The Amino Acid Sequence of Protein SCMK-B2A from the High-Sulphur Fraction of Wool Keratin

By T. C. ELLEMAN

*Division of Protein Chemistry, Commonwealth Scientific and Industrial Research Organization,
 Parkville (Melbourne), Vic. 3052, Australia*

(Received 4 July 1972)

1. The amino acid sequence of protein SCMK-B2A, a reduced and *S*-carboxymethylated protein from the high-sulphur fraction of wool, has been determined. 2. This protein of 171 amino acid residues displays both a high degree of internal homology and extensive external homology with other members of the SCMK-B2 group of proteins. 3. Evidence is presented which suggests that the SCMK-B2 group of proteins are produced by separate non-allelic genes.

The proteins of the high-sulphur fraction of α -keratins are derived from the non-filamentous matrix between the microfibrils. From this fraction a group of similar proteins, SCMK-B2,* has been isolated (Lindley & Elleman, 1972). The sequence of component SCMK-B2A is reported in the present paper; sequences of the two other major components

SCMK-B2B and SCMK-B2C have been reported (Elleman, 1972; Elleman & Dopheide, 1972).

These proteins, with their repetitive sequences of high cystine content, might mediate the thiol-disulphide interchange responsible for such properties of the wool fibre as creep and stress relaxation (Elleman, 1972).

* Abbreviations: SCMK, *S*-carboxymethylkeratine;

Cm
 |
 CmCys or Cys, *S*-carboxymethylcysteine; Ac-Ala or
 Ac
 |
 Ala, *N*-acetylalanine.

Fig. 1. Elution profile of the high-sulphur fraction of wool (3g) on a column (90 cm $\times$ 1.5 cm) of SE-Sephadex C-25

The flow rate was 0.8 ml/min with a linear gradient from 2% (w/v) citric acid–5.0M-urea (720 ml) to 4% (w/v) tricine–5.0M-urea (720 ml); 6 ml fractions were collected.

Vol. 130

Fig. 2. Elution profile of a tryptic digest of wool protein SCMK-B2A on DEAE-cellulose (Cl^- form)

The column was equilibrated with 0.1M-tris-HCl buffer, pH 7.0, and elution of peptides was effected with a linear gradient from equilibration buffer (1300 ml) to 0.1M-tris-HCl buffer (pH 7.0)–0.6M-NaCl (1300 ml); the flow rate was 0.8 ml/min.

Materials and Methods

Preparation of protein

Protein SCMK-B2A was prepared from Lincoln wool as described by Lindley & Elleman (1972). The elution profile obtained on fractionating high-sulphur protein on SE-Sephadex C-25 is shown in Fig. 1. The conditions of elution are described in the legend. Pool A was further fractionated on DEAE-cellulose in 2.0M-urea by using a formic acid gradient (Lindley & Elleman, 1972), to yield material which

gave essentially a single band on gel electrophoresis at pH9.5 in 15% polyacrylamide (Davis, 1964) (Plate 1). This was designated protein SCMK-B2A.

Digestion of protein SCMK-B2A with trypsin, and fractionation of the tryptic peptides

The protein (360mg) was digested at 37°C in 0.1M-NH₄HCO₃ with 2mg of trypsin [treated with L-(tosylamido-2-phenyl)ethyl chloromethyl ketone; Worthington, Freehold, N.J., U.S.A.]. After 1h,

Table 1. Amino acid composition of protein SCMK-B2A and tryptic peptides obtained from it
The results are expressed as residues/mol. The values in parentheses are inferred from sequence studies.

Hydrolysis conditions	5.7M-HCl		3M-Toluene- <i>p</i> -sulphonic acid		Peptides				Total peptides
	24h	48h	24h	48h	T1	T2	T3	T4	T1+T3+T4
Residue									
Arg	5.6	6.1	5.9	5.8	3.0	2.0	1.0	2.0	6
Asp	1.3	1.3	1.5	1.4	1.0	1.0	—	—	1
Thr	18.0	16.5	18.0	17.6	1.0	—	12.7	4.4	18
Ser	22.4	20.4	24.2	22.0	—	—	16.7 (18)	6.6	25
Glu	20.0	19.6	19.0	19.6	—	—	12.0	8.2	20
Pro	18.0	18.0	17.9	17.9	1.0	1.0	7.0	9.6	18
Gly	14.6	14.8	15.0	15.0	—	—	12.6	2.2	15
Ala	4.8	4.9	5.2	4.9	—	—	2.0	3.0	5
Val	4.6	5.0	4.3	4.7	—	—	2.0	2.6	5
Ile	8.6	8.7	7.7	8.5	—	—	6.7	1.8	9
Leu	2.9	3.1	2.9	3.0	—	—	1.0	1.8	3
Tyr	4.0	4.0	4.2	3.6	—	—	0.9	2.7	4
Phe	2.0	2.0	2.0	2.0	—	—	2.0	—	2
CmCys	38.5	39.0	36.6	37.6	2.0	2.0	17.5 (19)	18.0	39
Trp	—	—	1.0	1.0	(1)	(1)	—	—	1

Table 2. Composition and sequence of chymotryptic peptides derived from tryptic peptide T3

—→ indicates residues determined by the dansyl-Edman procedure; —→ indicates C-terminal residues identified as the Dns derivatives after complete Edman degradation of peptides; — indicates C-terminal residues determined by selective tritiation (Matsuo *et al.*, 1966); — indicates residues released by carboxypeptidase A.

Peptide	
T3c ₁	Composition... CmCys(16.4);Thr(12.0);Ser(11.5);Glu(12.0); Pro(7.0);Gly(9.2);Ile(6.4);Leu(0.9);Tyr(0.6); Phe(0.8)
	Sequence... $\begin{matrix} \text{Cm} & & \text{Cm} \\ \text{Cys-Gly-Phe-Pro-Ile-Cys-Ser-} \end{matrix}$
T3c ₂	Composition... CmCys(2.0);Thr(1.0);Ser(2.0);Ala(1.0);Phe(1.0)
	Sequence*... $\begin{matrix} & \text{Cm} & \text{Cm} \\ \text{Ac-Ala-Cys-Cys-Ser-Thr-Ser-Phe} \end{matrix}$
T3c ₃	Composition... Ser(3.7);Glu(1.0);Gly(3.0);Ala(1.0);Val(1.9); Arg(1.0)
m=0.30	Sequence... $\begin{matrix} \text{Gly-Gln-Val-Gly-Ser-Ser-Gly-Ala-Val-Ser-Ser-Arg} \\ \text{→ → → → → → → → → → → →} \end{matrix}$

* Determined by Lindley & Elleman, 1972.

EXPLANATION OF PLATE I

Polyacrylamide-gel electrophoretograms of fractions from chromatography of the high-sulphur fraction of wool on SE-Sephadex C-25, and of protein SCM-K-B2A

The fractions are those indicated in Fig. 1.

Table 3. Composition and sequence of thermolytic peptides derived from peptide T3c₁
For abbreviations see Table 2.

Peptide	m	Composition	Sequence	Remarks
T3c ₁ th ₁	0.00	Thr(1.0);Glu(1.0);Leu(1.0)	<u>Leu-Gln-Thr</u>	
T3c ₁ th ₂	0.00	Ser(1.0);Ile(1.0)	<u>Ile-Ser</u>	
T3c ₁ th ₃	0.00	Ser(1.0);Gly(2.0);Ile(1.0)	<u>Ile-Gly-Gly-Ser</u>	
T3c ₁ th ₄	0.00	Gly(1.0);Ile(1.0);Tyr(1.0)	<u>Ile-Gly-Tyr</u>	
T3c ₁ th ₅	-0.50	CmCys(0.8);Gly(1.0);Phe(0.7);Pro(1.0)	<u>Cm</u> <u>Cys-Gly-Phe-Pro</u>	
T3c ₁ th ₆	-0.50	CmCys(2.1);Thr(2.0);Ser(1.8);Glu(2.0);Pro(1.3);Ile(1.0)	<u>Ile-Gln-Thr-Ser-Cys-Gly-Gln-Pro-Thr-Ser</u>	70% yield
T3c ₁ th ₇	-0.54	CmCys(2.1);Thr(1.0);Ser(0.9);Glu(2.0);Pro(1.0);Ile(0.9)	<u>Ile-Gln-Thr-Ser-Cys-Gln-Pro</u>	80% yield
T3c ₁ th ₈	-0.70	CmCys(2.1);Thr(2.0);Ser(1.1);Glu(2.0);Gly(3.2);Leu(1.0)	<u>Leu-Glx-Thr-Ser-Gly-Cys-Glx-Thr-Gly-Cys-Gly</u>	50% yield
T3c ₁ th ₉	-0.70	CmCys(2.9);Thr(1.9);Ser(1.1);Glu(2.0);Pro(1.1);Ile(0.9)	<u>Ile-Gln-Thr-Ser-Cys-Gln-Pro-Thr-Cys</u>	70% yield
T3c ₁ th ₁₀	-0.77	CmCys(6.5);Thr(4.8);Ser(4.5);Glu(3.3);Pro(2.9);Gly(3.1);Ile(1.0)	<u>Ile-Cys-Ser-Thr-Gly-Gly-Thr-Cys-Gly-Ser-Ser-Pro-</u> <u>Cys-Gln-Pro-Thr-Cys-Gln-Thr-Ser-Cys-Gly-</u> <u>Gln-Pro-Thr-Ser</u>	See Table 4
T3c ₁ th ₁₁	-0.87	CmCys(1.8);Thr(1.0);Ser(0.9);Glu(1.0);Gly(2.9)	<u>Ser-Gly-Cys-Glu-Thr-Gly-Cys-Gly</u>	30% yield

2 mg of soya-bean trypsin inhibitor was added and the digestion was continued for a further $\frac{1}{2}$ h. The solution was then dried by rotary evaporation. The digest was redissolved in 0.1 M-tris-HCl buffer, pH 7.0, and applied to a column (60 cm $\times$ 2.5 cm) of DEAE-cellulose. Elution of peptides was effected with a linear gradient from 0–0.6 M-NaCl, as described in the legend to Fig. 2. Peptides were desalted on a column (40 cm $\times$ 2.5 cm) of Sephadex G-10 in 0.5 % NH_3 solution.

Amino acid analysis and sequence determination

Peptides were hydrolysed *in vacuo* at 110°C for 24 h in 200 μ l of 5.7 M-HCl containing 1 μ l of 2-mercaptoethanol. Protein was hydrolysed both under these conditions for 24 h and 48 h and also, to estimate tryptophan, in 500 μ l of 3 M-*p*-toluene sulphonic acid containing 1 mg of tryptamine (Liu & Chang, 1971). The hydrolysates were analysed on a Beckman 120 C amino acid analyser.

Peptide sequences were determined from the *N*-terminal end by the dansyl-Edman procedure (Gray, 1967); *C*-terminal residues were identified by the selective tritiation method of Matsuo *et al.* (1966) in conjunction with either acid hydrolysis or carboxypeptidase digestion. Enzymic digestions, excepting peptic digestions, were performed in 0.1 M- NH_4HCO_3 at 37°C by using an enzyme/substrate ratio of 1:200 (w/w). Peptic digestions were performed in aq. 5% (v/v) formic acid. Partial acid hydrolysis was carried out at 37°C in 11 M-HCl for 3 days. Cathepsin C, a gift from Dr. H. Lindley, was used at pH 6.0 (16 mM-HCl, 0.8% pyridine, 50 mM-mercaptoethanol).

High-voltage paper electrophoresis was performed as described by Ambler (1963). Peptides were located on guide-strips with 0.5% ninhydrin, 2% (v/v) collidine in ethanol or by the chlorination method of Reindel & Hoppe (1954). Elution of peptide bands was performed with aq. 1% (w/v) NH_3 solution. Amide groups were assigned to peptides on the basis of electrophoretic mobility at pH 6.5 (Offord, 1966) relative to aspartic acid ($m = -1.0$).

Nomenclature

All tryptic peptides isolated from protein SCMK-B2A are designated by the letter T, followed by a number indicating the fraction in which they were eluted from DEAE-cellulose. The peptides obtained by further degradation of tryptic peptides by thermolysin, chymotrypsin and pepsin are indicated by th, c and p respectively.

Results

The amino acid composition of wool protein SCMK-B2A (Table 1) is based on a mol. wt. of approx. 20000 (Lindley & Elleman, 1972). No free *N*-terminal residue was detected, whereas carboxypeptidase A released *S*-carboxymethylcysteine from the *C*-terminus of the molecule.

Tryptic peptides from protein SCMK-B2A

The elution profile of the tryptic digest of protein SCMK-B2A is shown in Fig. 2. Peptides T1 and T2 comprised small tryptophan-containing peptides

Table 4. Composition and sequence of peptides derived from T3c₁th₁₀ by further enzymic digestion

 indicates radioactive peptides isolated by paper electrophoresis after partial acid hydrolysis. For other abbreviations see Table 2.

Peptic peptides

T3c₁th₁₀P₁ Composition... CmCys(5.0);Thr(2.9);Ser(2.8);Glu(1.2);Pro(2.1);Gly(3.0);Ile(0.9)

m = -0.80 Sequence... Ile-Cys(Cys₂,Thr₂,Ser₃,Glx₁,Pro₂,Gly₃)Thr-Cys-Cys

T3c₁th₁₀P₂ Composition... CmCys(1.9);Thr(2.0);Ser(1.9);Glu(2.0);Pro(1.2)

m = -0.55 Sequence... Gln-Thr-Ser-Cys-Cys-Gln-Pro-Thr-Ser

Residual peptide
after cathepsin C
digestion of
peptide T3c₁th₁₀

Sequence... Ser-Pro-Cys-Gln-Pro-Thr-Cys-Cys-Glx-Thr(Cys₂,Thr₁,Ser₂,Glx₁,Pro₁)

m = -0.82 *m* after three Edman degradations = -0.74

m after four Edman degradations = -0.77

Table 5. Composition and sequence of some peptic peptides derived from peptide T3c₁
For abbreviations see Table 2.

Peptide	m	Composition	Sequence
T3c ₁ P ₁	-0.52	CmCys(2.1); Thr(1.7); Ser(1.8); Glu(2.2); Pro(1.0); Ile(1.1)	<u>Cm</u> <u>Cm</u> <u>Ser-Cys-Cys-Gln-Pro-Thr-Ser-Ile-Gln-Thr</u>
T3c ₁ P ₂	-0.54	CmCys(2.0); Thr(0.5); Ser(1.8); Glu(2.0); Pro(1.0); Ile(1.6)	<u>Cm</u> <u>Cm</u> (a) <u>Ser-Cys-Cys-Gln-Pro-Ile-Ser-Ile-Gln</u> (b) <u>Thr</u>
T3c ₁ P ₃	-0.58	CmCys(2.0); Thr(2.0); Ser(2.0); Glu(2.2); Gly(5.5); Ile(2.0)	<u>Cm</u> <u>Cm</u> <u>Glx-Thr-Ser-Gly-Cys-Glx-Thr-Gly-Cys-Gly-Ile-</u> <u>Gly-Gly-Ser-Ile-Gly</u>
T3c ₁ P ₄	-0.58	CmCys(2.0); Thr(1.6); Ser(1.2); Glu(2.0); Pro(1.0); Ile(1.0)	<u>Cm</u> <u>Cm</u> <u>Cys-Cys-Gln-Pro-Thr-Ser-Ile-Gln-Thr</u>
T3c ₁ P ₅	-0.66	CmCys(1.7); Ser(1.8); Glu(1.1); Pro(1.0); Ile(1.0)	<u>Cm</u> <u>Cm</u> <u>Ser-Cys-Cys-Gln-Pro-Ile-Ser</u>
T3c ₁ P ₆	-0.73	CmCys(3.4); Thr(2.6); Ser(2.5); Glu(1.8); Pro(2.3); Gly(4.0); Ile(1.1); Phe(0.4)	<u>Cm</u> <u>Cm</u> <u>Cys-Gly-Phe-Pro-Ile-Cys-Ser-Thr-Gly-Gly-</u> <u>Thr-Cys-Gly-Ser-Ser-Pro-Cys-Glx-Pro-Thr-Cys</u>
T3c ₁ P ₇	-0.78	CmCys(3.0); Thr(1.0); Ser(0.9); Glu(1.0); Pro(1.0); Leu(1.0)	<u>Cm</u> <u>Cm</u> <u>Cm</u> <u>Ser-Cys-Cys-Gln-Pro-Thr-Cys-Leu</u>
T3c ₁ P ₈	-0.80	CmCys(5.5); Thr(3.1); Ser(2.8); Glu(1.5); Pro(2.9); Gly(3.4); Ile(1.0); Phe(0.9)	<u>Cm</u> <u>Cm</u> <u>Cm</u> <u>Cm</u> <u>Cys-Gly-Phe-Pro-Ile-Cys-Ser-Thr-Gly-Gly-Thr-Cys-</u> <u>Gly-Ser-Ser-Pro-Cys-Gln-Pro-Thr-Cys-Cys</u>

related by a partial cleavage, which have been previously characterized as Thr-Arg-Trp-CmCys-Arg-Pro-Asp-CmCys-Arg and Trp-CmCys-Arg-Pro-Asp-CmCys-Arg respectively (Gillespie *et al.*, 1968; Lindley & Elleman, 1972); peptides T3 and T4 contained the remainder of the molecule (Table 1).

Peptide T3 (residues 1–98). This peptide had no free *N*-terminal group and so constituted the *N*-terminal sequence of protein SCMK-B2A. A short (1h) chymotryptic digest of peptide T3 produced three major fragments which were separated on a column (100cm×1.5cm) of Sephadex G-50 in 1% NH₄HCO₃, and by paper electrophoresis. These peptides are shown in Table 2.

Peptide T3c₁ was digested with thermolysin (16h) and the digest was fractionated by paper electrophoresis. The thermolytic peptides are shown in Table 3. Thermolytic peptide T3c₁th₁₀ was further digested with pepsin (6h) and cathepsin C. [Cathepsin C is an amino-dipeptidase which will not cleave a peptide containing proline in position 2 or 3 (Lindley, 1972).] (16h) to produce smaller fragments more amenable to sequence determination (Table 4). Peptides from peptic digestion (6h) of peptide T3c₁ were also isolated by paper electrophoresis and are shown in Table 5.

Insufficient information was available from these thermolytic (16h) and peptic digests to permit deduction of a unique sequence for peptide T3c₁, owing to the extensive repetition of sequence within this fragment. The isolation of peptide T3th₃ however, from a short (2h) thermolytic digest of peptide T3 (Table 6) permitted a unique solution to the sequence of peptide T3 (Fig. 3).

Two peptides present in very low yield were isolated from a heavily loaded thermolytic digest (16h) of peptide T3 (Table 7). The specificity of thermolysin suggests that these peptides occupy consecutive positions, namely residues 9–22 and 23–38, in the sequence of a minor variant of protein SCMK-B2A (Fig. 6).

Peptide T4 (residues 108–171). The *N*-terminal residue of this peptide was valine, and the *C*-terminal residue was *S*-carboxymethylcysteine. A sample of trypsin (Calbiochem, Los Angeles, Calif., U.S.A., Lot no. 801619) with well-demonstrated chymotrypsin-like activity (Lindley & Elleman, 1972) was used (6h) to cleave the peptide at a Tyr–Tyr bond. The resulting fragments (Table 8) were separated on a column (150cm×2cm²) of Sephadex G-50 in 1% NH₄HCO₃.

Peptide T4a had an *N*-terminal valine residue. Carboxypeptidase A (6h) released leucine and tyrosine from the *C*-terminus. Peptides from thermolytic digestion enabled the sequence of peptide T4a to be deduced (Table 9).

Peptide T4b had an *N*-terminal tyrosine residue. Selective tritiation and acid hydrolysis showed *S*-

Table 6. Composition and sequence of some thermolytic fragments found after 2 h of thermolytic digestion of tryptic peptide T3

For abbreviations see Table 2.

Peptide	m	Composition	Sequence	Remarks
T3th ₁	0.00	Glu(1.0); Gly(2.0); Ile(1.0); Tyr(0.9)	Ile-Gly-Tyr-Gly-Gln	
T3th ₂	-0.39	CmCys(1.0); Pro(1.0); Gly(1.0); Phe(1.9)	^{Cm} Phe-Cys-Gly-Phe-Pro	
T3th ₃	-0.45	CmCys(2.1); Thr(1.9); Ser(2.9); Glu(2.0); Pro(1.0); Ile(1.9)	Ile-Ser-Ile-Gln-Thr-Ser-Cys-Gln-Pro-Thr-Ser ^{Cm} ^{Cm}	35% yield
T3th ₄	-0.52	CmCys(2.0); Thr(1.8); Ser(1.8); Glu(2.0); Pro(1.1); Ile(0.8)	Ile-Gln-Thr-Ser-Cys-Gln-Pro-Thr-Ser ^{Cm} ^{Cm}	30% yield
T3th ₅	-0.60	CmCys(2.0); Thr(1.9); Ser(1.9); Glu(2.0); Gly(4.8); Ile(1.0); Leu(0.9)	Leu-Gln-Thr-Ser-Gly-Cys-Glu-Thr-Gly-Cys-Gly-Ile-Gly-Ser ^{Cm} ^{Cm} ^{Cm}	45% yield
T3th ₆	-0.68	CmCys(2.0); Thr(2.0); Ser(0.9); Glu(2.0); Gly(3.0); Leu(1.0)	Leu-Gln-Thr-Ser-Gly-Cys-Glu-Thr-Gly-Cys-Gly ^{Cm} ^{Cm}	15% yield

Fig. 3. Amino acid sequence of tryptic peptide T3
----- represents sequence not determined.

Table 7. Composition and sequence of minor peptides from thermolytic digestion of peptide T3

For abbreviations see Table 2.

Peptide		
T3th ₇	Composition...	CmCys(2.0);Asp(1.0);Thr(3.0);Ser(2.2);Pro(1.2);Gly(4.0);Phe(1.0)
m = -0.46	Sequence...	<u>Gly-Phe-Pro-Thr-Cys-Ser-Thr-Gly-Gly-Thr-Cys-Gly-Ser-Asn</u>
T3th ₈	Composition...	CmCys(4.5);Thr(2.7);Ser(2.0);Glu(3.0);Pro(2.1);Phe(0.7)
m = -0.80	Sequence...	<u>Phe-Cys-Gln-Pro-Thr-Cys-Cys-Glx-Thr-Ser-Cys-Cys-Gln-Pro-Thr-Ser</u>
Peptic peptides from T3th ₈		
T3th ₈ P ₁	Composition...	CmCys(2.6);Thr(1.0);Glu(1.0);Pro(0.9);Phe(0.9)
m = -0.78	Sequence...	<u>Phe-Cys-Gln-Pro-Thr-Cys-Cys</u>
T3th ₈ P ₂	Composition...	CmCys(1.9);Thr(1.9);Ser(2.0);Glu(2.0);Pro(1.6)
m = -0.53	Sequence...	<u>Gln-Thr-Ser-Cys-Cys-Gln-Pro-Thr-Ser</u>

carboxymethylcysteine was C-terminal. Fragments from a peptic digest of peptide T4b, fractionated by paper electrophoresis, and two large chymotryptic fragments isolated by fractionation on both Sephadex G-25 (superfine grade; column size 150cm×2cm²) and paper electrophoresis (Table 10) permitted the sequence of peptide T4b to be deduced. The amino acid sequence of peptide T4 is shown in Fig. 4.

The sum of the amino acid compositions of tryptic peptides T1, T3 and T4 is identical with the composition of protein SCMK-B2A as derived from acid hydrolysis, and expressed to the nearest integer. The order of the three fragments is apparent from their end groups, and no attempt has been made to isolate peptides overlapping the sites of tryptic cleavage since the juxtapositioning of the fragments is apparent from the complete amino acid sequences of the closely homologous proteins SCMK-B2B and SCMK-B2C (Fig. 6). The complete amino acid sequence of protein SCMK-B2A is shown in Fig. 5.

Discussion

Protein SCMK-B2A is one of a group of closely similar proteins present in the high-sulphur fraction of wool (Lindley & Elleman, 1972). It displays, along with the other members of this group, extensive internal homology based on a repeating sequence of ten residues. In the N-terminal half of the molecule a ten-residue sequence is repeated five times consecutively with few residue changes. The high specificity of thermolytic digestion with respect to this

Table 8. *Amino acid compositions of tryptic peptide T4 and the peptides derived from it by digestion with trypsin (Calbiochem Lot no. 801619)*

Results are expressed as residues/mol.

Residue	Peptide		
	T4	T4a	T4b
Arg	2.0	—	2.0
Thr	4.4	1.9	2.0
Ser	6.6	3.0	3.7
Glu	8.2	2.0	5.9
Pro	9.6	4.0	6.4
Gly	2.2	1.0	1.0
Ala	3.0	—	2.8
Val	2.6	2.5	—
Ile	1.8	—	1.9
Leu	1.8	2.0	—
Tyr	2.7	1.0	1.6
CmCys	18.0	5.1	13.4
N-terminus ...	Val	Val	Tyr
C-terminus ...	CmCys	Tyr	CmCys

Table 9. *Composition and sequence of thermolytic peptides from peptide T4a*

For abbreviations see Table 2.

Peptide	m	Composition	Sequence
T4ath ₁	-0.68	CmCys(1.9); Thr(1.0); Ser(1.0); Glu(1.0); Pro(1.9); Gly(1.9); Val(1.0); Leu(1.0)	Val-Glu-Gly-Thr-Ser-Leu-Pro-Cys-Gly Cm
T4ath ₂	-0.64	CmCys(3.1); Thr(1.0); Ser(2.0); Glu(1.0); Pro(1.9); Val(1.2)	Val-Val-Ser-Cys-Thr-Pro-Pro-Ser-Cys-Gln Cm
T4ath ₃	-0.80	CmCys(5.0); Thr(1.8); Ser(2.8); Glu(1.9); Pro(3.7); Gly(0.8); Val(2.5); Leu(1.0)	Val-Glx-Gly-Thr-Ser-Leu-Pro-Pro-Cys-Gly Cm
T4ath ₄	-0.00	Leu(1.0); Tyr(1.0)	Val-Val(Cys ₃ , Thr ₁ , Ser ₂ , Glx ₁ , Pro ₂) Leu-Tyr

Table 10. Composition and sequence of peptides derived from peptide T4b by further enzymic degradation
For abbreviations see Table 2.

Peptide	m	Composition	Sequence
T4bp ₁	-0.25	CmCys(2.1);Ser(2.0);Glu(1.0);Pro(0.9);Ala(2.0);Tyr(1.6);Arg(0.9)	<u>Tyr-Ala-Gln-Ala-Ser-Cys-Cys-Arg-Pro-Ser-Tyr</u> Cm Cm
T4bp ₂	-0.41	CmCys(3.0);Ser(2.8);Glu(2.0);Pro(1.1);Gly(1.1);Ala(2.0);Tyr(1.5);Arg(1.0)	<u>Tyr-Ala-Gln-Ala-Ser-Cys-Cys-Arg-Pro-Ser-Tyr</u> Cm Cm <u>Cys-Gly-Gln-Ser</u> Cm
T4bp ₃	-0.50	CmCys(1.0);Ser(1.0);Glu(1.0);Gly(1.0)	<u>Cys-Gly-Gln-Ser</u> Cm
T4bp ₄	-0.64	CmCys(2.8);Pro(1.0);Ala(1.0);Arg(1.1)	<u>Cys-Cys-Arg-Pro-Ala-Cys</u> Cm Cm
T4bp ₅	-0.91	CmCys(3.0);Thr(0.9);Glu(2.0);Pro(1.0);Ile(1.0)	<u>Cys-Cys-Gln-Pro-Thr-Cys-Ile-Glu</u> Cm Cm
T4bp ₆	-0.95	CmCys(4.2);Thr(1.0);Glu(3.2);Pro(2.0);Ile(1.8)	<u>Cys-Cys-Glx-Pro-Thr-Cys-Ile-Glx-Pro-Ile-Cys-Glu</u> Cm Cm Cm
T4bp ₇	-0.75	CmCys(1.0);Glu(1.0);Pro(1.0);Ile(1.1)	<u>Pro-Ile-Cys-Glu</u> Cm
T4bp ₈	-0.87	CmCys(1.0);Glu(2.0);Pro(1.0);Ile(1.8)	<u>Ile-Glu-Pro-Ile-Cys-Glu</u> Cm
T4bp ₉	-1.00	CmCys(3.1);Thr(1.0);Ser(0.8);Glu(1.0);Pro(2.0)	<u>Pro-Ser-Cys-Cys-Glu-Pro-Thr-Cys</u> Cm Cm
		Mobility (m) of product after four cycles of Edman reaction = -0.88	
T4bc ₁	-0.97	CmCys(1.0);Thr(2.0);Ser(1.9);Glu(4.8);Pro(5.2);Gly(1.0);Ala(1.0);Ile(1.7);Arg(1.1)	<u>Cys-Gly-Glx-Ser-Cys-Cys-Arg-Pro-Ala-Cys-Cys</u> Cm Cm Cm <u>Glx-Pro(Cys₄,Thr₂,Ser₁,Glx₃,Pro₃,Ile₂)Cys</u> Cm
T4bc ₂	-0.97	CmCys(10.4);Thr(2.0);Ser(2.0);Glu(3.7);Pro(5.0);Ala(1.1);Ile(1.8);Arg(0.9)	<u>Ser-Cys-Cys-Arg-Pro-Ala-Cys-Cys-Glx</u> Cm Cm Cm <u>Pro(Cys₄,Thr₂,Ser₁,Glx₃,Pro₃,Ile₂)Cys</u> Cm

Fig. 4. Amino acid sequence of tryptic peptide T4

----- represents sequences not determined. For abbreviations see Table 2.

Fig. 5. Amino acid sequence of wool protein SCMK-B2A

The sequence has been arranged to show maximum internal homology. — denotes homology present throughout the molecule; denotes homology confined to a particular region.

[illegible]

Fig. 6—(continued on next page)

Fig. 6. Comparison of the amino acid sequences of wool proteins *SCMK-B2A*, *SCMK-B2B* and *SCMK-B2C*

Part of a minor variant found with *SCMK-B2A* is also shown in square brackets. Residues differing between the components have been enclosed in boxes.

sequence, produced near stoichiometric yields of several very similar peptides, and so permitted the assignment of these peptides to a single unique sequence. In the C-terminal half of the molecule a series of similar repeating units is present, although the homology between these units is less extensive than that in the N-terminal region (Fig. 5). The possible implications of this sequence with regard to the structure and properties of the wool fibre have been considered by Elleman (1972).

Minor variability (approx. 10%) in the sequence of protein SCMK-B2A has been observed at positions 12, 22 and 23 and, since these replacements are linked, quite probably arises from the expression of a separate structural gene. The characteristic differences of each of proteins SCMK-B2A, SCMK-B2B and SCMK-B2C, isolated from a single fleece, are likewise linked (Fig. 6), indicating several separate structural genes for the SCMK-B2 group of proteins. Fleeces from seven sheep have been studied to date, and components corresponding with each of proteins SCMK-B2A, SCMK-B2B and SCMK-B2C, in both position of elution from SE-Sephadex and mobility on polyacrylamide-gel electrophoresis, have been observed in each fleece, suggesting the proteins are the expression of multiple duplicated genes rather than polymorphic alleles.

Of the sequence variations found between the SCMK-B2 type proteins many are confined to particular regions of the molecule (Fig. 6). This might indicate relatively less-stringent functional constraints at these positions so permitting the accumulation of random mutations, or a relatively higher mutability in the corresponding regions of the precursor genes. The observations of Benzer (1961) on the widely differing stability towards spontaneous mutations in parts of the r II region of bacteriophage T4 certainly show the latter to be feasible. Stable base-paired structures in mRNA might selectively dictate conservation of certain regions of a gene against mutations. Mäkelä & Cross (1970) have proposed that a low stability towards mutational events might generate diversity in immunoglobulin variable region genes (V-genes).

The repetitive sequences in the SCMK-B2 group of proteins indicate the major influence of tandem partial gene duplications in the evolution of the group.

McLachlan (1972) disputes series duplication of short segments of polypeptide chain as the most probable method of protein evolution since he failed to find repeating sequences of statistical significance in those proteins that he examined. However, as McLachlan (1972) states, 'any weak repeat that does not correspond to a structural repeat is far less likely to be a long-standing ancestral feature than one that does'. Repeats might then be more likely to persist in proteins subject to stringent structural constraints. The silk fibroin of *Bombyx mori* has alternate glycine and alanine (or serine) residues. The glycine residue permits the close approach (0.35nm) of adjacent faces of β -pleated sheets where only hydrogen atoms protrude. Collagen has a glycine residue located at every third position in the triple helix, only glycine can be accommodated in this position near the axis of the triple helix. The function of the SCMK-B2 group of proteins might likewise depend on a rigidly maintained repetitive structure, so explaining the persistence of the repeating sequences.

I am grateful to Mr. R. E. Guthrie for his skilled technical assistance.

References

- Ambler, R. P. (1963) *Biochem. J.* **89**, 349–378
- Benzer, S. (1961) *Proc. Nat. Acad. Sci. U.S.A.* **47**, 403–415
- Davis, B. J. (1964) *Ann. N.Y. Acad. Sci.* **121**, 404–427
- Elleman, T. C. (1972) *Biochem. J.* **128**, 1229–1239
- Elleman, T. C. & Dopheide, T. A. A. (1972) *J. Biol. Chem.* **247**, 3900–3909
- Gillespie, J. M., Haylett, T. & Lindley, H. (1968) *Biochem. J.* **110**, 193–200
- Gray, W. R. (1967) *Methods Enzymol.* **11**, 139–151
- Lindley, H. (1972) *Biochem. J.* **126**, 683–688
- Lindley, H. & Elleman, T. C. (1972) *Biochem. J.* **128**, 859–867
- Liu, T.-Y. & Chang, Y. H. (1971) *J. Biol. Chem.* **246**, 2842–2848
- Mäkelä, O. & Cross, A. M. (1970) *Progr. Allergy* **14**, 145–207
- Matsuo, H., Fujimoto, Y. & Tatsuno, T. (1966) *Biochem. Biophys. Res. Commun.* **22**, 69–74
- McLachlan, A. D. (1972) *J. Mol. Biol.* **64**, 417–437
- Offord, R. E. (1966) *Nature (London)* **211**, 591–593
- Reindel, R. & Hoppe, W. (1954) *Chem. Ber.* **87**, 1103–1107