The Amino Acid Sequence of Protein SCMK-B2C from the High-Sulphur Fraction of Wool Keratin

By T. C. ELLEMAN

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

(Received 21 February 1972)

1. The amino acid sequence of a protein from the reduced and carboxymethylated high-sulphur fraction of wool has been determined. 2. The sequence of this *S*-carboxymethylkerateine (SCMK-B2C) of 151 amino acid residues displays much internal homology and an unusual residue distribution. Thus a ten-residue sequence occurs four times near the *N*-terminus and five times near the *C*-terminus with few changes. These regions contain much of the molecule's half-cystine, whereas between them there is a region of 19 residues that are mainly small and devoid of cystine and proline. 3. Certain models of the wool fibre based on its mechanical and physical properties propose a matrix of small compact globular units linked together to form beaded chains. The unusual distribution of the component residues of protein SCMK-B2C suggests structures in the wool-fibre matrix compatible with certain features of the proposed models.

The high-sulphur proteins of α -keratins, which constitute the non-filamentous matrix between the microfibrils, consist of several family groups of proteins, each group comprising a number of closely similar components. These proteins are conveniently studied in a soluble form by reduction of the disulphide bonds and reaction of the resulting thiol groups with iodoacetic acid to form *S*-carboxymethylkerateines. A previous paper (Lindley & Elleman, 1972) has dealt with the purification and partial characterization of a group of three such proteins from wool. The present paper describes the determination of the first complete amino acid sequence of a member of this group, component SCMK-B2C (cf. Elleman, 1971).

Materials and Methods

Component SCMK-B2C* was prepared from Lincoln wool as described by Lindley & Elleman (1972). Trypsin, obtained from Calbiochem, Los Angeles, Calif., U.S.A. (B grade, lot no. 801619, treated with diphenylcarbamoyl chloride) was used at an enzyme/substrate ratio of 1:100 (w/w) and incubations were performed at 37°C in 1% (w/v) NH_4HCO_3 buffer, pH 8.0, for the indicated times. Chymotrypsin and trypsin treated with L-1-chloro-4-phenyl-3-tosylamidobutan-2-one (TPCK) (Worthington Biochemical Corp., Freehold, N.J., U.S.A.), thermolysin (Calbiochem) and carboxypeptidase A or B (treated with di-isopropyl phos-

phorofluoridate, Sigma Chemical Co., St. Louis, Mo., U.S.A.) were used under similar conditions for fragmentation of large peptides; pepsin digestions were performed in 5% (v/v) formic acid. These latter digestions were of 8 h duration, or the time stated in the text.

Preparation of peptides

Chromatography of tryptic digests of protein SCMK-B2C on DEAE-cellulose, and fractionation of peptides by high-voltage paper electrophoresis were performed as described by Lindley & Elleman (1972). Peptides on paper were located from guide-strips by the ninhydrin (0.25%)–collidine (2%) stain or by the chlorination method of Reindel & Hoppe (1954).

Several arginine-containing peptides were isolated from a chymotryptic digest of the protein by maleylation of the digest (Butler *et al.*, 1969) and specific retention of arginine-containing peptides on a column (20 cm $\times$ 1.5 cm) of the sulphonic resin Dowex 50 (X2; H^+ form) after elution of non-basic material with water. Arginine-containing peptides were subsequently eluted with 1.0 M NH_3 , and maleyl groups were removed by incubating overnight at pH 3.5, 60°C.

Peptide sequence determination

Sequential analysis of peptides from the *N*-terminus was performed by the dansyl-Edman procedure (Gray, 1967). Large peptides (>12 residues) were fragmented by further enzymic digestion and the

* Abbreviations: SCMK, *S*-carboxymethylkerateine; CmCys, *S*-carboxymethylcysteine.

products, after purification by paper electrophoresis, were then subjected to the dansyl-Edman procedure.

The C-terminal residues released by carboxypeptidase A or B were identified either by paper electrophoresis at pH1.9 or by the amino acid analyser. The selective tritiation method of Matsuo *et al.* (1966) was also used in conjunction with acid hydrolysis and paper electrophoresis at pH1.9 for identification of C-terminal residues.

Amino acid analyses were performed on a Beckman 120C amino acid analyser. Peptides were hydrolysed in 200 μ l of 5.7 M-HCl containing 1 μ l of β -mercaptoethanol for 24h at 110°C *in vacuo*.

Peptides were assigned amide groups on the basis of their electrophoretic mobilities (*m*) relative to aspartic acid (*m* = -1.00) at pH6.5 (Offord, 1966). When both free and amidated carboxyl groups occurred in the same peptide, the mobilities of peptides were measured both before and after removal of the ambiguous residues. Occasionally the mobility of the Dns-peptide is shown [*m*(Dns-OH)], and this is expressed relative to Dns-sulphonic acid, *m*(Dns OH) = -1.00 at pH6.5 (Gray, 1967).

Nomenclature

All tryptic peptides isolated from protein SCMK-B2C are designated by the letter C, followed by a number indicating the fraction in which they were eluted from DEAE-cellulose. The further degradation of tryptic peptides by thermolysin, chymo-

trypsin and pepsin are indicated by the suffixes th, ch and p respectively. The chymotryptic peptides used to obtain overlap between tryptic peptides are designated Ch.

Results

The molecular weight of protein SCMK-B2C was found to be approx. 18000 by the high-speed-equilibrium-ultracentrifugation method of Yphantis (1964). The amino acid analysis of the protein shown in Table 1 is based on this value. No free amino-terminal residue was detectable, and carboxypeptidase A released *S*-carboxymethylcysteine from the C-terminus of the chain.

Tryptic peptides from protein SCMK-B2C

The elution patterns obtained by fractionation on DEAE-cellulose of 3 and 8h tryptic digests of protein SCMK-B2C are shown in Fig. 1. Six major peptides were derived from the 3h digest, the amino acid compositions of which are shown in Table 1.

Fraction 1. This fraction contained much urea and a basic peptide C1a, which was threonylarginine (*m* = 0.75). In the 3h digest a small quantity only of a second basic peptide (*m* = 0.33), which stained yellow with ninhydrin-collidine reagent, was also present; the yield of this peptide (C1b) became near stoichiometric in the 8h digest. The dansyl-Edman

Table 1. Amino acid composition of protein SCMK-B2C and several tryptic peptides

Values are taken from 24h acid hydrolysates, and are expressed as residues/mol.

Amino acid	Protein SCMK-B2C	Peptide						Total tryptic peptides
		C1a	C2	C3	C4	C6a	C6b	
Arg	7.0	1.0	2.0	1.0		1.7	1.0	7
CmCys	32.8 (32.0*)		1.9	7.8	4.6	11.0	7.0	34
Asp	1.0		1.0					1
Thr	15.2 (14.6*)	1.0		6.5 (6.0*)	2.1	2.8	3.0	16
Ser	21.9 (19.0*)			9.8 (8.0*)	3.8	3.9	4.6	24
Glu	15.1			7.0	2.2	5.8		15
Pro	13.0		1.1	2.2	3.1	6.0	1.2	13
Gly	13.8			8.6	1.0	1.2	3.0	14
Ala	6.0			1.0		3.0	2.0	6
Val	5.5 (6.2*)			1.9	2.3	0.9		6
Ile	4.8			3.0		1.1	1.0	5
Leu	3.2			1.0	1.8			3
Tyr	3.7			1.0	1.0	1.4		4
Phe	1.9						1.7	2
Trp			1†					1

* Values obtained after 48h of acid hydrolysis.

† Value obtained from sequence studies.

Fig. 1. Elution profiles from DEAE-cellulose (acetate form) of tryptic digests of protein SCMK-B2C

DE-52 was first precycled according to the manufacturer's instructions, then equilibrated with 3% pyridine-30% acetic acid for 30min and finally washed with 0.15% pyridine until the effluent was near neutral. A column (100cm×1.5cm diam.) was used, flow rate 1.0ml/min, and a linear gradient from 0.15% pyridine (1.6 litre) to 3% pyridine-30% acetic acid (1.6 litre) was applied. Elution profiles of protein SCMK-B2C after 3h (—) and 8h (····) digestion are shown. Peptides were detected after alkaline hydrolysis by reaction with ninhydrin.

procedure established the sequence of this peptide as Gly-Gln-Val-Gly-Ser-Ser-Gly-Ala-Val-Ser-Ser-Arg.

Fraction 2. This fraction contained only a single peptide (C2), which was acidic (*m* = -0.27) and gave a positive Ehrlich reaction. The residual peptide after one Edman degradation was Ehrlich negative, indicating that in all probability an *N*-terminal tryptophan residue had been removed. The complete sequence as determined by the dansyl-Edman procedure was Trp-CmCys-Arg-Pro-Asp-CmCys-Arg.

Fraction 3. This fraction contained only one peptide. Peptide C3 (residues 26-78) stained yellow with the ninhydrin-collidine reagent. The *N*-terminal residue was serine, and the *C*-terminal residue was arginine. Chymotryptic digestion of this peptide produced three fragments (Table 2). Peptide C3ch₁ had *N*-terminal serine, but since amino acid analysis

Table 3. Amino acid compositions of peptides C3, C5 and C1b

Amino acid	Peptide		
	C3	C5	C1b
Arg	1.0		1.0
CmCys	7.8	7.5	
Thr	6.5	6.6	
Ser	9.8	6.4	3.8
Glu	7.0	6.0	1.0
Pro	2.2	2.3	
Gly	8.6	6.0	3.0
Ala	1.0		1.0
Val	1.9		1.9
Ile	3.0	3.0	
Leu	1.0	1.0	
Tyr	1.0	1.0	
<i>N</i> -Terminus ...	Ser	Ser	Gly
<i>C</i> -Terminus ...	Arg	Tyr	Arg

Table 2. Composition and sequence of chymotryptic peptides from peptide C3

The extent of the dansyl-Edman procedure is indicated by —; residues determined as the Dns-derivative both with and without acid hydrolysis are shown by ⇌.

Peptide		
C3ch ₁		<i>N</i> -Terminal serine, sequence not determined
<i>m</i> = -0.75		
C3ch ₂	Composition...	CmCys(2.1);Thr(2.0);Ser(2.0);Glu(2.0);Gly(5.7);Ile(2.0);Tyr(0.9)
<i>m</i> = -0.56	Sequence...	Gln(CmCys ₂ ,Thr ₂ ,Ser,Glu,Gly ₃)Ile(Gly ₂ ,Ser)Ile-Gly-Tyr
		←C3ch ₂ th ₁ → ←C3ch ₂ th ₂ → ←C3ch ₂ th ₃ →
C3ch ₃	Composition...	Ser(3.8);Glu(1.0);Gly(3.0);Ala(1.0);Val(1.9);Arg(1.0)
<i>m</i> = +0.33	Sequence...	Gly-Gln-Val-Gly-Ser-Ser-Gly-Ala-Val-Ser-Ser-Arg

indicated it was heterogeneous, owing to only partial cleavage at a leucine residue, it was not further investigated. Peptide C3ch₂ had *N*-terminal glutamine. It was further fragmented with thermolysin and produced three fragments, the complete sequences of which were not determined, although much could be inferred from enzyme specificity. Peptide C3ch₃ ($m = 0.33$) stained yellow with

ninhydrin-collidine reagent. The sequence was determined by the dansyl-Edman procedure (Table 2) and was identical with that of peptide C1b.

When protein SCM-K-B2C was subjected to prolonged digestion with trypsin (8h), peptide C3 disappeared and was replaced by two fragments (Table 3 and Fig. 1): peptide C1b ($m = 0.33$), which eluted in fraction 1 and was identical in sequence with

Table 4. *Composition and sequence of thermolytic peptides from peptide C5*

For abbreviations see Table 2.

Peptide	Composition	Sequence
C5th ₁ $m = -0.45$	CmCys(1.0);Thr(0.9);Ser(1.6);Glu(0.9)	Ser-Thr-CmCys-Ser-Gln → → → → →
C5th ₂	CmCys(2.0);Thr(1.5);Ser(1.5);Glu(1.0);Pro(0.9) Dns-peptide after four Edman steps; $m(\text{Dns-OH}) = -0.62$	Thr-Ser-CmCys-CmCys-Gln-Pro-Thr-Ser → → → → →
C5th ₃ $m = -0.67$	CmCys(2.9);Thr(2.0);Ser(1.2);Glu(2.0);Pro(1.2);Ile(1.0)	Ile-Gln-Thr-Ser-CmCys-CmCys-Gln-Pro-Thr-CmCys → → → → →
C5th ₄ $m = -0.70$	CmCys(2.0);Thr(2.0);Ser(1.2);Glu(2.3);Gly(2.8)	Leu-Glx-Thr-Ser-Gly-CmCys-Glx-Thr-Gly-CmCys-Gly → → → → →
C5th ₅ $m = 0.00$	Thr(0.9);Glu(1.0);Leu(0.9)	Leu-Gln-Thr → → →
C5th ₆ $m = 0.87$	CmCys(2.1);Thr(1.0);Ser(1.0);Glu(0.9);Gly(3.0)	Ser-Gly-CmCys-Glu-Thr-Gly-CmCys-Gly → → → → →
C5th ₇ $m = 0.00$	Thr(1.0);Ser(1.0);Gly(2.8);Ile(1.0);Tyr(1.0)	Ile-Gly-Gly-Ser-Thr-Gly-Tyr → → → → →
C5th ₈ $m = 0.00$	Ser(0.9);Gly(1.8);Ile(1.0)	Ile-Gly-Gly-Ser → → → →
C5th ₉ $m = 0.00$	Gly(0.9);Ile(1.0);Tyr(0.9)	Ile-Gly-Tyr → → →

Table 5. *Composition and sequence of peptic peptides from peptide C5*

For abbreviations see Table 2.

Peptide	Composition	Sequence
C5p ₁ $m = -0.40$	CmCys(0.9);Thr(1.6);Ser(1.7);Glu(1.0)	Ser-Thr-CmCys-Ser-Gln-Thr → → → → →
C5p ₂ $m = -0.64$	CmCys(2.0);Thr(1.4);Ser(1.9);Glu(1.2);Pro(0.8)	Ser-CmCys-CmCys-Gln-Pro-Thr-Ser → → → → →
C5p ₃ $m = 0.00$	Thr(1);Gln(1);Ile(1) (exact values not recorded)	Ile-Gln-Thr → → →
C5p ₄ $m = -0.80$	CmCys(3.0);Thr(1.0);Ser(0.8);Glu(1.0);Pro(1.0);Leu(0.8)	Ser-CmCys-CmCys-Gln-Pro-Thr-CmCys-Leu → → → → →
C5p ₅ $m = -0.54$	CmCys(2.0);Thr(2.0);Ser(2.0);Glu(2.0); Gly(5.6);Ile(2.0);Tyr(1.0)	Glx-Thr(CmCys ₂ ,Thr ₁ ,Ser ₂ ,Glx ₁ ,Gly ₆ ,Ile ₂)Tyr → → →

peptide C3ch₃, and peptide C5, which eluted in fraction 5 and was a large acidic fragment staining yellow with the ninhydrin-collidine reagent. These additional fragments had arisen from peptide C3 by chymotrypsin-like activity in the trypsin sample (Calbiochem). Further sequence studies were performed on peptide C5, since this was easily obtained in a homogeneous state, unlike peptide C3ch₁. Thermolytic (Table 4) and peptic (Table 5) digestions

of peptide C5 produced two complete sets of overlapping fragments. From the sequences of these peptides (as determined by the dansyl-Edman method), together with that of peptide C1b, the chymotryptic fragments from peptide C3, and the *N*-terminal and *C*-terminal residues of peptide C3, the complete sequence of peptide C3 could be deduced (Fig. 2).

Fraction 4. This fraction contained only one

Fig. 2. Amino acid sequence of tryptic peptide C3

Sequences determined by the dansyl-Edman procedure are shown by $\rightarrow$; identification of the *C*-terminal residue after the Edman degradation procedure has been completed, performed both with and without acid hydrolysis of the Dns-derivative, is shown by $\Rightarrow$; residues released by carboxypeptidase B are shown by $\leftarrow$; sequences not determined are shown by -----.

Table 6. Amino acid sequence of peptide C4 and the composition and sequence of its thermolytic and peptic peptides

For abbreviations see Table 2.

Sequence of peptide C4...	Val-Glu-Gly-Thr-Ser-Leu-Pro-Pro-CmCys-CmCys-Val-Val-Ser-CmCys-Thr-Ser-Pro-Ser-CmCys-CmCys-Gln-Leu-Tyr
	<-----C4th ₁ -----<-----C4th ₂ -----<-----C4th ₄ -----<-----
	<-----C4th ₃ -----<-----
	<-----C4p ₁ -----<-----
Release of amino acids by carboxypeptidase A...	Leu(+); Tyr(+)
Thermolytic peptides	
C4th ₁ Composition...	CmCys(1.8); Thr(0.9); Ser(0.9); Glu(1.0); Pro(2.0); Gly(1.0); Val(1.0); Leu(1.0)
<i>m</i> = -0.72 Sequence...	Val-Glu-Gly-Thr-Ser-Leu-Pro-Pro-CmCys-CmCys
C4th ₂ Composition...	CmCys(2.4); Thr(1.0); Ser(2.3); Glu(1.0); Pro(1.0); Val(1.1)
<i>m</i> = -0.65 Sequence...	Val-Val-Ser-CmCys-Thr-Ser-Pro-Ser-CmCys-CmCys-Gln
C4th ₃ Composition...	CmCys(4.8); Thr(2.0); Ser(3.5); Glu(2.1); Pro(3.0); Gly(1.1); Val(2.5); Leu(1.0)
<i>m</i> = -0.80 Sequence...	Val-Glx-Gly-Thr-Ser-Leu-Pro-Pro-CmCys-CmCys-Val-Val (CmCys ₃ , Thr ₁ , Ser ₃ , Glx ₁ , Pro ₁)
C4th ₄ Composition...	Leu(0.9); Tyr(1.0)
<i>m</i> = -0.00 Sequence...	Leu-Tyr
Pepsin peptides	
C4p ₁ Composition...	CmCys(1.0); Glu(1.0); Leu(1.0); Tyr(1.0)
<i>m</i> = -0.40 Sequence...	CmCys-Gln-Leu-Tyr

peptide. Peptide C4 (residues 88–110) had *N*-terminal valine and C-terminal tyrosine and presumably had arisen from chymotrypsin-like activity in the trypsin. Thermolytic digestion of this peptide produced four fragments, one of these (C4th₃) had apparently been partially hydrolysed to peptides C3th₁ and C3th₂.

These peptides and a peptic fragment which demonstrates the Gln-Leu bond between fragments C4th₃ and C4th₄ are shown in Table 6.

Fraction 5. This fraction as obtained from the 3 h digest contained peptides C4, C5 and C6a, all in very low yield. Prolonged digestion times (8h)

Table 7. *Composition and sequence of thermolytic peptides from peptide C6a*

For abbreviations see Table 2.

Peptide	Composition	Sequence
C6ath ₁ <i>m</i> = 0.00	Glu(1.0);Ala(1.0);Tyr(1.0)	<u>Tyr-Ala-Gln</u>
C6ath ₂ <i>m</i> = -0.31	CmCys(2);Ser(2);Pro(1);Ala(1);Arg(1) (exact analysis lost)	<u>Ala-Ser-CmCys-CmCys-Arg-Pro-Ser</u>
C6ath ₃ <i>m</i> = -0.50	CmCys(1.0);Tyr(1.0)	<u>Tyr-CmCys</u>
C6ath ₄ <i>m</i> = -0.31	Not determined because of slight contamination from Ala-Ser-CmCys-CmCys-Arg-Pro	<u>Gly-Gln-Ser-CmCys-CmCys-Arg-Pro</u>
C6ath ₅ <i>m</i> = -0.90	CmCys(3.8);Thr(2.0);Glu(2.0);Pro(2.3);Ala(1.0)	<u>Ala-CmCys-CmCys-CmCys-Glx-Pro</u> (CmCys ₁ , Thr ₂ , Glx ₁ , Pro ₁)
C6ath ₆ <i>m</i> = -0.70	CmCys(2.0);Thr(0.9);Ser(0.9);Glu(1.8);Pro(2.0);Val(0.8) Dns-peptide after two Edman steps, <i>m</i> (Dns-OH) = -1.18 Dns-peptide after three Edman steps, <i>m</i> (Dns-OH) = -0.83	<u>Val-CmCys-Glu-Pro-Thr-CmCys-Ser-Gln-Pro</u>
C6ath ₇ <i>m</i> = -0.90	CmCys(6.0);Thr(2.8);Ser(1.1);Glu(4.0) Pro(4.3);Ala(1.0);Val(0.9)	<u>Ala-CmCys-CmCys-CmCys-Glx-Pro-Thr-CmCys-Thr-Glx</u> (CmCys ₂ , Thr ₁ , Ser ₁ , Glx ₂ , Pro ₃ , Val ₁)
C6ath ₈ <i>m</i> = -0.60	CmCys(1.0);Ile(0.8)	<u>Ile-CmCys</u>

Table 8. *Composition and sequence of peptic peptides obtained from peptide C6a*

For abbreviations see Table 2.

Peptide	Composition	Sequence
C6ap ₁ <i>m</i> = -0.24	CmCys(1.8);Ser(2.0);Glu(1.1);Pro(0.9);Ala(1.8);Tyr(0.9);Arg(+)	<u>Ala-Gln-Ala-Ser-CmCys-CmCys-Arg-Pro-Ser-Tyr</u>
C6ap ₂ <i>m</i> = -0.50	CmCys(1.0);Ser(1.0);Glu(1.0);Gly(1.0)	<u>CmCys-Gly-Gln-Ser</u>
C6ap ₃ <i>m</i> = -0.66	CmCys(3.0);Pro(1.2);Ala(1.0);Arg(1.0)	<u>CmCys-CmCys-Arg-Pro-Ala-CmCys</u>
C6ap ₄ <i>m</i> = -1.00	CmCys(3.0);Thr(1.8);Glu(2.0);Pro(1.0) Dns-peptide after two Edman steps, <i>m</i> (Dns-OH) = -1.21 Dns-peptide after three Edman steps, <i>m</i> (Dns-OH) = -1.26	<u>CmCys-CmCys-Gln-Pro-Thr-CmCys-Thr-Glu</u>
C6ap ₅ <i>m</i> = -0.80	CmCys(1.0);Glu(0.9);Pro(1.1);Val(0.7)	<u>Pro-Val-CmCys-Glu</u>
C6ap ₆ <i>m</i> = -0.60	CmCys(2.0);Thr(0.9);Ser(0.9);Glu(1.0);Pro(2.0);Ile(0.9)	<u>Pro-Thr-CmCys-Ser-Gln-Pro-Ile-CmCys</u>

increased the yield of peptide C5 in this fraction at the expense of peptide C3. Peptide C5 was separated from small amounts of C4 and C6a by high-voltage paper electrophoresis at pH 3.5, and the sequence was determined as indicated above.

Fraction 6. This fraction contained two peptides, which were separated by paper electrophoresis at pH1.9. Peptide C6a (residues 111–151) had *N*-terminal tyrosine and *C*-terminal *S*-carboxymethylcysteine, and so constitutes the *C*-terminal sequence of the protein. Complete overlapping sets of thermolytic (Table 7) and peptic (Table 8) peptides were obtained from this fragment, from which the complete amino acid sequence of peptide C6a could be deduced (Fig. 3). Peptide C6b (residues 1–25) was ninhydrin-negative. Digestion with chymotrypsin produced two fragments (Table 9), a ninhydrin negative heptapeptide C6bch₁ characteristic of the SCMK-B2 family of proteins (Lindley & Elleman, 1972) and a second large fragment C6bch₂.

The *C*-terminal residue of the ninhydrin-negative

fragment C6bch₁ was determined, by the selective tritiation method of Matsuo *et al.* (1966), as phenylalanine. Carboxypeptidase A released both phenylalanine and serine after 1 h of digestion, phenylalanine, serine and trace amounts of threonine after 2 h of digestion and equimolar amounts of each after 24 h of digestion. Isolation of the residual ninhydrin-negative fragment after 24 h of carboxypeptidase digestion yielded a peptide of amino acid composition *S*-carboxymethylcysteine (2.0), serine (0.8), alanine (1.0), with a *C*-terminal serine residue, as determined by the selective tritiation method of Matsuo *et al.* (1966). The acetylated alanine residue from the chymotryptic fragment has been identified by mass spectrometry (Lindley & Elleman, 1972).

The sequence of peptide C6bch₂ was elucidated by the dansyl-Edman technique applied to both the fragment and to the peptides produced by thermolytic digestion of the fragment (Table 10).

Peptide C7 (residues 88–151). This fragment, the amino acid composition of which is shown in Table

Fig. 3. Amino acid sequence of tryptic peptide C6a

Symbols used are described in the legend to Fig. 2, except ← denotes residues released by carboxypeptidase A.

Table 9. Amino acid sequence of peptide C6b and the composition and sequence of its chymotryptic and thermolytic peptides

Residues determined by release with carboxypeptidase A are shown by ←, residues determined by selective tritiation are shown by ≡; residues determined by the dansyl-Edman procedures are shown by →.

Sequence of C6b...	Ac-Ala-CmCys-CmCys-Ser-Thr-Ser-Phe-CmCys-Gly-Phe-Pro-Ile-CmCys-Ser-Thr-Ala-Gly-Thr-CmCys-Gly-Ser-Ser-CmCys-CmCys-Arg
	←C6bch ₁ →←C6bch ₂ →
	←C6bth ₁ →
Chymotryptic peptides	
C6bch ₁	Composition... CmCys(2.0);Thr(1.0);Ser(1.8);Ala(1.1);Phe(1.0)
<i>m</i> = -0.87	Sequence... Ac-Ala-CmCys-CmCys-Ser-Thr-Ser-Phe
C6bch ₂	Composition... CmCys(5.0);Thr(2.0);Ser(2.8);Pro(1.2);Gly(3.0);Ala(1.0);Ile(1.0);Phe(1.0);Arg(1.0)
<i>m</i> = -0.62	Sequence... CmCys-Gly-Phe-Pro-Ile-CmCys-Ser-Thr-Ala-Gly-Thr-CmCys-(CmCys ₂ ,Ser ₂ ,Gly)-Arg
	←C6bch ₂ th ₁ →←C6bch ₂ th ₂ →←C6bch ₂ th ₃ →
Thermolytic peptide	
C6bth ₁	Composition... CmCys(1.0);Pro(1.0);Gly(1.0);Phe(2.0)
<i>m</i> = -0.40	Sequence... Phe-CmCys-Gly-Phe-Pro

Table 10. *Composition and sequence of thermolytic peptides from chymotryptic peptide C6bch₂*

For abbreviations see Table 2.

Peptide	Composition	Sequence
C6bch ₂ th ₁ <i>m</i> = -0.45	CmCys(1.0);Pro(1.0);Gly(1.0);Phe(1.0)	<u>CmCys-Gly-Phe-Pro</u>
C6bch ₂ th ₂ <i>m</i> = -0.45	CmCys(1.1);Thr(0.9);Ser(0.9);Ile(0.9)	<u>Ile-CmCys-Ser-Thr</u>
C6bch ₂ th ₃ <i>m</i> = -0.50	CmCys(3.0);Thr(1.2);Ser(1.7);Gly(2.0);Ala(0.9);Arg(1.0)	<u>Ala-Gly-Thr-CmCys-Gly-Ser-Ser-CmCys-CmCys-Arg</u>

11, was not present in digests of protein SCMK-B2C obtained by the use of trypsin treated with diphenylcarbamoyl chloride (Calbiochem), but was later isolated after digestion of the protein with trypsin treated with L-1-chloro-4-phenyl-3-tosylamidobutan-2-one (TPCK) (Worthington). It eluted later than peptide C6b from the DEAE-cellulose column, and peptides C4 and C6a were notably absent from this digest. Digestion of peptide C7 with the trypsin treated with diphenylcarbamoyl chloride (Calbiochem) produced two fragments which were separated on a column (150cm×2cm) of Sephadex G-50 in 1% (w/v) NH₄HCO₃ solution. These peptides were identical in amino acid composition with peptides C4 and C6a, which had evidently arisen from peptide C7 by cleavage of a Tyr-Tyr bond (Table 11).

Arginine-containing chymotryptic peptides from protein SCMK-B2C

Several small arginine-containing peptides were isolated from a chymotryptic digest of protein SCMK-B2C by high-voltage paper electrophoresis at pH 6.5 and pH 3.5 (Ch₁ to Ch₅ of Table 12). Other larger peptides retained by the Dowex 50 (X2) column after maleylation of the digest were separated from the smaller material by a column (100cm×2cm diam.) of Sephadex G-50 in 1% (w/v) NH₄HCO₃ solution and were further fractionated on a column (30cm×1.5cm) of DEAE-cellulose bicarbonate by using a linear gradient from 0.005M- to 1.0M-NH₄HCO₃. Peptides Ch₆ and Ch₇ were further purified by paper electrophoresis at pH 3.5 from two fractions which eluted from the DEAE-cellulose at approx. 0.3M- and approx. 0.4M-NH₄HCO₃ respectively (Table 12).

The absence of the remaining arginine-containing peptide from these studies indicated that it was probably too large or too acidic to be retained by the Dowex 50 (X2). Consequently a chymotryptic digest of protein SCMK-B2C was fractionated on a column (150cm×2cm) of Sephadex G-75 in 1% (w/v) NH₄HCO₃ solution and the large material was ex-

Table 11. *Amino acid compositions of peptides C7, C4 and C6a*

Residue	Peptide		
	C7	C4	C6a
Arg	2.0		1.7
CmCys	16.1	4.6	11.0
Thr	5.0	2.1	2.8
Ser	8.0	3.8	3.9
Glu	8.1	2.2	5.8
Pro	9.1	3.1	6.0
Gly	2.0	1.0	1.2
Ala	3.0		3.0
Val	3.0	2.3	0.9
Ile	0.9	—	1.1
Leu	2.0	1.8	
Tyr	2.9	1.0	1.4
N-Terminus ...	Val	Val	Tyr
C-Terminus ...	CmCys	Tyr	CmCys

amined for a trypsin-sensitive peptide. Peptide Ch₈ was thus found (Table 13) and was purified by paper electrophoresis at pH 3.5.

The tryptic peptides and the chymotryptic arginine-containing peptides allow the unequivocal construction of the sequence of protein SCMK-B2C (Fig. 4) since the protein lacks lysine.

Discussion

Protein SCMK-B2C displays an unusual amino acid sequence. Half-cystine comprises 23% of the molecule and serine and threonine another 26%. Charged groups are scarce except near the centre of the chain; residues 78 to 88 contain four of the seven arginines and two of the five acidic residues. The central region has a preponderance of small residues (residues 59 to 77 contain seven of the 14 glycines) and the chain is devoid of half-cystine and proline from residues 59 to 81. This could lend flexibility to this region, since glycine has no β-carbon

Table 12. Composition and sequence of arginine overlap peptides from a chymotryptic digest of protein SCMK-B2C
For abbreviations see Table 2.

Peptide	Composition	Sequence
Ch ₁	Thr(1.0);Ser(1.0);Arg(2.0);Trp(+)	<u>Ser-Arg-Thr-Arg</u> -(Trp)
m=+0.65		
Ch ₂	Arg(+);Trp(+)	<u>Arg-Trp</u>
m=+0.65		
Ch ₃	Thr(1.0);Ser(3.5);Glu(1.0);Gly(3.0); Ala(1.1);Val(2.1);Arg(2.0);Trp(+)	<u>Gly-Glx-Val-Gly-Ser-Ser-Gly-Ala-Val-Ser-Ser-Arg-Thr-Arg-Trp</u> *
Ch ₄	CmCys(1.6);Ser(1.8);Glu(1.0);Pro(1.0); Ala(1.7);Tyr(0.7);Arg(1.0)	<u>Ala-Gln-Ala-Ser-CmCys-CmCys-Arg-Pro-Ser-Tyr</u>
m=-0.25		
Ch ₅	CmCys(3.0);Ser(1.6);Glu(2.0);Pro(1.3); Gly(0.9);Ala(2.0);Tyr(0.9);Arg(1.0)	<u>Ala-Gln-Ala-Ser-CmCys-CmCys-Arg-Pro-Ser-Tyr-CmCys-Gly-Gln</u>
m=-0.40		
Ch ₆	CmCys(7.1);Asp(1.0);Thr(2.0);Ser(3.3);Glu(2.1); Pro(4.2);Gly(1.0);Val(2.6);Leu(2.0);Tyr(1.0);Arg(2.0)	<u>CmCys-Arg-Pro-Asx-CmCys-Arg-Val-Glx</u> -(CmCys, Thr ₂ ,Ser ₄ ,Glx,Pro ₃ ,Gly,Val ₂ ,Leu ₂ ,Tyr)
m=-0.64		
Ch ₇ †	CmCys(12.0);Thr(2.8);Ser(3.2);Glu(5.8);Pro(6.1); Gly(1.1);Ala(2.8);Val(1.1);Ile(0.9);Tyr(1.0);Arg(2.0)	

* Sequence deduced from composition and tryptic peptides.
† N-terminus: alanine, sequence not determined; the peptide is equivalent to peptide C6a less the N-terminal tyrosine residue.

Table 13. Amino acid sequence and composition of chymotryptic peptide Ch₈ and its tryptic peptides
Residues determined by release with carboxypeptidase A are shown by ←, residues determined by the dansyl-Edman procedures are shown by →.

Composition...	CmCys(8.0);Thr(4.5);Ser(5.4);Glu(3.0);Pro(2.4);Gly(2.7);Ala(1.0);Ile(1.7);Phe(0.8);Arg(1.0)	
Sequence...	CmCys-Gly-Phe-Pro-Ile-CmCys(CmCys ₃ ,Thr ₂ ,Ser ₃ ,Gly ₂ ,Ala ₁)Arg-Ser-Thr-CmCys-Ser-Glx-Thr(CmCys ₂ ,Ser ₁ ,Glx ₁ ,Pro ₁)Thr(Ser ₁ ,Ile ₁ ,Gln)	
	← Ch _{8t1}	→ Ch _{8t2}
Tryptic peptides	Composition	Sequence
Ch _{8t1}	CmCys(5.0);Thr(1.9);Ser(2.7);Pro(1.1);Gly(3.0); Ala(1.0);Ile(1.0);Phe(1.0);Arg(0.9)	<u>CmCys-Gly-Phe-Pro-Ile-CmCys</u> (CmCys ₃ ,Thr ₂ ,Ser ₃ ,Gly ₂ ,Ala ₁)Arg
m=-0.60		
Ch _{8t2}	CmCys(3.0);Thr(3.0);Ser(3.3);Glu(2.8);Pro(0.9); Ile(0.9)	<u>Ser-Thr-CmCys-Ser-Glx-Thr</u> (CmCys ₂ ,Ser ₁ ,Glx ₁ ,Pro)Thr(Ser,Ile,Gln)
m=-0.52		
Amino acids released by carboxypeptidase A action on peptide		
Ch _{8t2} ...	Thr(0.5);Ser(1.0);Gln(0.7);Ile(1.0)	

atom and can rotate particularly easily, whereas proline is the ‘stiffest’ of the residues, and cystine is restrictive by its disulphide bonding.
The N- and C-terminal regions of high cystine content (more than one residue in every four), are of similar size and both are largely composed of a similar repeating segment. A group of smaller proteins from the high-sulphur fraction of wool has been studied by Haylett *et al.* (1971). The proteins of this group also contain a region of high cystine content of similar size to the corresponding regions in protein SCMK-B2C.

Models of the wool-fibre-matrix structure must be compatible with observations on the mechanical properties of the wool fibre, such as (a) the extensive swelling of matrix in water (Fraser *et al.*, 1959, 1971) and formic acid (Bendit, 1966; Gillespie, 1970), (b) the resistance of wool to proteolysis despite the large water uptake (Geiger *et al.*, 1941), (c) the sudden increase in resistance to extension of the wool fibre at about 30% strain and the change in position of this transition point to higher strains at slow rates of stress, higher temperatures and on disulphide reduction, (d) a post-yield region only

1	Ac-Ala	Cys	Cys	Ser	Thr	Ser	Phe	Cys	Gly	Phe
11	Pro	Ile	Cys	Ser	Thr	Ala	Gly	Thr	Cys	Gly
21	Ser	<u>Ser</u>	<u>Cys</u>	<u>Cys</u>	<u>Arg</u>	Ser	<u>Thr</u>	<u>Cys</u>	Ser	<u>Gln</u>
31	<u>Thr</u>	<u>Ser</u>	<u>Cys</u>	<u>Cys</u>	<u>Gln</u>	Pro	<u>Thr</u>	Ser	<u>Ile</u>	<u>Gln</u>
41	<u>Thr</u>	<u>Ser</u>	<u>Cys</u>	<u>Cys</u>	<u>Gln</u>	Pro	<u>Thr</u>	<u>Cys</u>	<u>Leu</u>	<u>Gln</u>
51	<u>Thr</u>	<u>Ser</u>	Gly	<u>Cys</u>	Glu	Thr	Gly	<u>Cys</u>	Gly	Ile
61	Gly	Gly	Ser	Ile Thr	Gly	Tyr	Gly	Gln	Val	Gly
71	Ser	Ser	Gly	Ala	Val	Ser	Ser	Arg	Thr	Arg
81	Trp	Cys	Arg	Pro	Asp	Cys	Arg	Val	Glu	Gly
91	Thr	Ser	Leu							
94	Pro	Pro	<u>Cys</u>	<u>Cys</u>	Val	Val	<u>Ser</u>	<u>Cys</u>	Thr	Ser
104	Pro	<u>Ser</u>	<u>Cys</u>	<u>Cys</u>	<u>Gln</u>	Leu	Tyr	<u>Tyr</u>	Ala	<u>Gln</u>
114	Ala	<u>Ser</u>	<u>Cys</u>	<u>Cys</u>	<u>Arg</u>	Pro	<u>Ser</u>	<u>Tyr</u>	Cys	Gly
124	Gln	<u>Ser</u>	<u>Cys</u>	<u>Cys</u>	<u>Arg</u>	Pro	Ala	<u>Cys</u>		
132			<u>Cys</u>	<u>Cys</u>	<u>Gln</u>	Pro	<u>Thr</u>	<u>Cys</u>	Thr	Glu
140		Pro	Val	<u>Cys</u>	Glu	Pro	<u>Thr</u>	<u>Cys</u>	Ser	<u>Gln</u>
149	Pro	Ile	Cys							

Fig. 4. Complete amino acid sequence of protein SCMK-B2C, arranged to show maximum internal homology. Homologous residues are underlined. Homology confined only to a small region of the molecule is shown by ...

slightly affected by complete disruption of the microfibrillar helices (Crewther, 1972; Feughelman, 1968).

Crewther (1965, 1972) favours a model for the wool fibre which incorporates these features on the basis of a matrix composed of globular units highly stabilized by intramolecular cross-links, but with few intermolecular bonds. On this model, extension to the transition point is resisted by strain and by rupture of the secondary bonds of the microfibrillar helices and by rupture of the bonds between the side-chains of helices and the closely coupled matrix molecules. Partial unfolding of the helices and concomitant alignment of the beaded chain-like matrix molecules occur in this region. Beyond this point opposition to extension is offered by the now aligned matrix molecules, which must stretch or open out, a process facilitated by thiol-disulphide interchange (Burley, 1956). X-ray-diffraction studies (Fraser, 1961) are in accord with the view that the matrix consists of more or less randomly arranged globules with about 2.1 nm (21 Å) between centres. This value would indicate the globules were much smaller than whole high-sulphur protein molecules.

The reported sequence with regions high in cystine content, separated by a flexible chain devoid of cystine, suggests a structure of two compact globular units, which might move relatively freely with respect to each other, an arrangement compatible with both the proposed model and the X-ray information.

The low elastic modulus of the matrix in the presence of water, 0.076 times that of the microfibrils

(Feughelman, 1959), further emphasizes the lack of intermolecular bonding within the matrix and the relative ease of matrix deformation. A matrix exhibiting such properties would easily be deformed under load, distributing the stress along the microfibrils of high tensile strength. Breaks in the microfibrils would be isolated by the matrix, its yielding viscoelastic properties transferring the stress to the other intact microfibrils and so preventing the propagation of a break. Such composite structures are associated with high tensile strength and yet remarkable flexibility (Slayter, 1962).

Crewther & Dowling (1960) have suggested that thiol-disulphide interchanges involving little conformational energy might be mediated through structures in the matrix. The repetitive sequences in protein SCMK-B2C are suggestive of an ordered structure within the matrix units, which might be involved in facilitating this interchange and so endowing the wool fibre with its characteristic properties of creep and stress relaxation.

I am grateful to Dr. H. Lindley and Dr. W. G. Crewther for their interest and advice, to Mr. R. E. Guthrie for technical assistance, and to Mrs. M. J. Pont for the molecular weight determination.

References

- Bendit, E. G. (1966) *Nature (London)* **211**, 1257
- Burley, R. W. (1956) *Proc. Int. Wool Text. Res. Conf.* 1st. 1955, D, 88

- Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1969) *Biochem. J.* **112**, 679
- Crewther, W. G. (1965) *Text. Res. J.* **35**, 867
- Crewther, W. G. (1972) *Text. Res. J.* **42**, 77
- Crewther, W. G. & Dowling, L. M. (1960) *Text. Res. J.* **51**, T775
- Elleman, T. C. (1971) *Nature New Biol. (London)* **234**, 148
- Feughelman, M. (1959) *Text. Res. J.* **29**, 223
- Feughelman, M. (1968) *Symp. Fibrous Proteins, Int. Conf.*, 3rd, 1967, 397
- Fraser, R. D. B. (1961) *Structure de la Laine*, p. 25, Institut Textil de France, Paris
- Fraser, R. D. B., MacRae, T. P. & Rogers, G. E. (1959) *Nature (London)* **183**, 592
- Fraser, R. D. B., MacRae, T. P., Millward, G. R., Parry, D. A. D., Suzuki, E. & Tulloch, P. A. (1971) *Appl. Polym. Symp.* **18**, 65
- Geiger, W. B., Patterson, W. I., Mizell, L. R. & Harris, M. (1941) *J. Res. Nat. Bur. Stand.* **27**, 459
- Gillespie, J. M. (1970) *Text. Res. J.* **40**, 853
- Gray, W. R. (1967) *Methods Enzymol.* **11**, 139
- Haylett, T., Swart, L. S. & Parris, D. (1971) *Biochem. J.* **123**, 191
- Lindley, H. & Elleman, T. C. (1972) *Biochem. J.* **128**, 859
- Matsuo, M., Fujimoto, Y. & Tatsuno, T. (1966) *Biochem. Biophys. Res. Commun.* **22**, 69
- Offord, R. E. (1966) *Nature (London)* **211**, 591
- Reindel, R. & Hoppe, W. (1954) *Chem. Ber.* **87**, 1103
- Slayter, G. (1962) *Sci. Amer.* **206**, 124
- Yphantis, D. A. (1964) *Biochemistry* **3**, 297