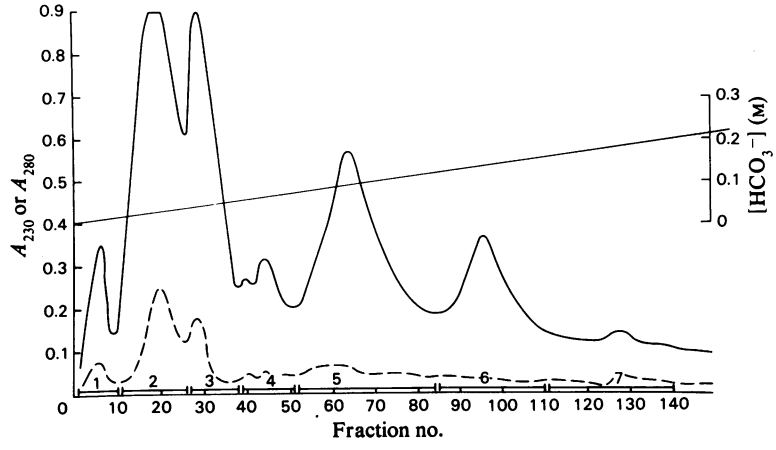


Fig. 2. *Elution profile of the tryptic digest on a column* (0.8 cm × 20 cm) *DEAE-cellulose*
A buffer of $NH_4HCO_3$ (0.005 to 0.3 M; 350 ml of each) was used at a flow rate of 7 ml/h; 3.5 ml fractions were collected and pooled as shown. The eluate was monitored at $A_{230}$ (——) and $A_{280}$ (----).

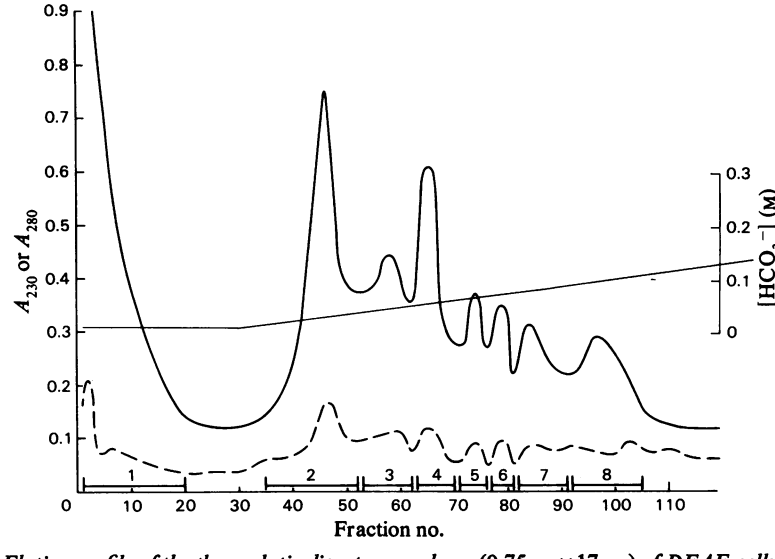


Fig. 3. *Elution profile of the thermolytic digest on a column* (0.75 cm × 17 cm) *of DEAE-cellulose*
The buffer was $NH_4HCO_3$ (0.005 to 0.30 M; 1000 ml of each); flow rate was 10.5 ml/h; the fraction size was 7.5 ml. The eluate monitored at $A_{230}$ (——) and $A_{280}$ (----).

Peptic peptides were obtained by digesting 90 mg (approx. 7.5 μmol) of the helical segment with 0.9 mg of pepsin in 9 ml of 5% formic acid at 37°C for 2 h. The digest was adjusted to pH 8.0 with aq. 8 M-$NH_3$ and freeze-dried. The resultant peptides were dissolved in 5 ml of 0.01 M-$NH_4HCO_3$ at pH 8.1 for loading on a column (1.4 cm × 23 cm) of DEAE-cellulose (Fig. 4).

*Peptide fractionation*

Peptides obtained by proteolysis of the helical segment were first fractionated by ion-exchange chromatography on DEAE-cellulose at pH 8.0 with $NH_4HCO_3$ both as buffer salt and to provide the gradient. Usually both $A_{230}$ and $A_{280}$ of the eluate were monitored. The peptides obtained in this way were further purified by high-voltage paper electro-

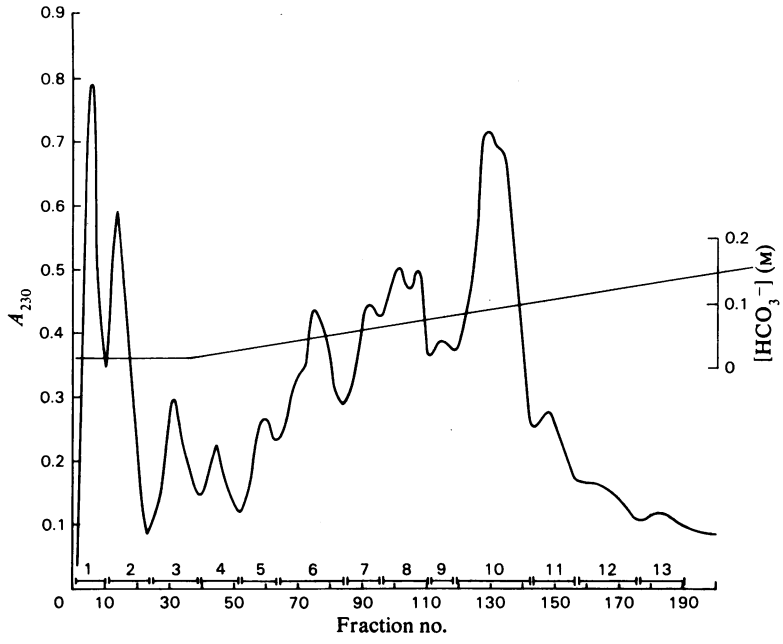


Fig. 4. *Elution profile of the peptic digest on a column* (1.4 cm × 23 cm) *of DEAE-cellulose*
A buffer of $NH_4HCO_3$ (0.01 M to 0.3 M; 1000 ml of each) was used at a flow rate of 10 ml/h; 5 ml fractions were collected. The eluate was only monitored at $A_{230}$.

phoresis on Whatman 3MM paper in toluene- or Varsol-cooled tanks (Ambler, 1963). The peptides were detected on side strips by spraying with 0.25% (w/v) ninhydrin, 2% (v/v) collidine in 95% ethanol and drying in a stream of air at about 60°C. Sometimes specific stains were used (Easley, 1965). Paper strips containing the peptides were excised, the peptides were eluted with aq. 0.1 M-$NH_3$ and the solutions were freeze-dried.

*Sequence determination*

The amino acid sequence of the first 45 residues was determined in a sequenator built with the collaboration of the late Dr. P. Edman and by using the procedures previously described (Edman & Begg, 1967). The 3-phenyl-2-thiohydantoins were identified by t.l.c. (Inglis & Nicholls, 1973) or by amino acid analysis after hydrolysis (Inglis *et al.*, 1971) when appropriate.

Manual sequencing of peptides was performed by the Edman method with dansylation essentially as described by Hartley (1970). The dansyl-amino acids were identified by t.l.c. on polyamide coated sheets (15 cm × 15 cm) by the methods of Woods & Wang 1967). A third solvent, ethyl acetate/methanol/acetic acid (20:1:1, by vol.) (Hartley, 1970), was used to distinguish between dansyl-threonine and dansyl-serine and between dansyl-glutamate and dansyl-aspartate. To identify dansyl-histidine a fourth solvent, $Na_3PO_4$/ethanol, run, after and in the same dimension as the third solvent, was used (Hartley, 1970).

*Amides*

The glutamine and asparagine residues in the first 45 residues were identified from the sequenator analysis. Other amide residues were identified either by subjecting peptides to sequenator analysis (Crewther & Inglis, 1975) or, if insufficient material (below 200 nmol) was available, by using the electrophoretic mobility relative to aspartic acid, at pH 6.5 (Offord, 1966).

*Carboxypeptidase digestion*

Approx. 100 nmol of the complete segment or about 30 nmol of a constituent peptide was dissolved in 200 $\mu$l of 0.1 M-*N*-ethylmorpholine adjusted to pH 8.0 with acetic acid. To this was added 20 $\mu$l of a 0.07% solution of carboxypeptidase A and 2 $\mu$l of a suspension of carboxypeptidase B, and the mixture was incubated at 37°C. Samples were taken at suitable intervals, acidified with acetic acid, freeze-dried and analysed on an automatic amino acid

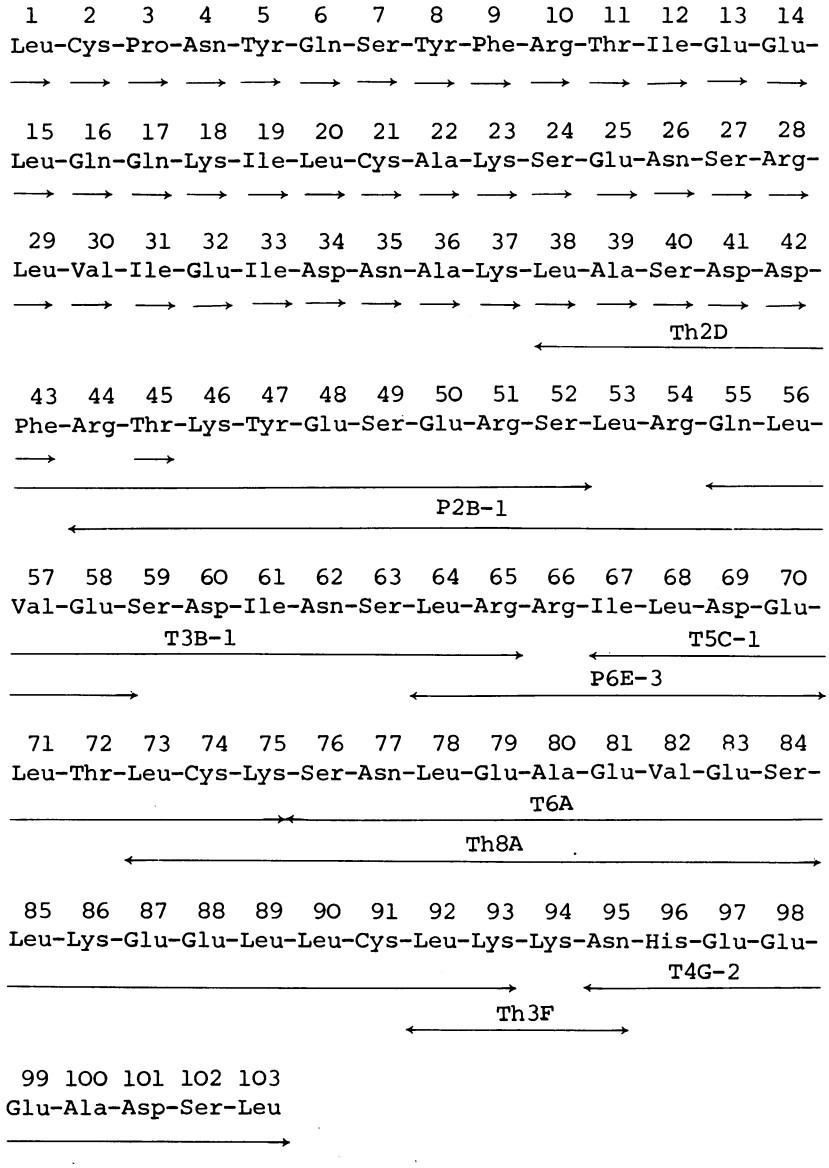


Fig. 5. *Complete amino acid sequence of a type-I helical segment*

The residues determined on the whole molecule by the automated sequenator are shown by arrows (⟶). Tryptic thermolytic and peptic peptides are identified by the prefix T, Th and P respectively and only the peptides necessary to reconstruct the sequence are shown.

analyser. In this way leucine was positively identified as the *C*-terminal residue of the whole segment.

## Results

Residues 1–45 of the helical segment were identified by using the automatic sequenator (Fig. 5), except for residue 44 (arginine), which could be unequivocally identified from the analysis of six peptides spanning arginine (44) (Tables 1–3). The data from the sequenator were confirmed throughout by analysis of several peptides from thermolytic (Table 1), tryptic (Table 2) and peptic digestions (Table 3). There was evidence from the duplicate sequenator analysis for substitution of both threonine and alanine for serine at residue 40.

The amino acid sequence of residues 1–45

Table 1. *Details of the thermolytic peptides*

→, — and ←, Residues determined by the sequenator, manual sequencing and use of carboxypeptidase respectively.

| Residues | Peptide | Analysis (residues/molecule of peptide) | Sequence |
|---|---|---|---|
| 1-8 | 4D | CmCys(0.9) Asp(1.0) Ser(0.9) Glu(1.1) Pro(1.0) Leu(0.9) Tyr(1.9) | Leu-CmCys-Pro-Asn-Tyr-Gln-Ser-Tyr |
| 1-5 | 3B | CmCys(0.8) Asp(1.2) Pro(1.0) Leu(1.0) Tyr(1.1) | Leu-CmCys-Pro-Asn-Tyr |
| 9-11 | 1D-2 | Arg(0.9) Thr(0.9) Phe(1.0) | Phe-Arg-Thr |
| 12-14 | 7A | Glu(2.2) Ile(1.0) | Ile-Glu-Glu |
| 20-21 | 2A | CmCys(0.9) Leu(1.0) | Leu-CmCys |
| 22-29 | 1B-3 | Lys(1.0) Arg(1.1) Asp(1.0) Ser(1.5) Glu(1.2) Ala(1.0) Leu(1.0) | Ala-Lys-Ser-Glu-Asn-Ser-Arg-Leu |
| 22-28 | 1B-4 | Lys(1.0) Arg(1.0) Asp(1.0) Ser(1.6) Glu(1.0) Ala(1.0) | Ala-Lys-Ser-Glu-Asn-Ser-Arg |
| 30-37 | 1A-2A-1 | Lys(1.3) Asp(2.2) Glu(1.2) Ala(1.0) Val(0.9) Ile(2.0) | Val-Ile-Glu-Ile-Asp-Asn-Ala-Lys |
| 33-37 | 1A-2B-2 | Lys(1.2) Asp(2.0) Ala(1.0) Ile(0.8) | Ile-Asp-Asn-Ala-Lys |
| 38-52 | 2D | Lys(1.0) Arg(2.0) Asp(1.8) Thr(1.1) Ser(1.9) Glu(1.9) Ala(1.0) Leu(1.0) Tyr(0.9) Phe(1.0) | Leu-Ala-Ser-Asp-Asp-Phe-Arg-Thr<br>Lys-Tyr-Glu-Ser-Glu-Arg-Ser |
| 53-55 | 1D-2 | Arg(1.0) Glu(1.0) Leu(1.0) | Leu-Arg-Gln |
| 56-63 | 5A | Asp(2.3) Ser(1.4) Glu(1.1) Val(0.9) Ile(0.9) Leu(1.0) | Leu-Val-Glu-Ser-Asp-Ile-Asn-Ser |
| 67-70 | 7B | Asp(1.1) Glu(1.1) Ile(0.9) Leu(1.0) | Ile-Leu-Asp-Glu |
| 71-72 | 1A-1C | Thr(0.9) Leu(1.2) | Leu-Thr |
| 73-84 | 8A | Lys(1.0) CmCys(0.7) Asp(1.1) Ser(1.5) Glu(2.7) Ala(1.0) Val(1.0) Leu(1.8) | Leu-CmCys-Lys-Ser-Asn-Leu-Glu-Ala-<br>Glu-Val-Glu-Ser |
| 85-89 | 2C-1 | Lys(1.2) Asp(0.9) Glu(3.6) Ile(1.0) | Ile/Leu-Lys-Glu-Glu-Leu |
| 90-91 | 2A | CmCys(0.9) Leu(1.3) | Leu-CmCys |
| 92-103 | 3F | Lys(1.9) His(1.0) Asp(2.0) Ser(0.9) Glu(3.1) Ala(1.1) Leu(1.6) | Leu-Lys-Lys-Asx- -Glx-Glx-Glx-<br>Ala-Asx-Ser-Leu |

Table 2. *Details of the tryptic peptides*

→, — and ←, Residues determined by the sequenator, manual sequencing and carboxypeptidases respectively.

| Residues | Peptide | Analysis (residues/molecule of peptide) | Sequence |
|---|---|---|---|
| 1-10 | 3A | Arg(1.2) CmCys(0.8) Asp(1.7) Ser(1.3) Glu(1.8) Pro(1.0) Leu(1.5) Tyr(1.9) Phe(1.0) | Leu-CmCys-Pro-Asn-Tyr-Gln-Ser-Tyr-Phe-Arg |
| 11-18 | 3C-1 | Lys(1.0) Thr(0.8) Glu(3.8) Ile(0.9) Leu(1.2) | Thr-Ile-Glu-Glu-Leu-Gln-Gln-Lys |
| 19-23 | 2D-1 | Lys(1.0) CmCys(1.0) Ala(0.9) Ile(0.9) Leu(1.1) | Ile-Leu-CmCys-Ala-Lys |
| 19-23 | 1C-1 | Lys(0.8) Asp(1.0) Ala(1.0) Ile(1.0) Leu(1.0) | Ile-Leu-Ala-Asp-Lys |
| 24-28 | 2D-1 | Arg(0.9) Asp(1.5) Ser(1.2) Glu(1.1) | Ser-Glu-Asn-Ser-Arg |
| 29-37 | 2A-2 | Lys(0.9) Asp(2.1) Glu(1.2) Ala(1.1) Val(0.9) Ile(1.5) Leu(1.1) | Leu-Val-Ile-Glu-Ile-Asp-Asn-Ala-Lys |
| 38-44 | 4 | Arg(1.0) Asp(2.1) Ser(0.8) Glu(0.7) Ala(1.2) Leu(1.3) Phe(0.9) | Leu-Ala-Ser-Asp-Asp-Phe-Arg |
| 45-46 | 1E | Lys(1.1) Thr(0.8) | Thr-Lys |
| 45-51 | 2E-2 | Lys(1.0) Arg(1.0) Thr(0.8) Ser(0.8) Glu(1.9) Tyr(1.0) | Thr-Lys-Tyr-Glu-Ser-Glu-Arg |
| 52-54 | 1D-1 | Arg(1.0) Ser(0.9) Leu(1.1) | Ser-Leu-Arg |
| 55-65 | 3B-1 | Arg(1.1) Asp(1.8) Ser(1.6) Glu(2.1) Val(0.9) Ile(0.9) Leu(2.1) | Gln-Leu-Val-Glu-Ser-Asp-Ile-Asn-Ser-Leu-Arg |
| 66 | 1F-1 | Arg | Arg |
| 67-75 | 5C-1 | Lys(1.0) CmCys(0.8) Asp(1.4) Thr(1.0) Glu(1.7) Ile(0.5) Leu(3.1) | Ile-Leu-Asp-Glu-Leu-Thr-Leu-CmCys-Lys |
| 76-93 | 6A | Lys(2.0) CmCys(0.5) Asp(1.0) Ser(1.7) Glu(4.6) Ala(1.0) Val(0.9) Leu(4.5) | Ser-Asn-Leu-Glu-Ala-Glu-Val-Glu-Ser-Leu(-$Lys_2$,$Cys_1$,$Glu_2$,$Leu_3$) |
| 76-86 | 6A-1 | Lys(1.0) Asp(1.0) Ser(1.5) Glu(3.1) Ala(1.1) Val(0.9) Leu(2.4) | Ser-Asn-Leu-Glu-Ala-Glu-Val-Glu-Ser-Leu-Lys |
| 87-93 | 6A-2 | Lys(0.9) CmCys(0.6) Glu(2.4) Leu(2.8) | Glu-Glu-Leu-Leu-CmCys-Leu-Lys |
| 95-103 | 4G-2 | His(1.0) Asp(2.0) Ser(0.9) Glu(2.9) Ala(1.0) Leu(1.0) | Asn-His-Glu-Glu-Glu-Ala-Asp-Ser-Leu |

obtained from the sequenator analyses can be extended progressively with good overlaps by using the amino acid sequences for peptides Th2D (38–52), P2B-1 (44–58), T3B-1 (55–65), P6E-3 (64–70), T5C-1 (67–75), Th8A (73–84), 55 T6A (76–93), Th3F (92–103 incomplete) and T4G-2 (95–103) (Fig. 5).

The initial manual sequence analysis of T6A (76–93; Table 2) identified only ten residues. However, the peptide contained two lysine residues, one of which proved to be an internal residue at which the trypsin had failed to cleave; this was probably due to the proximity of the two adjacent glutamic acid residues. Redigestion of the peptide with trypsin (300 nmol of peptide + 10 $\mu$g of treated trypsin at pH 8.0, 37°C for 4h) and subsequent high-voltage paper electrophoresis of the digest at pH 6.5 provided a pure sample of the *N*-terminal section (T6A-1), but the

Table 3. *Details of the peptic peptides*

→, — and ←, Residues determined by the sequenator, manual sequencing and carboxypeptidases respectively.

| Residues | Peptide | Analysis (residues/molecule of peptide) | Sequence |
|---|---|---|---|
| 1-8 | 10B | CmCys(0.9) Asp(1.0) Ser(0.8) Glu(1.0) Pro(1.1) Val(0.3) Leu(0.7) Tyr(1.9) | Leu/Val-CmCys-Pro-Asn-Tyr-Gln-Ser-Tyr |
| 1-4 | 9B | CmCys(0.7) Asp(1.0) Pro(0.9) Val(0.5) Leu(0.8) | Leu/Val-CmCys-Pro-Asn |
| 1-5 | 10C-2 | CmCys(0.8) Asp(1.1) Pro(1.0) Val(0.3) Leu(0.7) Tyr(0.9) | Leu/Val-CmCys-Pro-Asn-Tyr |
| 6-8 | 6C | Ser(0.9) Glu(1.1) Tyr(1.0) | Gln-Ser-Tyr |
| 9-15 | 9C | Arg(1.0) Thr(1.0) Glu(2.2) Ile(1.0) Leu(1.0) Phe(1.1) | Phe-Arg-Thr-Ile-Glu-Glu-Leu |
| 9-14 | 9D | Arg(1.0) Thr(0.9) Glu(2.1) Ile(1.0) Phe(1.2) | Phe-Arg-Thr-Ile-Glu-Glu |
| 10-15 | 8F-3 | Arg(1.0) Thr(0.9) Glu(2.2) Ile(1.0) Leu(0.6) | Arg-Thr-Ile-Glu-Glu-Leu |
| 15-22 | 4E-2 | Lys(1.0) CmCys(0.8) Glu(2.0) Ala(0.9) Ile(0.9) Leu(2.3) | Leu-Gln-Gln-Lys-Ile-Leu-CmCys-Ala |
| 15-20 | 1E-5 | Lys(1.0) Glu(1.8) Ile(0.9) Leu(1.7) | Leu-Gln-Gln-Lys-Ile-Leu |
| 21-29 | 3C-3 | Lys(1.0) Arg(0.9) CmCys(0.8) Asp(1.0) Ser(1.3) Glu(1.0) Ala(1.1) Leu(1.0) | CmCys-Ala-Lys-Ser-Glu-Asn-Ser-Arg-Leu |
| 23-29 | 1D-3 | Lys(1.0) Arg(1.0) Asp(1.0) Ser(1.4) Glu(1.0) Leu(1.0) | Lys-Ser-Glu-Asn-Ser-Arg-Leu |
| 30-38 | 2E-1 | Lys(0.9) Asp(1.7) Glu(1.0) Ala(1.1) Val(1.1) Ile(1.4) Leu(1.0) | Val-Ile-Glu-Ile-Asp-Asn-Ala-Lys-Leu |
| 30-36 | 7D-1 | Asp(2.1) Glu(1.0) Ala(0.8) Val(1.0) Ile(1.6) | Val-Ile-Glu-Ile-Asp-Asn-Ala |

| | | | |
|---|---|---|---|
| 36-38 | 1E-8 | Lys(0.9) Ala(1.1) Leu(1.0) | Ala-Lys-Leu |
| 39-43 | 12B | Asp(1.9) Ser(0.7) Ala(1.0) Phe(1.0) | Ala-Ser-Asp-Asp-Phe |
| 39-42 | 11D | Asp(2.1) Ser(0.9) Ala(1.0) | Ala-Ser-Asp-Asp |
| 43-50 | 6D | Lys(0.9) Arg(1.1) Thr(1.4) Ser(0.3) Glu(2.1) Tyr(1.0) Phe(0.8) | Phe-Arg-Thr-Lys-Tyr-Glu-Thr/Ser-Glu |
| 44-53 | 2B-2 | Lys(1.0) Arg(2.0) Thr(0.8) Ser(1.6) Glu(2.2) Leu(0.8) Tyr(0.9) | Arg-Thr-Lys-Tyr-Glu-Ser-Glu-Arg-Ser-Leu |
| 44-55 | 1B-2 | Lys(1.0) Arg(2.7) Thr(0.8) Ser(1.8) Glu(2.8) Leu(1.1) Tyr(0.9) | Arg-Thr-Lys-Tyr-Glu-Ser-Glu-Arg-Ser-Leu-Arg-Gln |
| 44-58 | 2B-1 | Lys(0.9) Arg(3.0) Thr(0.9) Ser(1.9) Glu(3.7) Val(1.0) Leu(2.0) Tyr(0.9) | Arg-Thr-Lys-Tyr-Glu-Ser-Glu-Arg-Ser-Leu-Arg-Gln-Leu-Val-Glu |
| 53-58 | 4F-1 | Arg(1.0) Glu(2.2) Val(0.9) Leu(1.9) | Leu-Arg-Gln-Leu-Val-Glu |
| 54-58 | 4G-1 | Arg(1.0) Glu(1.9) Val(1.0) Leu(1.0) | Arg-Gln-Leu-Val-Glu |
| 56-58 | 8D-3 | Glu(1.0) Val(0.9) Leu(1.0) | Leu-Val-Glu |
| 59-64 | 7C-1 | Asp(1.8) Ser(1.6) Ile(1.0) Leu(1.0) | Ser-Asp-Ile-Asn-Ser-Leu |
| 60-64 | 7A | Asp(1.8) Ser(1.1) Ile(0.9) Leu(1.0) | Asp-Ile-Asn-Ser-Leu |
| 64-70 | 6E-3 | Arg(1.7) Asp(1.3) Glu(1.4) Ile(1.0) Leu(1.0) | Leu-Arg-Arg-Ile-Leu-Asp-Glu |
| 65-70 | 5B | Arg(1.8) Asp(0.9) Glu(0.9) Ile(0.9) Leu(1.0) | Arg-Arg-Ile-Leu-Asp-Glu-→ |
| 74-78 | 7E-3 | Lys(1.0) CmCys(0.9) Asp(1.1) Ser(0.9) Leu(1.0) | CmCys-Lys-Ser-Asn-Leu |
| 79-85 | 11B | Ser(0.9) Glu(2.9) Ala(0.9) Val(1.0) Leu(1.0) | Glu-Ala-Glu-Val-Glu-Ser-Leu |
| 81-85 | 8B-2 | Ser(0.9) Glu(1.9) Val(1.0) Leu(1.0) | Glu-Val-Glu-Ser-Leu |
| 89-100 | 10F-1 | Lys(1.9) His(1.0) CmCys(0.8) Asp(2.1) Ser(0.8) Glu(3.2) Ala(1.0) Leu(2.8) | Leu-CmCys-Leu-Lys-Lys-Asx-Glx-Glx-Glx-(Asx, Ser, Ala)-Leu |

*C*-terminal section (T6A-2) did not separate from residual uncleaved material. The sequence of the *C*-terminal section, however, was determined by the manual procedure on the mixture, by subtracting the known *N*-terminal sequence of the original peptide. The sequence of peptide T6A-2 was confirmed by the isolation and sequencing of a pure peptide corresponding to peptide T6A-2 from a later tryptic digestion of the complete segment. Confirmation of the sequences of the above peptides is to be found in the analyses of other thermolytic, tryptic and peptic peptides given in Tables 1–3.

The thermolytic 'peptide' Th1D-2 (Table 1) was a mixture of two peptides. Sequence analysis of the mixture indicated that a peptide Phe-Arg-Thr (9–11) corresponded to a portion of peptic peptide 9D (Table 3), the other peptide being identified as Leu-Arg-Gln (53–55). Thermolytic peptide Th2C-1 (85–89) also was impure and only small yields could be achieved. Its sequence, Ile-Lys-Glu-Glu-Leu, corresponded to a portion of tryptic peptide T6A (Table 2), except for the *N*-terminal residue, which is apparently substituted for leucine in some protein chains.

Apart from peptide T6A mentioned above the only tryptic peptides deserving comment are peptides T4G-2 and T1C-1. Peptide T4G-2 (residues 95–103) was devoid of any lysine and arginine and was thus considered to be *C*-terminal. This peptide stained poorly with ninhydrin, giving a yellow colour that turned greyish after approx. 30min; the consequent uncertainty in locating the peptide probably accounts for the very poor yields. But, on using the Pauly stain (Easley, 1965), the peptide was readily visible. Peptide T1C-1 was isolated in very small amounts and its analysis and end group indicated that it was a variant of peptide T2D-1 with a sequence of Ile-Leu-Ala-Asp-Lys. This sequence is known (Crewther *et al.*, 1978*a*) to occur at residues 19–23 in a homologous type-I segment.

All peptic peptides derived from the *N*-terminus

| a | b | c | d | e | f | g |
|---|---|---|---|---|---|---|
| | | | Leu | Cys | Pro | Asn |
| Tyr | Gln | Ser | Tyr | Phe | Arg | Thr |
| Ile | Glu | Glu | Leu | Gln | Gln | Lys |
| Ile | Leu | Cys | Ala | Lys | Ser | Glu |
| Asn | Ser | Arg | Leu | Val | Ile | Glu |
| Ile | Asp | Asn | Ala | Lys | Leu | Ala |
| Ser | Asp | Asp | Phe | Arg | Thr | Lys |
| Tyr | Glu | Ser | Glu | Arg | Ser | Leu |
| Arg | Gln | Leu | Val | Glu | Ser | Asp |
| Ile | Asn | Ser | Leu | Arg | Arg | Ile |
| Leu | Asp | Glu | Leu | Thr | Leu | Cys |
| Lys | Ser | Asn | Leu | Glu | Ala | Glu |
| Val | Glu | Ser | Leu | Lys | Glu | Glu |
| Leu | Leu | Cys | Leu | Lys | Lys | Asn |
| His | Glu | Glu | Glu | Ala | Asp | Ser |
| Leu | | | | | | |

Fig. 6. *Complete amino acid sequence arranged as a series of heptads and phased so that the hydrophobic residues occur at positions a and d*

Columns a and d are boxed. The helical regions as determined by Chou & Fasman (1974) (····) and by Robson & Suzuki (1976) (——) are shown. Cys was determined as CmCys.

of the helical segment contained a mixed *N*-terminal residue Leu/Val (Table 3).

## Discussion

The complete amino acid sequence of a type-I helical segment from the low-sulphur fraction of wool supports the earlier evidence (Crewther & Dowling, 1971) for a highly helical molecule. The method of Chou & Fasman (1974) was applied to examine the sequence for potentially helical regions, and much of the polypeptide had this character but with indications of possible breaks in the α-configuration. The method of Robson & Suzuki (1976) confirms this conclusion, although there are some differences in the assignments by the two methods (Fig. 6). The predictions by the method of Robson & Suzuki (1976) agree with the helix content (85%) of the original helical fragments as measured by o.r.d. (Crewther & Dowling, 1971).

Of the six asparagine residues in the sequence only one occurs in a region predicted to be helical by the Robson & Suzuki (1976) method. Three occur in association with a serine residue at discontinuities between adjacent helical regions. Comparison of successive heptapeptides constituting the sequence (Fig. 6; see also Elleman *et al.*, 1978) showed that positions a and d of each heptapeptide were filled predominantly with hydrophobic residues. Hence, as with the type-II segment (Crewther *et al.*, 1978*b*), these generate a line of hydrophobic residues along one side of an α-helical model formed from the sequence. These positions could correspond to the inner region of a coiled-coil, thus lending further support to the view that the ordered regions of α-keratin assume a coiled-coil configuration (Crick, 1952, 1953; Pauling & Corey, 1953). The repeating character of the sequence and the obvious homology (approx. 32%) between sequences of type-I and type-II segments (Fig. 7) are examined statistically in the following paper (Elleman *et al.*, 1978).

The molecular weight calculated from the sequence

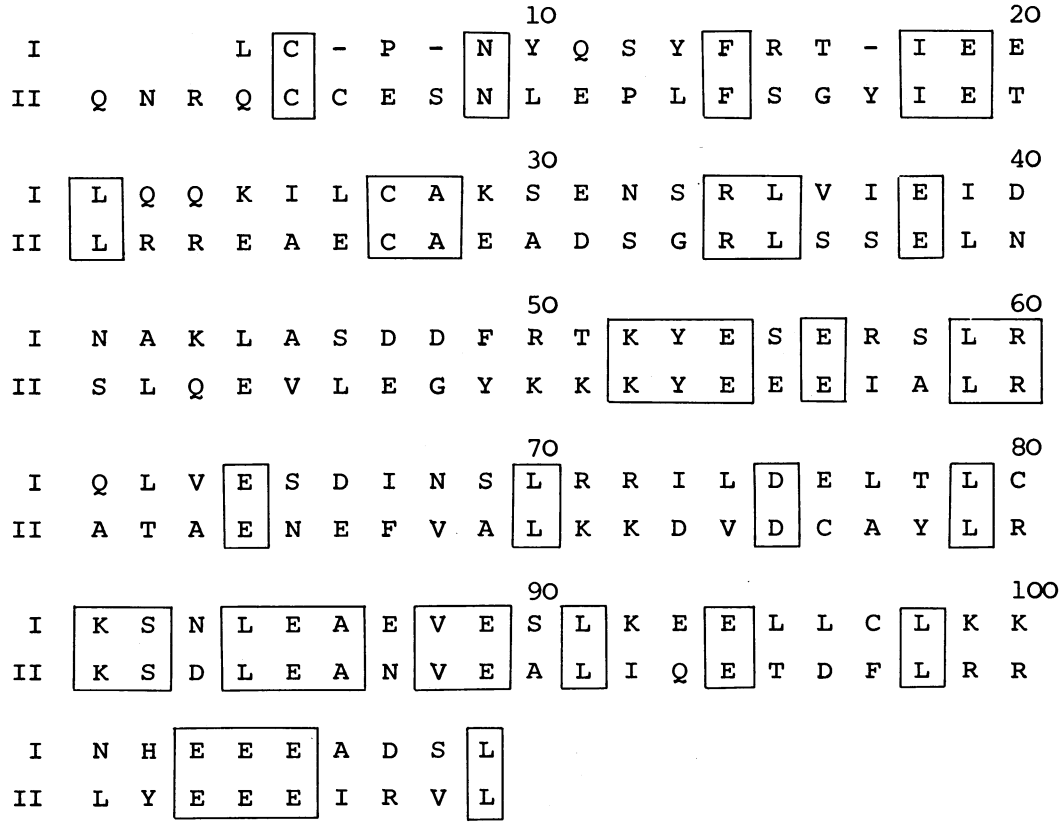


Fig. 7. *Comparison of the type-I and -II helical segments*
These are aligned to obtain maximum homology.

Table 4. *Amino acid composition of the type-I segment* Analytical data are the means of two analyses (uncorrected values).

| Amino acid | Sequence data (residues found) | Analytical data (residues/ molecule) |
|---|---|---|
| Lys | 8 | 7.7 |
| His | 1 | 0.9 |
| Arg | 7 | 7.6 |
| CmCys | 4 | 3.4 |
| Asp | 6 | 11.9 |
| Asn | 6 | |
| Thr | 3 | 3.4 |
| Ser | 11 | 8.9 |
| Glu | 16 | 19.9 |
| Gln | 4 | |
| Pro | 1 | 1.5 |
| Gly | 0 | 0.8 |
| Ala | 5 | 5.9 |
| Val | 3 | 3.2 |
| Ile | 6 | 6.1 |
| Leu | 17 | 16.7 |
| Tyr | 3 | 2.9 |
| Phe | 2 | 1.7 |
| Trp* | 0 | 0 |
| Total no. of residues | 103 | 103 |

* From a separate toluene-*p*-sulphonic acid hydrolysate.

Table 5. *Comparison of the peptides from the type-I sequence with those found in wool by other workers*

| Tryptic and chymotryptic peptides isolated by other workers* | Type-I peptide |
|---|---|
| C15$\gamma$B8 | Residues 16–20 |
| Glx-Glx-Lys-Ile-Leu | Gln-Gln-Lys-Ile-Leu |
| IT/6$\gamma$4 | Residues 19–23 |
| Ile-Leu-Cya-Ala-Lys | Ile-Leu-Cys-Ala-Lys |
| C25$\beta$B1 | Residues 21–29 |
| Cya-Ala-Lys-Ser-Glx-Asx-Ala-Arg-Leu | Cys-Ala-Lys-Ser-Glu-Asn-Ser-Arg-Leu |
| IT/12R | Residues 29–37 |
| Leu-Val-Val-Glu-Asp-Ile-Asp-Ala-Lys | Leu-Val-Ile-Glu-Ile-Asp-Asn-Ala-Lys |
| C15$\beta$B1 | Residues 39–43 |
| Ala-Ser-Asx-Asx-Phe | Ala-Ser-Asp-Asp-Phe |
| C15$\beta$B2 | Residues 39–43 |
| Ala-Ala-Asx-Asx-Phe | Ala-Ser-Asp-Asp-Phe |
| IT/13$\alpha$4 | Residues 45–46 |
| Thr-Lys | Thr-Lys |
| IT/30$\beta$1 | Residues 45–51 |
| Thr-Lys-Tyr-Ser-Glu-Arg | Thr-Lys-Tyr-Glu-Ser-Glu-Arg |
| C15$\delta$ B5H3 | Residues 72–73 |
| Thr-Leu | Thr-Leu |

* Corfield *et al.* (1967); Corfield & Fletcher (1969).

data is 12258, which compares favourably with the value of 12500 obtained for the combined ChC fraction by gel filtration in 8M-urea (Crewther & Dowling, 1971) The amino acid sequence agrees well with the amino acid analysis (Table 4). The analytical values for threonine, proline and alanine are somewhat high and some glycine was present, although it is absent from the sequence. Uncorrected values for serine and *S*-carboxymethylcysteine are low by comparison with the sequence data, but this is partly attributable to destruction of these amino acids during acid hydrolysis. This segment is one of three very similar chains comprising the type-I class of helical segments (Crewther *et al.*, 1978*a*) and some discrepancy in amino acid composition would result from the presence of small amounts of the homologous segments. Apart from the hydrophobic repeats discussed above, the sequence is characterized by regions rich in charged residues.

As was noted in the amino acid sequence of the type-II helical segment from *S*-carboxymethylkerateine-A, anionic and cationic residues occur frequently displaced by three or four residues along the sequence. This characteristic, which gives the opportunity for interaction of oppositely charged groups and consequent stabilization of the helical structure, is even more apparent in the type-I segment than in the type-II segment and may account in part for the greater helicity of the type-I as compared with the type-II segment (Crewther & Dowling, 1971).

It is also apparent that, whereas 10 of the 14 cationic residues are situated adjacent to a residue of like charge in the type-II sequence, 11 of the 22 anionic residues occur in a similar manner in pairs or triplets in the type-I sequence. Taken in conjunction with the thermal stability of the $\alpha$-helix when the type-I and type-II segments are mixed in solution (Crewther & Dowling, 1971), this suggests that ionic interactions between groups of ionized residues may contribute to the self-assembly of the ordered regions in the $\alpha$-keratin. In fact, even a cursory examination of possible ionic interactions between residues 18, 23, 37, 60, 65, 79, 88 and 93 of type-I and residues 29, 24, 38, 71, 66, 80, 99 and 94 respectively of type-II sequences when assuming an $\alpha$-helical conformation indicates that an association of the segments approximately in register and in a parallel rather than an anti-parallel arrangement would be favoured. This is considered in detail elsewhere (Parry *et al.*, 1977).

The sequence (Fig. 5) includes a number of regions with sequences identical with or similar to those of tryptic and chymotryptic peptides (Table 5) isolated from oxidized wool by Corfield *et al.* (1967) and Corfield & Fletcher (1969). The data obtained by these workers represent about 40% of the type-I segment and 32% of the type-II segment. Our experience suggests that many of the peptides that were not recovered by Corfield *et al.* (1967) and Corfield &

Fletcher (1969) may have been precipitated in the loading buffer used for ion-exchange chromatography. The loss of these peptides nullifies their conclusion that all the protein chains of wool could be derived from a single protein of mol.wt. 74000.

We thank Dr. C. M. Roxburgh and Mr. R. W. Cranston for amino acid analyses, Mr. E. Suzuki for the analysis of helical areas, and Mr. P. McK. Strike for skilled technical assistance.

## References

Ambler, R. P. (1963) *Biochem. J.* **89**, 349–378
Ambler, R. P. (1967) *Methods Enzymol.* **11**, 156
Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
Corfield, M. C. & Fletcher, J. C. (1969) *Biochem. J.* **115**, 323–334
Corfield, M. C., Fletcher, J. C. & Robson, A. (1967) *Biochem. J.* **102**, 801–804
Crewther, W. G. & Dowling, L. M. (1971) *Appl. Polym. Symp. no. 18*, 1–20
Crewther, W. G. & Harrap, B. S. (1967) *J. Biol. Chem.* **242**, 4310–4319
Crewther, W. G. & Inglis, A. S. (1975) *Anal. Biochem.* **68**, 572–585
Crewther, W. G., Gough, K. H., Inglis, A. S. & McKern, N. M. (1978*a*) *Text. Res. J.* **48**, 160–162
Crewther, W. G., Inglis, A. A. & McKern, N. M. (1978*b*) *Biochem. J.* **173**, 365–371
Crick, F. H. C. (1952) *Nature (London)* **170**, 882–883
Crick, F. H. C. (1953) *Acta Crystallogr.* **6**, 689–697
Easley, C. W. (1965) *Biochim. Biophys. Acta* **107**, 386–388
Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–91
Elleman, T. C., Crewther, W. G. & Van der Touw, J. (1978) *Biochem. J.* **173**, 387–391
Fraenkel-Conrat, H., Harris, J. I. & Levy, A. L. (1955) *Methods Biochem. Anal.* **2**, 359–425
Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
Hogg, D. McC., Dowling, L. M. & Crewther, W. G. (1978) *Biochem. J.* **173**, 353–363
Inglis, A. S. & Nicholls, P. W. (1973) *J. Chromatogr.* **79**, 344–346
Inglis, A. S., Nicholls, P. W. & Roxburgh, C. M. (1971) *Aust. J. Biol. Sci.* **24**, 1235–1240
Moffitt, W. & Yang, J. T. (1956) *Proc. Natl. Acad. Sci. U.S.A.* **42**, 596–603
Offord, R. E. (1966) *Nature (London)* **211**, 591–593
Parry, D. A. D., Crewther, W. G., Fraser, R. D. B. & MacRae, T. P. (1977) *J. Mol. Biol.* **113**, 449–454
Pauling, L. & Corey, R. B. (1953) *Nature (London)* **171**, 59–61
Robson, B. & Suzuki, E. (1976) *J. Mol. Biol.* **107**, 327–356
Sparrow, L. G. & Crewther, W. G. (1972) *J. Text. Inst.* **63**, 619–621
Woods, K. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **133**, 369–370