THE COMPLETE AMINO ACID SEQUENCE OF A FEATHER KERATIN FROM EMU (*DROMAIUS NOVAE-HOLLANDIAE*)

By I. J. O'DONNELL*

WITH AN ADDENDUM BY E. SUZUKI

[Manuscript received 31 August 1972]

Abstract

The complete amino acid sequence of one of the two major components comprising the rachis and calamus of emu (*D. novae-hollandiae*) feather keratin has been determined. The molecule consists of 102 amino acids and has a molecular weight (in the *S*-carboxymethyl form) of 10,459. The half-cystine residues are located towards either end of the molecule. There is an insoluble tryptic peptide of 65 residues, T3, in which the hydrophobic residues Val, Leu, and Ile are preferentially placed and this section also contains most of the Ser and Gly residues of the molecule. It is the T3 section which contains the crystalline part of the molecule. Comparison of the sequence with peptides isolated by Schroeder *et al.* (1957) from calamus from turkey feather suggests that the T3 section in both feathers could be very similar. The proline residues are not regularly distributed. The *N*-terminal sequence, NAc-Ser-SCMCys-Tyr is the same as that found previously in goose (*Anser domesticus*) feather calamus.

I. INTRODUCTION

There are four predominant bands (two major ones and two minor ones) present in the patterns obtained when reduced and carboxymethylated (SCM) proteins from the calamus or rachis from emu feather (*Dromaius novae-hollandiae*) are examined by electrophoresis on acrylamide gels in buffers containing 8M urea (see Fig. 1, O'Donnell 1973). Results in that paper showed that the two major protein bands (bands 2 and 3) are almost identical in amino acid composition, and that there is probably only one non-common peptide in the peptide map pattern of SCM-peptides produced from these two bands by the action of trypsin plus chymotrypsin.

The work in the present paper describes the determination of the amino acid sequence of one of these two major proteins (band 3). The only previous work in the field of amino acid sequence of feather proteins is the isolation and sequencing of some peptides from a partial acid hydrolysate of the calamus from white turkey feather (Schroeder *et al.* 1957), and the finding of a terminal peptide, NAc-Ser-SCMCys-Tyr, present in approximately molar amount in the protein extracts of goose (*Anser domesticus*) feather calamus (O'Donnell 1971*a*).

* Division of Protein Chemistry, CSIRO, Parkville, Vic. 3052.

The two immediate aims of the work were:

(1) To find out if there are any obvious principles or features of structure of keratin molecules common to α-keratins (e.g. wool, hair, stratum corneum) and natural β-keratins (e.g. feather, avian beak and claw, reptile claw).

(2) To help in the determination of the structure of the feather keratin molecule. The amino acid sequence of the polypeptide chain is necessary for combination with data derived from physical measurements such as X-ray diffraction patterns to provide adequate information for the calculation of the structure of the molecule. The most recent model for the feather keratin molecule is that proposed by Fraser *et al.* (1971) from X-ray diffraction and infrared measurements, and such a model must be compatible with the amino acid sequence.

II. Materials and Methods

(*a*) *Preparation of Labelled SCM-keratin from Emu Feathers*

This was prepared from either rachis or calamus of emu feather as described previously (O'Donnell 1973), the *S*-carboxymethyl residues being labelled with [2-^{14}C]iodoacetic acid. There were 9×10^7 counts per minute per μmole of *S*-carboxymethylcysteine. Pure band 3 protein was isolated using DEAE-cellulose in buffers containing 8M urea as described previously. Emu beak was extracted in the same manner.

(*b*) *Digestion of SCM-keratin* (*Band* 3) *by Trypsin*

Two μmoles of freeze-dried SCM-keratin from emu feather (band 3) were dissolved in 0·85 ml 2M ammonia and brought to pH 8·5 with approximately 5 ml 0·2M acetic acid. It was digested for 20–24 hr at 37°C with 1% by weight of TPCK trypsin.

(*c*) *Fractionation of Tryptic Peptides*

After digestion, an insoluble gel was removed by centrifugation. This gel (peptide T3, Fig. 4) was dried in a vacuum desiccator, dissolved in 1 ml 98–100% formic acid, and further purified by passage through a column of Sephadex G50 in 50% formic acid. The required effluent fractions were dried using freeze-drying equipment.

The supernatant was fractionated either by paper electrophoresis at pH 6·5 or pH 3·5 or both or by gradient elution from a column of DEAE-cellulose using an ammonium bicarbonate gradient. The peptides were recovered by freeze-drying or by drying over P_2O_5 and NaOH in a vacuum desiccator, followed by the addition of a small amount of water and re-drying.

(*d*) *Enzymic Digestion of Tryptic Peptides*

Tryptic peptides were further digested at 37°C using an enzyme: peptide ratio of 1–2% (w/w). Soluble tryptic peptides were digested with chymotrypsin in 2% ammonium bicarbonate, or thermolysin in ammonium acetate (0·2N with respect to acetic acid) at pH 8·5 containing 5 mM $CaCl_2$ (Ambler and Meadway 1968). The insoluble peptide, T3, was difficult to digest, and all digestions of it were for 16 hr with vigorous stirring. When digesting with chymotrypsin or thermolysin the peptide was stood for 2–3 min in 0·85 ml 0·2N ammonia containing 25–50 μl triethylamine. This procedure usually dissolved it, and it was then brought to pH 8·5 with 2N acetic acid, and the enzyme added. With thermolysin this solubilizing procedure was not strictly necessary since the T3 in the insoluble state would slowly dissolve with vigorous stirring in the presence of thermolysin. Peptide T3 was also digested with papain (50 μg) in 1 ml pyridine acetate buffer at pH 6·5 (the one used for paper electrophoresis) in the presence of 2 μl mercaptoethanol. For digestion with pepsin the T3 was dissolved in 200 μl 98–100% formic acid, 3·8 ml water was then added to give a final acid concentration of approximately 5%, and 2% of pepsin was added. Digestion of whole band 3 protein with thermolysin and chymotrypsin to get overlap peptides was carried out conventionally.

(*e*) *Partial Acid Hydrolysis of Tryptic Peptide, T3*

Peptide T3 was dissolved in 0·37 ml 98–100% formic acid and 7·7 ml of water was added. The test tube was sealed and incubated at 105–110°C for 19 hr. The contents were then freeze-dried.

(*f*) *Fractionation of Enzymic Digests and Partial Acid Hydrolysates*

Enzymic digests of the soluble tryptic peptides were fractionated by high-voltage paper electrophoresis at pH 6·5 or pH 3·5.

In the case of enzymic and partial acid hydrolysates of the insoluble tryptic peptide T3 (and also of a thermolysin digest of the whole band 3 protein) the digests were initially fractionated on five connected columns (each 120 cm by 0·58 cm int. diam.) of Sephadex G25 Superfine in pyridine–1N ammonia (7 : 3 v/v) as described by Rees *et al.* (1970). The flow rate was 6 ml/hr and 3·0-ml fractions were collected. 100-μl samples were taken from each tube, dried, and subjected to paper electrophoresis (pH 3·5) to monitor the separation. Peptides were located by staining with ninhydrin and by radioautography, and appropriate fractions pooled and concentrated by rotary evaporation. Pyridine–acetate buffers were removed by repeated drying in a desiccator over P_2O_5 and NaOH. The peptides in these fractions were further purified, where necessary, by paper electrophoresis at pH 6·5, 3·5, or 1·9 and occasionally also by paper chromatography in butan-1-ol–acetic acid–water–pyridine (15 : 3 : 12 : 10 by volume).

(*g*) *Electrophoretic Mobilities of Peptides*

Acid fuchsin (red), xylene cyanol FF (blue), and ε-DNP-lysine were used as markers during paper electrophoresis. The mobilities at pH 6·5 were measured from the position of the neutral amino acids and are expressed relative to aspartic acid, its value being taken as −1. Those at pH 3·5 are measured from ε-DNP-lysine and are expressed relative to lysine, its value being taken as +1 (cf. Milstein *et al.* 1968).

(*h*) *Designation of Amide Groups*

Designation of side-chain carboxyl groups in a peptide as being in the acid or amide form was based on the mobility of the peptide at pH 6·5 and the graphs of Offord (1966) relating mobility at pH 6·5 to size and charge of a peptide.

(*i*) *Amino Acid Sequence Determinations*

These were performed by the "dansyl"–Edman methods (Gray 1967) with identification of the dansyl amino acids on polyamide sheets (Woods and Wang 1967).

(*j*) *Other Methods*

Liquid scintillation counting, paper electrophoresis, radioautography, and amino acid analyses were carried out as described previously (O'Donnell 1971*b*). The tritium-labelling method of labelling *C*-terminal amino acids was used as described by Matsuo *et al.* (1966). The labelled amino acids were identified after separation by paper electrophoresis at pH 1·9 followed by chromatography in 2-methoxyethanol–formic acid (98–100%)–butan-2-ol–water (2 : 1 : 8 : 3 v/v) (Eager and Savige 1963).

III. Results

The molecule of SCM-keratin (band 3) from emu feather contains four arginine residues and there are five tryptic peptides. They have been called in order T1, T2, T3, T4, and T5 starting from the amino terminal end of the molecule. Subfragments of these peptides, using other enzymes are then numbered from the *N*-terminus of that peptide. C, P, Pep, Th, and PF indicate peptides that have been produced by chymotrypsin, papain, pepsin, thermolysin, and partial acid hydrolysis with dilute formic acid respectively.

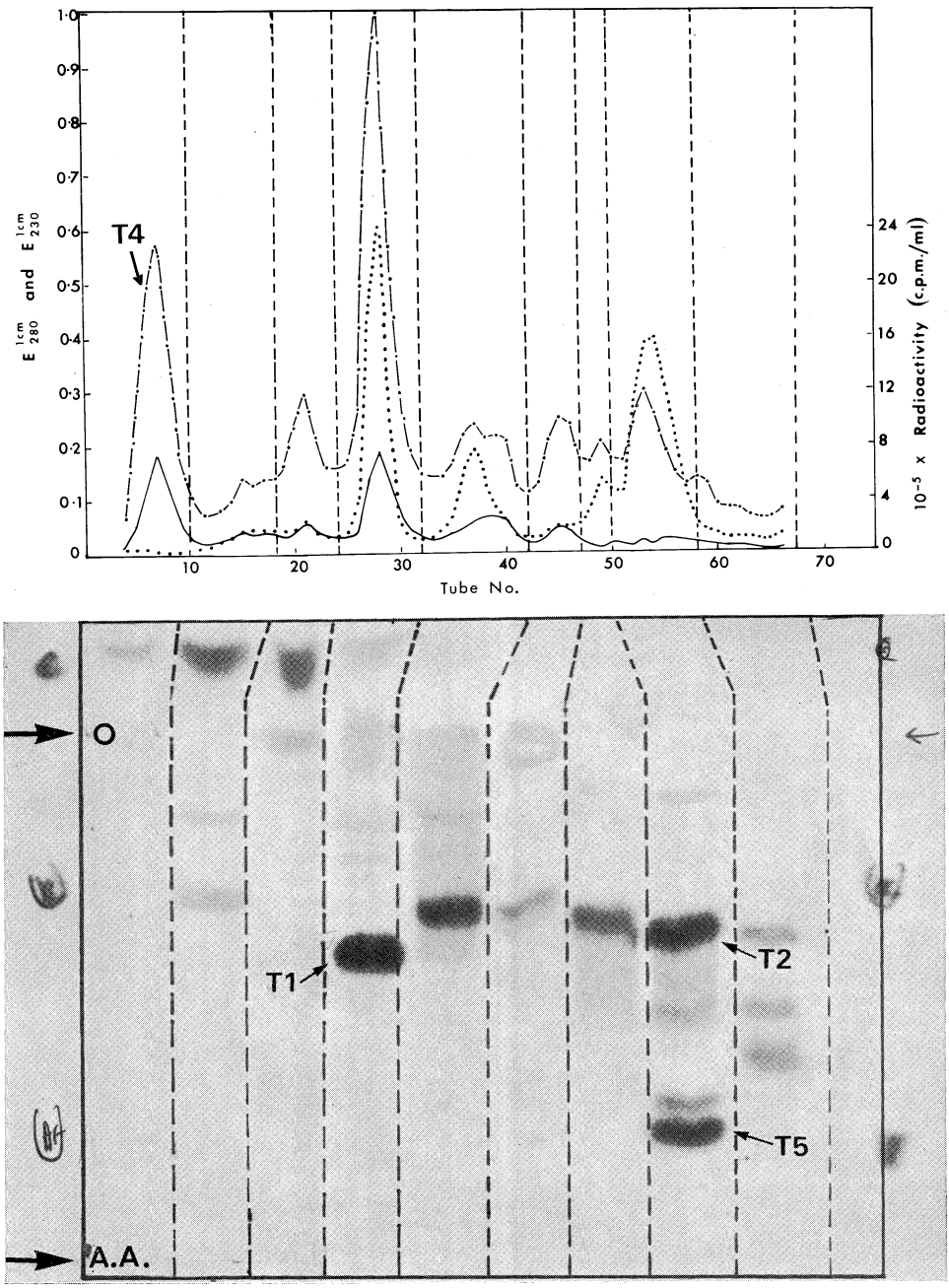


Fig. 1.—Elution diagram showing the fractionation on DEAE-cellulose of the supernatant from the tryptic digestion of SCM-keratin extracted from whole emu calamus. The column eluates were combined to give 10 separate fractions as indicated. An aliquot of each fraction was then subjected to paper electrophoresis at pH 6·5 and the locations of the peptides containing *S*-carboxymethylcysteine determined by radioautography. The column (15 cm by 0·9 cm int. diam.) was packed with DEAE-cellulose (Eastman–Kodak) and operated at 25°C. The column was equilibrated with 0·005M

Band 3 protein seemed to produce the same peptides whether isolated from the rachis or the calamus.

(*a*) *Tryptic Cleavage of SCM-keratin* (*Band* 3)

It was found that a time of 16–24 hr of digestion with trypsin was necessary to get complete breakage of the Arg-SCMCys bond at position 98–99 (see Fig. 4) of the SCM-keratin. The percentage of the original total radioactive counts present in the soluble peptides (T1, T2, T4, T5) was 88% and that in the precipitate (T3) was 10·4%. This is in agreement with the amino acid sequence which shows seven SCMCys residues in the soluble peptides and one SCMCys residue in the insoluble one.

The soluble tryptic peptides were separated and purified by paper electrophoresis at pH 6·5. They could also be purified on DEAE-cellulose using a salt gradient of ammonium bicarbonate as shown in Figure 1. The insoluble tryptic peptide fraction T3 when purified on Sephadex (Fig. 2) emerged as a predominantly single peak. This purification in 50% formic acid converted the *N*-terminal glutamine to pyrrolid-2-one-5-carboxylic acid (PCA).

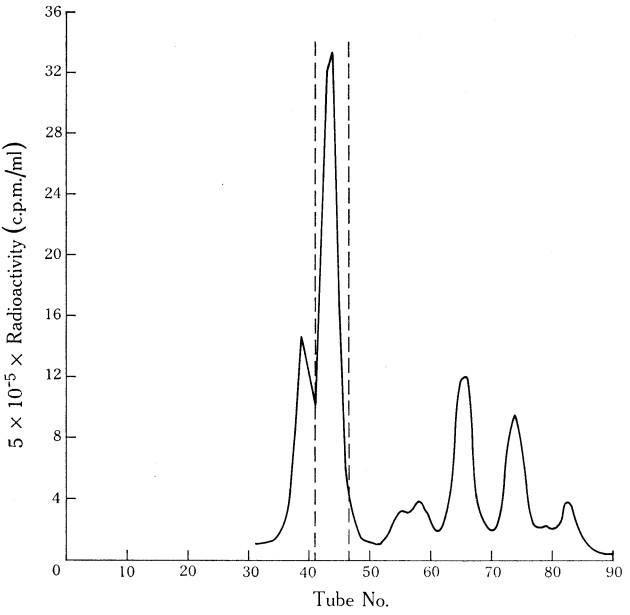


Fig. 2.—Elution diagram of T3 dissolved in 1 ml 98–100% formic acid from a column (145 cm by 0·92 cm int. diam.) of Sephadex G50 in 50% formic acid. 35 drops (*c.* 1·0 ml) were collected per fraction on an LKB Ultrorac collector. Flow rate was *c.* 3 ml/hr. 50 μl of each fraction was taken with 0·45 ml water and 5 ml scintillation mixture to count the radioactivity. The tubes between the dotted lines were collected for sequence studies on peptide T3.

(*b*) *Amino Acid Composition of Tryptic Peptides*

The amino acid composition of the purified peptides are shown in Table 1 together with values from the completed sequence.

NH_4HCO_3 and the peptide mixture loaded in this buffer. A linear gradient consisting of 150 ml each of 0·005M NH_4HCO_3 and 0·2M NH_4HCO_3 was applied after tube 10. Fractions of 80 drops (5·0 ml) were collected using an LKB Ultrorac collector. The unnamed band in the tube 30–40 area probably arose from band 2 protein. —E_{280}^{1cm}. — · — E_{230}^{1cm}. ··· Radioactivity. O, origin; A.A., aspartic acid.

(c) *Amino Acid Sequence of Tryptic Peptides T1–T5*

The tryptic peptides are considered in turn and the evidence for the assigned sequence described. Residues identified as dansyl derivatives are marked by arrows pointing to the right. Arrows pointing to the left indicate release by carboxypeptidase or identification by the tritium-labelling method.

TABLE 1

AMINO ACID COMPOSITIONS OF TRYPTIC PEPTIDES OF SCM-KERATIN (BAND 3)

The fragments were purified as described in the text and then hydrolysed in 6N HCl for 24 hr. Values are given as mole per mole of each fragment uncorrected for losses or incomplete hydrolysis (except for T3—see footnote). Also shown are the amino acid compositions obtained from the completed sequence. H, hydrolysate; S, sequence

Amino acid	T1		T2		T3*		T4		T5	
	H	S	H	S	H	S	H	S	H	S
Arginine	0·9	1	1·0	1	1·0	1	1·0	1		
SCM-cysteine	1·9	2	2·3	3	1·0	1			1·5	2
Aspartic acid	1·0		1·9		3·0	1				
Asparagine		1		2		2				
Threonine			0·9	1	3·6	4				
Serine	0·9	1	2·5	3	13·3	14	1·0	1		
Glutamic acid			1·0	1	4·7	1			0·3	
Glutamine						4				
Proline	2·3	2	3·2	3	6·9	6			0·9	1
Glycine			1·0	1	8·9	9	1·2	1		
Alanine			1·0	1	2·8	2	1·1	1		
Valine					9·3	9				
Isoleucine					3·9	4				
Leucine	1·0	1	1·5	2	5·6	5			1·0	1
Tyrosine	1·0	1					0·9	1		
Phenylalanine			0·9	1	2·0	2				
Total		9		19		65		5		4
Net No. of charges	2−		3− or 4−				1+		2−	
Mobility at pH 6·5	−0·53		−0·46		Insoluble		+0·42		−0·80	
Method of purification†	$E_{6·5}$		$E_{6·5}$		Sephadex in 50% formic acid		$E_{6·5}$		$E_{6·5}$	

* Hydrolyses were carried out for 24, 48, 72 hr and values for leucine, isoleucine, and valine were extrapolated to infinite time and those for serine and threonine to zero time.

† $E_{6·5}$, paper ionophoresis at pH 6·5.

(i) *Peptide T1*

This was ninhydrin-negative and therefore had a blocked amino terminus. Chymotrypsin released T1-C1 and T1-C2 (Tables 2 and 3). This together with the identification of NAc-Ser-SCMCys-Tyr as the amino terminal sequence of goose calamus suggested that the same amino terminal sequence exists here.

TABLE 2

AMINO ACID COMPOSITION OF SOME PEPTIDES PRODUCED FROM THE TRYPTIC PEPTIDES T1 AND T2 BY THE ACTION OF CHYMOTRYPSIN OR THERMOLYSIN

The peptides were separated by paper ionophoresis at pH 6·5 ($E_{6\cdot5}$) and pH 3·5 ($E_{3\cdot5}$) or both. Some had an additional purification of being oxidized on the paper after electrophoresis at pH 3·5 (Brown and Hartley 1966) followed by electrophoresis at pH 3·5 ($E_{ox.,\ 3\cdot5}$) or 1·9 ($E_{ox.,\ 1\cdot9}$). Values are not corrected for losses during hydrolysis and are given as moles per mole of peptide, with the final value from the sequence in parenthesis. Hydrolyses were carried out at 105–110°C for 24 hr

Peptide*	T1-C1	T1-C2	T1-Th1	T2-C1	T2-C2	T2-Th1	T2-Th2
Arginine		1·0 (1)			0·9 (1)		
SCM-cysteine	1·0 (1)	+ (1)	0·7 (1)	+ (1)	+ (2)	1·0 (1)	1·5 (2)
Aspartic acid Asparagine		0·9 (1)			2·0 (2)		1·9 (2)
Threonine				1·1 (1)		1·0 (1)	
Serine	0·9 (1)		1·0 (1)	1·7 (2)	1·0 (1)	2·0 (2)	1·3 (1)
Glutamic acid Glutamine					1·0 (1)		1·1 (1)
Proline		1·9 (2)		2·0 (2)	+ (1)	2·0 (2)	1·0 (1)
Glycine			0·3	1·0 (1)		1·1 (1)	0·8† (0)
Alanine					1·3 (1)		1·2 (1)
Valine							
Isoleucine							
Leucine		1·0 (1)		1·0 (1)	0·6 (1)		1·4 (1)
Tyrosine	0·9 (1)						
Phenylalanine					0·6 (1)		
Total	3	6	2	8	11	7	9
Mobility at pH 6·5	−0·82	0	−1·11	−0·30	−0·40	−0·40	−0·70
Net No. of charges	2−	0	2−	1−	2−	1−	3−
Method of purification	$E_{6\cdot5}$ $E_{3\cdot5}$	$E_{6\cdot5}$ $E_{3\cdot5}$	$E_{6\cdot5}$	$E_{6\cdot5}$ $E_{3\cdot5}$ $E_{ox.,\ 3\cdot5}$ $E_{ox.,\ 1\cdot9}$	$E_{6\cdot5}$ $E_{3\cdot5}$ $E_{ox.,\ 3\cdot5}$ $E_{ox.,\ 1\cdot9}$	$E_{6\cdot5}$	$E_{6\cdot5}$

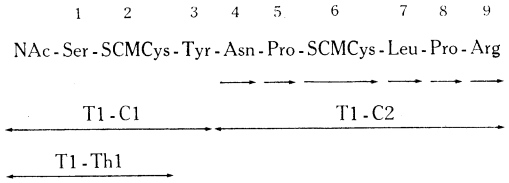
* See Table 3 for amino acid sequence in these peptides.

† Glycine probably arose from the breakdown of SCM-cysteine during hydrolysis.

The mobility at pH 6·5 of −0·53 for T1 and 0 for T1-C2 showed that Asx(4) is Asn(4).* These considerations together with data given in Tables 2 and 3 showed

* Numbers in parenthesis in running text indicate the residue number in the sequence.

the sequence of T1 to be:



(ii) *Peptide T2*

A direct dansyl–Edman determination gave Ser-Ser-SCMCys-Gly-Pro(Thr, Pro, Leu, Ala, Asx, Ser, SCMCys, Asx, Glx, Pro, SCMCys, Leu, Phe, Arg). Chymotrypsin gave two peptides T2-C1 and T2-C2 and thermolysin gave peptides T2-Th1 and T2-Th2 (Tables 2 and 3). When determining the sequence of T2-Th2 the mobility

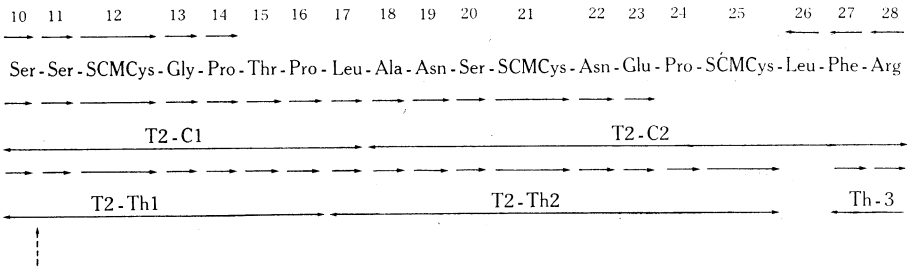
TABLE 3

AMINO ACID SEQUENCE OF PEPTIDES ISOLATED AS IN TABLE 2

In this table and also in Tables 5, 7, and 9, arrows above the residues pointing to the right indicate those residues identified by the "dansyl" method, whilst those pointing to the left indicate release by carboxypeptidase or identification by the tritium-labelling method

Peptide	Sequence
T1-C1	NAc-Ser-SCMCys-Tyr
T1-C2	Asn-Pro-SCMCys-Leu-Pro-Arg
T1-Th1	NAc-Ser-SCMCys
T2-C1	Ser-Ser-SCMCys-Gly-Pro-Thr-(Pro, Leu)
T2-C2	Ala-Asn-Ser-SCMCys-Asn-Glu-(Pro, SCMCys, Leu, Phe, Arg)
T2-Th1	Ser-Ser-SCMCys-Gly-Pro-Thr-Pro
T2-Th2	Leu-Ala-Asn-Ser-SCMCys-Asn-Glu-Pro-SCMCys

at pH 6·5 after the removal of Asx(19) was −0·90, and therefore Asx(19) is Asn(19). After removal of Asx(22) the mobility was −0·94. This shows that Asx(22) is Asn(22) and Glx(23) is Glu(23). A mixture of carboxypeptidases A and B on T2 released Leu, Phe, and Arg. It is shown later (Table 8) in the radioactive thermolysin peptide Th3 from the whole band 3 that Phe is next to Arg. The amino acid sequence of T2 is therefore:



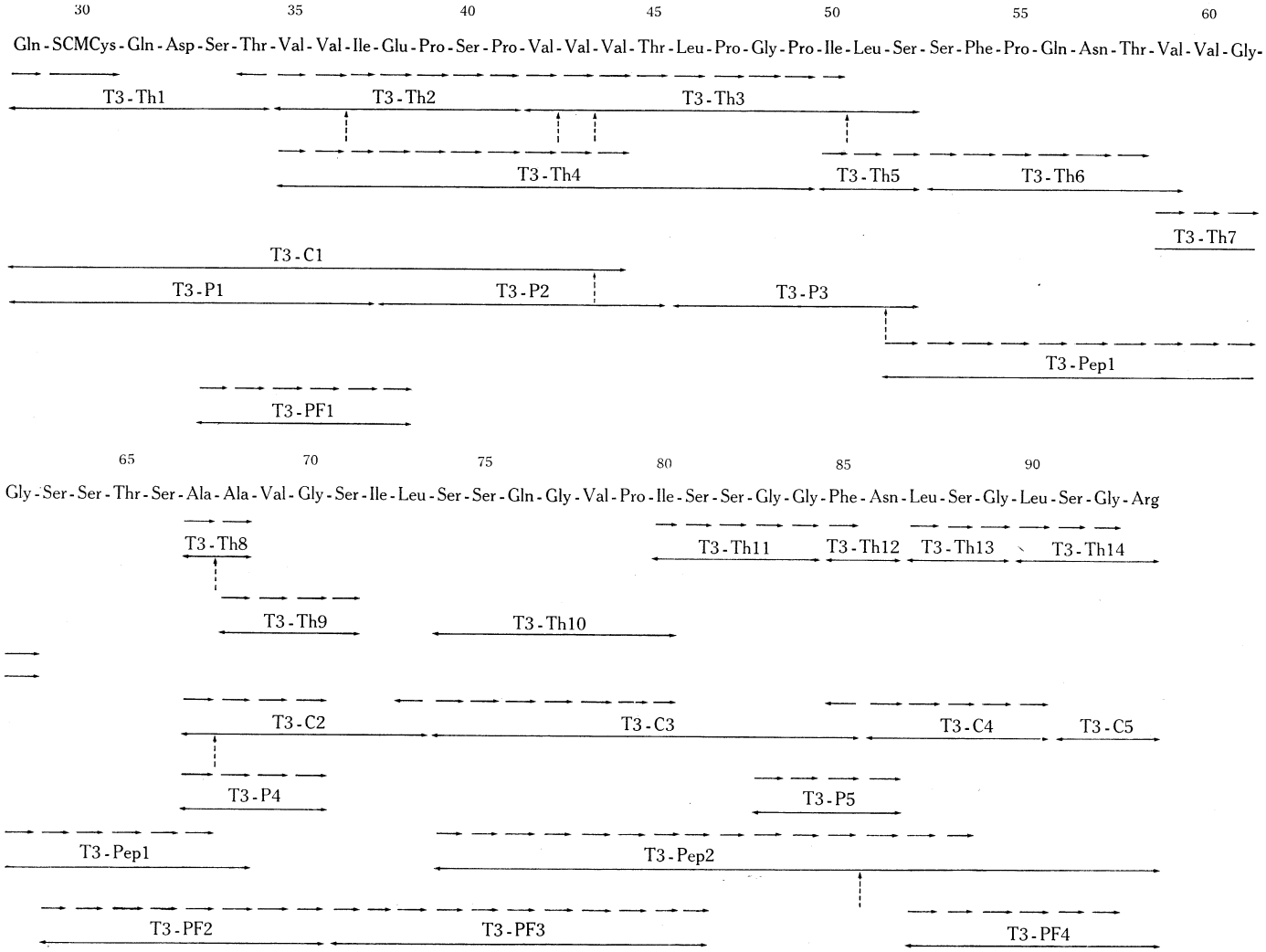


Fig. 3.—Amino acid sequences of peptides (Tables 4–7) obtained from tryptic peptide T3 (65 residues) by digestion with thermolysin (e.g. T3-Th1), chymotrypsin (e.g. T3-C1), papain (e.g. T3-P1), or pepsin (e.g. T3-Pep1) and also by partial acid hydrolysis with 5% formic acid (e.g. T3-PF1). The dotted arrows show some positions of partial splits.

(iii) *Peptide T3*

This peptide was insoluble in aqueous solutions in the pH range 1–11. It could be dissolved in solutions containing triethylamine and in 100% formic acid. It was purified by dissolving it in 100% formic acid and then passing it through a column of Sephadex G50 in 50% formic acid as shown in Figure 2. The elution volume of T3, compared with those of insulin and the A chain from insulin, suggested that T3 was bigger than either of these molecules and is consistent with a molecular weight of 6346 determined from the amino acid sequence of T3.

Examination of the completed sequence of T3 (Fig. 4) shows no distinguishing features which could have made its determination any easier. Furthermore the *N*-terminal glutamine ring-closed to PCA during passage through the Sephadex in 50% formic acid. The sequence was finally determined by the isolation and sequencing of peptides from many enzyme digests and also from peptides isolated from partial acid hydrolysates. The peptides which were found useful can be seen in Figure 3 and Tables 4–7. The peptides which did not contain SCMCys residues were not isolated from T3 prepared from pure band 3 protein but from T3 prepared by tryptic digest of the whole extract of emu rachis or calamus. The fact that no ambiguity was found shows that this region (35–93) is identical in the two major proteins (i.e. band 2 and band 3 proteins) of either calamus or rachis.

The peptides discussed in the following tabulation enable identification of the amides and carboxylic acid side-chains:

Peptide	Mobility at pH 6·5	Conclusions
T3-C3	0	Therefore Glx(76) is Gln(76). Confirmed by zero mobility at pH 6·5 of peptide T3-Th10
T3-C4	0	Therefore Asx(86) is Asn(86). Confirmed by zero mobility at pH 6·5 of peptide T3-P5
T3-P2	−0·34	Suggests it has one negative charge. Therefore Glx(38) is Glu(38)
T3-Pep1	0	Therefore Glx(56) and Asx(57) are Gln(56) and Asn(57)
Th3 [see Table 8 and Section III(*d*)]	−0·28	Sequence as follows: 27 28 29 30 31 32 33 34 Phe-Arg-Gln-SCMCys-Gln-Asp-Ser-Thr Mobility corresponds to one negative charge. After the Phe(27), Arg(28), and Gln(29) had been removed by Edman degradation the remaining peptide had a mobility at pH 6·5 of −0·75 which suggests the remaining peptide has two negative charges. Therefore Glx(29) is Gln(29). After the SCMCys(30) and Gln(31) had been removed, the remaining peptide had a mobility of −0·70 and this suggests that Asx(32) is Asp(32) and Glx(31) is Gln(31).

(iv) *Peptide T4*

A direct dansyl–Edman determination gave

$$\xrightarrow{}\ \xrightarrow{}\ \xrightarrow{}\ \xrightarrow{}\ \xrightarrow{}$$
Tyr-Ser-Gly-Ala-Arg

(v) *Peptide T5*

A direct dansyl–Edman determination gave

$$\xrightarrow{\qquad}\ \xrightarrow{}\ \xrightarrow{}\ \xrightarrow{\qquad}$$
SCMCys-Leu-Pro-SCMCys

TABLE 4

AMINO ACID COMPOSITION OF SOME PEPTIDES OBTAINED BY DIGESTION OF TRYPTIC PEPTIDE T3 WITH CHYMOTRYPSIN (C), PEPSIN (Pep), PAPAIN (P), OR BY PARTIAL ACID HYDROLYSIS IN DILUTE FORMIC ACID (PF)

The peptides were initially separated on a 600-cm column of Sephadex G25 and then by paper electrophoresis at pH 6·5 or pH 3·5. Sometimes they had an additional purification in butanol–acetic acid–water–pyridine. Hydrolyses were done in 6N HCl at 105–110°C for 24 hr. Values are not corrected for losses during hydrolysis and are given as moles per mole of peptide with the final value from the sequence in parenthesis

	T3-C1	T3-C2	T3-C3	T3-C4	T3-C5	T3-P1	T3-P2	T3-P3	T3-P4*	T3-P5	T3-Pep1	T3-Pep2	T3-PF1	T3-PF2	T3-PF3*	T3-PF4
Arginine					1·0 (1)							+ (1)				1·0 (1)
SCM-cysteine	1·0 (1)					0·8 (1)										
Aspartic acid	1·0 (1)		0·1	1·0		1·0 (1)				1·1	1·2	1·1				
Asparagine				(1)						(1)	(1)	(1)				
Threonine	0·9 (1)		0·2			0·9 (1)	0·8 (1)			0·2	1·9 (2)		0·9 (1)	0·8 (1)		
Serine	1·7 (2)	0·9 (1)	3·6 (4)	1·0 (1)	1·0 (1)	1·0 (1)	1·0 (1)	1·0 (1)		0·4	4·8 (5)	5·6 (6)	1·0 (1)	2·5 (3)	(4)	2·0 (2)
Glutamic acid	3·0 (1)		1·0			2·0	1·0 (1)				1·5	1·1	1·1 (1)			
Glutamine	(2)		(1)			(2)					(1)	(1)			(1)	
Proline	2·2 (2)		1·0 (1)				2·0 (2)	2·2 (2)			+ (1)	1·0 (1)			(1)	
Glycine		1·0 (1)	3·1 (3)	1·1 (1)	1·5 (1)		0·1	1·1 (1)	(1)	1·8 (2)	2·2 (2)	5·5 (5)	0·6	1·2 (1)	(1)	3·0 (2)
Alanine		1·4 (2)							(2)		1·7 (2)			2·0 (2)		
Valine	4·6 (5)	0·7 (1)	1·1 (1)			1·6 (2)	1·8† (3)		(1)		1·8 (2)	1·1 (1)	1·8 (2)	1·0 (1)	(1)	
Isoleucine	1·0 (1)	0·8 (1)	0·9 (1)			1·0 (1)		1·0 (1)				1·0 (1)	1·0 (1)		(2)	
Leucine		1·0 (1)		2·2 (2)				2·0 (2)				2·0 (2)			(1)	1·9 (2)
Tyrosine					0·3											
Phenylalanine			0·9 (1)							1·0 (1)	1·0 (1)	1·0 (1)				
Total	16	7	12	5	3	9	8	7	4	4	17	20	6	8	11	7

* Analysis lost. † Incomplete hydrolysis of Val-Val bond in 24 hr.

(*d*) *Overlap Peptides for the Tryptic Peptides T1–T5*

Isolation of the radioactive peptides (i.e. those containing SCMCys) in a thermolysin digest of the whole SCM-keratin (band 3) gave four of the five overlap peptides required to join the tryptic peptides T1–T5 in their correct order. These peptides and their sequences are listed in Tables 8 and 9. The overlap peptide required to join T3 and T4 was obtained from a chymotryptic digest of band 3 by paper electrophoresis at pH 6·5. It was

$$\overset{91\quad 92\quad 93\quad 94}{\longrightarrow\ \longrightarrow\ \longrightarrow\ \longrightarrow}$$
$$\text{Ser-Gly-Arg-Tyr}$$

TABLE 5

AMINO ACID SEQUENCES OF PEPTIDES FROM TRYPTIC PEPTIDE T3 ISOLATED AS IN TABLE 4

Peptide	Mobility at pH 6·5	Net No. of charges	Mobility at pH 3·5	Method of purification*	Sequences
T3-C1	−0·68†	4−	—	S, $E_{6\cdot5}$	Gln-SCMCys-Gln-Asp-Ser-Thr-Val-Val-Ile-Glu-Pro-Ser-Pro-Val-Val-Val
T3-C2	0	0	+0·28	S, $E_{3\cdot5}$	Ala-Ala-Val-Gly-Ser-Ile-Leu
T3-C3	0	0	+0·13	S, $E_{3\cdot5}$	Ser-Ser-Gln-Gly-Val-Pro-Ile-Ser-Ser-Gly-Gly-Phe
T3-C4	0	0	+0·26	S, $E_{6\cdot5}$, $E_{3\cdot5}$	Asn-Leu-Ser-Gly-Leu
T3-C5	+0·61	+1	—	S, $E_{6\cdot5}$	Ser-Gly-Arg
T3-P1	−0·80†	3−	−0·33	S, $E_{6\cdot5}$	Gln-SCMCys-Gln-Asp-Ser-Thr-Val-Val-Ile
T3-P2	−0·34	1−	+0·10	S, $E_{6\cdot5}$	Glu-Pro-Ser-Pro-Val-Val-Val-Thr
T3-P3	0	0	+0·17	S, $E_{3\cdot5}$, C_B	Leu-Pro-Gly-Pro-Ile-Leu-Ser
T3-P4	0	0	+0·31	S, $E_{3\cdot5}$	Ala-Ala-Val-Gly
T3-P5	0	0	+0·23	S, $E_{3\cdot5}$	Gly-Gly-Phe-Asn
T3-Pep1	0	0	+0·28	S, $E_{3\cdot5}$, C_B	Ser-Ser-Phe-Pro-Gln-Asn-Thr-Val-Val-Gly-Gly-Ser-Ser-Thr-Ser-Ala-Ala
T3-Pep2	—	—	+0·34	S, $E_{3\cdot5}$, C_B	Ser-Ser-Gln-Gly-Val-Pro-Ile-Ser-Ser-Gly-Gly-Phe-Asn-Leu-Ser-Gly-Leu-Ser-Gly-Arg
T3-PF1	—	—	+0·26	S, $E_{3\cdot5}$	Ser-Thr-Val-Val-Ile-Glu
T3-PF2	—	—	+0·29	S, $E_{3\cdot5}$	Ser-Ser-Thr-Ser-Ala-Ala-Val-Gly
T3-PF3	—	—	+0·23	S, $E_{1\cdot9}$	Ser-Ile-Leu-Ser-Ser-Gln-Gly-Val-Pro-Ile-Ser
T3-PF4	+0·41	1+	—	S, $E_{3\cdot5}$	Leu-Ser-Gly-Leu-Ser-Gly-Arg

* S, separation on 600-cm column of Sephadex G25; $E_{6\cdot5}$, $E_{3\cdot5}$, paper electrophoresis at pH 6·5 or 3·5 respectively; C_B, additional purification by paper chromatography in butanol–acetic acid–water–pyridine.

† The *N*-terminal glutamine is in the pyrrolidonyl carboxylyl form.

(*e*) *Complete Sequence of Band 3*

The complete amino acid sequence of band 3 protein from emu feather keratin is shown in Figure 4 and a comparison of this with the amino acid composition of the original band 3 protein is made in Table 10. The values are in good agreement. The only discrepancies found relate to serine and SCM-cysteine, which are subject

TABLE 6

AMINO ACID COMPOSITION OF SOME PEPTIDES OBTAINED BY DIGESTION OF TRYPTIC PEPTIDE T3 WITH THERMOLYSIN

The peptides were initially separated on a 600-cm column of Sephadex G25 and then by paper electrophoresis at pH 6·5 or pH 3·5. Sometimes they had an additional purification in butanol–acetic acid–water–pyridine. Hydrolyses were done in 6N HCl at 105–110°C for 24 hr. Values are not corrected for losses during hydrolysis and are given as moles per mole of peptide with the final value from the sequence in parenthesis

Amino acid	T3-Th1	T3-Th2	T3-Th3	T3-Th4	T3-Th5	T3-Th6	T3-Th7	T3-Th8	T3-Th9	T3-Th10	T3-Th11	T3-Th12	T3-Th13	T3-Th14
Arginine														0·8 (1)
SCM-cysteine	0·6 (1)													
Aspartic acid	1·0 (1)					1·0						1·0		
Asparagine						(1)						(1)		
Threonine	0·9 (1)		1·0 (1)	0·9 (1)		1·0 (1)								
Serine	1·0 (1)	1·2 (1)	1·1 (1)	1·0 (1)	0·9 (1)	1·7 (1)			1·3 (1)	2·6 (2)	1·9 (2)		0·9 (1)	1·0 (1)
Glutamic acid	1·8	1·0 (1)		0·9 (1)		1·1				1·0				
Glutamine	(2)					(1)				(1)				
Proline		2·1 (2)	2·2 (2)	2·1 (4)		1·3 (1)				1·0 (1)				
Glycine		0·3	1·2 (1)	1·0 (1)	0·3	0·7	1·0 (2)		1·4 (1)	1·5 (1)	2·0 (2)		1·1 (1)	1·2 (1)
Alanine								1·0 (2)	0·9 (1)	0·5			0·1	
Valine		1·0* (2)	2·0* (3)	3·0 (5)		0·6 (1)	1·1 (2)		1·0 (1)	1·3 (1)				
Isoleucine		0·9 (1)	1·0 (1)	1·0 (1)	0·9 (1)					0·9 (1)	1·0 (1)			
Leucine			2·0 (2)	1·0 (1)	1·1 (1)								1·0 (1)	1·0 (1)
Tyrosine					0·2									
Phenylalanine						1·0 (1)						1·0 (1)		
Total	6	7	11	15	3	7	4	2	4	7	5	2	3	4

* Low value due to difficulty of hydrolysis of Val–Val bond.

to destruction on hydrolysis, and to valine and isoleucine where corrections for incomplete hydrolysis may have been excessive. The two discrepancies which deserve comment are the presence of one more proline and 0·6 more alanine in the original analysis than in the final sequence. It is probable that these are partially substituted at one or more other residues in the chain to account for this fact. In particular there seems to be a partial alanine substitution in the neighbourhood of residue 75.

TABLE 7

AMINO ACID SEQUENCES OF PEPTIDES FROM TRYPTIC PEPTIDE T3 ISOLATED AS IN TABLE 6

Peptide	Mobility at pH 6·5	Net No. of charges	Mobility at pH 6·5	Method of purification*	Sequence
T3-Th1	−0·69†	2−	—	S, $E_{6\cdot5}$	Gln-SCMCys-Gln-Asp-Ser-Thr
T3-Th2	—	—	+0·14	S, $E_{3\cdot5}$	Val-Val-Ile-Glu-Pro-Ser-Pro
T3-Th3	—	—	+0·10	S, $E_{3\cdot5}$	Val-Val-Val-Thr-Leu-Pro-Gly-Pro-Ile-Leu-Ser
T3-Th4	—	—	+0·17	S, $E_{3\cdot5}$	Val-Val-Ile-Glu-Pro-Ser-Pro-Val-Val-Val-Thr-Leu-Pro-Gly-Pro
T3-Th5	—	—	+0·23	S, $E_{3\cdot5}$	Ile-Leu-Ser
T3-Th6	—	—	+0·18	S, $E_{3\cdot5}$	Ser-Phe-Pro-Gln-Asn-Thr-Val
T3-Th7	—	—	+0·33	S, $E_{3\cdot5}$	Val-Val-Gly-Gly
T3-Th8	—	—	+0·39	S, $E_{3\cdot5}$	Ala-Ala‡
T3-Th9	—	—	+0·17	S, $E_{3\cdot5}$	Ala-Val-Gly-Ser
T3-Th10	0	0	+0·21	S, $E_{6\cdot5}$,$E_{3\cdot5}$	Ser-Ser-Gln-Gly-Val-Pro-Ile
T3-Th11	—	—	+0·21	S, $E_{3\cdot5}$	Ile-Ser-Ser-Gly-Gly
T3-Th12	—	—	+0·32	S, $E_{3\cdot5}$	Phe-Asn
T3-Th13	—	—	+0·27	S, $E_{3\cdot5}$	Leu-Ser-Gly
T3-Th14	+0·52	+1	+0·65	S, $E_{6\cdot5}$	Leu-Ser-Gly-Arg

* Abbreviations as defined in Table 5.

† Value as −0·94 in the pyrrolidonyl carboxylyl form and then it had three negative charges.

‡ Had same mobility on pH 3·5 electrophoresis as Ala-Ala.

Schroeder *et al.* (1957) subjected the calamus of white turkey feather to partial acid hydrolysis and isolated several peptides, some of which they sequenced.

Possible positions of these peptides are marked on Figure 4 together with the numbers given to them by Schroeder *et al.* (1957). It is seen that they mostly came from the T3 region.

Table 11 lists the amino acids present in the insoluble T3 section and those present collectively in the soluble peptides T1, T2, T4, and T5. It can be seen that serine, threonine, and glycine are concentrated in the T3 section, and the hydrophobic amino acids valine and isoleucine reside exclusively there. On the other hand

SCM-cysteine exists almost exclusively in the smaller soluble peptides at either end of the molecule.

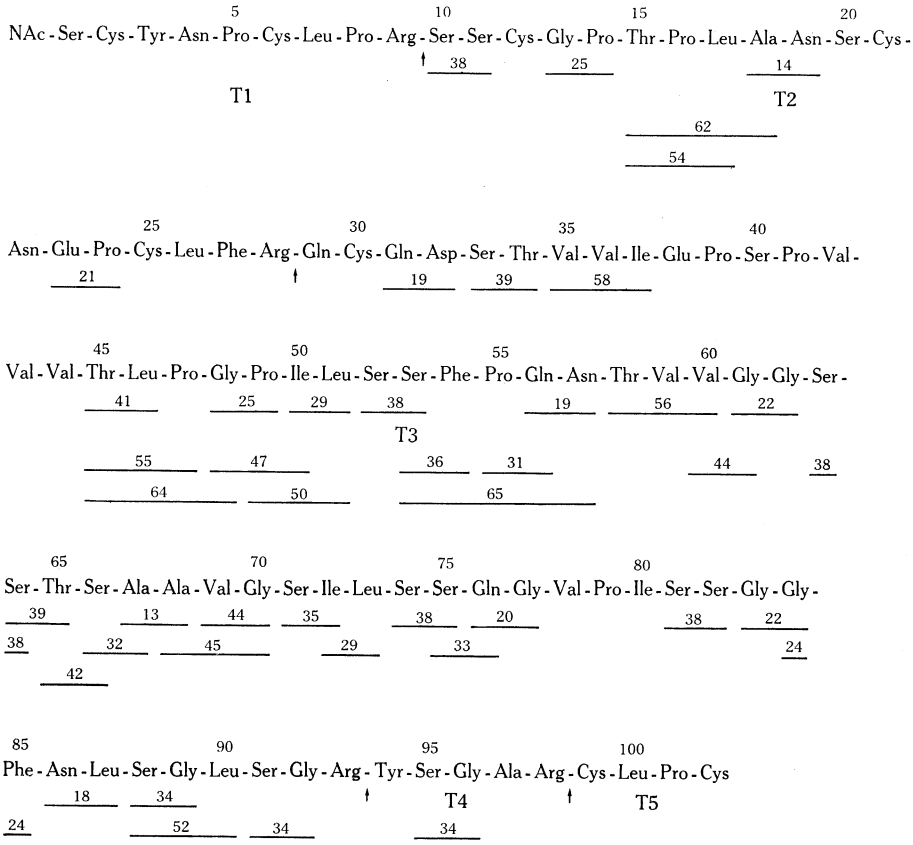


Fig. 4.—Complete amino acid sequence of band 3 of SCM-keratin from emu feather rachis. Residues are numbered from the *N*-terminal residue and the positions of cleavage by trypsin are indicated by arrows. The lines beneath the sequence indicate peptides isolated by Schroeder *et al.* (1957) from a partial acid hydrolysate of calamus from white turkey feather. They have been placed here into positions which may correspond in emu feather and turkey feather keratins. The SCMCys residues have been written as Cys as they would be in the parent molecule.

IV. Discussion

The amino acid sequence of one of the two major proteins (band 3) seen in an acrylamide-gel electrophoresis pattern of SCM-keratin from emu feather rachis (or calamus) shows it to contain 102 residues and to have a molecular weight of 10,459 (in the SCM-form). This value is in good agreement with values determined previously by hydrodynamic methods of 10,980 (Harrap and Woods 1964*b*) and 10,000 (Woodin 1954). The first-named authors used whole extracts from fowl feather rachis while the latter used extracts from whole fowl feathers. Jeffrey (1970) made molecular weight measurements in acrylamide gels of varying porosities and found that the six main bands in the acrylamide-gel patterns of SCM-keratin of duck feather rachis had molecular weights in the range $10{,}500 \pm 1500$. It is worth

noting that the method of Weber and Osborne (1969) using acrylamide gel and sodium dodecyl sulphate gave values some 30% higher than this (Woods and Pont, personal communication).

Band 3 protein has a very similar sequence to band 2 protein and the difference, responsible for the charge difference and consequent separation on acrylamide gel, probably occurs in the 20–30 area and involves a SCMCys residue (O'Donnell 1973). The other predominant bands (bands 1 and 4) seen in the acrylamide-gel pat-

TABLE 8

AMINO ACID COMPOSITION OF SOME PEPTIDES CONTAINING SCM-CYSTEINE IN A THERMOLYSIN DIGEST OF SCM-KERATIN FROM WHOLE RACHIS

The peptides were initially separated on a 600-cm column of Sephadex G25 and purified by paper electrophoresis at pH 6·5 or pH 3·5. Hydrolyses were done in 6N HCl at 105–110°C for 24 hr. Values are not corrected for losses during hydrolysis and are given as moles per mole of peptide with the final value from the sequence in parenthesis

Amino acid	Th1		Th2		Th3		Th4	
Arginine	0·9	(1)			1·0	(1)	1·0	(1)
SCM-cysteine	1·6	(2)	1·5	(2)	1·0	(1)	1·7	(2)
Aspartic acid	1·0		1·9		1·1	(1)		
Asparagine		(1)		(2)				
Threonine	0·9	(1)			0·9	(1)		
Serine	1·8	(2)	1·3	(1)	0·9	(1)		
Glutamic acid			1·1	(1)	2·0			
Glutamine						(2)		
Proline	3·6	(4)	1·0	(1)			+	(1)
Glycine	1·4	(1)	0·8				0·4	
Alanine			1·2	(1)			1·0	(1)
Valine								
Isoleucine								
Leucine	1·0	(1)	1·4	(1)			1·0	(1)
Tyrosine	0·9	(1)						
Phenylalanine					1·0	(1)		
Total		14		9		8		6

tern of the extracts (Fig. 1, O'Donnell 1973) have amino acid compositions very similar to bands 2 and 3, but no separate peptide maps have been done on them to compare their sequences with bands 2 and 3. Woods (1971) on the basis of fractionations of extracts of SCM-keratin from rachis from goose, fowl, and cockatoo feathers concluded that the many (greater than 10) protein components from each feather have no major differences in amino acid composition. So we see that the natural β-keratin from the feather rachis or calamus of one particular species of bird consists, when in the SCM-form, of molecules which are remarkably uniform in size and which also probably have only limited variations of amino acid sequence. This is in marked

TABLE 9

AMINO ACID SEQUENCES OF PEPTIDES FROM SCM-KERATIN ISOLATED AS IN TABLE 8

Peptide	Mobility at pH 6·5	Net No. of charges	Method of purification*	Sequence
Th1	−0·21	1−	S, $E_{3·5}$	Tyr-Asn-Pro-SCMCys-Leu-Pro-Arg-Ser-Ser-SCMCys-Gly-Pro-Thr-Pro
Th2	−0·70	3−	S, $E_{6·5}$	Leu-Ala-Asn-Ser-SCMCys-Asn-Glu-Pro-SCMCys
Th3	−0·28	1−	S, $E_{3·5}$, $E_{6·5}$	Phe-Arg-Gln-SCMCys-Gln-Asp-Ser-Thr
Th4	−0·32	1−	S, $E_{3·5}$, $E_{6·5}$	Ala-Arg-SCMCys-Leu-Pro-SCMCys

* Abbreviations as defined in Table 5.

TABLE 10

COMPARISON OF THE AMINO ACID COMPOSITION OF BAND 3 FROM SCM-KERATIN FROM EMU WITH THAT DERIVED FROM THE SEQUENCE

Analysis values are calculated from Table 2 of O'Donnell (1973) and have been corrected for hydrolytic losses or incomplete release of amino acids. Values are given as moles per mole of keratin (102 residues). The analyses (24 hr hydrolysis only) of unfractionated protein extracted from emu beak is also given and arbitrarily rationalized to 102 residues

Amino acid	Band 3		Beak protein analysis	Amino acid	Band 3		Beak protein analysis
	Analysis	Sequence			Analysis	Sequence	
Lysine	0	0	2·0	Proline	13·1	12	7·1
Histidine	0	0	1·4	Glycine	10·2	11	22·4
Arginine	3·9	4	5·1	Alanine	4·6	4	5·5
SCM-cysteine	7·3	8	3·7	Valine	9·5	9	5·1
Aspartic acid + Asparagine	6·3	1 + 5	5·3	Methionine	0	0	0·5
				Isoleucine	4·6	4	3·4
Threonine	4·9	5	3·1	Leucine	9·3	9	8·5
Serine	17·4	19	9·4	Tyrosine	1·9	2	7·6
Glutamic acid + Glutamine	6·0	2 + 4	7·2	Phenylalanine	2·8	3	2·9

TABLE 11

COMPARISON OF THE AMINO ACIDS PRESENT IN THE SOLUBLE (T1, T2, T4, T5) AND INSOLUBLE (T3) PEPTIDES OF BAND 3

Values are residues per mole (102 amino acids)

Amino acid	T1, T2, T4, T5	T3	Amino acid	T1, T2, T4, T5	T3
Arginine	3	1	Proline	6	6
SCM-cysteine	7	1	Glycine	2	9
Aspartic acid		1	Alanine	2	2
Asparagine	3	2	Valine		9
Threonine	1	4	Isoleucine		4
Serine	5	14	Leucine	4	5
Glutamic acid	1	1	Tyrosine	2	
Glutamine		4	Phenylalanine	1	2

contrast to α-keratins such as wool and hair. In these there are three broad groups of proteins (see Fraser *et al.* 1972) known as the "high-sulphur", "low-sulphur", and "high-glycine–tyrosine groups". Within each of these there is pronounced heterogeneity of size and sequence. The results reported here suggest that the natural β-keratin, feather, is less complicated in molecular make-up than the α-keratins.

The molecule as a whole has a net charge of zero when in the native form at neutral pH values. Of the 12 Glx and Asx residues three are present in the carboxyl form and nine as amides. The value of nine agrees with that found by Harrap and Woods (1964*a*) for whole calamus and rachis from feathers of White Leghorn hens.

The features of the sequence and groupings of amino acids within the sequence that are relevant to the aims set out in the Introduction are discussed below.

(1) Are there any obvious features of structure of keratin molecules common to α-keratins and natural β-keratins? The SCMCys residues (half-cystine groups in the original feather) occur towards either end of the molecule and are comparable with (*a*) the cystine-rich "tails" in the low-sulphur proteins (those proteins responsible for the crystallinity in the X-ray diffraction pattern) of the α-keratin wool (O'Donnell 1969; Crewther and Dowling 1970) and (*b*) the separate high-sulphur group of proteins in α-keratins (see Fraser *et al.* 1972). This finding agrees with the model of Fraser *et al.* (1971). It also explains why Harrap and Woods (1964*a*) and Woods (1971) did not find separate high-sulphur and low-sulphur molecules in the feather extracts. This polarity of half-cystine residues in the molecule shows that the microfibrillar-matrix structure seen in feather keratin in the electron microscope (Filshie and Rogers 1962) may be interpreted in the terms of areas of lower and higher half-cystine content, the higher one being in the matrix. Thus the two-phase structure of feather has a similarity to that of wool, but in the latter case much of the matrix protein is contributed by separate high-sulphur protein molecules.

The existence of cystine-rich sections towards the ends of the feather keratin molecules and the low-sulphur α-keratin molecules suggests that this is how keratin molecules join together. This would be similar to the telopeptide cross-links towards the end of collagen molecules. All of the high-sulphur proteins from wool which have been sequenced to date terminate in SCM-cysteine as does the feather protein sequenced here.

There are no repeating sequences (apart from Leu-Ser-Gly at positions 87–89 and 90–92) in the sequence of this feather keratin such as are found in some of the high-sulphur proteins from the α-keratin wool (Elleman 1971) where a sequence of approximately 10 amino acid residues is repeated up to eight times. Nor is there any obvious similarity between the sequences in the high-sulphur regions of feather and those of other high-sulphur wool proteins which have been sequenced (see Swart and Haylett 1971).

(2) How does knowledge of this sequence help in the design of a model for the feather keratin molecule to account for the features of the X-ray diffraction pattern? The latest molecular model for the crystalline section of feather keratin is that proposed by Fraser *et al.* (1971). The latter group of authors using infrared spectroscopy have shown that approximately 28% of the feather keratin molecule is in the crystalline form, i.e. there is a regular β-region and an amorphous region (matrix). Their structural unit is a pleated sheet with four antiparallel chains each

containing 8 amino acid residues. The sheets are arranged on two intertwined helical ruled surfaces. There are two extreme possibilities. One, the "compact" molecule, has all the chains in the one sheet belonging to the same molecule. The other, the "extended" molecule, has the chain spanning four crystallites up the axis.

It appears that the T3 section (residues 29–93) contains the crystalline part of the feather keratin molecule and this requires approximately half of the T3 section. This is suggested by its insolubility, its almost complete lack of SCM-cysteine, and the concentration in it of those hydrophobic amino acids, valine, leucine, and isoleucine found preferentially in β-sections of globular proteins (Ptitsyn and Finkelstein 1971). Furthermore there is a concentration and high proportion of serine (1 residue in 4·6) and glycine (1 residue in 7), amino acids which are pronounced in the composition of silk of *Bombyx mori*. This silk is composed of antiparallel chain pleated sheets. Suzuki (see addendum to this paper) has proved experimentally with infrared spectroscopy that the T3 section does in fact have a much increased percentage of β-material and that β-material is absent in other peptides. It is probably this section of the molecule which is responsible for the ready gel formation in solutions of feather proteins (Harrap and Woods 1967).

Feather keratin is unusually rich in proline; six residues of this amino acid are found in peptide T3 (approx. 1 residue in 11) and another six residues found elsewhere (approx. 1 residue in 6) and there has been speculation as to its role (cf. Schroeder *et al.* 1957; Schor and Krimm 1961*a*, 1961*b*). One aim of the work was to see if proline occurred regularly at every seventh or eighth residue in the crystalline section to allow the runs of β-chain to turn sharply as in the model of Fraser *et al.* (1971). Inspection of the sequence shows that proline is not so situated as would be the case with one of the extremes of an exact model of Fraser *et al.* However, if the real situation is only approximated by this model then the proline residues could be involved in sharp turns of the polypeptide chain at positions 39–41 (Pro-Ser-Pro), 47–49 (Pro-Gly-Pro), and 55 (Pro) with additional turning points in the vicinity of 61–64 (Gly-Gly-Ser-Ser) and 70–71 (Gly-Ser). Thus a multiple-type of folding of the single polypeptide chain to form a compact molecule could well be the real situation in feather keratin. The absence of regularly repeating features of amino acid sequence such as are found in silk and collagen and the fact that the degree of crystallinity of feather keratin is not sufficient for conventional crystallographic procedures to be used (Fraser *et al.* 1971) means that the detailed molecular structure of feather keratin cannot be determined at present.

Harrap and Woods (1964*a*, 1964*b*) found when using solvent mixtures of 2-chlorethanol–formic acid that SCM-proteins from feathers could be induced to develop an apparent α-helical content of 40%. The sequence (Fig. 4) does not show any obvious section in which this happens. The irregular pattern of proline residues suggests that the regions of α-helices are not continuous.

The presence of several double residues, e.g. Ser-Ser, Gly-Gly, Val-Val, Ala-Ala, in the sequence could mean that both sides of a section of β-pleated sheet have the same aspect at these positions.

(3) Is there much similarity in the amino acid sequence of molecules from the feather keratin from birds of different species? Schroeder *et al.* (1957) isolated di- to pentapeptides from partial acid hydrolysates of white turkey feather calamus.

Although the position of these peptides in the molecule is not known, sequences which are the same in emu feather keratin have been underlined in Figure 4. Some 34 out of 54, i.e. 63% (of those not containing half-cystine residues) listed by Schroeder *et al.* are found at least once in the emu keratin molecule. The percentage derived from the half-cystine residues is much less (approximately 2 out of 7). The peptides isolated by Schroeder *et al.* are concentrated in the T3 section of the molecule which is the one which contains the crystalline region. Hence one may conclude from a comparison of the feather proteins of these two birds that the crystalline section has been more conserved during the evolution of the two species of birds than has the non-crystalline, cystine-containing "tails" section.

That there are differences in the tails sections of the proteins of the feathers from other species of birds can be inferred from the differences in the peptides maps of peptides containing radioactively labelled SCMCys-peptides derived from proteins of feathers of emu, goose (*Anser domesticus*), and silver gull (*Larus novae-hollandiae*) (O'Donnell 1973). However, there is at least some similarity in the tails sections of proteins of the feathers from different birds and this is exemplified by the existence of the same amino terminal peptide, NAc-Ser-SCMCys-Tyr, in both calamus from goose (O'Donnell 1971*a*) and emu feather.

(4) How different are the amino acid sequences of β-keratins from different parts (beak, claw, scales, feathers) of a bird? Since the keratins extracted from corresponding parts of feathers of different species of birds show differences in the amino acid sequences in their cystine-rich tails sections, it is probable that another β-keratin, e.g. the beak, from the same bird would show differences in this area too since this is an amorphous region ("matrix"). The amino acid analysis of protein extracted from emu beak suggests that there must be differences in the tyrosine content (Table 10) of the tails region of the proteins from the two tissues (cf. Gillespie 1973, who found that the beaks of many birds contain proteins rich in glycine (30%) and the aromatic amino acids). But there must also be changes in the T3 section containing the crystalline region. In particular it can be inferred that the ratio of serine and glycine must be reversed in this section of the two tissues. However, the exact extent and nature of these changes must await the determination of the sequence of keratin from the beak. Such a determination may also, by observation of conservation or otherwise of particular areas of sequence, point to those residues which are involved in the crystalline regions of β-keratin.

Finally, the unique structure of this molecule of feather keratin, its small molecular weight, the ready separation of its components on DEAE-cellulose, and its ready availability suggest that it could be a useful tool, additional to the now standard proteins cytochrome *c*, haemoglobin, and the fibrinopeptides, to study avian evolution and differentiation.

V. References

AMBLER, R. P., and MEADWAY, R. J. (1968).—*Biochem. J.* **108**, 803.

BENDIT, E. G. ,(1967).—In "Symposium on Fibrous Proteins, Australia 1967". (Ed. W. G. Crewther.) p. 386. (Butterworths (Australia): Sydney.)

BROWN, J. R., and HARTLEY, B. S. (1966).—*Biochem. J.* **101**, 214.

CREWTHER, W. G., and HARRAP, B. S. (1967).—*J. biol. Chem.* **242**, 4310.

CREWTHER, W. G., and DOWLING, L. M. (1971).—*Appl. Polym. Symp.* No. 18. pp. 1–20.

EAGER, J. E., and SAVIGE, W. E. (1963).—*Photochem. Photobiol.* **2**, 25.

ELLEMAN, T. C. (1971).—*Nature New Biol.* **234**, 148.

FILSHIE, B. K., and ROGERS, G. E. (1962).—*J. Cell Biol.* **13**, 1.
FRASER, R. D. B., MACRAE, T. P., PARRY, D. A. D., and SUZUKI, E. (1971).—*Polymer* **12**, 35.
FRASER, R. D. B., MACRAE, T. P., and ROGERS, G. E. (1972).—"Keratins." (Charles C. Thomas: Springfield, Ill.)
FRASER, R. D. B., and SUZUKI, E. (1965).—*J. molec. Biol.* **14**, 279.
GILLESPIE, J. M. (1972).—Proc. 14th Int. Congr. Derm., Venice.
GRAY, W. R. (1967).—In "Methods in Enzymology." (Ed. C. H. W. Hirs.) Vol. XI. p. 469. (Academic Press: New York and London.)
HARRAP, B. S., and WOODS, E. F. (1964*a*).—*Biochem. J.* **92**, 19.
HARRAP, B. S., and WOODS, E. F. (1964*b*).—*Biochem. J.* **92**, 8.
HARRAP, B. S., and WOODS, E. F. (1967).—*Comp. Biochem. Physiol.* **20**, 449.
JEFFREY, P. D. (1970).—*Aust. J. biol. Sci.* **23**, 809.
MATSUO, H., FUJIMOTO, Y., and TATSUNO, T. (1966).—*Biochem. biophys. Res. Commun.* **22**, 69.
MILSTEIN, C., CLEGG, J. B., and JARVIS, J. M. (1968).—*Biochem. J.* **110**, 631.
MIYAZAWA, T., and BLOUT, E. R. (1961).—*J. Am. Chem. Soc.* **83**, 712.
OFFORD, R. E. (1966).—*Nature* **211**, 591.
O'DONNELL, I. J. (1969).—*Aust. J. biol. Sci.* **22**, 471.
O'DONNELL, I. J. (1971*a*).—*Aust. J. biol. Sci.* **24**, 179.
O'DONNELL, I. J. (1971*b*).—*Aust. J. biol. Sci.* **24**, 1219.
O'DONNELL, I. J. (1973).—*Aust. J. biol. Sci.* **26**, 401.
PTITSYN, O. B., and FINKELSTEIN, A. V. (1971).—*J. molec. Biol* **62**, 613.
REES, M. W., SHORT, M. N., and KASSANIS, B. (1970).—*Virology* **40**, 448.
SCHOR, R., and KRIMM, S. (1961*a*).—*Biophys. J.* **1**, 467.
SCHOR, R., and KRIMM, S. (1961*b*).—*Biophys. J.* **1**, 489.
SCHROEDER, W. A., KAY, L. M., MUNGER, N., MARTIN, N., and BALOG, J. (1957).—*J. Am. chem. Soc.* **79**, 2769.
SWART, L. S., and HAYLETT, T. (1971).—*Biochem. J.* **123**, 201.
WEBER, K., and OSBORNE, M. (1969).—*J. biol. Chem.* **244**, 4406.
WOODIN, A. M. (1954).—*Biochem. J.* **57**, 99.
WOODS, E. F. (1971).—*Comp. Biochem. Physiol.* **39A**, 325.
WOODS, K. R., and WANG, K.-T. (1967).—*Biochim. biophys. Acta* **133**, 369.

ADDENDUM

LOCALIZATION OF β-CONFORMATION IN FEATHER KERATIN

BY E. SUZUKI*

Electron-microscope studies of cross-sections of feather rachis (Filshie and Rogers 1962) suggest that feather keratin consists of a microfibril–matrix complex similar to that found in α-keratin. However, separate microfibril and matrix proteins have not been isolated as in α-keratin (Harrap and Woods 1964), and it has been suggested (Fraser *et al.* 1971) that the lightly and darkly stained areas are derived from different parts of the same molecule.

The infrared spectra of native and partially deuterated feather rachis were investigated by Fraser and Suzuki (1965). They showed that a proportion of the molecule was in the β-conformation and from the presence of characteristic absorption bands at 1636 and 1690 cm^{-1}, it was concluded that antiparallel chain pleated sheets were present. Fraser *et al.* (1971) carried out a quantative analysis of the polarized infrared spectra in the amide-I region of a protein extract from feather rachis from the silver gull and concluded that the content of β-conformation in this material is 28%. Using the value of 10,400 for the molecular weight (Harrap

* Division of Protein Chemistry, CSIRO, Parkville, Vic. 3052.

and Woods 1964*a*, 1964*b*), and a mean residue weight of 99·5 for the native protein, calculated from the amino acid analysis given by Harrap and Woods (1967), it was concluded that about 29 residues per molecule were in the β-conformation.

Having determined the amino acid sequence of the soluble extract from emu calamus (see Fig. 4), the infrared spectra of the tryptic peptides were recorded to establish their chain conformations. The specimens were dissolved in 100% formic acid and films were cast onto BaF_2 plates by evaporating the formic acid at room temperature. The films were heated at 120°C *in vacuo* (10^{-5} Torr) for 4 hr to remove unbound formic acid. The infrared spectra were recorded using a Beckman IR-9 infrared spectrophotometer with a spectral slitwidth of 2·0 cm^{-1} at 1600 cm^{-1}.

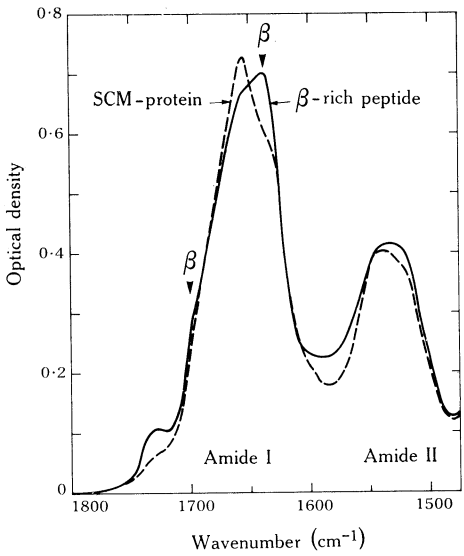


Fig. 5.—The infrared spectra of the peptide T3 (solid line) and SCM-protein from emu calamus (broken line).

The infrared spectra of SCM-protein extract from calamus and the peptide T3 in the amide-I and amide-II regions are shown in Figure 5. In the amide-I region the peptide T3 exhibits a strong absorption band at 1637 cm^{-1} and a weak shoulder at *c.* 1700 cm^{-1} indicating that antiparallel chain pleated sheets are present. A broad absorption band centred at *c.* 1655 cm^{-1} may be assigned to a component with a random-chain conformation. The broadness of this band precludes the possibility that it is associated with α-helical material. The spectrum of SCM-protein from calamus shows a maximum at 1655 cm^{-1}, indicating that the major conformation in this material is a random one with the antiparallel β-conformation as the minor component.

Attempts to obtain oriented films of the peptides were not successful and therefore the least-squares procedure employed to determine the β-content in silver gull feather rachis (Fraser *et al.* 1971) could not be used. The spectrum of the peptide T3 in Figure 5 was scaled by a factor

$$\frac{\text{area of the amide-I bands in SCM-protein}}{\text{area of the amide-I bands in the peptide T3}}$$

to compensate for the difference in the specimen thicknesses, and then multiplied

by the ratio of the number of residues in two specimens (65/102). The difference between the SCM-protein spectrum (Fig. 5) and the scaled T3 spectrum [Fig. 6(*a*)] is shown in Figure 6(*b*) and indicates that the β-content of the peptide T3 is more than enough to account for all the β-conformation present in SCM-protein from emu calamus.

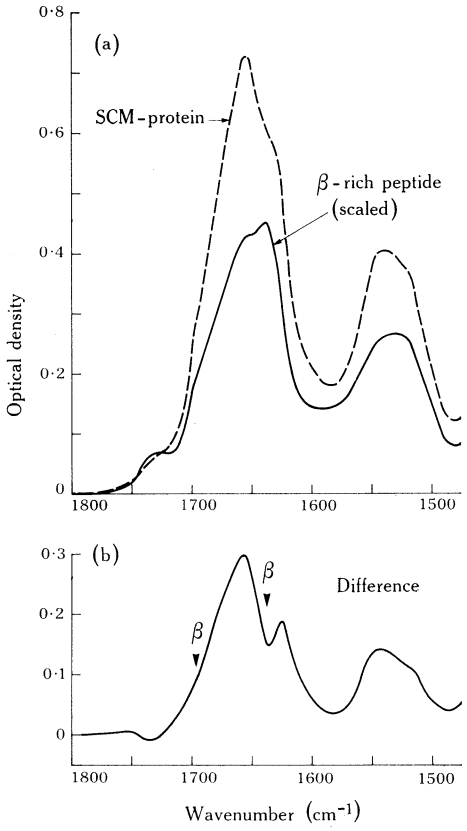


Fig. 6.—(*a*) Spectrum of the peptide T3 scaled to the ratio of residues present (solid line) compared with the spectrum of SCM-protein shown in Fig. 5 (broken line). (*b*) Difference between the two spectra.

The ambiguity in assessing the areas of the amide-I bands and the normal dependence of the β-content on specimen preparation procedure precludes any quantitative treatment of the difference spectrum. However, it seems likely that the β-portion of the feather molecule is localized in the peptide T3. This conclusion is further supported by the fact that the spectrum of the soluble fraction of the tryptic digested SCM-emu (a mixture of T1, T2, T4, and T5) exhibits no indication of the presence of the β-conformation.

In the amide-II region, protein spectra are heavily overlapped with strongly absorbing side-chain bands. For example, COO^- absorbs at 1575 cm^{-1}, the arginyl group at 1598 and 1572 cm^{-1}, and the tyrosyl group at 1516 cm^{-1} (Bendit 1967). Thus the frequency-conformation criterion developed for synthetic polypeptides (Miyazawa and Blout 1961) cannot readily be applied to interpret the protein spectra in this region.