Studies on the High-Sulphur Proteins of Reduced Merino Wool

AMINO ACID SEQUENCE OF PROTEIN SCMKB-IIIB3

By T. HAYLETT, L. S. SWART AND D. PARRIS

*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 wool protein SCMKB-IIIB3 was determined. The peptides used for the sequence work were obtained by peptic and thermolysin digestions and were fractionated by chromatography on DEAE-cellulose, paper chromatography and electrophoresis. The peptides were analysed by dansyl-Edman degradation, mass spectrometry and tritium-labelling of *C*-terminal residues. The protein consists of 98 residues and has acetylalanine as *N*-terminal residue and carboxymethylcysteine as *C*-terminus. It is homologous with protein SCMKB-IIIB2 (Haylett & Swart, 1969). A salient feature of the sequence of protein SCMKB-IIIB3 is three consecutive cysteine residues.

The complete amino acid sequence of a high-sulphur protein, SCMKB-IIIB2, of Merino wool has been described by Haylett & Swart (1969). This protein contains 97 residues, has an acetylated *N*-terminal alanyl residue and *S*-carboxymethylcysteine as *C*-terminal amino acid. A notable feature of the molecule is the concentration of most of the sulphur-containing residues in the first half of the sequence.

Protein SCMKB-IIIB2 is one of a group of three proteins with molecular weight 12000 and very similar amino acid compositions (Swart, Haylett & Joubert, 1969). The present paper describes the determination of the amino acid sequence of the second protein, SCMKB-IIIB3, from this group.

MATERIALS AND METHODS

Protein IIIB3* was prepared as described by Swart *et al.* (1969).

Enzymes. Pepsin and papain were twice-crystallized preparations from Worthington Biochemical Corp., Freehold 2, N.J., U.S.A.; thermolysin was from Daika Kasei K.K., Osaka, Japan.

Enzymic digestions. Peptic digestion was carried out on 10 μ mol of protein IIIB3 in 10 ml of 5% (v/v) formic acid (enzyme/substrate ratio 1:100, w/w) for 2 h at 40°C. The digestion was stopped by addition of pyridine and the product was freeze-dried.

Thermolysin digestion was performed on 10 μ mol of protein IIIB3 in 12 ml of 2% (w/v) NH_4HCO_3 (enzyme/

substrate ratio 1:100, w/w) for 1 h at 40°C. The digestion was acidified to pH 3 with acetic acid and the product was freeze-dried.

Papain digestion was carried out on 2 μ mol of peptide 3L7 in 2 ml of buffer (0.2M-pyridine adjusted to pH 5.5 with acetic acid); 20 μ l of papain suspension (13.6 mg/ml) and 4 μ l of β -mercaptoethanol were added. The mixture was incubated under N_2 for 16 h at 40°C and then freeze-dried.

Fractionation of peptides. The initial fractionations of the peptic and thermolysin peptides were carried out by chromatography on DEAE-cellulose as described by Haylett & Swart (1969). The number of peptides in each fraction was assessed by descending paper chromatography in solvent BAWP and solvent BAW and by paper electrophoresis at pH 4.5 and 1.9. Details of electrophoretic techniques and buffers are given by Haylett & Swart (1969). The peptides were purified or separated in the most suitable system. Peptides were located on paper by the chlorine-*o*-tolidine method of Reindel & Hoppe (1954). Electrophoretic mobilities of the peptides at pH 1.9 were measured from the origin relative to lysine and at pH 4.5 from the neutral amino acids relative to lysine (negative) and to aspartic acid (positive).

Amino acid analysis. Amino acid compositions were determined on Beckman 120B and C automatic amino acid analysers coupled to Beckman 125 integrators. Samples of proteins and peptides (0.1–0.3 μ mol) were hydrolysed with 0.25 ml of constant-boiling HCl in the presence of 50 μ g of phenol (Sanger & Thompson, 1963) at 110°C for 24 h in sealed evacuated (about 50 μ m) tubes. The destruction of serine and threonine during hydrolysis was compensated for by correction factors of 1.075 and 1.042 respectively. Tryptophan was determined quantitatively on the amino acid analyser after complete enzymic hydrolysis of the proteins (Hill & Schmidt, 1962) and qualitatively detected in peptides with the Ehrlich reagent (Easley, 1965).

C-Terminal analysis. The *C*-terminal residue of protein

* Abbreviations: IIIB3, SCMKB-IIIB3; IIIB2, SCMKB-IIIB2; 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.

IIIB3 and the papain peptides were determined by the tritium-labelling method of Matsuo, Fujimoto & Tatsuno (1966). The protein or peptide to be analysed for amino acid composition was dissolved in 0.4 ml of dry pyridine and 0.1 ml of $^3\text{H}_2\text{O}$ (200 mCi/ml) in a hydrolysis tube. The solution was cooled, 100 μl of acetic anhydride carefully added without mixing and the mixture incubated at 2°C for 16 h. The solution was then dried *in vacuo*. The tritium was removed from exchangeable sites in the molecule by addition of 250 μl of water, incubation at 40°C for 5 min and drying *in vacuo*. This exchange process was repeated at least six times. The sample was hydrolysed and an amino acid analysis performed. The effluent of the analyser was recovered as 2.5 ml fractions in a fraction collector. A 1 ml sample was taken at the peak of each amino acid and dissolved in 15 ml of a mixture of scintillant solution, Triton X-100 and water (7:6:2, by vol.). The scintillant solution consisted of 0.4% 2,5-diphenyloxazole and 0.01% 1,4-bis-(5-phenyloxazol-2-yl)benzene dissolved in toluene. The radioactivities of the amino acid solutions were counted for 1 min in a Packard Tri-Carb model 574 scintillation spectrometer.

N-Terminal analyses and sequence determinations of peptides. These were carried out by the dansyl-Edman method of Gray (1967). After coupling with DNS-chloride the peptides were hydrolysed with 0.025 ml of constant-boiling HCl at 110°C for 4 h in sealed evacuated tubes. DNS-amino acids were identified by paper electrophoresis at pH 4.38 (pyridine-acetic acid-water, 1:2:250, by vol.) on a Shandon flat-plate apparatus. A voltage gradient of 46 V/cm for 3 h was used. DNS-alanine, DNS-glycine, DNS-proline and DNS-serine were identified by t.l.c. on precoated silica-gel plates without fluorescent indicator (Merck) with the benzene-pyridine-acetic acid (40:10:1, by vol.) system (Morse & Horecker, 1966).

The Edman degradation procedure was based on the

methods of Doolittle (1965) and Edman & Begg (1967) as described by Haylett & Swart (1969). The peptide (0.5–2.0 μmol) was dissolved in 0.5 ml of pH 9.6 buffer (*N*-ethylmorpholine-pyridine-water, 1:12:8, by vol., adjusted to pH 9.6 with trifluoroacetic acid). A 10 μl sample was removed for examination by the dansyl procedure. The peptide was made to react with 50 μl of phenyl isothiocyanate at 50°C for 30 min under N_2 . After

Table 1. *Amino acid composition of protein BIII3*

	Amino acid composition (mol of residue/mol)		
	Analysis values	Integral values	Sequence values
Trp	0.75	1	1
Lys	1.12	1	1
His	2.09	2	2
Arg	3.79	4	3
CMCys	16.44	16	17
Asp	5.70	6	5
Thr	11.12	11	11
Ser	11.63	12	12
Glu	5.93	6	6
Pro	14.53	15	15
Gly	2.94	3	2
Ala	3.11	3	3
Val	5.36	5	5
Ile	3.46	3	3
Leu	8.19	8	8
Tyr	1.82	2	2
Phe	2.26	2	2
Total		100	98

Table 2. *Amino acid analysis of*

Peptide ...	Amino acid composition (mol of residue/mol)									
	3P1	3P2	3P3	3P4	3P5	3P6	3P7	3P8	3P9	3P10
Trp								+		
Lys					1.02	0.99				
His								0.93		0.98
Arg		1.01					0.98			1.04
CMCys	1.91		1.85	2.79	0.80		2.77	0.82	2.69	0.80
Asp					1.06	1.01				1.73
Thr			2.90	2.83				1.98	0.97	1.07
Ser			2.09	2.01	2.13	2.02		1.00		0.95
Glu									0.88	1.15
Pro			2.06	1.97				2.03	1.15	5.98
Gly							1.07			
Ala	1.00	0.92	0.95	1.03						
Val			0.91	0.99			0.86	1.00		1.95
Ile				0.96	0.85					
Leu		1.06					0.95		1.78	
Tyr										0.89
Phe					0.98	0.96				1.17
Yield (μmol)	6.5	6.5	6.3	2.3	5.8	1.4	6.9	2.2	4.1	2.5
R_F^*			0.25		0.38	0.31		0.51		0.36
$m_{\text{Lys}}^\dagger$										

* Chromatography in solvent BAWP.

† Electrophoretic mobilities at pH 1.9 relative to lysine.

extracting the mixture four times with benzene the aqueous phase was freeze-dried. The phenylthiocarbamoyl derivative of the peptide was cyclized in 50μl of trifluoroacetic acid under N₂ for 3 min at 50°C. The resulting thiazolinone was extracted with ethylene chloride and converted into the phenylthiohydantoin with 25μl of m-HCl at 80°C for 10 min under N₂.

The phenylthiohydantoin derivatives of the amino

acids were only used to identify the amides of aspartic and glutamic acid. This was done by t.l.c. on precoated silica-gel plates with fluorescent indicator (Merck) with chloroform-methanol (9:1, v/v) as developer (Brenner, Niederwieser & Pataki, 1961).

Mass spectrometry of the N-terminal peptide. The peptide was permethylated by the method of Lenard & Gallop (1970) and desulphurized by the method of Kiryushkin,

Fig. 1. Elution pattern of the peptic peptides of protein IIIB3 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 0.02M-pyridine to 0.35M-pyridine-acetic acid buffer (pyridine-acetic acid-water, 7:24:219, by vol.) at pH4.1 (flow rate, 200ml/h; fraction volume 20 ml).

peptic peptides of protein IIIB3

		Amino acid composition (mol of residue/mol)									
Peptide	...	3P11	3P12	3P13	3P14	3P15	3P16	3P17	2P8	2P15	2P18
Trp											
Lys											
His									0.85	0.80	
Arg											0.93
CMCys						0.80	3.47	2.75	0.83		1.83
Asp		1.06	1.00	1.00	1.17				0.99	1.75	0.99
Thr		0.96	1.00			2.75	2.76		0.90	0.90	
Ser		1.81	2.81	1.07		1.82	3.01	1.00	2.08	2.94	
Glu		1.03	2.07	1.00		2.00	2.02				1.05
Pro		2.12	2.09				1.97	2.10	1.94	1.96	1.03
Gly		0.99	1.05							2.10	
Ala											
Val											0.86
Ile				0.94	0.98		0.80	0.87	1.01	0.91	
Leu		3.03	3.00	1.00	1.02				1.08	2.73	
Tyr						0.76	0.86				
Phe											
Yield (μmol)		1.9	2.1	2.7	1.0	5.0	1.2	8.2	1.1	0.8	2.7
R _F *		0.66		0.59	0.66						0.15
m _{Lys} †		0.24			0.41						

Gorlenko, Rosinov, Ovchinnikov & Shemyakin (1969). Mass spectrometry was performed at 105°C with an AEI MS9 mass spectrometer.

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

RESULTS

Analysis of protein IIIB3

The amino acid composition of protein IIIB3 is given in Table 1. The nearest integral values are in reasonable agreement with those calculated from the complete sequence. The discrepancies in content of arginine, carboxymethylcysteine, aspartic acid and glycine can be ascribed to the presence of protein IIIB2 in the preparation of protein IIIB3 as found during the analysis of the peptic and thermolytic peptides. No *N*-terminal residue could be found for protein IIIB3 by the dansyl-Edman method, whereas the tritium-labelling method

showed that carboxymethylcysteine was the *C*-terminal residue of the protein.

Peptides from peptic digestion

Purification of 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. The peptides contained in the peaks are enumerated on the graph.

Peptides 3P2, 3P4, 3P9 and 3P17 were pure enough for sequence studies. Peptides 3P1 and 2P8 were purified by paper electrophoresis at pH 1.9 and peptide 2P15 was purified by electrophoresis at pH 4.5. Peptides 3P7, 3P10, 3P12, 3P15 and 3P16 were purified by paper chromatography in solvent BAWP. The impurities resulting from overlapping peptides were discarded. The peptides contained in peaks 1, 2 and 3 (see Fig. 1) were separated by paper chromatography in solvent BAWP. The R_F values of these peptides are given in Table 2. Peptides 3P6 and 3P8 were purified by paper electrophoresis at pH 1.9. Peptides 3P11 and 3P14 were separated by the same system and their

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

Peptide no.	Residue no.	Sequence results
3P1	1-3	(Ala, CMCys, CMCys)
3P2	4-6	<u>Ala-Arg-Leu</u>
3P3	7-17	<u>CMCys-CMCys-Ser-Val-Pro-Thr-Ser-Pro-Ala-Thr-Thr</u>
3P4	7-19	<u>CMCys-CMCys-Ser</u> (Val, Pro, Thr, Ser, Pro, Ala, Thr, Thr, Ile, CMCys)
3P5	18-24	<u>Ile-CMCys-Ser-Ser-Asp</u> (Lys, Phe)
3P6	20-24	<u>Ser</u> (Ser, Asx, Lys, Phe)
3P7	25-31	<u>CMCys-Arg-CMCys-Gly-Val-CMCys-Leu</u>
3P8	32-40	<u>Pro-Ser-Thr-CMCys-Pro-His-Thr-Val-Trp</u>
3P9	41-48	<u>Leu-Leu-Gln-Pro-Thr-CMCys-CMCys-CMCys</u>
3P10	49-66	<u>Asp</u> (Asx, Arg, Pro, Pro, Pro, Tyr, His, Val, Pro, Glx, Pro, Ser, Val, Pro, Thr, CMCys, Phe)
3P11	67-77	<u>Leu-Leu-Asn-Ser-Ser-Gln-Pro-Thr-Pro-Gly-Leu</u>
3P12	67-79	<u>Leu</u> (Leu, Asx, Ser, Ser, Glx, Pro, Thr, Pro, Gly, Leu, Glx, Ser)
3P13	78-82	<u>Glu-Ser-Ile-Asn-Leu</u>
3P14	80-82	<u>Ile</u> (Asx, Leu)
3P15	83-91	<u>Thr</u> (Thr, Tyr, Thr, Glx, Ser, Ser, CMCys, Glx)
3P16	83-99	<u>Thr-Thr-Tyr-Thr-Gln-Ser-Ser-CMCys-Glu</u> (Pro, CMCys, Ile, Pro, Ser, CMCys, CMCys)
3P17	93-99	<u>Pro-CMCys-Ile-Pro-Ser-CMCys-CMCys</u>
2P8	32-41	<u>Pro-Ser-Thr-CMCys-Pro</u> (His, Asx, Ile, Ser, Leu)
2P15	67-81	<u>Leu-Leu-Asn-Ser</u> (Ser, His, Pro, Thr, Pro, Gly, Leu, Ser, Gly, Ile, Asx)
2P18	92-98	<u>Asn-Val-CMCys-Glu-Pro-Arg-CMCys</u>

mobilities relative to free lysine are given in Table 2. The amino acid compositions of the purified peptides are given in Table 2.

Sequence of peptic peptides. The sequences of the peptic peptides, as determined by the dansyl-Edman procedure, are given in Table 3.

No *N*-terminal residue could be detected for peptide 3P1 by the dansyl-Edman method. It was assumed that the *N*-terminal group was acetylated, and the peptide was examined by mass spectrometry. It was permethylated and the carboxymethylcysteinyl residues were converted into [β - 3 H]alanine. The mass spectrum (Fig. 2) ob-

tained for the peptide unambiguously proved the sequence of the peptide to be acetyl-Ala-CMCys-CMCys. This peptide must be the *N*-terminus of protein IIIB3.

Peptide 3P4, which has the composition of peptide 3P3 plus isoleucine and carboxymethylcysteine provides an overlap between peptides 3P3 and 3P5. Similarly, peptide 3P12, which is peptide 3P11 plus the residues glutamic acid and serine, provides the overlap between peptides 3P11 and 3P13. Peptide 3P16 is the overlapping peptide between peptides 3P15 and 3P17.

When peptide 2P8 is compared with peptide 3P8 and peptide 2P15 with peptide 3P11, it is obvious that these peptides are homologous. Fig. 5 shows that peptides 2P8 and 2P15 are derived from protein SCMKB-IIIB2, i.e. consist of residues 32-41 and 67-81 respectively. Peptide 2P18 corresponds to the *C*-terminus of protein IIIB2. From the yields of the peptides (Table 2) there is approx. 25% of protein IIIB2 included with 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. 3. The contents of the tubes under each peak were pooled as shown by the solid bars. The peptides contained in the peaks are enumerated on the Figure.

All the peaks except those numbered 1-6 in Fig. 3 contained only one major peptide. These major peptides were purified by using the systems

Fig. 2. Mass spectrum obtained for permethylated and desulphurized peptide 3P1 at 105°C. Abbreviations: Ac, acetyl; Me, methyl.

Fig. 3. Elution pattern of the thermolysin peptides of protein IIIB3 from DEAE-cellulose by using a Technicon Peptide AutoAnalyzer. Conditions are as for Fig. 1.

Table 4. *Amino acid analysis of thermolysin peptides of protein IIIB3*

Peptide ...	Amino acid composition (mol of residue/mol)																	
	3L1	3L2	3L3	3L4	3L5	3L6	3L7	3L8	3L9	3L10	3L11	3L12	3L13	3L14	3L15	3L16	3L17	3L18
Trp			1.02	1.00														
Lys				0.87	0.82		0.91	0.87										
His				1.00	0.95		1.18	1.02										
Arg	0.92	1.63	0.67	4.61	3.42		3.61	3.50									3.81	3.56
CMCys	1.87		0.98	0.98			1.89	1.80		0.98	1.02	0.81		1.00	1.99	2.02	1.09	0.93
Asp		3.06	1.76	1.95	1.04		2.19	1.97		1.00	1.03	1.21	0.93				2.76	2.67
Thr		1.91	2.73	1.04			1.19	1.04		1.02	1.89	2.61	1.08				2.11	2.00
Ser							2.04	2.19		1.02	1.10	1.72					2.12	1.93
Glu							7.34	7.26		1.95	2.17	2.12					2.12	
Pro		2.14		1.89	2.06					1.07	1.07	1.13						
Gly				0.88	1.12													
Ala	2.00	1.09		0.92	0.90	1.00	1.99	1.88										
Val		1.01	0.86	0.92			2.01	1.03	1.00	1.79	0.92	2.00	1.00	0.92	1.00	0.97	1.04	0.91
Ile		0.93		0.99	1.04		1.00	0.87										
Leu																		
Tyr				0.92	0.91			0.93										
Phe			2.0	1.8	1.3	2.9	2.5	1.5	13.4	1.8	3.3	3.1	5.8	8.0	4.3	2.5	4.9	2.4
Yield (μmol)	8.4	4.6	1.9	1.9	BAWP	BAWP	1.9	1.9	BAWP	BAW	BAW	BAW	1.9	BAW	BAWP	BAWP	1.9	BAW
System*	BAWP	1.9			BAWP	BAWP			BAWP	BAW	BAW	BAW		0.38	0.57	0.72		0.19
R_F						0.79		0.91	0.91	0.29	0.15							
$m_{Lys}^\dagger$							0.34	0.86										

* System used to purify peptide: 1.9, electrophoresis at pH1.9; BAWP and BAW, chromatography in solvent BAWP and BAW respectively.

† Electrophoretic mobility relative to lysine.

Table 5. *Amino acid analysis of thermolysin peptides of protein IIIB2*

Peptide ...	Amino acid composition (mol of residue/mol)															
	2L1	2L2	2L4	2L5	2L7	2L8	2L10	2L11	2L14	2L15	2L16					
Lys			1.06													
His			0.90	0.96			0.89									
Arg	1.00	0.85	1.11	1.00						1.00	1.00					
CMCys	3.76		4.65	3.51		3.73			1.01	2.81	1.55					
Asp			1.88	1.17		2.89	1.13		1.25	0.90						
Thr		3.07	0.90	1.10		3.15	1.00									
Ser	1.01		2.57	1.06	0.89	0.93	1.89	0.98								
Glu						1.00										
Pro	1.11	1.09	2.04	1.78		5.18	1.84		2.04	2.85	1.10					
Gly		1.03	0.88	1.02			1.13	1.13	1.19	2.12	1.01					
Ala	2.00	1.00							0.95	0.85						
Val		0.90	0.83	0.85		2.62										
Ile			0.79		0.93				0.89	0.89	0.77					
Leu			0.93	1.07	1.00	0.85	0.81	1.00								
Tyr						1.66										
Phe			0.89	0.85												
Yield (μ mol)	5.6	2.4	1.7	0.4	2.6	2.5	2.2	3.1	1.7	3.0	1.4					
System*	—	1.9	BAWP	1.9	BAWP	BAW	BAWP	BAWP	1.9	—	1.9					
R_F					0.75	0.26		0.48								
m_{Lys}									0.30							

* Abbreviations are as defined in Table 4.

Table 6. *Sequence of thermolysin peptides by the dansyl-Edman procedure*

Peptide no.	Residue no.	Sequence results
3L1	1-5	(Ala, CMCys, CMCys, Ala, Arg)
3L2	6-17	Leu-CMCys-CMCys-Ser(Val, Pro, Thr, Ser, Pro, Ala, Thr, Thr)
3L3	18-23	Ile-CMCys-Ser(Ser, Asx, Lys)
3L4	18-38	Ile-CMCys(Ser, Ser, Asx, Lys, Phe, CMCys, Arg, CMCys, Gly, Val, CMCys, Leu, Pro, Ser, Thr, CMCys, Pro, His, Thr)
3L5	24-38	Phe-CMCys-Arg(CMCys, Gly, Val, CMCys, Leu, Pro, Ser, Thr, CMCys, Pro, His, Thr)
3L6	39-40	Val-Trp
3L7	41-65	Leu-Leu-Gln-Pro-Thr(CMCys, CMCys, CMCys, Asx, Asx, Arg, Pro, Pro, Pro, Tyr, His, Val, Pro, Glx, Pro, Ser, Val, Pro, Thr, CMCys)
3L8	42-65	Leu-Gln(Pro, Thr, CMCys, CMCys, CMCys, Asx, Asx, Arg, Pro, Pro, Pro, Tyr, His, Val, Pro, Glx, Pro, Ser, Val, Pro, Thr, CMCys)
3L9	66-67	Phe-Leu
3L10	67-76	Leu-Leu-Asn-Ser(Ser, Glx, Pro, Thr, Pro, Gly)
3L11	68-76	Leu-Asn-Ser(Ser, Glx, Pro, Thr, Pro, Gly)
3L12	68-79	Leu-Asn-Ser(Ser, Glx, Pro, Thr, Pro, Gly, Leu, Glx, Ser)
3L13	77-79	Leu-Glu-Ser
3L14	80-81	Ile-Asn
3L15	82-84	Leu-Thr-Thr
3L16	82-85	Leu-Thr-Thr-Tyr
3L17	85-99	Tyr-Thr-Gln-Ser-Ser-CMCys-Glu-Pro(CMCys, Ile, Pro, Ser, CMCys, CMCys)
3L18	86-99	Thr-Gln(Ser, Ser, CMCys, Glx, Pro, CMCys, Ile, Pro, Ser, CMCys, CMCys)

Table 7. *Amino acid sequence of papain peptides of peptide 3L7*

Peptide	m*	R _F †	Amino acid composition	Sequence results‡
Q1	0.51	0.54	CMCys 1.9; Thr 1.0; Glu 1.2; Pro 1.0; Leu 2.1	Leu(Leu, Glx, Pro, Thr, CMCys)CMCys
Q2	0.73	0.38	CMCys 2.8; Asp 1.0; Thr 0.9; Glu 1.0; Pro 0.9; Leu 1.8	Leu(Leu, Glx, Pro, Thr, CMCys, CMCys, CMCys)Asx
Q3	0	0.21	His 0.9; Arg 1.2; CMCys 1.0; Asp 2.1; Pro 2.8; Tyr 0.7	CMCys-Asp-Asn-Arg(Pro, Pro, Pro, Tyr)His
Q4	-0.24	0.32	His 1.1; Arg 1.1; Asp 0.9; Pro 3.1; Tyr 0.8	Asn(Arg, Pro, Pro, Pro, Tyr)His
Q5	-0.32	0.50	Arg 1.1; Pro 2.9; Tyr 1.0	Arg(Pro, Pro, Pro)Tyr
Q6	-0.46	0.38	His 0.8; Arg 1.0; Pro 3.0; Tyr 0.9	Arg(Pro, Pro, Pro, Tyr)His
Q7	0	0.35	Ser 0.9; Glx 1.0; Pro 2.0; Val 1.0	Val(Pro, Glx, Pro)Ser
Q8	0.24	0.38	CMCys 0.9; Thr 1.0; Ser 1.0; Glu 1.2; Pro 3.0; Val 2.2	Val-Pro-Gln-Pro-Ser-Val-Pro-Thr-CMCys
Q9	0.39	0.37	CMCys 0.9; Thr 1.0; Pro 1.0; Val 1.0	Val(Pro, Thr)CMCys

* Electrophoretic mobilities at pH 4.5 relative to lysine (negative) and aspartic acid (positive).

† Chromatography in solvent BAWP.

‡ N-Terminal sequences by dansyl-Edman degradation and C-terminal residues by the tritium-labelling method.

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

indicated in Tables 4 and 5. Peptides 2L1 and 2L15 were pure enough for sequence analysis. The peptides contained in peaks 1-6 were separated in the systems indicated in Tables 4 and 5 and are identified by their mobilities or R_F values. The amino acid compositions of the purified peptides are also given in Tables 4 and 5.

Sequence of thermolysin peptides. The sequences

of the thermolysin peptides of protein IIIB3, as determined by the dansyl-Edman procedure, are given in Table 6.

A papain digestion of peptide 3L7 was used to complete its sequence. The papain peptides were separated by paper electrophoresis at pH 4.5 and further purified by paper chromatography in solvent BAWP. The mobilities and R_F values as well as

Table 8. *Sequence of thermolysin peptides of protein IIIB2 by the dansyl-Edman procedure*

Peptide no.	Residue no.	Sequence results
2L1	1-9	(Ala, CMCys, CMCys, Ala, Pro, Arg, CMCys, CMCys, Ser)
2L2	10-17	<u>Val-Arg-Thr-Gly</u> (Pro, Ala, Thr, Thr)
2L4	18-38	<u>Ile-CMCys-Ser</u> (Ser, Asx, Lys, Phe, CMCys, Arg, CMCys, Gly, Val, CMCys, Leu, Pro, Ser, Thr, CMCys, Pro, His, Asx)
2L5	24-38	<u>Phe-CMCys</u> (Arg, CMCys, Gly, Val, CMCys, Leu, Pro, Ser, Thr, CMCys, Pro, His, Asx)
2L7	39-41	<u>Ile-Ser-Leu</u>
2L8	42-65	<u>Leu-Gln-Pro</u> (Thr, CMCys, CMCys, Asx, Asx, Ser, Pro, Val, Pro, CMCys, Val, Tyr, Pro, Asx, Thr, Tyr, Val, Pro, Thr, CMCys)
2L10	68-76	<u>Leu-Asn</u> (Ser, Ser, His, Pro, Thr, Pro, Gly)
2L11	77-79	<u>Leu-Ser-Gly</u>
2L14	86-92	<u>Ile-Gln</u> (Pro, Gly, CMCys, Glx, Asx)
2L15	86-98	<u>Ile</u> (Glx, Pro, Gly, CMCys, Glx, Asx, Val, CMCys, Glx, Pro, Arg, CMCys)
2L16	93-98	<u>Val-CMCys</u> (Glx, Pro, Arg, CMCys)

amino acid composition and sequence results of the purified peptides are given in Table 7. The papain peptides established the sequence of peptide 3L7 as shown in Fig. 4.

The sequence of peptide Q1 followed from the *N*-terminal sequence of peptide 3L7. The complete sequence of peptide Q3 was established from the *C*-terminal residues and amino acid compositions of peptides Q6 and Q5. Peptide Q2 provided the overlap between peptides Q1 and Q3. Peptide Q8 completed the sequence of peptide 3L7.

The sequence and amino acid composition of the thermolysin peptides given in Table 8 establish that the peptides are derived from protein IIIB2 (Haylett, 1969). The yields of these peptides (Table 5) in comparison with those of the thermolysin peptides of protein IIIB3 (Table 4) show that there is approx. 30% of protein IIIB2 included with protein IIIB3.

Alignment of peptic peptides

The thermolysin peptides of protein IIIB3 provided overlaps between the peptic peptides and completed the amino acid sequence of protein IIIB3 as shown in Fig. 5.

No *N*-terminal residue could be detected from peptide 3L1 with the dansyl-Edman procedure. Its amino acid composition showed that this peptide was the *N*-terminus of protein IIIB3 and that it provided an overlap between peptides 3P1 and 3P2.

Peptides 3L3 and 3P5 have the same *N*-terminal sequence but peptide 3L3 is a phenylalanyl residue shorter than peptide 3P5. Thus the sequence of the latter peptide is Ile-CMCys-Ser-Ser-Asp-Lys-Phe.

From its sequence and amino acid composition peptide 3L4 combines peptides 3P5, 3P7 and 3P8 in that order. Peptide 3L4 has the same *N*-terminal sequence as peptide 3P5 and lacks the

residues valine and tryptophan, which appear as the *C*-terminus of peptide 3P8.

The sequence of peptide 3L7 overlaps peptides 3P9 and 3P10. The combination of peptides 3P9 and 3P10 is a phenylalanyl residue longer than peptide 3L7. Hence this residue must be the *C*-terminus of peptide 3P10. Peptide 3L9 with the sequence Phe-Leu provides an overlap between peptides 3P10 and 3P11.

Peptic peptides 3P1, 3P9 and 3P16 have carboxymethylcysteine as *C*-terminal residue. Since peptides 3P1 and 3P9 are the *N*-terminal sequences of thermolysin peptides 3L1 and 3L7 respectively, peptide 3P16 must be the *C*-terminal peptide of protein IIIB3.

The sequence of peptide 3L16 provides an overlap between peptides 3P13 and 3P16 since peptide 3P2, which also has a *C*-terminal leucyl residue, is connected to the *N*-terminal peptide 3P1. It is now evident that the *N*-terminal sequence of peptide 3L2 overlaps peptides 3P2 and 3P3. The only peptic split in the protein for which there is no overlap is between peptides 3P8 and 3P9. The sequence of protein IIIB3 must therefore be as shown in Fig. 5.

DISCUSSION

Although the preparation of protein IIIB3 contained approx. 30% of protein IIIB2, it was possible to assign the correct peptides to protein IIIB3 because the complete amino acid sequence of protein IIIB2 was known (Haylett & Swart, 1969). All the peptic and thermolysin peptides could be fitted into the sequence of either protein IIIB3 or IIIB2.

Protein IIIB3 is the second wool protein of which the complete amino acid sequence has been determined. It consists of 98 residues and has acetylalanine and carboxymethylcysteine as *N*- and *C*-terminal residues respectively. A salient feature of

Excluding deletions, there are 25 changes between the two proteins; 18 substitutions can be accounted

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

REFERENCES

- Brenner, M., Niederwieser, A. & Pataki, G. (1961). *Experientia*, **17**, 145.
- Doolittle, R. F. (1965). *Biochem. J.* **94**, 742.
- Easley, C. W. (1965). *Biochim. biophys. Acta*, **107**, 386.
- Edman, P. & Begg, G. (1967). *Eur. J. Biochem.* **1**, 80.
- 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. (1969). Ph.D. Dissertation: University of South Africa.
- Haylett, T. & Swart, L. S. (1969). *Text. Res. J.* **39**, 917.
- Hill, R. L. & Schmidt, W. R. (1962). *J. biol. Chem.* **237**, 389.
- 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.
- Morse, D. & Horecker, B. L. (1966). *Analyt. Biochem.* **14**, 429.
- Reindel, F. & Hoppe, W. (1954). *Chem. Ber.* **87**, 1103.
- Sanger, F. & Thompson, E. O. P. (1963). *Biochim. biophys. Acta*, **71**, 468.
- Swart, L. S., Haylett, T. & Joubert, F. J. (1969). *Text. Res. J.* **39**, 912.