Biochem. J. (1971) **123**, 201–210
Printed in Great Britain

Studies on the High-Sulphur Proteins of Reduced Merino Wool

AMINO ACID SEQUENCE OF PROTEIN SCMKB-IIIB4

BY L. S. SWART AND T. HAYLETT
South African Wool Textile Research Institute, South African Council for Scientific and Industrial Research, Port Elizabeth, South Africa

(*Received* 30 *December* 1970)

The complete amino acid sequence of protein SCMKB-IIIB4 is presented. It is closely related to the sequence of protein SCMKB-IIIB3 (Haylett, Swart & Parris, 1971) differing in only four positions. The peptic and thermolysin peptides of protein SCMKB-IIIB4 were analysed by the dansyl–Edman method (Gray, 1967) and by tritium-labelling of *C*-terminal residues (Matsuo, Fujimoto & Tatsuno, 1966). This protein is the third member of a group of high-sulphur wool proteins with molecular weight of about 11400. It consists of 98 residues and has acetylalanine and carboxymethylcysteine as *N*- and *C*-terminal residues respectively.

The complete amino acid sequences of two high-sulphur proteins, SCMKB-IIIB2 (Haylett & Swart, 1969) and SCMKB-IIIB3 (Haylett, Swart & Parris, 1971) have been described. These proteins are homologous and have 28 mutational differences between them. Although the preparation of protein SCMKB-IIIB3 contained approx. 30% of protein SCMKB-IIIB2 as impurity, the sequence of the former protein could still be determined unequivocally.

Protein SCMKB-IIIB4 has the same molecular weight and very similar amino acid composition as the proteins of which the sequence has been determined. This paper describes the elucidation of the sequence of protein SCMKB-IIIB4.

MATERIALS AND METHODS

Protein IIIB4* was prepared as described by Swart, Haylett & Joubert (1969) except that after rechromatography on DEAE-cellulose only the material eluted after the maximum of the peak was used for sequence determination.

Enzymes. Pepsin and papain were twice crystallized preparations from Worthington Biochemical Corp., Freehold, N.J., U.S.A.; thermolysin was from Daika Kasei K.K., Osaka, Japan.

Enzymic digestions. Peptic digestion was carried out on 30 μmol of protein IIIB4 in 35ml of 5% (v/v) formic acid (enzyme/substrate ratio 1:100, w/w) for 2h at 40°C. At this stage all the protein was not in solution and the digestion was continued for a further 16h at 35°C. The digestion was stopped by addition of pyridine and the material was freeze-dried.

Thermolysin digestion was performed on 20 μmol of the protein in 24ml of 2% (w/v) NH_4HCO_3 (enzyme/substrate ratio 1:100, w/w) for 1h at 40°C. The digestion mixture was acidified to pH3 with acetic acid and freeze-dried.

Papain digestion was carried out on 4 μmol of peptide 4L7 in 4ml of buffer (0.2M-pyridine adjusted to pH5.5 with acetic acid); 40 μl of papain suspension (13.6mg/ml) and 8 μl of β-mercaptoethanol were added. The mixture was incubated under N_2 for 16h at 40°C and freeze-dried.

Fractionation of peptides. The peptides were separated by chromatography on DEAE-cellulose and by paper electrophoresis and chromatography as described by Haylett & Swart (1969) and Haylett *et al.* (1971).

Analysis of protein and peptides. Amino acid analyses, dansyl–Edman degradations and amide determinations were carried out as described by Haylett *et al.* (1971). *C*-Terminal residues were tritium-labelled and determined as described by Haylett *et al.* (1971). Mass spectrometry was performed at 105°C with an AEI MS9 mass spectrometer after permethylation (Lenard & Gallop, 1970) and desulphurization (Kiryushkin, Gorlenko, Rosinov, Ovchinnikov & Shemyakin, 1969) of the peptide.

Nomenclature of peptides. Capital letters indicate enzymic digestions: P, L and Q for pepsin, thermolysin and papain respectively. Arabic numerals preceding these letters indicate the protein from which the peptides are derived: 2, 3 and 4 for proteins IIIB2, IIIB3 and IIIB4 respectively. Arabic numerals following the capital letters indicate sequential numbering of the peptides.

RESULTS

Terminal analysis of protein IIIB4

No *N*-terminal residue could be found for protein IIIB4 by the dansyl–Edman method. Only carboxymethylcysteine was found as the *C*-terminal

* Abbreviations: IIIB4, SCMKB-IIIB4; IIIB2, SCMKB-IIIB2; IIIB3, SCMKB-IIIB3; BAWP, butan-1-ol–acetic acid–water–pyridine (15:10:12:3, by vol.); BAW, butan-1-ol–acetic acid–water (40:6:15, by vol.); CMCys (in sequences and tables), *S*-carboxymethylcysteine.

Table 1. *Amino acid composition of the peptic peptides of protein III B4*

Peptide ...	4P1	4P2	4P3	4P4	4P5	4P6	4P7	4P8	4P9
	Amino acid composition (mol of residue/mol)								
Trp								+	
Lys					0.99	1.06			
His								0.97	
Arg		0.99					0.89		
CMCys	1.88		1.99	2.57	0.96		2.97	1.07	2.65
Asp					1.13	0.99			
Thr			3.12	2.94				2.02	0.96
Ser			1.77	1.95	2.02	2.01		0.92	
Glu									1.12
Pro			1.89	2.12				1.78	1.00
Gly							0.99		
Ala	1.00	0.97	0.83	0.84					
Val			1.00	1.17			0.97	1.01	
Ile				0.79	1.07				
Leu		1.04					1.04		0.85
Tyr									
Phe					0.99	0.80			0.83
Yield (μmol)	17.5	13.4	13.2	1.4	15.7	3.6	16.0	10.8	6.1
Peak no. (from Fig. 1)	13	1	6	10	5	3	9	3	11
System*	—	1.9	BAW	1.9	BAW	BAW	—	BAW	BAWP
R_F	—	—	0.70	—	0.23	0.14	—	0.22	0.50
m_{Lys}†	—	0.89	—	0.29	—	—	—	—	—
N-Terminal	—	Ala	CMCys	CMCys	Ile	Ser	CMCys	Pro	Phe
C-Terminal	CMCys	Leu	Thr	CMCys	Phe	Phe	Leu	Trp	CMCys

Peptide ...	4P10	4P11	4P12	4P13	4P14	4P15	4P16	4P17	4P18	4P19
	Amino acid composition (mol of residue/mol)									
Trp										
Lys										
His	0.92									
Arg	0.97									
CMCys	1.88							0.98	1.03	2.82
Asp	1.85	1.01	0.99	0.99	1.01	1.01	1.02			
Thr	1.04	1.16	0.99	0.99				3.05	1.02	
Ser	0.90	1.82	1.80	2.82	0.95	0.96		0.93	0.91	0.95
Glu	1.03	1.02	2.09	2.04	1.03			2.10	1.89	
Pro	5.86	1.78	2.25	2.14				0.96	0.96	2.07
Gly		1.19	1.01	0.98						
Ala										
Val	1.03									
Ile	0.91				0.98	0.98	1.00			0.98
Leu		2.78	3.13	3.09	1.02	1.04	0.99			
Tyr								0.70	0.83	
Phe	0.91									
Yield (μmol)	4.7	5.6	5.3	2.8	5.0	5.8	4.7	8.6	1.3	18.9
Peak no. (from Fig. 1)	5	2	4	4	5	3	2	7	8	12
System*	BAW	BAWP	BAW	BAW	BAW	BAW	BAWP	BAW	BAW	—
R_F	0.10	0.54	0.36	0.26	0.42	0.41	0.61	0.10	0.14	—
m_{Lys}†	—	—	—	—	—	—	—	—	—	—
N-Terminal	Asp	Leu	Leu	Leu	Glu	Ser	Ile	Thr	Tyr	Pro
C-Terminal	Phe	Leu	Glx	Ser	Leu	Leu	Leu	Glx	Glx	CMCys

* System used to purify peptides: 1.9, paper electrophoresis at pH 1.9; BAW and BAWP, chromatography in solvents BAW and BAWP respectively.

† Electrophoretic mobilities at pH 1.9 relative to lysine.

residue of the protein by the tritium-labelling method.

Peptides from peptic digestion

Purification of the peptic peptides. The elution pattern of the peptic peptides from a DEAE-cellulose column is shown in Fig. 1. The contents of the tubes under each peak were pooled as shown by the solid bars. Peaks 9, 12 and 13 each contained a pure peptide. The peptides contained in the other peaks were separated by paper chromatography and electrophoresis. The system used as well as R_F values or mobilities are given in Tables 1 and 2. These tables also give the amino acid compositions and *N*- and *C*-terminal residues of the purified peptides.

Sequences of the peptic peptides of protein IIIB4

The sequences of the peptic peptides of protein IIIB4 as determined by dansyl–Edman degradation and the tritium-labelling method are given in Table 3. The sequence of peptide 4P1 was confirmed as acetyl-Ala-CMCys-CMCys by mass spectrometry.

Assignment of the peptic peptides. The amino acid composition and sequence of peptides 4P9, 4P10, 4P17 and 4P18 must be unique to protein IIIB4 since they cannot be positioned in proteins IIIB2 or IIIB3 (see Fig. 4).

Peptide 4P9 is homologous with residues 41–48 in protein IIIB3 with leucine-41 replaced by phenylalanine. The partial sequence and amino acid composition of peptide 4P10 correspond to residues 49–66 in protein IIIB3 except that a tyrosyl and a valyl residue have been replaced by carboxymethylcysteine and isoleucine respectively. Peptide 4P17 is homologous with residues 83–91 of protein IIIB3 with serine-88 exchanged for proline. Peptide 4P18 is evidently the *C*-terminal portion of peptide 4P17.

Peptides 3P9, 3P10 and 3P15 (Table 2) from protein IIIB3, which are homologous to the peptides unique to protein IIIB4, were also isolated from the peptic digestion. When the yields of the corresponding peptides from protein IIIB4 and IIIB3 are compared (Tables 1 and 2 respectively), it is evident that these proteins are present in the preparation of protein IIIB4 in a ratio of approx. 3:1.

The amino acid composition and sequence of all the peptides allocated to protein IIIB4 (Table 3) except peptides 4P9, 4P10, 4P17 and 4P18 are identical with those obtained from protein IIIB3 (Haylett *et al.* 1971). The high yields of these peptides justify their assignment to protein IIIB4.

Peptides 2P14, 2P15, 2P16 and 2P17 were assigned to protein IIIB2 from their amino acid composition and terminal residues (Table 2). Their yields showed that the preparation of protein IIIB4

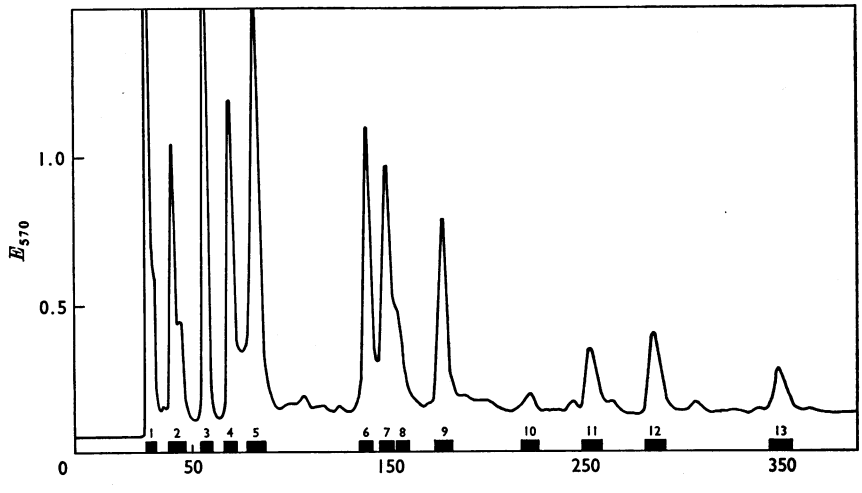

Fig. 1. Elution pattern of peptic peptides of protein IIIB4 by using a Technicon Peptide AutoAnalyzer. Chromatography was performed on a DEAE-cellulose column (150 cm × 1.9 cm) at 30°C with a 10-litre linear gradient of 0.02 M-pyridine to 0.35 M-pyridine–acetic acid buffer (pyridine–acetic acid–water, 7:24:219, by vol.), pH 4.1 (flow rate, 200 ml/h; fraction volume, 20 ml).

Table 2. *Amino acid composition of contaminating peptic peptides in protein IIIB4*

Peptide ...	Amino acid composition (mol of residue/mol)						
	2P14	2P15	2P16	2P17	3P9	3P10	3P15
Trp							
Lys							
His	1.02	1.01				0.86	
Arg						0.92	
CMCys				0.96	2.71	0.74	1.01
Asp	1.01	2.02				2.02	
Thr	1.33	1.11	1.98		1.00	1.01	3.04
Ser	2.69	2.82				0.92	1.84
Glu				2.01	0.99	1.03	2.12
Pro	2.11	2.02		1.10	1.00	5.75	
Gly	1.08	2.00		1.04			
Ala							
Val						2.13	
Ile		0.98		1.02			
Leu	2.76	3.05	1.02		1.80		
Tyr						0.73	0.95
Phe				0.96		0.93	
Yield (μmol)	1.2	1.0	1.4	1.3	2.2	1.2	1.6
Peak no. (from Fig. 1)	1	1	2	7	1	3	8
System*	1.9	1.9	BAWP	BAW	BAWP	BAW	BAW
R_F	—	—	0.45	0.37	0.28	0.08	0.09
m_{Lys}*	0.53	0.48	—	—	—	—	—
N-Terminal	Leu	Leu	Leu	Phe	Leu	Asp	Thr
C-Terminal	Glx	Asx	Thr	Glx	CMCys	Phe	Glx
Residue nos.	67–78	67–81	82–84	85–91	41–48	49–66	83–91

* Abbreviations are defined in Table 1.

Table 3. *Sequence of peptic peptides by the dansyl–Edman procedure*

Peptide no.	Residue no.	Sequence results
4P1	1-3	(Ala,CMCys)CMCys
4P2	4-6	Ala-Arg-Leu
4P3	7-17	CMCys-CMCys-Ser-Val-Pro-Thr-Ser-Pro-Ala-Thr-Thr
4P4	7-19	CMCys(CMCys,Ser,Val,Pro,Thr,Ser,Pro,Ala,Thr,Thr,Ile)CMCys
4P5	18-24	Ile-CMCys-Ser-Ser-Asp-Lys-Phe
4P6	20-24	Ser(Ser,Asx,Lys)Phe
4P7	25-31	CMCys-Arg-CMCys-Gly-Val-CMCys-Leu
4P8	32-40	Pro-Ser-Thr-CMCys-Pro-His(Thr,Val)Trp
4P9	41-48	Phe-Leu-Gln-Pro-Thr-CMCys-CMCys-CMCys
4P10	49-66	Asp-Asn-Arg-Pro(Pro,Pro,CMCys,His,Ile,Pro,Glx,Pro,Ser,Val,Pro,Thr,CMCys)Phe
4P11	67-77	Leu-Leu-Asn-Ser-Ser-Gln-Pro-Thr-Pro-Gly-Leu
4P12	67-78	Leu(Leu,Asx,Ser,Ser,Glx,Pro,Thr,Pro,Gly,Leu)Glx
4P13	67-79	Leu(Leu,Asx,Ser,Ser,Glx,Pro,Thr,Pro,Gly,Leu,Glx)Ser
4P14	78-82	Glu-Ser-Ile-Asn-Leu
4P15	79-82	Ser(Ile,Asx)Leu
4P16	80-82	Ile-Asx-Leu
4P17	83-91	Thr-Thr-Tyr-Thr-Gln-Pro-Ser-CMCys-Glu
4P18	85-91	Tyr(Thr,Glx,Pro,Ser,CMCys)Glx
4P19	93-99	Pro-CMCys-Ile-Pro-Ser-CMCys-CMCys

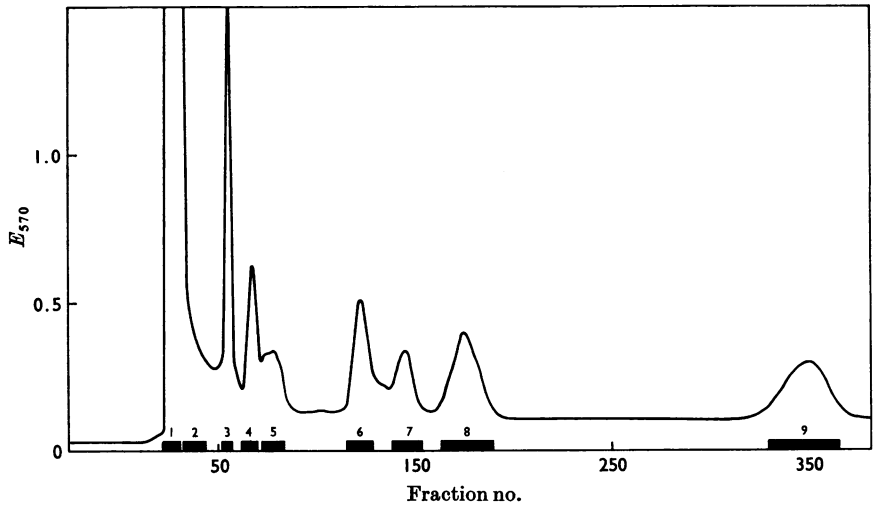

Fig. 2. Elution pattern of thermolysin peptides of protein IIIB4 from DEAE-cellulose obtained with the Technicon Peptide AutoAnalyzer. Conditions are as for Fig. 1.

contained approx. 14% of protein IIIB2 and therefore approx. 22% of protein IIIB3.

Peptides from thermolysin digestion

Purification of the thermolysin peptides. The elution pattern of the thermolysin peptides from a DEAE-cellulose column is shown in Fig. 2. The peptides contained in the enumerated peaks were separated by paper chromatography or electrophoresis as given in Tables 4 and 5. These tables also list the amino acid compositions and *N*- and *C*-terminal residues of the purified peptides.

Sequences of the thermolysin peptides. The sequences of the peptic peptides given in Table 3 covered all the residues of protein IIIB4 except those contained in peptide 4P10. From its amino acid composition and *N*-terminal residue, thermolysin peptide 4L7 provides an overlap between peptides 4P9 and 4P10. Peptide 4L7 was digested with papain and the resulting peptides separated by paper electrophoresis at pH 4.5 and paper chromatography in solvent BAWP. Table 6 contains the mobilities, R_F values, amino acid compositions and sequences of the purified peptides necessary to establish the sequence of peptide 4L7 as shown in Fig. 3.

The sequence of peptic peptide 4P9 (Table 3) provides an overlap between peptides Q1 and Q2, and the *N*-terminal sequence of peptic peptide 4P10 (Table 3) connects peptides Q2 and Q3. Histidine at the *C*-terminus of peptide Q4 and at the *N*-terminus of peptide Q5 couples peptides Q3 and Q5. Peptide Q6 provides the *C*-terminal sequence of peptide Q5. Peptide Q7 completes the sequence of peptide 4L7.

Assignment of the thermolysin peptides. The thermolysin peptides were assigned to proteins IIIB2, IIIB3 and IIIB4 as given in Tables 4 and 5. The amino acid compositions and sequences of peptides 4L7, 4L8 and 4L17 show that they are unique to protein IIIB4. Peptide 4L7 is homologous to residues 41–65 of protein IIIB3 (Fig. 4) with leucine-41, tyrosine-55 and valine-57 replaced by phenylalanine, carboxymethylcysteine and isoleucine respectively. Peptide 4L8 consists of the *C*-terminal sequence of peptide 4L7. Peptide 4L17 is homologous to residues 86–99 of protein IIIB3, with serine-88 replaced by proline.

The yields of the unique peptides from each of the three proteins (Tables 4 and 5) agree with the percentage contamination of proteins IIIB2 and IIIB3 in the preparation of protein IIIB4 as found from the peptic peptides. Thus the assignment of the thermolysin peptides ascribed to protein IIIB4 is supported by their high yields.

The similarity of the peptides of protein IIIB4 to those of protein IIIB3 allows them to be aligned in the same manner. This completes the sequence of protein IIIB4, as given in Fig. 4.

Amino acid composition of protein IIIB4

The amino acid composition of protein IIIB4 is given in Table 7. Comparison with the calculated composition shows that the experimental values are low for carboxymethylcysteine, serine and proline and high for arginine, glycine and valine.

Table 4. *Amino acid composition of the thermolysin peptides of protein III B4*

	Amino acid composition (mol of residue/mol)																
Peptide ...	4L1	4L2	4L3	4L4	4L5	4L6 +	4L7	4L8	4L9	4L10	4L11	4L12	4L13	4L14	4L15	4L16	4L17
Trp																	
Lys			1.06	0.97													
His				0.95	0.77		0.90	1.00									
Arg	0.98			1.08	1.00		1.10	1.00									
CMCys	1.76	1.71	0.75	4.66	3.43		4.76	4.85									3.74
Asp			1.02	1.03			2.15	2.11		1.08	1.07	0.99		1.00			
Thr		2.77		1.85	1.87		2.03	1.96		0.97	0.92	1.00			1.85	2.01	1.01
Ser		1.89	2.04	2.79	1.02		1.00	1.14		1.81	1.78	2.92	0.76				1.77
Glu							2.12	2.06		1.16	1.18	1.97	1.11				2.13
Pro		2.01		1.95	1.93		6.98	7.36		2.01	2.18	2.03					2.99
Gly				0.93	0.90					1.01	1.07	1.03					
Ala	2.02	1.00															
Val		1.02		0.96	1.11	1.00	1.17	1.07									
Ile			0.91	0.82			1.00	1.08						1.00			1.00
Leu		0.92		0.82	1.10		1.02	1.00	1.00	1.88	0.98	1.77	0.90		1.00	0.98	
Tyr																0.58	
Phe				0.81	0.85		0.95		0.95								
Yield (μmol)	8.76	13.30	6.03	4.15	3.02	8.80	5.51	2.48	11.12	0.92	3.06	2.71	5.81	11.65	6.86	3.34	4.43
Peak no. (from Fig. 2)	8	6	5	7	6	3	8	8	3	2	2	4	5	3	3	3	9
System*	BAW	BAWP	BAW	1.9	BAWP	BAW	BAW	BAW	BAW	BAW	BAW	BAW	BAW	BAW	BAW	BAW	BAWP
R_F	0.28	0.36	0.11	—	0.27	0.71	0.22	0.12	0.83	0.31	0.19	0.25	0.41	0.37	0.48	0.57	0.24
m_{Lys}*	—	—	—	0.76	—	—	—	—	—	—	—	—	—	—	—	—	—
N-Terminal	—	Leu	Ile	Ile	Phe	Val	Phe	Leu	Phe	Leu	Leu	Leu	Leu	Ile	Leu	Leu	Thr
C-Terminal	Arg	Thr	—	Thr	Thr	Trp	CMCys	CMCys	Leu	Gly	Gly	Ser	Ser	Asx	Thr	Tyr	CMCys
Residue nos.	1–5	6–17	18–23	18–38	24–38	39–40	41–65	42–65	66–67	67–76	68–76	68–79	77–79	80–81	82–84	82–85	86–99

* Abbreviations are defined in Table 1.

Table 5. *Amino acid composition of contaminating thermolysin peptides in protein IIIB4*

	Amino acid composition (mol of residue/mol)						
Peptide ...	2L2	2L8	2L10	2L11	2L14	3L7	3L18
Trp							
Lys							
His			0.99			0.98	
Arg	1.03					1.04	
CMCys		3.96			0.79	3.76	3.55
Asp		2.80	1.02		0.93	1.92	
Thr	2.83	3.04	0.98			2.10	1.01
Ser		0.88	2.06	1.06		1.10	2.69
Glu		1.00			2.02	2.02	2.10
Pro	0.97	4.65	1.86		1.00	7.14	2.02
Gly	0.73		0.94	1.01	0.98		
Ala	1.10						
Val		2.95				1.99	
Ile	0.83				0.87		0.89
Leu		0.89	0.78	0.93		2.00	
Tyr		1.51				0.93	
Phe							
Yield (μmol)	0.69	0.46	1.24	1.89	0.62	1.14	1.98
Peak no. (from Fig. 2)	1	9	2	3	7	7	9
System*	BAW	BAWP	BAW	BAW	1.9	1.9	BAWP
R_F	0.26	0.52	0.11	0.41	—	—	0.17
m_{Lys}*	—	—	—	—	0.48	0.53	—
N-Terminal	Val	Leu	Leu	Leu	Ile	Leu	Thr
C-Terminal	Thr	CMCys	Gly	Gly	Asx	CMCys	CMCys
Residue nos.	10–17	42–65	68–76	77–79	86–92	41–65	86–99

* Abbreviations are defined in Table 1.

Table 6. *Amino acid sequence of papain peptides of peptide 4L7*

Peptide	m*	R_F†	Amino acid composition	Sequence results‡
Q1	0.40	0.54	CMCys 0.81, Thr 0.94, Glu 1.04, Pro 0.94, Leu 1.02, Phe 0.96	Phe-Leu(Glx,Pro,Thr)CMCys
Q2	0.86	0.15	CMCys 1.56, Asp 2.00	CMCys-CMCys-Asx-Asx
Q3	0	0.19	Arg 1.04, CMCys 0.87, Pro 2.96	Arg-Pro-Pro-Pro-CMCys
Q4	-0.32	0.14	His 0.94, Arg 1.03, CMCys 0.72, Pro 3.02	Arg(Pro,Pro,Pro,CMCys)His
Q5	-0.14	0.29	His 0.81, Ser 0.82, Glu 1.12, Pro 2.07, Ile 1.00	His-Ile(Pro,Glx,Pro)Ser
Q6	0	0.31	Ser 0.94, Glu 1.10, Pro 1.92, Ile 0.99	Ile-Pro-Gln-Pro-Ser
Q7	0.40	0.35	CMCys 0.69, Thr 0.91, Pro 1.02, Val 1.06	Val-Pro-Thr-CMCys

* Electrophoretic mobilities at pH4.5 relative to lysine (negative) and aspartic acid (positive).
† Chromatography in BAWP.
‡ *N*-Terminal sequences by dansyl–Edman degradation and *C*-terminal residues by the tritium-labelling method.

The discrepancies can be explained from the calculated amino acid compositions of proteins IIIB2 and IIIB3 (Table 7), taking into account the contamination of these proteins in protein IIIB4. The molecular weights of the proteins calculated on the basis of the carboxymethylated and the native form are given in Table 7. These are in good agreement with the molecular weight of 10800 determined for the Sephadex fraction IIIB (Swart *et al.* 1969).

DISCUSSION

The *C*-terminal tritium-labelling method was extensively used in this study and proved to be a valuable tool. In most cases the *C*-terminal residue was well defined by its radioactivity. The exception was lysine-23 preceded by aspartic acid at the *C*-terminus of peptide 4L3. This lysyl residue could be labelled, however, if the ϵ-amino group is first

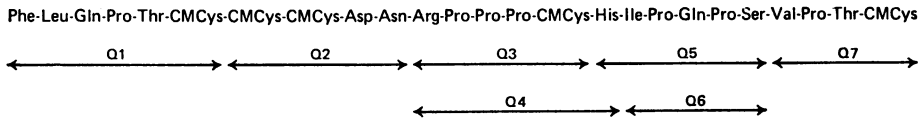

Fig. 3. Amino acid sequence of peptide 4L7 from papain peptides.

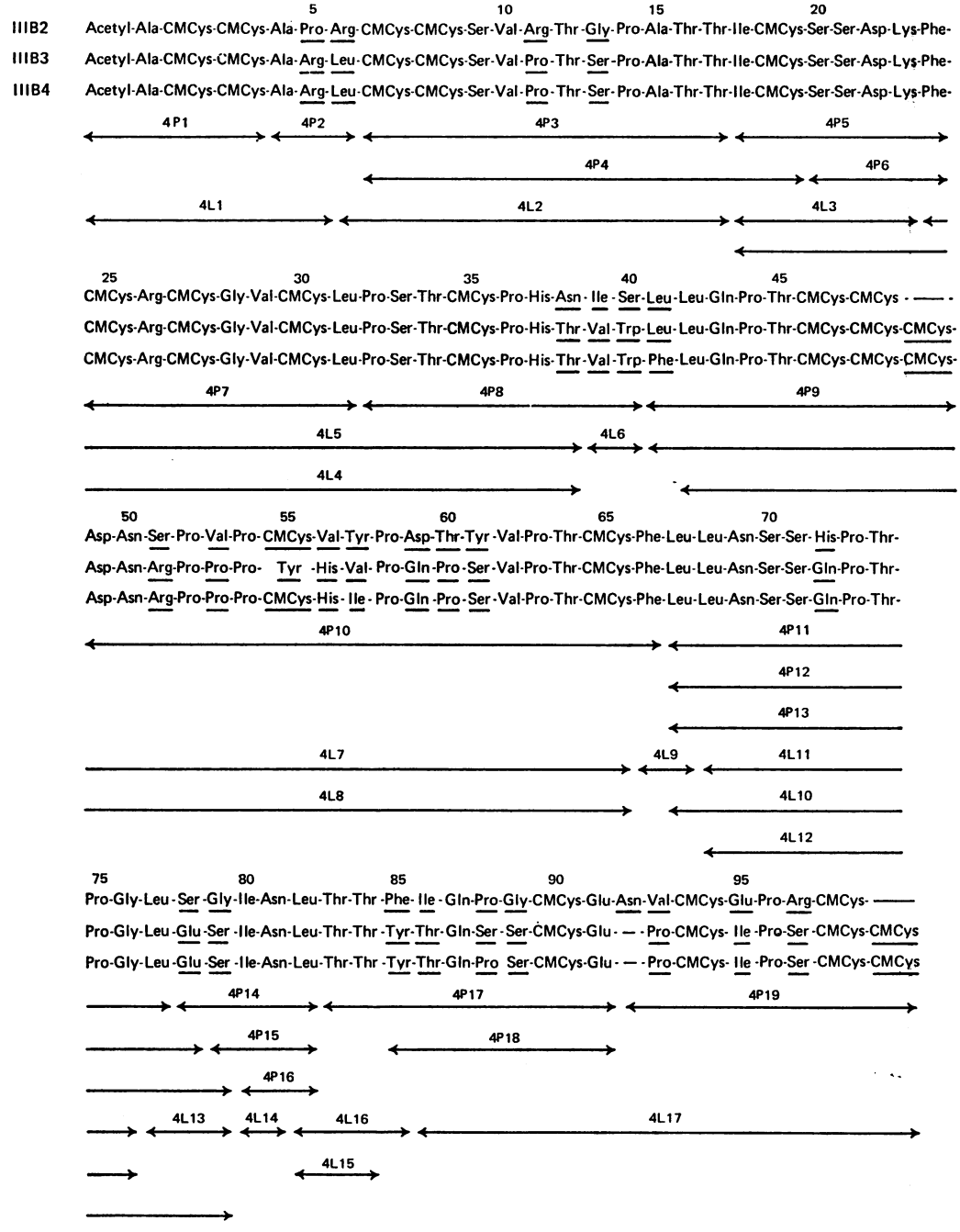

Fig. 4. Complete amino acid sequence of protein IIIB4 together with the sequences of protein IIIB2 (Haylett & Swart, 1969) and IIIB3 (Haylett *et al.* 1971). Peptic (P) and thermolysin (L) peptides of protein IIIB4 are indicated. Differences between proteins are shown underlined.

Table 7. *Amino acid composition of protein IIIB4, IIIB3 and IIIB2*

	Amino acid composition (mol of residue/mol)			
Protein ...	IIIB4 Analysis values	IIIB4 Sequence	IIIB3 Sequence	IIIB2 Sequence
Trp	0.83	1	1	—
Lys	0.97	1	1	1
His	1.89	2	2	2
Arg	3.42	3	3	4
CMCys	17.42	18	17	16
Asp	5.21	2	2	3
Asn		3	3	5
Thr	11.15	11	11	10
Ser	10.32	11	12	9
Glu	6.16	2	2	2
Gln		4	4	2
Pro	15.36	16	15	13
Gly	2.45	2	2	5
Ala	2.94	3	3	3
Val	4.46	4	5	6
Ile	3.61	4	3	4
Leu	6.92	7	8	7
Tyr	1.15	1	2	2
Phe	2.74	3	2	3
Total		98	98	97
Molecular weight				
Carboxymethylated		11559	11503	11260
Native form		10496	10499	10315

blocked with isopropenyl acetate (D. Parris, personal communication).

The preparation of protein IIIB4 contained appreciable amounts of protein IIIB2 and IIIB3. This resulted from asymmetrical elution of the proteins from DEAE-cellulose (Swart *et al.* 1969). Because the sequences of the contaminating proteins were known, the deduction of the complete amino acid sequence of protein IIIB4 was possible.

Protein IIIB4 is closely related to protein IIIB3, differing in four positions. According to the genetic code these differences constitute single nucleotide base changes. Compared with protein IIIB2 the sequence of protein IIIB4 contains 27 differences. Excluding three deletions, these differences can be attributed to 17 single and seven double base changes. The three proteins have a unique difference between them at position 57.

In the free-thiol form, the proteins have the same total charge. As *S*-carboxymethylated proteins they increase by one carboxymethylcysteinyl residue from protein IIIB2 to IIIB3 to IIIB4. These charge differences mainly account for their separation on DEAE-cellulose.

A total of 70 of the 99 residues and 15 of the 18 half-cystine positions are identical in all three proteins. This strongly suggests that the proteins have similar tertiary structures. Some of the disulphide bonds must be interchain and thereby contribute to the quaternary structure of wool. The deletions of half-cystine-48 and half-cystine-99 in protein IIIB2 and the exchange of half-cystine-55 in protein IIIB3 show that half-cystine residues in these positions are probably involved in interchain bridges.

Protein IIIB1 is a minor component of the Sephadex fraction IIIB (Swart *et al.* 1969) and could not be isolated in sufficient quantity for sequence work.

The primary structures of the three main components of the Sephadex fraction SCMKB-IIIB of the high-sulphur proteins of merino wool have thus been established. It is appropriate to summarize the properties of these components. The *S*-carboxymethylated proteins have a molecular weight of 11409 ± 150. They consist of single chains containing 97 or 98 residues and are homologous. Their *N*-terminal sequence is acetyl-Ala-CMCys-CMCys and they have carboxymethylcysteine as *C*-terminus. These terminal properties are shared with other high-sulphur proteins (Gillespie, Haylett & Lindley, 1968) but there is no other correspondence in sequence with any known wool proteins. The molecules can arbitrarily be divided at position 49 into high- and low-sulphur portions but the amino acids that prevent helix formation, such as

proline, serine and threonine, are fairly evenly distributed throughout the molecules. These high-sulphur proteins therefore appear to have no helical content and are ideally suited for the function of matrix proteins in wool.

We thank Mr S. J. van der Walt for the mass-spectrometry experiment and Miss P. G. Das for technical assistance.

REFERENCES

Gillespie, J. M., Haylett, T. & Lindley, H. (1968). *Biochem. J.* **110**, 193.

Gray, W. R. (1967). In *Methods in Enzymology*, vol. 11, p. 469. Ed. by Colowick, S. P. & Kaplan, N. O. New York: Academic Press Inc.

Haylett, T. & Swart, L. S. (1969). *Text. Res. J.* **39**, 917.

Haylett, T., Swart, L. S. & Parris, D. (1971). *Biochem. J.* **123**, 191.

Kiryushkin, A. A., Gorlenko, V. A., Rosinov, B. V., Ovchinnikov, Yu. A. & Shemyakin, M. M. (1969). *Experientia*, **25**, 913.

Lenard, J. & Gallop, P. M. (1970). *Analyt. Biochem.* **34**, 286.

Matsuo, H., Fujimoto, Y. & Tatsuno, T. (1966). *Biochem. biophys. Res. Commun.* **22**, 69.

Swart, L. S., Haylett, T. & Joubert, F. J. (1969). *Text. Res. J.* **39**, 912.