Biochem. J. (1973) **133**, 641–654
Printed in Great Britain

Studies on the High-Sulphur Proteins of Reduced Merino Wool

AMINO ACID SEQUENCE OF PROTEIN SCMKB-IIIA3

By LOUIS S. SWART and THOMAS HAYLETT
National Chemical Research Laboratory, South African Council for Scientific and Industrial Research, Pretoria, Republic of South Africa

(*Received* 27 *November* 1972)

The complete amino acid sequences of wool protein SCMKB-IIIA3 (131 residues) and a minor component SCMKB-IIIA3A (130 residues) have been determined. The proteins are mutually homologous and have free threonine as the *N*-terminal residue and carboxymethylcysteine as the *C*-terminus. The peptides used for the sequence work were obtained by trypsin, thermolysin, pepsin and chymotrypsin digestions and were fractionated by chromatography on DEAE-cellulose, gel filtration on Sephadex G-25 and G-50, paper chromatography and electrophoresis. The Edman degradation method (employing both the Beckman Sequencer and a non-automatic procedure) was used to obtain the sequences of the peptides.

The high-sulphur proteins of reduced and carboxymethylated wool can be divided into four main groups, with molecular weights of approx. 23000, 19000, 17000 and 11000 (Haylett *et al.*, 1971*b*). Each group consists of a number of closely related proteins as shown by the amino acid sequences of some of these proteins. The group with a molecular weight of 11000 yielded three amino acid sequences: SCMKB-IIIB2 (Haylett & Swart, 1969), SCMKB-IIIB3 (Haylett *et al.*, 1971*a*) and SCMKB-IIIB4 (Swart & Haylett, 1971). The complete amino acid sequences of two proteins, SCMK-B2B (Elleman & Dopheide, 1972) and SCMK-B2C (Elleman, 1972), and the partial sequence of protein SCMK-B2A (Lindley & Elleman, 1972), from the 19000-molecular-weight group, have been determined.

The group with a relatively high arginine content, of molecular weight 17000, contains at least eight proteins with very similar amino acid compositions (Swart *et al.*, 1973). The present paper describes the determination of the amino acid sequences of protein SCMKB-IIIA3 and a minor component, SCMKB-IIIA3A, from this group.

Materials and Methods

Materials

Protein. Protein IIIA3* was prepared as described by Swart *et al.* (1973).

Enzymes. Trypsin was a twice-crystallized, diphenylcarbamyl chloride-treated, salt-free preparation (batch 336A) from Seravac Laboratories (Cape Town, Republic of South Africa) and had an activity of 3200 trypsin units/mg and 5 chymotrypsin units/mg; thermolysin was from Daika Kasei K.K., Osaka, Japan; pepsin and chymotrypsin (batch CD1 6100-1) were twice-crystallized, salt-free preparations from Worthington Biochemical Corp., Freehold, N.J., U.S.A.

* Abbreviations: IIIA3, protein SCMKB-IIIA3; 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; >PhNCS, 3-phenyl-2-thiohydantoin derivative.

Methods

Enzymic digestions. Tryptic digestion was carried out on 15 μmol of protein IIIA3 in 24ml of 2% (w/v) NH_4HCO_3 (enzyme/substrate ratio 1:100, w/w) for 2h at 40°C. The digestion was stopped by addition of acetic acid and the material was freeze-dried.

Thermolysin digestions were performed on 20 μmol of protein IIIA3 and 10 μmol of peptide C7 in 24ml of 2% (w/v) NH_4HCO_3 (enzyme/substrate ratio 1:100, w/w) for 1h at 40°C. The digestion mixtures were acidified to pH3 with acetic acid and freeze-dried.

Peptic digestion was carried out on 10 μmol of protein IIIA3 in 16ml of 5% (v/v) formic acid (enzyme/substrate ratio 1:100, w/w) for 2h at 40°C. The digestion was stopped by addition of pyridine and the material was freeze-dried.

Chymotryptic digestion was performed on 10 μmol of protein IIIA3 in 16ml of 2% (w/v) NH_4HCO_3. The chymotrypsin was treated in 2% (w/v) NH_4HCO_3 with an equal weight of soya-bean trypsin inhibitor for 30min at 40°C and then added to the protein (enzyme/substrate ratio 2.5:100, w/w) for 2h at 40°C. The digestion was stopped with di-isopropylphosphorofluoridate and the material was freeze-dried.

Fractionation of peptides. The peptides were separated by chromatography on DEAE-cellulose, with the Technicon Peptide AutoAnalyzer, and by paper electrophoresis and paper chromatography as described by Haylett & Swart (1969) and Haylett *et al.* (1971*a*). Chromatography on DEAE-cellulose was

also carried out in the presence of NH_4HCO_3 solutions with a 2-litre linear gradient of 0.05M–0.60M-NH_4HCO_3 solutions, on a column (150cm × 1.9cm) at 20°C. Fractions (5ml) were collected at a flow rate of 50ml/h and a Beckman DB spectrophotometer was used to monitor the effluent at 230nm.

Gel-filtration separation of peptides was performed on Sephadex G-50 or G-25 in 0.05M-ammonium acetate by using three columns (150cm × 1.9cm) connected in series. The peptides in the effluent were monitored at 230nm on a Beckman DB spectrophotometer.

Amino acid analysis. Amino acid compositions of the protein and peptides were determined as described by Haylett *et al.* (1971*a*).

Sequence analysis. Edman degradations were carried out on small peptides by the non-automatic method described by Haylett *et al.* (1971*a*).

The protein and some peptides were sequentially analysed (Edman & Begg, 1967) on the Beckman Sequencer (model 890). The reagents and solvents were purchased from Beckman Instruments or Pierce Chemical Co., Rockford, Ill., U.S.A. Quadrol buffer was used for the protein and large peptides (48 and 62 residues), and dimethylaminopropyne buffer (5% solution in propan-1-ol–water, 3:4, v/v) (Braunitzer & Schrank, 1970) was used for shorter peptides. The yield of serine was improved by addition of 1,4-dithiolbutane to the 1-chlorobutane (1:10,000, v/v). The procedures employed are described in the Beckman Sequencer Manual.

The thiazolinone derivatives of the amino acids were converted into >PhNCS amino acids by standard procedures (Haylett *et al.*, 1971*a*). The acid solution for preparing the >PhNCS derivative contained mercaptoethanol (1:1000, v/v) to improve the yield of serine (Beckman Sequencer Manual).

Gas chromatography (Pisano & Bronzert, 1969) with SP400 as stationary phase was employed to identify >PhNCS amino acids, before and after silylation with *NO*-bis(trimethylsilyl)trifluoroacetamide. Identifications were confirmed by t.l.c. on Merck aluminium t.l.c. plates precoated with silica gel F254. The systems of Brenner *et al.* (1961) were employed to identify the >PhNCS derivatives of glutamic acid, aspartic acid, their amides and carboxymethylcysteine with chloroform–methanol (9:1, v/v) as developer. The >PhNCS derivative of arginine was identified with chloroform–methanol–formic acid (70:30:1, by vol.) as developer. The >PhNCS derivatives of the other amino acids were identified by the system of Cherbuliez *et al.* (1963) with chloroform–ethyl acetate (9:1, v/v) as developing solvent.

Nomenclature of peptides

Capital letters are used to indicate enzymic digestions: T, L, P and C for trypsin, thermolysin, pepsin and chymotrypsin respectively. The capital letter A occasionally preceding these letters indicates minor component peptides. Arabic numerals following the capital letters indicate sequential numbering of the peptides. For unresolved peptides the minor component is given in parentheses.

Results

N-Terminal sequence of protein IIIA3

The protein has a free *N*-terminal amino acid, and 0.5μmol of protein was subjected to sequential analysis in the Beckman Sequencer. The *N*-terminal sequence shown in Fig. 1 was obtained.

Peptides from tryptic digestion

Purification of the tryptic petides. The elution pattern of the tryptic peptides from a Sephadex G-25 column is shown in Fig. 2(*a*). The contents of the tubes under each peak were pooled as shown by the solid bars. Peaks I, II and III were each re-run on a DEAE-cellulose column and the elution patterns are given in Fig. 2(*b*), 2(*c*) and 2(*d*) respectively. The peptides were collected as shown by the solid bars and were found to be homogeneous. The amino acid compositions of the tryptic peptides are given in Table 1 together with the analysis of protein IIIA3.

Sequences of the tryptic peptides of protein IIIA3. The sequences of the tryptic peptides of protein IIIA3 as determined by the non-automatic Edman degradation method and the partial sequence of peptide T4 as determined in the Beckman Sequencer are given in Table 2.

Assignment of the tryptic peptides. The sum of the analyses of peptides T1–T7 corresponds exactly to the analysis of protein IIIA3, which indicates that only six of the 14 arginine residues are susceptible to

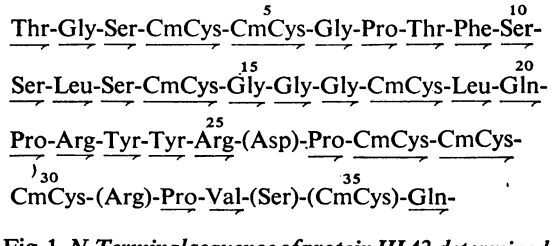


Fig. 1. *N-Terminal sequence of protein IIIA3 determined by sequential analysis in the Beckman Sequencer*

Arrows indicate positive identification of residues. Parentheses indicate loci where the cycle was performed but no identification was made; the residues shown were inferred from other experiments.

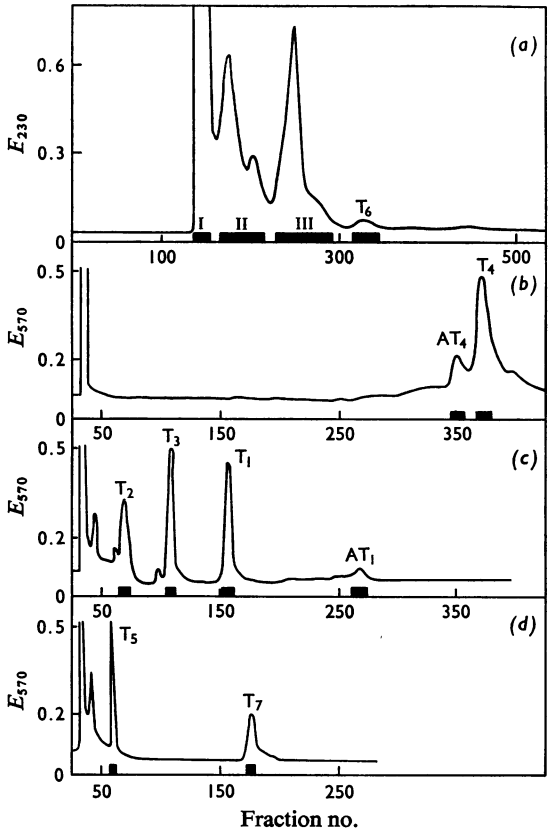


Fig. 2. *Fractionation of tryptic peptides of protein IIIA3*

(*a*) Gel-filtration pattern of tryptic peptides of protein IIIA3 on Sephadex G-25 (three columns, 150cm×1.9cm, connected in series) at room temperature in 0.05M-ammonium acetate solution (flow rate, 50ml/h; fraction volume, 5ml). (*b*)–(*d*) Elution patterns of fraction I (*b*), fraction II (*c*) and fraction III (*d*) obtained with a Technicon Peptide Auto-Analyzer. Chromatography was performed on a DEAE-cellulose column (150cm×1.9cm) 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.), pH4.1 (flow rate, 200ml/h; fraction volume, 20ml).

tryptic hydrolysis. The amino acid compositions and lower yields of peptides AT1 and AT4, when compared with peptides T1 and T4 respectively, show that the former peptides can be assigned to a contaminating minor component of protein IIIA3.

Peptides from thermolysin digestion

Purification of the thermolysin petides. 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 enumerated peaks were separated by paper chromatography or electrophoresis as given in Table 3. This table also lists the amino acid compositions, R_F values and mobilities of the purified peptides.

Sequence of the thermolysin peptides of protein IIIA3. The sequences of the thermolysin peptides of protein IIIA3 as determined by the non-automatic Edman degradation method are given in Table 4. Edman degradation studies on peptides L1 (L15) and L16 (L17) proved that they were mixtures and it was impossible to separate them by conventional methods.

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. 4. The contents of the tubes under each peak were pooled as shown by the solid bars. The peptides contained in the enumerated peaks were separated by paper chromatography or electrophoresis as given in Table 5. This table also lists the R_F values, mobilities and amino acid compositions of the purified peptides.

Sequence of the peptic peptides of protein IIIA3. The sequences of the peptic peptides of protein IIIA3 as determined by the Edman degradation method are given in Table 6.

Peptide P7 (AP7) yielded both a prolyl and a seryl residue at the second residue from the *N*-terminal end. This indicates that the peptide is a mixture. The amino acid composition of the peptide supports this observation.

Peptides from chymotryptic digestion

Purification of the chymotryptic peptides. The elution pattern of the chymotryptic peptides from a DEAE-cellulose column is shown in Fig. 5. The contents of the tubes under each peak were pooled as shown by the solid bars. The peptides contained in the enumerated peaks were separated and purified on Sephadex G-25 and/or G-50. The amino acid compositions of the peptides are given in Table 7.

Sequence of the chymotryptic peptides of protein IIIA3. The partial sequences of peptides C6 (AC6) and AC4,5, as determined in the Beckman Sequencer, are given in Table 8. The second residue of peptide C6 (AC6) yielded both proline and serine. To determine the relative yields of the two components present in peptide C6 (AC6), the peptide (2μmol) was digested with trypsin and the two tripeptides, Val-Pro-Arg and Val-Ser-Arg, were eluted as a single peak from Sephadex G-25.

Table 1. *Amino acid composition of protein IIIA3 and its tryptic peptides*

+ indicates the presence of the amino acid.

Protein or peptide ...	Amino acid composition (mol of residue/mol)										
	IIIA3	T1	T2	T3	T4	T5	T6	T7	Total T1–T7	AT1	AT4
Trp	+				+				1		+
Arg	13.95	0.87	3.86	1.95	5.08	0.98	1.04		14	1.13	4.94
CmCys	32.37	4.02	3.88	2.97	18.23	1.08		2.09	32	5.88	17.22
Asp	1.96		0.96		0.94				2		1.11
Thr	14.14	2.02	1.86	1.03	4.87	2.02	1.00	0.99	14	1.93	4.91
Ser	12.07	3.67	1.94		4.78			0.98	12	4.13	5.21
Glu	9.20	1.10	1.06	1.08	6.10				9	1.13	5.01
Pro	19.02	2.10	3.72	2.00	9.51			1.03	19	2.12	10.13
Gly	5.68	4.79			1.08				6	5.10	1.16
Ala	2.88				2.69				3		2.71
Val	8.81		3.53		4.30				8/9		4.94
Ile	2.14			0.94	0.98				2		0.77
Leu	2.71	1.65			0.95				3	1.90	0.85
Tyr	1.93		1.48						2	0.81	
Phe	2.88	0.91	0.77				0.92		3	0.71	
Yield (μmol from 15 μmol)	—	7.82	4.93	6.99	5.56	6.69	8.05	6.31	—	0.91	2.33

Table 2. *Sequence of tryptic peptides by the Edman procedure*

Peptide no.	Residue no.	Sequence results
T3	49–58	CmCys-Thr-Arg-Pro-Ile-CmCys-Glu-Pro-CmCys-Arg
T4	59–121	Arg-Pro-Val-CmCys-CmCys-Asp-Pro-CmCys-Ser-Leu-Gln-Glu-Gly-CmCys-CmCys-Arg-Pro-Ile-Thr-CmCys-CmCys-Pro-Thr-Ser-CmCys-Gln-Ala-Val-Val-CmCys-Arg-Pro-CmCys-CmCys-Trp-Ala-Thr-Thr-CmCys-CmCys-Gln-Pro-Val(Ser)-Val-(Gln,CmCys,Pro,CmCys,CmCys,-Arg,Pro,Thr,Ser,CmCys,Gln,Pro,Ala,Pro,CmCys,-Ser,Arg)
T5	122–125	Thr-Thr-CmCys-Arg
T6	126–128	Thr-Phe-Arg
T7	129–133	Thr-Ser-Pro-CmCys-CmCys

Table 3. *Amino acid composition of the thermolysin peptides of protein IIIA3*

+ indicates the presence of the amino acid.

Peptide ...	Amino acid composition (mol of residue/mol)																		
	L1 (L15)	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L16 (L17)	L18	AL6	AL9	AL16 (AL17)
Trp														+					
Arg				1.13	1.93		1.00		2.86		0.85	0.95		1.00	2.88	0.77		1.85	3.09
CmCys	3.66		1.61		2.84	1.03			1.74	2.53	4.19	2.08	2.61	2.67	5.60	1.72		0.93	4.71
Asp					1.12					1.04	0.96								
Thr	3.70					1.18		1.00	1.05				1.87		4.15	0.93	1.95	0.97	3.74
Ser	1.28	1.61	1.01			1.08	0.94			0.83	0.99		1.03		2.76	1.01	1.97	1.12	2.38
Glu	1.16			1.03		1.00			1.03		1.77	1.92	1.07		2.19		1.08		1.12
Pro	2.08			1.00	2.11		1.13		3.06	1.10	1.78	1.20	1.03	1.11	4.23	1.03		1.14	3.79
Gly	2.28		3.02								1.13	1.10							
Ala	0.68												0.99		1.18				1.02
Val						0.89	0.93	1.00		0.86	0.77			1.93	1.69		1.01	0.88	1.28
Ile									0.92				0.87						
Leu			0.97	0.97							0.76	0.83							
Tyr				0.92	0.94														
Phe		1.00						1.00								0.81			
Yield (μmol from 20 μmol)	4.38	8.09	6.77	8.06	4.76	8.27	9.17	3.01	6.18	5.44	2.51	4.02	5.57	6.57	5.06	7.66	3.10	1.02	1.59
Peak no. (from Fig. 3)	7	2	8	1	7	3	1	2	2	11	11	6	9	9	4	5	2	1	2
System*	1.9	BAWP	BAWP	BAWP	1.9	—	BAWP	BAWP	BAWP	1.9	1.9	—	BAW	BAW	—	BAW	BAWP	BAWP	BAWP
R_F	—	0.46	0.23	0.45	—	—	0.30	0.69	0.20	—	—	—	0.41	0.68	—	0.25	0.31	0.21	0.06
m_{Lys}†	0.21	—	—	—	0.34	—	—	—	—	0.45	0.55	—	—	—	—	—	—	—	—

* System used to purify peptides: 1.9, electrophoresis at pH1.9; BAWP and BAW, chromatography in solvent BAWP and BAW respectively.
† Electrophoretic mobility relative to lysine.

Table 4. *Sequence of thermolysin peptides by the Edman procedure*

Peptide no.	Residue no.	Sequence results
L4	19–23	Leu-Gln-Pro-Arg-Tyr
L5	24–32	Tyr-Arg-Asp-Pro-CmCys-CmCys-CmCys-Arg-Pro
L7	39–42	Val-Ser-Arg-Pro
L8	43–45	Val-Thr-Phe
L9	50–60	Thr-(Arg)-Pro-Ile-CmCys-Glu-Pro-CmCys-(Arg)-Arg-(Pro)
L10	61–67	Val-CmCys-CmCys-Asp-Pro-CmCys-Ser
L12	68–75	Leu-Gln-Glu-Gly-CmCys-CmCys-Arg-Pro
L13	76–85	Ile-Thr-CmCys-CmCys-(Pro,Thr,Ser,CmCys,Gln,Ala)
L14	86–93	Val-Val-CmCys-Arg-Pro-CmCys-CmCys-Trp
L18	127–133	Phe-Arg-Thr-Ser-Pro-CmCys-CmCys
AL6	33–38	Val-Ser-Ser-Gln-Thr-Thr
AL9	46–52	Val-Ser-Arg-CmCys-Thr-Arg-Pro

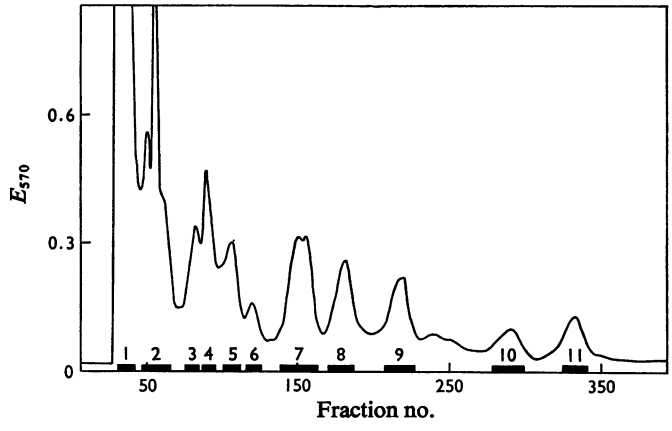


Fig. 3. *Elution pattern of thermolysin peptides of protein IIIA3*

DEAE-cellulose was used, with a Technicon Peptide AutoAnalyzer. Conditions are as given for Fig. 2(*b*).

Amino acid analysis gave the following composition: Arg 1.02, Ser 0.34, Pro 0.68, Val 0.98.

This result proves that the peptide containing proline in the second residue (C6) is the major component and is present in a ratio of approx. 2:1.

Reinvestigation of peptides L16 (L17)

Glutaminyl residues in positions 99 and 104 of protein IIIA3 were partly converted into pyrrolidonecarboxylic acid during the Edman degradation of peptides P12 and AP12, and peptides L16 (L17)

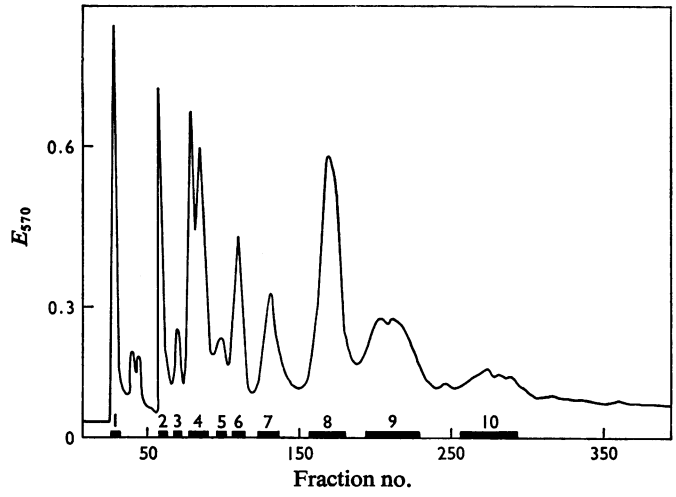


Fig. 4. *Elution pattern of peptic peptides of protein IIIA3*

DEAE-cellulose was used with a Technicon Peptide AutoAnalyzer. Conditions are as given for Fig. 2(*b*).

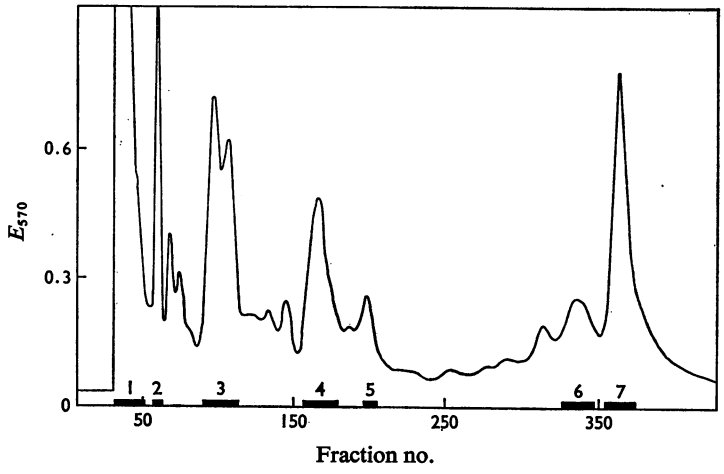


Fig. 5. *Elution pattern of chymotryptic peptides of protein IIIA3*

DEAE-cellulose was used with a Technicon Peptide AutoAnalyzer. Conditions are as given for Fig. 2(*b*).

could not be purified. In addition, the presence of peptides AL16 (AL17) and AP12 showed that a minor component was complicating the sequence of protein IIIA3 from residues 105–120. The result was that no sequence data could be obtained from these peptides for this part of the molecule.

Examination of the results obtained from the different enzymic digestions of protein IIIA3 suggested the use of the following procedure to elucidate the sequence from residues 105–120 of the protein.

Protein IIIA3 was digested with chymotrypsin and the peptides were separated by Sephadex G-50 gel filtration. The second largest peptide had a composition corresponding to peptide C7 and was further digested with thermolysin. The digest was separated on Sephadex G-25. The material eluted at the exclusion volume of the column had a composition corresponding to peptide L16 (L17) with an additional phenylalanyl residue.

This peptide, C7L1 (8 μmol), was chromatographed

Table 5. *Amino acid composition of the peptic peptides of protein IIIA3*

	Amino acid composition (mol of residue/mol)																
Peptide ...	P1	P2	P3	P4	P5	P6	P7 (AP7)	P8	P9	P10	P11	P12	P13	P14	AP5	AP5,6	AP12
Arg			0.95	0.98	1.99	0.99	1.98	0.78	0.97	0.97		2.96	2.86	0.92	1.95	3.06	3.11
CmCys	2.02		1.78	2.00	4.05		1.87	4.53	2.59	3.08		7.64	5.66	1.74	3.01	3.07	6.59
Asp					0.93										1.02	1.06	
Thr	1.98					2.16	1.01	1.60			1.96	3.86	3.92	1.09	1.02	2.97	3.91
Ser	0.99	1.89	2.62	0.95	1.03	0.94	0.44	0.97				3.04	2.77	1.05	2.12	2.86	3.07
Glu			1.18	1.28	1.09		1.06	1.18				3.05	2.18		1.09	1.18	2.17
Pro	0.97		0.94	0.92	2.12	1.18	1.58	2.00	1.03	1.08		5.08	4.26	1.03	2.01	3.06	5.02
Gly	2.05		2.99	3.12				0.73									
Ala									0.86		1.04	1.20	0.88				1.06
Val					0.95	1.92	0.97		1.65	1.54		2.08	1.01		0.77	2.88	1.83
Ile							0.81	0.74									
Leu		1.00	2.17	1.04													
Tyr			0.94	0.98	0.79										0.80	0.92	
Phe	0.98					1.00								1.00		0.92	
Yield (μmol from 10 μmol)	5.81	4.03	1.13	4.31	5.47	5.83	6.44	4.17	6.47	1.55	5.34	5.11	2.22	6.47	1.63	0.68	1.08
Peak no. (from Fig. 4)	8	2	4	4	9	1	4	10	8	9	2	8	5	6	7	3	7
System*	BAWP	1.9	BAWP	BAWP	BAWP	BAWP	BAWP	BAWP	BAWP	BAWP	1.9	BAWP	BAWP	BAW	BAWP	BAWP	BAWP
R_F	0.55	—	0.61	0.50	0.11	0.65	0.34	0.13	0.35	0.43	—	0.05	0.04	0.25	0.18	0.28	0.03
m_{Lys}	—	0.35	—	—	—	—	—	—	—	—	0.26	—	—	—	—	—	—

* Abbreviations are as defined in Table 3.

Table 6. *Sequence of peptic peptides by the Edman procedure*

Peptide no.	Residue no.	Sequence results
P1	1–9	Thr-Gly-Ser-CmCys-CmCys-Gly-Pro-Thr-Phe
P2	10–12	Ser-Ser-Leu
P4	13–23	Ser-CmCys-Gly-Gly-Gly-CmCys-Leu-Gln-Pro-Arg-Tyr
P5	24–37	Tyr-Arg-Asp-Pro-CmCys-CmCys-CmCys-Arg-Pro-Val-Ser-CmCys-Gln
P6	38–45	Thr-Val-Ser-Arg-Pro-Val-Thr-Phe
P7(AP7)	46–55	Val-(Pro/Ser)-Arg-CmCys-Thr-Arg-Pro-Ile-CmCys-Glu
P8	71–84	Gly-CmCys-CmCys-Arg-Pro-Ile-Thr-CmCys(CmCys,Pro,Thr,Ser,CmCys,Gln)
P9	85–92	Ala-Val-Val-CmCys-Arg-Pro-CmCys-CmCys
P10	86–93	Val-Val-CmCys-Arg(Pro,CmCys,CmCys)
P11	94–96	Ala-Thr-Thr
P12	97–126	CmCys-CmCys-Gln-Pro-Val-Ser-Val-Gln(CmCys,Pro,CmCys,-CmCys,Arg,Pro,Thr,Ser,CmCys,Gln,Pro,Ala,Pro,CmCys,Ser,-Arg,Thr,Thr,CmCys,Arg,Thr)
AP12	97–126	CmCys-CmCys-Gln-Pro-Val-Ser-Val-Gln(CmCys,Pro,CmCys,-CmCys,Arg,Pro,Thr,Ser,CmCys,Pro,Ser,Ala,Pro,Arg,Thr,-Thr,CmCys,Arg,Thr)

on DEAE-cellulose with a linear gradient of 0.05M–0.60M-NH_4HCO_3 and the elution pattern is given in Fig. 6. Components a to d were collected as shown by the solid bars and their amino acid compositions are given in Table 9.

Sequence of the components of peptide C7L1. The sequence of components a, c and d of peptide C7L1, as determined in the Beckman Sequencer, are given in Table 10.

It is evident from the results that component d resulted from a partial thermolysin split of component c and that component b is a fragment of component a. The latter components, a and b, were obtained in lower yields than their corresponding components c and d respectively. Components a and b contain a carboxymethylcysteinyl and a glutaminyl residue less than components c and d and must therefore originate from the same protein as peptide AP12.

Alignment of protein IIIA3

The *N*-terminal sequence of protein IIIA3 connects peptides T1 and T2 and the partial sequence of peptide C6 connects peptides T2, T3 and T4. Peptide

Table 7. *Amino acid composition of the chymotryptic peptides of protein IIIA3*

+ indicates the presence of the amino acid.

Peptide ...	Amino acid composition (mol of residue/mol)									
	C1	C2	C3	C4	C5	C6 (AC6)	C7	C8	AC3	AC4,5
Trp						+				
Arg			1.03	1.75	1.17	5.95	2.94	0.71		2.80
CmCys	2.02		1.95	3.86		14.62	7.59	2.00	3.89	2.85
Asp				0.95		1.01				1.00
Thr	1.86				2.09	3.03	5.64	0.95		2.94
Ser	1.10	1.98	1.02	1.05	1.17	2.46	3.18	1.08	1.11	3.04
Glu			1.02	1.10		4.11	3.22		0.98	1.10
Pro	0.97		0.93	2.08	1.01	7.51	5.20	1.02	1.11	3.30
Gly	2.05		2.68			1.13			2.95	
Ala						1.03	2.03			
Val				1.01	1.75	3.69	2.10			2.81
Ile						1.84				
Leu		1.00	0.91			0.98			0.90	
Tyr			1.85						0.89	
Phe	0.92				0.81		0.80			0.76
Yield (μmol from 10 μmol)	6.60	5.12	6.11	5.30	6.90	6.76	5.14	7.46	1.09	1.99
Peak no. (from Fig. 5)	4	2	3	5	1	7	4	3	6	3

Table 8. *Sequence of chymotryptic peptides by the Edman procedure*

Peptide no.	Residue no.	Sequence results
C6 (AC6)	46–93	Pro Val-Ser-Arg-CmCys-Thr-Arg-Pro-Ile- CmCys-Glu-Pro-CmCys-Arg-Arg-Pro-Val- CmCys-CmCys-Asp-Pro-CmCys-Ser-Leu-Gln- Glu-Gly-CmCys-CmCys-Arg-Pro(Ile,Thr,- CmCys,CmCys,Pro,Thr,Ser,CmCys,Gln,Ala,- Val,Val,CmCys,Arg,Pro,CmCys,CmCys,Trp)
AC4,5	25–45	Arg-Asp-Pro-CmCys-CmCys-CmCys-Arg-Pro- Val-Ser-Ser-Gln-(Thr,Thr)-Val(Ser,Arg,Pro,- Val,Thr,Phe)

T5 is connected to peptide T6 by peptide C7L1c and peptide L18 connects peptides T6 and T7.

The sequence of peptide T2 is provided by peptides P5, P6 and P7. The complete sequence of peptide T4 follows from the overlap of its partial sequence with the partial sequences of peptides P12 and C7L1c.

The unequivocal sequence of protein IIIA3 can now be constructed as presented in Fig. 7.

Sequence of minor component

The proteolytic enzymes yielded minor peptides with compositions or sequences that could not be fitted into the sequence of protein IIIA3. The yields of these peptides were, however, lower than those of peptides assigned to protein IIIA3. The minor peptides could be used to deduce the sequence of a component designated IIIA3A.

Comparison of peptides AT1 and AC3 with peptides T1 and C3 respectively indicates that an arginine and a tyrosine of protein IIIA3 have been exchanged for carboxymethylcysteinyl residues. Tryptic peptide AT1 is three residues longer than peptide T1, and both contain only one arginyl residue, proving that

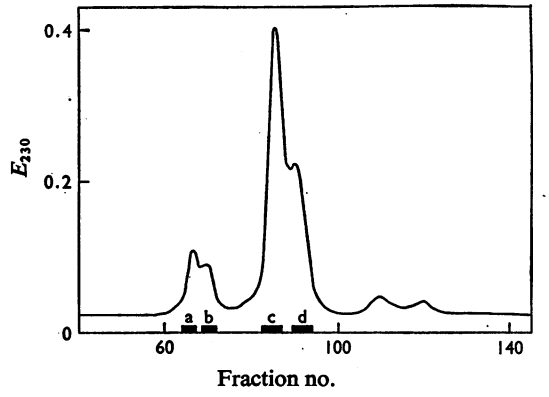


Fig. 6. *Elution pattern of components of peptide C7L1*

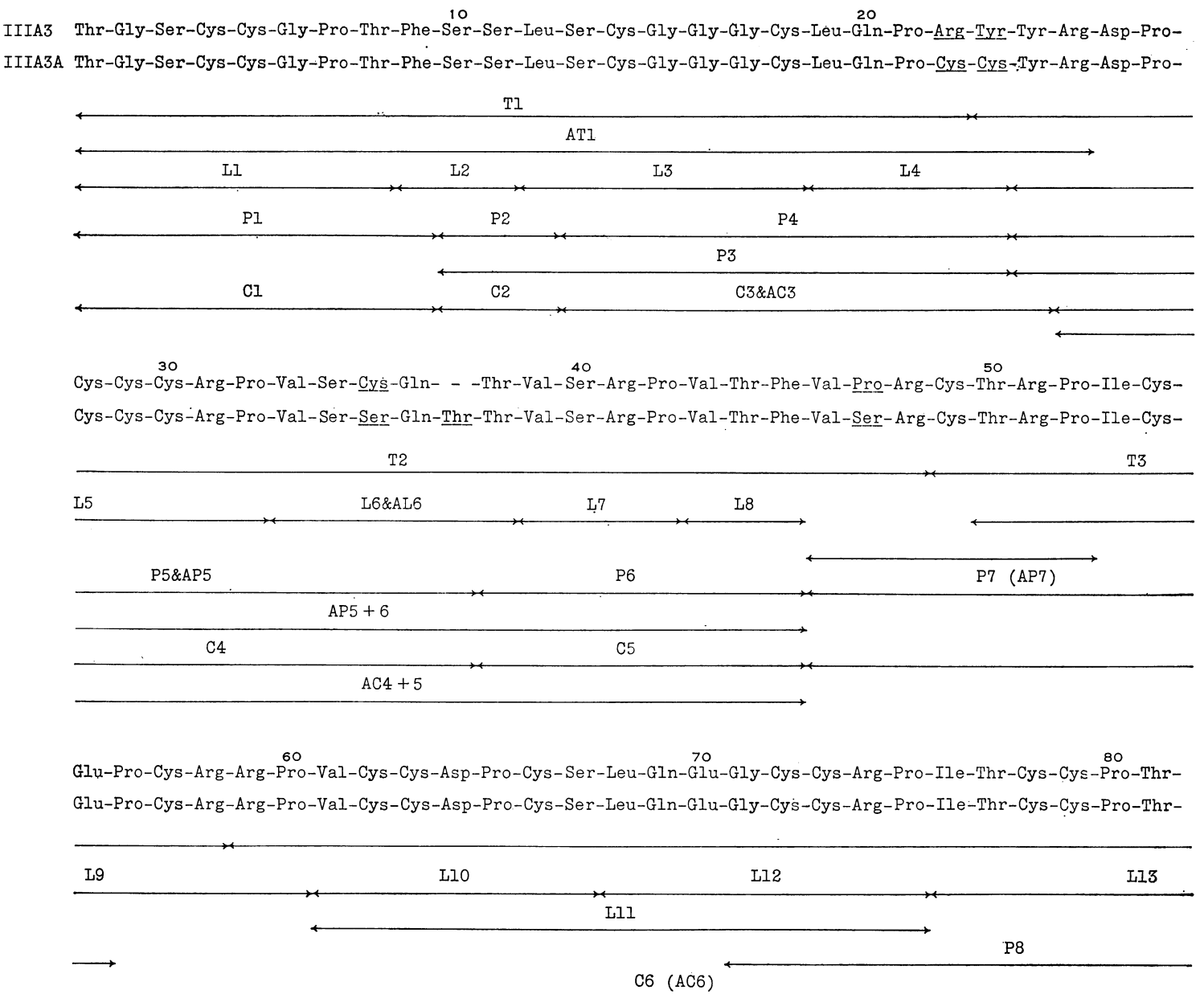
A Beckman DB spectrophotometer was used. Chromatography was performed on a DEAE-cellulose column (150cm×1.9cm) at 20°C with a 2 litre linear gradient of 0.05M–0.60M-NH_4HCO_3 solution (flow rate, 50ml/h; fraction volume, 5ml).

Table 9. *Amino acid composition of the components of peptide C7L1*

Peptide ...	Amino acid composition (mol of residue/mol)			
	C7L1 a	C7L1 b	C7L1 c	C7L1 d
Arg	2.85	2.83	2.98	2.89
CmCys	4.67	4.62	5.57	5.35
Thr	3.70	3.78	3.94	3.79
Ser	3.00	1.95	3.08	2.14
Glu	1.12	1.03	2.22	2.16
Pro	3.72	4.17	4.11	3.96
Ala	0.95	0.88	1.13	1.01
Val	1.66	0.82	2.01	1.08
Phe	0.84	0.87	0.79	0.68
Yield (μmol)	0.70	0.50	3.28	2.02

Table 10. *Sequence of major components of peptide C7L1 by the Edman procedure*

Peptide no.	Residue no.	Sequence results
C7L1 a	101–127	Val-Ser-Val-Gln-CmCys-Pro-CmCys-CmCys-Arg-Pro-Thr-Ser-CmCys-Pro-Ser-Ala-Pro-(Arg,Thr,Thr,CmCys,Arg,Thr)
C7L1 c	101–127	Val-Ser-Val-Gln-CmCys-Pro-CmCys-CmCys-Arg-Pro-Thr-Ser-CmCys-Gln-Pro-Ala-Pro-CmCys-Ser-Arg-Thr-Thr-CmCys(Arg,Thr,Phe)
C7L1 d	103–127	Val-Gln-CmCys-Pro-CmCys-CmCys-Arg-Pro-Thr-Ser-CmCys-Gln-Pro-Ala-Pro-CmCys-Ser-Arg-Thr(Thr,CmCys,Arg,Thr,Phe)


IIIA3 Thr-Gly-Ser-Cys-Cys-Gly-Pro-Thr-Phe-Ser-Ser-Leu-Ser-Cys-Gly-Gly-Gly-Cys-Leu-Gln-Pro-Arg-Tyr-Tyr-Arg-Asp-Pro-
IIIA3A Thr-Gly-Ser-Cys-Cys-Gly-Pro-Thr-Phe-Ser-Ser-Leu-Ser-Cys-Gly-Gly-Gly-Cys-Leu-Gln-Pro-Cys-Cys-Tyr-Arg-Asp-Pro-
10
20
T1
AT1
L1
L2
L3
L4
P1
P2
P4
P3
C1
C2
C3&AC3
Cys-Cys-Cys-Arg-Pro-Val-Ser-Cys-Gln- - -Thr-Val-Ser-Arg-Pro-Val-Thr-Phe-Val-Pro-Arg-Cys-Thr-Arg-Pro-Ile-Cys-
Cys-Cys-Cys-Arg-Pro-Val-Ser-Ser-Gln-Thr-Thr-Val-Ser-Arg-Pro-Val-Thr-Phe-Val-Ser-Arg-Cys-Thr-Arg-Pro-Ile-Cys-
30
40
50
T2
T3
L5
L6&AL6
L7
L8
P5&AP5
P6
P7 (AP7)
AP5 + 6
C4
C5
AC4 + 5
Glu-Pro-Cys-Arg-Arg-Pro-Val-Cys-Cys-Asp-Pro-Cys-Ser-Leu-Gln-Glu-Gly-Cys-Cys-Arg-Pro-Ile-Thr-Cys-Cys-Pro-Thr-
Glu-Pro-Cys-Arg-Arg-Pro-Val-Cys-Cys-Asp-Pro-Cys-Ser-Leu-Gln-Glu-Gly-Cys-Cys-Arg-Pro-Ile-Thr-Cys-Cys-Pro-Thr-
60
70
80
L9
L10
L12
L13
L11
P8
C6 (AC6)

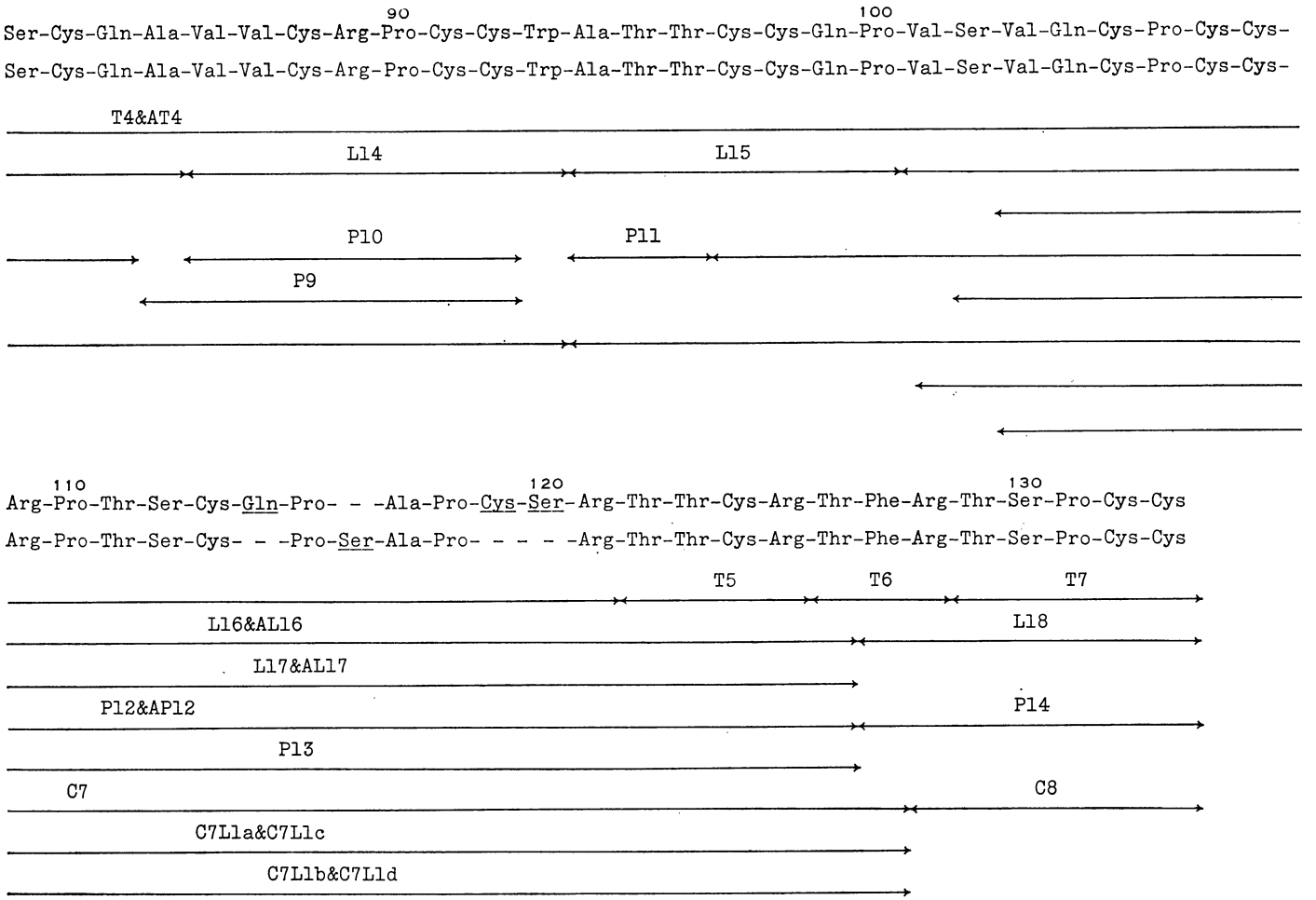


Fig. 7. ***Complete amino acid sequences of proteins IIIA3 and IIIA3A***

Tryptic (T), thermolysin (L), peptic (P) and chymotryptic (C) peptides are indicated. Where protein IIIA3A is the origin A precedes the peptide designation. Names coupled by '&' indicate resolved peptides from a common region. Bracketed names indicate unresolved peptides. Differences between proteins are underlined.

arginine in position 22 has been replaced. Peptides AC3 and AP5 both contain a single tyrosyl residue, which must be common. This proves that tyrosine at residue 23 has been exchanged.

Peptide AL6, with the unique sequence Val-Ser-Ser-Gln-Thr-Thr, corresponds to peptide L6, with the sequence Val-Ser-CmCys-Gln-Thr (positions 33–38). This correspondence is supported by the amino acid compositions of peptides AP5, AP5,6 and AC4,5 and the partial sequence of peptide AC4,5. The carboxymethylcysteinyl residue at position 35 of protein IIIA3 has therefore been exchanged for a seryl residue and an extra threonyl residue inserted between residues 36 and 38.

The exchange of proline at position 47 of protein IIIA3 with a seryl residue is established by the amino acid sequences of peptides AL9, P7 (AP7) and C6 (AC6).

The sequence of peptide C7L1a compared with the sequence of peptide C7L1c proves that the glutaminyl residue at position 114, the carboxymethylcysteinyl residue at position 119 and the seryl residue at position 120 have been deleted and that serine has been inserted between residues 115 and 117.

The complete amino acid sequence of the minor component, IIIA3A, in the preparation of protein IIIA3, can therefore be presented (Fig. 7).

Discussion

Chromatography on DEAE-cellulose with NH_4HCO_3 solutions partially separated two peptides that were eluted as a single peak with pyridinium buffers. NH_4HCO_3 solutions offer a further advantage because peptides can be monitored at 230nm, eliminating the use of the elaborate Peptide AutoAnalyzer.

In the latter part of the project gel filtration on Sephadex G-50 and G-25, with a total column length of 450cm, was used to separate and purify peptides obtained by chromatography on DEAE-cellulose. This method offers an advantage over paper-separation techniques because yields of the peptides are higher and contamination with extraneous material extracted from paper is avoided.

The cleavage of the proteins with trypsin followed general tryptic specificity except for arginine in position 26, followed by an aspartyl residue, which did not split in protein IIIA3, but did in protein IIIA3A. Thermolysin digestion of protein IIIA3 hydrolysed the bond between carboxymethylcysteine and threonine at positions 49 and 50, but not between proline and isoleucine at positions 52 and 53. The opposite result was obtained for protein IIIA3A. Chymotrypsin unexpectedly did not cleave the bond between leucine and glutamine at positions 19 and 20.

It has been possible to elucidate the complete amino acid sequences of protein IIIA3 and a minor component IIIA3A, although these proteins were present in an approximate ratio of 2:1. The proteins are closely related and differ in only nine positions. Excluding the five deletions, these differences can be attributed to single nucleotide-base changes. The proteins have calculated molecular weights of 16133 and 15848 and in their native state in wool have 10 and 9 net basic charges respectively.

Proteins IIIA3 and IIIA3A originate from the group of high-sulphur proteins with a relatively high arginine content (Swart *et al.*, 1973). The proteins are void of lysine, histidine and methionine and have a relatively high content of half-cystine (25%) and arginine (11–10%). A major feature of their sequences is the presence of threonine as the *N*-terminal amino acid, whereas all the other high-sulphur proteins with known amino acid sequences have acetylalanine (Haylett *et al.*, 1971*b*). These proteins follow the common feature of all high-sulphur proteins by having a half-cystine as *C*-terminus (Haylett *et al.*, 1971*b*).

The various amino acids in the sequences of protein IIIA3 and IIIA3A are fairly evenly distributed. This is contrary to the finding for the groups of 11000 and 19000 molecular weight, where long regions without half-cystinyl residues are found (Swart & Haylett, 1971; Elleman, 1972).

We thank Mrs. A. Ruelle, Miss C. M. Potgieter and Mrs. P. G. Drijfhout for technical assistance.

References

Braunitzer, G. & Schrank, B. (1970) *Hoppe-Seyler's Z. Physiol. Chem.* **351**, 417

Brenner, M., Niederweiser, A. & Pataki, G. (1961) *Experientia* **17**, 145–153

Cherbuliez, E., Baehler, B. R., Marszaleck, J., Sussman, A. R. & Rabinowitz, J. (1963) *Helv. Chim. Acta* **46**, 2446–2452

Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–91

Elleman, T. C. (1972) *Biochem. J.* **128**, 1229–1239

Elleman, T. C. & Dopheide, T. A. (1972) *J. Biol. Chem.* **247**, 3900–3909

Haylett, T. & Swart, L. S. (1969) *Text. Res. J.* **39**, 917–929

Haylett, T., Swart, L. S. & Parris, D. (1971*a*) *Biochem. J.* **123**, 191–200

Haylett, T., Swart, L. S., Parris, D. & Joubert, F. J. (1971*b*) *Appl. Polym. Symp.* **18**, 37–44

Lindley, H. & Elleman, T. C. (1972) *Biochem. J.* **128**, 859–867

Pisano, J. J. & Bronzert, T. J. (1969) *J. Biol. Chem.* **244**, 5597–5607

Swart, L. S. & Haylett, T. (1971) *Biochem. J.* **123**, 201–210

Swart, L. S., Parris, D. & Joubert, F. J. (1973) *J. Chromatogr.* in the press