Biochem. J. (1978) **173**, 353–363
Printed in Great Britain

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

TRYPTIC AND CHYMOTRYPTIC PEPTIDES FROM A TYPE-II SEGMENT

By DAVID McC. HOGG,* LYNDSAY M. DOWLING and W. GORDON CREWTHER
Division of Protein Chemistry, CSIRO, Parkville (Melbourne), Vic. 3052, Australia

(*Received* 2 *November* 1977)

1. Amino acid-sequence studies were done on a peptide of mol.wt. approx. 12500 that was isolated from the highly helical fragments obtained by partial chymotryptic digestion of the low-sulphur proteins (*S*-carboxymethylkerateine-A) from wool. 2. The peptides obtained by tryptic and chymotryptic digestion of this large peptide were separated by ion-exchange chromatography on DEAE-cellulose at pH8.5 with an $(NH_4)_2CO_3$ concentration gradient and, where necessary, purified further by paper electrophoresis. 3. Determination of the sequences of many of these peptides showed that a high proportion of the cationic residues occurs in pairs. 4. Although two of the four *S*-carboxymethylcysteine residues are located in what appears to be a non-helical region near the *N*-terminus the other two *S*-carboxymethylcysteine residues occur in or near sequences suggesting a helical conformation. 5. Some peptides were obtained, in low yields, that appeared to be homologues of more major ones. These suggest either homologies in the helical portions of the low-sulphur proteins or the presence of closely related amino acid sequences in helical regions of completely different origins. 6. A partial sequence of the complete peptide is proposed.

Since Consden *et al.* (1947) first separated peptides from acid hydrolysates of wool and studied their composition, considerable effort has been expended in an attempt to obtain amino acid sequences for the three clases of protein chains constituting the wool fibre. The chief difficulties have been the great complexity of the mixture of proteins extracted from wool after rupture of the disulphide bonds, and the close similarity between sequences for different protein chains (Crewther *et al.*, 1965).

Two of the three major classes of protein constituting the extractable proteins of wool, the high-sulphur proteins (Haylett & Swart, 1969; Haylett *et al.*, 1971; Swart & Haylett, 1971; Lindley & Elleman, 1972; Elleman, 1972*a*,*b*; Elleman & Dopheide, 1972; Swart & Haylett, 1973) and the proteins rich in glycine and tyrosine (Dopheide, 1973), have now yielded complete sequence information for a few chains. Sequence studies on the low-sulphur proteins, which appear to constitute the partly helical filamentous structures in α-keratin (Crewther *et al.*, 1965; Jones, 1975; Fraser *et al.*, 1976), have been less successful. Three main factors contribute to this: the size of the protein chains (mol.wt. 40000–56000), the presence of a large number of similar chains within this relatively narrow range of molecular weights and the lack of adequate means of identifying these chains. As a result, attempts have been made to obtain sequence data on fractions that contained several chain species, fraction U.S.3 from oxidized wool (Corfield, 1963; Cole *et al.*, 1965; Corfield *et al.*, 1965, 1967, 1968; Corfield & Fletcher, 1969) and 'Component 8' from *S*-carboxymethylated wool (Thompson & O'Donnell, 1967; O'Donnell & Thompson, 1968; Hosken *et al.*, 1968; O'Donnell, 1969).

The highly helical rod-like fragments obtained by partial proteolysis of the low-sulphur protein fraction, *S*-carboxymethylkerateine-A, from wool (Crewther & Harrap, 1965, 1967) provide a potential source of large segments of protein chain from this fraction. Gel filtration on Sephadex G-75 in 8M-urea of the helical fragments obtained by partial chymotryptic digestion of *S*-carboxymethylkerateine-A (Crewther & Dowling, 1971) leads to the recovery of two major fractions, one corresponding to a mol.wt. of about 12500 (fraction ChC) and the other, containing less material, to a mol.wt. of about 25000 (fraction ChB). Fraction ChC has been further fractionated by ion-exchange chromatography into some 17 subfractions (Crewther & Dowling, 1971). These were shown by amino acid analysis, peptide 'mapping' and o.r.d. measurements to contain two types of chain segments (I and II).

The present paper describes the isolation of tryptic and chymotryptic peptides derived from one of the

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

* Present address: National Capital Development Commission, Canberra, A.C.T. 2061, Australia.

type-II helical chain segments (fraction ChC 9; Crewther & Dowling, 1971) and the determination of their amino acid sequences. An incomplete tentative sequence is proposed for this segment.

Experimental

Materials

α-Chymotrypsin and trypsin (1-chloro-4-phenyl-3-L-toluene-*p*-sulphonamido-2-one-treated) were from Worthington Biochemicals Corp., Freehold, NJ, U.S.A. Lima-bean trypsin inhibitor was from Koch–Light Laboratories, Colnbrook, Bucks., U.K., DEAE-cellulose was Whatman DE-52, Whatman, Maidstone, Kent, U.K., and Sephadex G-75 was from Pharmacia Fine Chemicals, Uppsala, Sweden. Phenyl isothiocyanate was from Fluka A.G., Buchs, Switzerland. Trifluoroacetic acid from Koch–Light was redistilled before use, pyridine (AnalaR; BDH Chemicals, Poole, Dorset, U.K.) was refluxed with ninhydrin for 4h and distilled, and butyl acetate (Unilab Ajax Chemicals, Cheltenham, Vic., Australia) was redistilled. Constant-boiling HCl was prepared by distillation of BDH AnalaR HCl diluted to approx. 6M. Dansyl chloride solution (0.25%) was prepared by dissolving dansyl chloride (BDH Biochemicals) in hot acetone (AnalaR) and filtering when cold. Glass-distilled water was used to prepare all solutions for sequence determination by the dansyl–Edman procedure. All other chemicals were of Analytical Reagent quality. Urea solutions were purified by passage through a mixed-bed ion-exchange resin immediately before addition of buffer salts (Thompson & O'Donnell, 1965).

Preparation of polypeptide chain segments

The preparation of helix-rich fragments from the low-sulphur protein fraction, *S*-carboxymethyl-kerateine-A, by chymotryptic digestion, the fractionation of the constituent polypeptide chain segments by gel filtration on Sephadex G-75 in 8M-urea to yield fractions ChB (mol.wt. approx. 25000) and ChC (mol.wt. approx. 12500), and the further fractionation of the latter in 8M-urea by gradient elution from DEAE-cellulose to yield type-I and type-II helical segments has been described elsewhere (Crewther & Dowling, 1971). Fraction ChC 9 was one of the type-II fractions obtained in these studies.

Starch-gel electrophoresis of fraction ChC 9 at pH8.9 in 8M-urea, as described by Thompson & O'Donnell (1964), gave a single band of material near the front.

Proteolysis of helical segments

Tryptic digestion of fraction ChC 9 (50mg) was carried out at 37°C in a pH-stat at pH8.5. The enzyme was added in solution in amounts of 0.250, 0.125 and 0.125mg at 0, 1.5 and 3h respectively. The total digestion time was 4.5h. The chymotrypsin was incubated for 15min at 37°C with trypsin inhibitor (enzyme/inhibitor; 4:1, w/w) before mixing (0.5mg) with the substrate (50mg) and digesting for 7h at the same temperature and pH.

High-voltage paper electrophoresis

Peptide fractions were examined by high-voltage electrophoresis on paper by the method of Michl (1951) with buffers specified by Naughton *et al.* (1960) or 1% $(NH_4)_2CO_3$ at pH9.0. Samples were dissolved in 0.4ml of 50% (v/v) pyridine and 20μl was applied over a distance of 1cm on Whatman 3MM paper. After electrophoresis at the desired pH the papers were dried, sprayed with 0.5% ninhydrin in ethanol, heated in a stream of air at about 60°C for 1h and the positions of peptides marked or recorded photographically. The papers were then stained by the chlorine/starch/iodine method of Rydon & Smith (1952).

Similar electrophoretic procedures were used for purifying peptides, the sample usually being applied over a distance of 10cm. After electrophoresis, strips from the terminal portions were excised and stained to locate peptides that were then eluted from the paper with 0.01M-NH_3 and freeze-dried.

Amino acid analysis

The amino acid compositions of peptides were determined by hydrolysing the peptide *in vacuo* at 108°C for 24h in constant-boiling HCl containing 1 drop of 0.1M-phenol per ml and analysing the hydrolysate by using a Beckman 120C automatic amino acid analyser. Tryptophan was determined by the method of Goodwin & Morton (1946).

Isolation of peptides

Digests of fraction ChC 9 were adsorbed directly on a column (2cm×15cm) of DEAE-cellulose equilibrated with 0.01M-$(NH_4)_2CO_3$ at pH8.5 and eluted with a linear concentration gradient of the same buffer salt. The gradient ranges for the fractionations are shown in the Figures. The fractionations were monitored by measuring the A_{230} in a Beckman DB spectrophotometer. The fractions were pooled, then dried *in vacuo* to remove $(NH_4)_2CO_3$ and their peptide contents checked by high-voltage electrophoresis at pH6.5 and pH9.0. Where necessary, final purification was carried out by electrophoresis on paper at the pH value showing optimum separation of bands in the initial experiments. In most instances a pH of 9.0 was used.

Peptides were eluted, freeze-dried and checked for purity by electrophoresis at other pH values.

Sequence determination

The dansyl–Edman method as detailed by Hartley (1970) was used. Sequences of a few peptides were confirmed by the method of Dowling & Stark (1969).

Table 1. *Amino acid composition of fraction ChC 9*
Analytical values are the means of duplicate 24h hydrolyses carried out as indicated in the Experimental section. They are uncorrected for losses by decomposition.

Amino acid	Analysis (residues/ 100 residues)	Analysis (residues/ 109 residues)	Proposed sequence (residues/ 109 residues)
Ala	8.4	9.2	12
Arg	8.6	9.4	10
Asp	9.9	10.8	10
CmCys	3.9	4.3	4
Glu	21.4	23.3	24
Gly	2.9	3.2	3
Ile	3.6	3.9	3.5
Leu	13.8	15.0	14
Lys	5.1	5.6	5
Phe	2.7	2.9	3
Pro	0.8	0.9	1
Ser	6.1	6.6	6
Thr	2.8	3.1	3
Trp	0.0	0.0	—
Tyr	4.4	4.8	5
Val	5.7	6.2	5.5

Results

Amino acid analysis of fraction ChC 9

The amino acid analysis of fraction ChC 9 is given in Table 1.

Peptides from fraction ChC 9

(*a*) *Tryptic peptides.* Preliminary experiments showed that adequate separation of peptides in the pH range 3.5–6.0 was difficult owing to the relatively low solubility of certain tryptic peptides in this pH range and their consequent tendency to be precipitated with change in pH or salt concentration. Corfield *et al.* (1967) observed a similar precipitation of tryptic peptides from acidified digests of their wool-protein fraction U.S.3. Good separations, however, were obtained by DEAE-cellulose chromatography at pH8.5 with a linear gradient of $(NH_4)_2CO_3$. The elution pattern for tryptic peptides from fraction ChC 9 is shown in Fig. 1. Fractions T1–T23 were obtained in this way and after removal of the $(NH_4)_2CO_3$ *in vacuo* were subjected to electrophoresis at pH6.5 and 9.0 (Figs. 2*a* and 2*b*). Staining with chlorine/starch/iodide revealed a major peptide in fraction T15, and minor peptides in fractions T8 and T17, which did not stain with ninhydrin. Since the original fraction, ChC 9, did not yield an *N*-terminal derivative by either the Edman or the dansyl techniques these peptides were assumed to be *N*-terminal. The analyses and assumed compositions of purified peptides are listed in Table 2. The amino

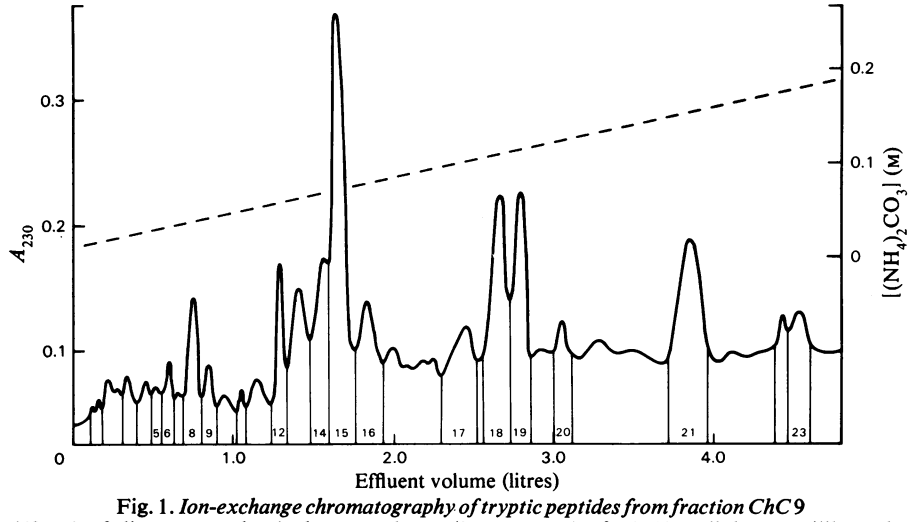

Fig. 1. *Ion-exchange chromatography of tryptic peptides from fraction ChC 9*
A portion (50mg) of digest was adsorbed on a column (2cm×15cm) of DEAE-cellulose equilibrated with 0.01M-$(NH_4)_2CO_3$, pH8.5. Peptides were eluted with a linear concentration gradient (0.01–0.30M, ----) of the same buffer salt at a flow rate of 30ml/h (10ml fractions were collected). The total volume of eluent (8 litres) was not used as all significant peptides had been collected at a carbonate concentration of approx. 0.19M. Only the fractions containing peptides in recoverable amounts have been numbered.

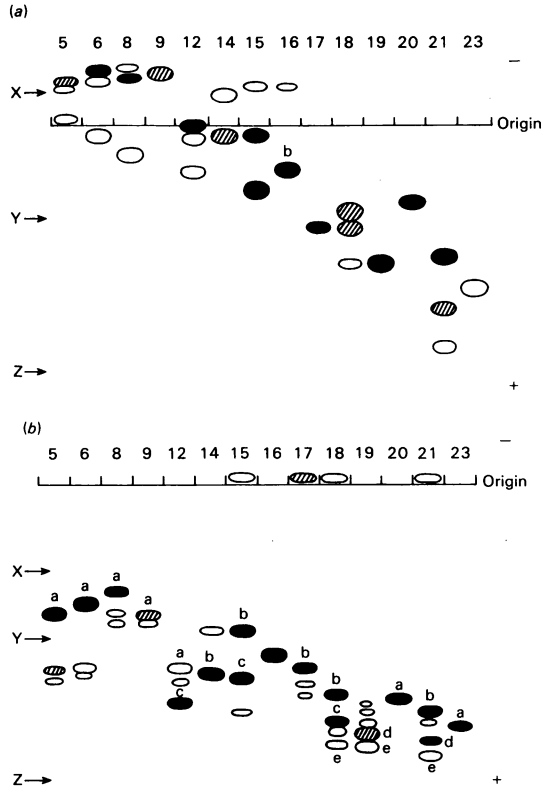

Fig. 2. *High-voltage paper electrophoresis of tryptic peptides from fraction ChC 9*

Diagrammatic representation of the patterns obtained (*a*) at pH6.5 and (*b*) at pH9.0 from DEAE-cellulose fractions that contained significant amounts of one or more peptides. Letters a–e indicate peptides that were purified further. Arrows X, Y and Z indicate the positions of the marker dyes ε-dinitrophenyl-lysine, Xylene Cyanol Blue FF and Carbol Fuchsin respectively. The intensity of peptide spots is indicated by: ●, strong; ◍, moderate; ○, weak.

acid sequences of the peptides as determined by the dansyl–Edman procedure are listed in Table 3.

(*b*) *Chymotryptic peptides.* The chymotryptic digest of fraction ChC 9, when chromatographed on DEAE-cellulose by the same procedure used for the tryptic digest, gave the complex elution pattern shown in Fig. 3. Paper electrophoresis of the fractions obtained in this way showed that many contained no major peptides. Those fractions containing major peptides were submitted to electrophoresis under the conditions shown to give optimum fractionation by the preliminary experiment (usually at pH6.5) and the peptides were excised and eluted.

Table 2. *Amino acid compositions of the major tryptic peptides obtained from fraction ChC9*

Numerals in parentheses represent the numbers (*n*) of residues per molecule assumed in determining the sequence. The letter P is used to indicate a partial residue resulting from substitution. The experimental values were obtained from the expression $n_i = mc_i/\sum(c_i/n_i')$, where c_i is the content (μmol) of amino acid *i* in a sample of the hydrolysate of a peptide containing *m* different types of residue, and *n*, and n_i' are the experimental and assumed values respectively for the number of residues of type *i* in the peptide.

Peptide no.	Composition (residues per molecule)													
	Ala	Arg	Asx	Glx	Gly	Ile	Leu	Lys	Phe	Pro	CmCys	Ser	Thr	Tyr	Val
T6a	2.8 (3)		1.0 (1)	1.9 (2)			1.1 (1)	2.1 (2)	1.1 (1)				0.9 (1)		1.0 (1)
T14b		1.9 (2)		3.1 (3)		0.9 (1)	2.0 (2)							1.1 (1)	1.0 (1)
T15b		3.1 (3)	2.0 (2)	4.9 (5)	1.2 (1)	0.9 (1)	3.2 (3)		1.1 (1)	1.1 (1)	1.7 (2)	1.9 (2)	0.9 (1)	1.1 (1)	
T15c	0.9 (1)	1.8 (2)		3.0 (3)		0.4(P)	1.1 (1)	0.2(P)						1.0 (1)	0.6(P)
T16b	1.6 (2)	2.1 (2)	3.0 (3)	4.2 (4)		1.1 (1)	3.0 (3)	1.1 (1)	1.0 (1)			1.1 (1)	1.1 (1)		0.9 (1)
T17b		3.2 (3)	2.4 (2)	5.0 (5)	1.1 (1)	1.0 (1)	3.1 (3)		1.0 (1)	0.9 (1)	1.4 (2)	1.2 (2)	0.9 (1)	1.0 (1)	
T18b	2.0 (2)	1.4 (1)	3.0 (3)	4.2 (4)		0.9 (1)	3.1 (3)	0.8 (1)	0.9 (1)			1.1 (1)	0.8 (1)		1.0 (1)
T18c		1.0 (1)		3.1 (3)		1.0 (1)	2.0 (2)							1.1 (1)	0.9 (1)
T19d	1.1 (1)	0.9 (1)	1.4 (1)	1.1 (1)			1.0 (1)				0.6 (1)			1.3 (1)	0.8 (1)
T19e	1.1 (1)	1.0 (1)	1.9 (2)				1.1 (1)				0.8 (1)			1.1 (1)	1.0 (1)
T21b	1.8 (2)	0.9 (1)	1.4 (1)	4.4 (4)	1.2 (1)		2.5 (3)	1.3 (1)				2.1 (2)		0.6 (1)	0.8 (1)
T21d			1.1 (1)	3.1 (3)		1.9 (2)	1.9 (2)							1.0 (1)	
T21e	2.8 (3)	1.0 (1)	1.0 (1)	3.0 (3)	1.0 (1)						0.6 (1)	0.9 (1)			
T23a	1.8 (2)	0.7 (1)	1.3 (1)	4.4 (4)	1.4 (1)		2.5 (3)	1.1 (1)				2.1 (2)		0.7 (1)	0.9 (1)

Table 3. *Amino acid sequences and yields of tryptic peptides from fraction ChC 9*
Related peptides are bracketed.

Peptide no.	Yield (mol/mol)	Sequence
T6a	0.18	Ala-Thr-Ala-Glx-Asx-Glx-Phe-Val-Ala-Leu-Lys-Lys
T14b	0.06	Arg-Leu-Tyr-Glx-Glx-Glx-Ile-Arg-Val-Leu
T18c	0.71	Leu-Tyr-Glx-Glx-Glx-Ile-Arg-Val-Leu
T21d	0.04	Leu-Tyr-Asx-Glx-Glx-Ile-Glx-Ile-Leu
T15b	0.20	No reactive end group
T17b	0.05	No reactive end group
T15c	0.34	$\frac{Lys}{Arg}$-Tyr-Glx-Glx-Glx-$\frac{Ile}{Val}$-Ala-Leu-Arg
T16b	0.17	Lys-Ser-Asx-Leu-Glx-Ala-Asx-Val-Glx-Ala-(Arg,Asx,Glx_2,Ile,Leu_2,Phe,Thr)-Arg
T18b	0.10	Lys$\frac{Leu}{Ser}$-Asx-Leu-Glx-Ala-$\frac{Glx}{Asx}$-(Ala,Asx,Glx_3,Ile,Leu_2,Phe,Thr,Val)-Arg
T19d	0.05	Asx-Val-Glx-CmCys-Ala-Tyr-Leu-Arg
T19e	0.17	Asx-Val-Asx-CmCys-Ala-Tyr-Leu-Arg
T21b	0.20	No clear assessment of *N*-terminal sequence possible
T23a	0.10	No clear assessment of *N*-terminal sequence possible
T21e	0.05	Glx-Ala-(Ala_2,Asx,CmCys,Glx_2,Gly,Ser)-Arg
*		Glx-Ala-Glx-CmCys-Val-Glx-Ala-(Asx,Gly,Ser)-Arg

* Isolated from a different preparation of type-II helical segment and partially sequenced by the method of Dowling & Stark (1969). The sequence of peptide 1T/4γ3 from Corfield *et al.* (1967) indicates that the *C*-terminal sequence is -Asx-Ser-Gly-Arg.

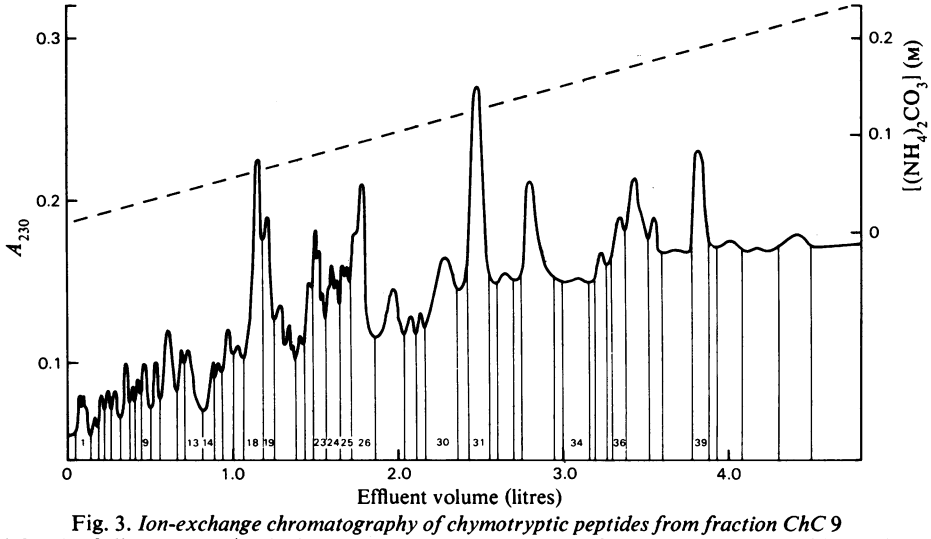

Fig. 3. *Ion-exchange chromatography of chymotryptic peptides from fraction ChC 9*
A sample (50mg) of digest was adsorbed on a column (2cm×50cm) of DEAE-cellulose equilibrated with 0.01M-$(NH_4)_2CO_3$, pH8.5. Peptides were eluted with a linear concentration gradient (0.01–0.30M, ----) of the same buffer salt at a flow rate of 30ml/h (10ml fractions were collected). The total volume of eluent (8 litres) was not used as all significant peptides had been collected at a carbonate concentration of approx. 0.22M. Only the fractions containing peptides in recoverable amounts have been numbered.

Table 4 records the compositions and Table 5 the sequences of those peptides that were sufficiently pure to permit at least partial sequence determination.

Many of the chymotryptic peptides had one, or more often two, cationic residues at the *N*-terminus. This observation is in accord with the presence (Table 3) of a leucine residue immediately preceding the basic residues in several of the tryptic peptides.

Further degradation of N-terminal peptides

Peptide T15b, the major *N*-terminal peptide from the tryptic digest of fraction ChC 9, was divided into

Table 4. *Amino acid compositions of peptides isolated from a chymotryptic digest of fraction ChC 9*

Numerals in parentheses represent the numbers (n) of residues per molecule assumed in determining the sequence. The letter P is used to indicate a partial residue resulting from substitution. The experimental values were obtained from the expression $n_i = mc_i/\Sigma_1^m(c_i/n_i')$, where c_i is the content (μmol) of amino acid, i, in a sample of the hydrolysate of a peptide containing m different types of residue, and n_i and n_i' are the experimental and assumed values respectively for the number of residues of type i in the peptide. Abbreviation: N.D., not determined.

Peptide no.	Composition (residues per molecule)														
	Ala	Arg	Asx	Glx	Gly	Ile	Leu	Lys	Phe	Pro	CmCys	Ser	Thr	Tyr	Val
C1e		1.8 (2)					1.0 (1)							1.1 (1)	
C13b		2.6 (3)		2.9 (3)		0.9 (1)	2.4 (2)							0.9 (1)	1.2 (1)
C14a		2.4 (2+P)		3.0 (3)		0.9 (1)	2.1 (2)	0.4 (P)						1.0 (1)	1.0 (1)
C18a	1.1 (1)	0.6 (P)		2.9 (3)		0.4 (P)	1.1 (1)	2.0 (2+P)						1.0 (1)	0.7 (P)
C19c	2.8 (3)	1.0 (1)	1.0 (1)	2.1 (2)			1.1 (1)		1.0 (1)				0.9 (1)		1.0 (1)
C23a	4.2 (4)	1.2 (1+P)	1.2 (1)	5.0 (5)		0.4 (P)	2.3 (2)	2.2 (2+P)	0.9 (1)				1.0 (1)	N.D.	1.6 (1+P)
C24a		0.7 (1)	1.1 (1)	2.5 (2)		1.0 (1)	1.4 (1)	0.7 (1)						0.8 (1)	
C25a	1.0 (1)	0.6 (P)	1.2 (1)	1.9 (2)		0.3 (P)	0.8 (1)	0.9 (1+P)						1.2 (1)	0.7 (P)
C26c	1.4 (1)		2.4 (2)				0.5 (1)	2.2 (2)			0.6 (1)			1.2 (1)	1.3 (1)
C26e		0.9 (1)		3.0 (3)		0.9 (1)	1.1 (1)								1.0 (1)
C30a	2.1 (2)	3.6 (4)	3.1 (3)	7.0 (7)		1.7 (2)	5.0 (6)	1.1 (1)	0.9 (1)			1.3 (1)	1.0 (1)	1.0 (1)	1.9 (2)
C31c		0.9 (1)	2.4 (2)	4.1 (4)	0.8 (1)		2.5 (2)		0.7 (1)	1.0 (1)	1.6 (2)	1.7 (2)		1.0 (1)	
C34a	3.0 (3)	1.7 (2)	4.1 (4)	6.4 (6)		1.3 (1)	5.5 (5)	2.1 (2)	0.6 (1)		0.6 (1)	1.8 (2)	1.0 (1)	0.9 (1)	2.1 (2)
C36b	2.5 (2+P)	2.4 (3)	2.5 (2)	6.6 (7)	1.6 (2)		3.8 (4)				0.9 (1)	2.8 (3)		1.0 (1)	1.7 (1+P)
C39c	2.8 (3)	3.4 (3)	2.4 (2)	8.3 (8)	1.9 (2)	0.9 (1)	5.1 (5)				0.6 (1)	3.5 (4)	0.9 (1)	1.0 (1)	1.1 (1)
C39d	3.0 (3)	3.5 (3)	2.3 (2)	8.4 (8)	2.0 (2)	0.9 (1)	5.3 (5)				0.9 (1)	3.8 (4)	0.9 (1)	0.9 (1)	1.2 (1)

Table 5. *Amino acid sequences and yields of chymotryptic peptides from fraction ChC 9*
Related peptides are bracketed.

Peptide no.	Yield (mol/mol)	Sequence
C1e	0.01	Arg-Arg-Leu-Tyr
C13b	0.01	Arg-(Arg_2,Glx_3,Ile,Leu_2,Tyr,Val)
C14a	0.02	Arg-?-Leu-Tyr-($Arg_{(1+P)}$,Glx_3,Ile,Leu,$Lys_{(P)}$,Val)
C24a	0.03	Lys-Arg-Leu-Tyr-Asx-Glx-Glx-Ile
C26e	0.06	Glx-Glx-Glx-Ile-Arg-Val-Leu
C18a	0.12	Lys-$\frac{Arg}{Lys}$-?-Tyr-Glx-Glx-Glx-$\frac{Ile}{Val}$-Ala-Leu
C23a	0.03	Lys-$\frac{Arg}{Lys}$-?-Tyr-Glx-Glx-Glx-$\frac{Ile}{Val}$-(Ala_4,Arg,Asx,Glx_2,Leu_2,Phe,Thr,Val)
C25a	0.06	Lys-?-Leu-Tyr-Glx-Asx-Glx-$\frac{Ile}{Val}$-Ala
C19c	0.22	Arg-Ala-Thr-Ala-Glx-Asx-Glx-Phe-Val-Ala-Leu
C26c	0.10	Lys-Lys-Asx-Val-Asx-CmCys-Ala-Tyr-Val
C34a	0.01	Arg-$\frac{Arg}{Lys}$-Asx-(Ala_3,$Arg_{(P)}$,Asx_3,CmCys,Glx_6,Ile,Leu_5,$Lys_{(1+P)}$,Ser_2,Phe,Thr,Tyr,Val_2)
C30a	0.03	Leu-Arg-(Ala_2,Arg_3,Asx_3,Glx_7,Ile_2,Leu_5,Lys,Phe,Ser,Thr,Tyr,Val_2)
C31c	0.05	No reactive end group
C36b	0.20	Arg-?-Glx-Ala-Glx-($Ala_{(1+P)}$,Arg,Asx_2,CmCys,Glx_5,Gly_2,Leu_4,Ser_3,Tyr,$Val_{(1+P)}$)
C39c	0.07	Ile-Glx-Thr-Leu-(Ala_3,Arg_4,Asx_2,CmCys,Glx_8,Gly_2,Leu_4,Ser_3,Tyr,Val)
C39d	0.04	Ile-Glx-Thr-Leu-Arg-(Ala_3,Arg_3,Asx_2,CmCys,Glx_8,Gly_2,Leu_4,Ser_4,Tyr,Val)

Table 6. *Analyses of peptides derived from peptide T15b*

Peptide	Composition	Sequence
Tryptic		
T15b-T1	Arg	Arg
T15b-T2	Arg, Asx, Glx	(Asx,Glx)-Arg No end group
T15b-T3	Arg, Asx, Glx_4, Gly, Ile, Leu_3, Phe, Pro, $CmCys_{1-2}$, Ser, Thr, Tyr	No end group
Chymotryptic		
T15b-C1	Arg	Arg-Arg
T15b-C2	Gly, Ser, Phe, Tyr	Phe-(Gly,Ser,Tyr)
T15b-C3	Glx, Ile, Leu, Thr	Ile-Glx-Thr-Leu
T15b-C4	Arg, Asx_2, $CmCys_2$, Glx_4, Gly, Leu_2, Pro, Ser_2, Phe, Tyr	No end group
T15b-C5	Arg, Asx, Glx, Leu, Pro, Ser	No end group

two equal portions, one of which was digested with the treated trypsin (1% by wt. of the peptide) in 1.0ml of 0.1M-$(NH_4)_2CO_3$ solution at pH8.5 for 5h at 37°C. A further equal amount of trypsin was then added and incubation continued for a further 16h. The second portion of the peptide T15b was digested with chymotrypsin under identical conditions. The chymotrypsin was again incubated for 15min at 37°C with lima-bean trypsin inhibitor (enzyme/inhibitor, 4:1, w/w) before mixing with the substrate.

Electrophoresis of the digests on paper at pH6.5 yielded three chymotryptic peptides, T15b-C3, T15b-C4 and T15b-C5, and two tryptic peptides, T15b-T2 and T15b-T4, which gave faint or no staining with ninhydrin and relatively strong staining with the chlorine/starch/iodide treatment. It seemed probable that some of these were *N*-terminal. Analytical and sequence data for all peptides derived from peptide T15b are summarized in Table 6. These data indicate that peptide T15b takes the form shown in Fig. 4.

Discussion

The tryptic peptide T15b, like the entire segment ChC 9, does not have a reactive *N*-terminal amino group and is therefore assumed to occur at the *N*-terminus. The *C*-terminal sequence of peptide T15b, -Ile-Glx-Thr-Leu-Arg-Arg, is identical with the *N*-terminal sequence of peptides C39c and C39d (Fig. 5). The remaining sections of these chymotryptic

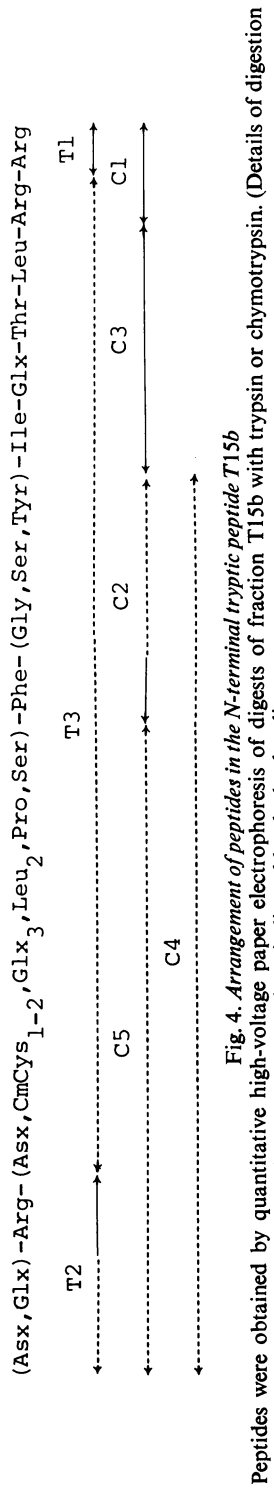

Fig. 4. *Arrangement of peptides in the N-terminal tryptic peptide T15b*

Peptides were obtained by quantitative high-voltage paper electrophoresis of digests of fraction T15b with trypsin or chymotrypsin. (Details of digestion conditions given in the text.) Residues not sequenced are indicated by the broken lines.

peptides correspond closely in amino acid composition to the combined amino acid composition of peptides T21e and T21b (or T23a) if terminal basic residues are excluded. Since no other combination of tryptic peptides could be considered to correspond to this portion of peptide C39c, the *N*-terminal sequence of segment ChC 9 can be considered to consist of peptide T15b–(T21e+T21b) or the equivalent peptides C31c–C39c.

The chymotryptic peptides C23a (50–70; Fig. 5) and C18a (50–59) have identical *N*-terminal sequences (Table 5) and the difference in composition of these peptides is identical with the composition of peptide C19c. Peptide C23a therefore corresponds to peptides C18a–C19c, and this in turn, apart from terminal basic residues, corresponds to peptides T15c–T6a (Fig. 5). The *N*-terminal arginine residue in peptide C19c is one of the few basic residues in segment ChC 9 that is not adjacent to a second basic residue. Attempts to place peptide C19c at any of the other three single basic residues (positions 3, 34 or 107; Fig. 5) conflicts with analytical data for other peptides.

The *C*-terminal tryptic peptide can be identified as peptide T14b, since it lacks a *C*-terminal basic residue and this corresponds to chymotryptic peptides C13b, C14a and Cle–C26e (Fig. 5). The chymotryptic peptide C30a has Leu-Arg- as its *N*-terminus and the composition of the remainder of the peptide corresponds closely to that of peptides T18b+T14b or T16b+T18c. Further, the remaining unassigned sequence Leu-Arg (found also in peptides T15b and T15c) occurs at the *C*-terminus of peptides T19e and T19d, indicating that the peptides occur in the order T19e–T18b–T14b or T19e–T16b–T18c at the *C*-terminus of segment ChC 9.

The whole chain can then be arranged as shown in Fig. 5, peptide C26c overlapping peptides T6a and T19e, an arrangement confirmed by the isolation of a peptide with the composition of peptide T6a+T19e by tryptic digestion of a type-II helical segment that had been citraconylated by the method of Dixon & Perham (1968). Although no other arrangement of peptides gives such a satisfactory correspondence of tryptic and chymotryptic sequences, it is apparent that, owing to the presence of hydrophobic residues and consequent chymotryptic cleavage immediately preceding the basic residues, the overlaps obtained were small and not readily determined with certainty. The rather low yield of peptide T21e and the high content of alanine and low content of leucine in peptides T21e+T23a as compared with peptides C39c or C39d also need explanation.

In a preliminary study of the tryptic peptides from segment ChC 9, a peptide was isolated with electrophoretic properties similar to those of peptide T21e

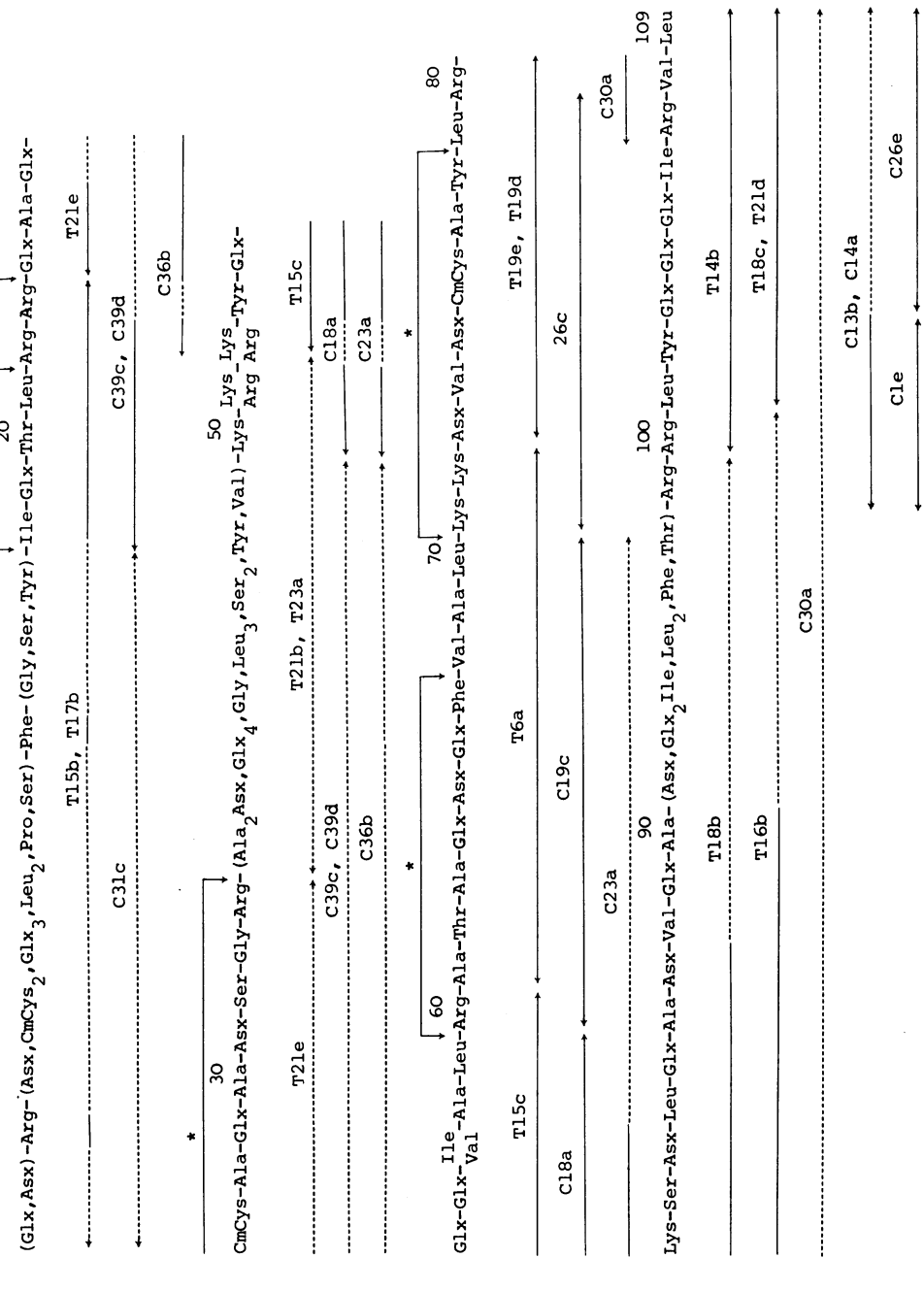

Fig. 5. *Tentative partial amino acid sequence of the major component in fraction ChC 9*
Residues not sequenced are indicated by the broken lines. Peptides identified by Corfield & Fletcher (1969) are indicated by asterisks.

and with the same *N*-terminal sequence, Glx-Ala, but with a leucine residue replacing one alanine residue. A third variant of this peptide was observed in which one alanine residue of peptide T21e is substituted by a valine residue. Finally, peptide 1T/4γ2, isolated by Corfield *et al.* (1967) from oxidized wool, had the sequence Gln-Ala-Asp-$CySO_3H$-Glu-Ala-Ser-Gly-Arg and their peptide 1T/4γ3 with similar terminal sequences contained an additional valine residue. These data suggest that the region represented by peptide T21e is highly variable; the low yield of peptide T21e and the discrepancy in amino acid content are probably in part due to this circumstance. On the basis of the sequences obtained by Corfield *et al.* (1967) we can tentatively assign the sequence Glx-Ala-Glx-CmCys-Ala-Glx-Ala-Asx-Ser-Gly-Arg to peptide T21e.

The sequence in Fig. 5 does not include a number of other peptides isolated in very low yields. Some of these may be homologues of major peptides or form part of these peptides arising from non-specific hydrolysis by trypsin or chymotrypsin. Others may have been associated with contaminants in the original preparation of segment ChC9 and may be derived from other wool proteins. None fall obviously into the pattern of the major peptides. When overlapping sequences within the group of major tryptic peptides are taken into account, the yield of peptide material ranges between 20 and 35%, except for the portion of the chain represented by peptide T21e.

The sequence of Fig. 5 corresponds to a mol.wt. of 12829, which agrees satisfactorily with the sedimentation and gel-filtration data indicating a mean mol.wt. of about 12500 for fraction ChC (Crewther & Dowling, 1971). The amino acid composition of the proposed sequence is compared with the amino acid analysis in Table 1. Differences of more than 10% are evident for alanine and valine. These and lesser discrepancies in aspartic acid, glutamic acid, lysine and arginine are partly attributable to the impurity of peptides T21e, T23a and C36b but mainly to partial substitutions involving these residues. The data (Tables 3 and 5) provide evidence for several homologous substitutions throughout the segment apart from those indicated in Fig. 5: Val for Ala at position 28, Glx for Asx at positions 75 and 87, Lys for Arg at position 99, Asx for Glx at position 103, Glx for Arg at position 107 and Ile for Val at position 108.

The presence of repeating basic residues in the sequence in several positions confirms the findings of Fell *et al.* (1960) that pairs of basic residues are present in the low-sulphur proteins of wool. Fig. 5 also shows that four peptides identified by Corfield & Fletcher (1969) in their keratose fraction U.S.3 occur in the type-II helical segment. However, most of the sequence of segment ChC 9 is not represented in their peptides and hence conclusions about the size of protein chains present in the fraction U.S.3 were premature. It seems probable that the precipitation of peptides under the conditions used by Corfield *et al.* (1967) and Corfield & Fletcher (1969) for ion-exchange chromatography of peptide preparations was responsible for the loss of a considerable proportion of this material.

An important fact arising from the sequence data is the observation that, although two of the *S*-carboxymethylcysteine residues are near the *N*-terminus and associated with residues that may well constitute a non-helical section of chain, the other two *S*-carboxymethylcysteine residues are at positions 27 and 76 and are probably in or near α-helical sections of chain. The associated disulphide bonds would be expected to have a major effect on the stability of the helical structure in the wool fibre.

References

Cole, M., Corfield, M. C., Fletcher, J. C. & Robson, A. (1965) *Proc. Int. Wool Text. Res. Conf. 3rd, Sect. 1*, 335–343

Consden, R., Gordon, A. H. & Martin, A. J. P. (1947) *Biochem. J.* **41**, 590–596

Corfield, M. C. (1963) *Biochem. J.* **86**, 125–129

Corfield, M. C. & Fletcher, J. C. (1969) *Biochem. J.* **115**, 323–334

Corfield, M. C., Fletcher, J. C., Myers, C. & Robson, A. (1965) *Proc. Int. Wool Text. Res. Conf. 3rd, Sect. 1*, 345–353

Corfield, M. C., Fletcher, J. C. & Robson, A. (1967) *Biochem. J.* **102**, 801–814

Corfield, M. C., Fletcher, J. C. & Robson, A. (1968) in *Symposium on Fibrous Proteins, Australia, 1967* (Crewther, W. G., ed.), pp. 289–298, Butterworths (Australia), Sydney

Crewther, W. G. & Dowling, L. M. (1971) *Appl. Polym. Symp. no. 18*, 1–20

Crewther, W. G. & Harrap, B. S. (1965) *Nature (London)* **207**, 295

Crewther, W. G. & Harrap, B. S. (1967) *J. Biol. Chem.* **242**, 4310–4319

Crewther, W. G., Fraser, R. D. B., Lennox, F. G. & Lindley, H. (1965) *Adv. Protein Chem.* **20**, 191–346

Dixon, H. B. F. & Perham, R. N. (1968) *Biochem. J.* **109**, 312–314

Dopheide, T. A. A. (1973) *Eur. J. Biochem.* **34**, 120–124

Dowling, L. M. & Stark, G. R. (1969) *Biochemistry* **8**, 4728–4734

Elleman, T. C. (1972*a*) *Biochem. J.* **128**, 1229–1239

Elleman, T. C. (1972*b*) *Biochem. J.* **130**, 833–845

Elleman, T. C. & Dopheide, T. A. A. (1972) *J. Biol. Chem.* **247**, 3900–3909

Fell, M., La France, N. H. & Ziegler, K. (1960) *J. Text. Inst.* **51**, T797

Fraser, R. D. B., Jones, L. N., MacRae, T. P., Rowlands, R. J. & Tulloch, P. A. (1976) *Proc. Int. Wool. Text. Res. Conf. 5th* **2**, 130–138

Goodwin, T. W. & Morton, R. A. (1946) *Biochem. J.* **40**, 628–632

Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
Haylett, T. & Swart, L. S. (1969) *Text. Res. J.* **39**, 917–929
Haylett, T., Swart, L. S. & Parris, D. (1971) *Biochem. J.* **123**, 191–200
Hosken, R., Moss, B. A., O'Donnell, I. J. & Thompson, E. O. P. (1968) *Aust. J. Biol. Sci.* **21**, 593–595
Jones, L. N. (1975) *Biochim. Biophys. Acta* **412**, 91–98
Lindley, H. & Elleman, T. C. (1972) *Biochem. J.* **128**, 859–867
Michl, H. (1951) *Monatsh. Chem.* **82**, 489–493
Naughton, M. A., Sanger, F., Hartley, B. S. & Shaw, D. C. (1960) *Biochem. J.* **77**, 149–163
O'Donnell, I. J. (1969) *Aust. J. Biol. Sci.* **22**, 471–488
O'Donnell, I. J. & Thompson, E. O. P. (1968) *Aust. J. Biol. Sci.* **21**, 385–393
Rydon, A. N. & Smith, P. W. G. (1952) *Nature (London)* **169**, 922–923
Swart, L. S. & Haylett, T. (1971) *Biochem. J.* **123**, 201–210
Swart, L. S. & Haylett, T. (1973) *Biochem. J.* **133**, 641–654
Thompson, E. O. P. & O'Donnell, I. J. (1964) *Aust. J. Biol. Sci.* **17**, 277–281
Thompson, E. O. P. & O'Donnell, I. J. (1965) *Aust. J. Biol. Sci.* **18**, 1207–1225
Thompson, E. O. P. & O'Donnell, I. J. (1967) *Aust. J. Biol. Sci.* **20**, 1001–1024