THE POSSIBILITY OF COMMON AMINO ACID SEQUENCES IN HIGH-SULPHUR PROTEIN FRACTIONS FROM WOOL

By R. L. Darskus,* J. M. Gillespie,* and H. Lindley*

[*Manuscript received March* 3, 1969]

Summary

S-Carboxymethyl derivatives of the high-sulphur components of reduced Merino wool have been subdivided by chromatography into 17 fractions, the amino acid compositions of which are reported. Tryptic, chymotryptic, and thermolysin digests of each fraction have been studied by high-voltage paper electrophoresis at pH 3·5 and 6·5. The results suggest that the high-sulphur proteins consist of families of proteins probably containing common structural features. Evidence is presented that the heterogeneity of high-sulphur proteins is not artefactual.

I. Introduction

Hair keratins of the α-type are thought to be made up of microfibrils containing α-helical proteins of relatively low cystine content, embedded in a matrix of high-sulphur proteins. The low-sulphur protein is believed to be responsible for the characteristic X-ray diffraction pattern of α-keratins. Compared to the high-sulphur fraction it appears to be relatively homogeneous and so far as is known its biosynthesis follows classical lines. The high-sulphur proteins, on the other hand, are extremely heterogeneous in composition and molecular weight, which ranges from 10,000 to 30,000 or even higher, and they pose many unsolved problems regarding their structure and synthesis. The review by Crewther *et al.* (1965) can be consulted for further details on the low-sulphur proteins and that by Gillespie (1965) for the high-sulphur proteins.

The high-sulphur proteins are probably synthesized higher up the follicle than the fibrillar low-sulphur material (Downes, Sharry, and Rogers 1963; Downes *et al.* 1966). Coincident with the appearance of matrix proteins in the follicle, the cell nuclei begin to show signs of changes and thereafter break down, and their contents are largely resorbed (Mercer 1961). Shortly after this the keratinization process sets in and the cell is rapidly converted to a solid mass of protein. Obviously this is a unique milieu in which to carry out protein synthesis and it would not be

* Division of Protein Chemistry, CSIRO Wool Research Laboratories, Parkville, Vic. 3052.

surprising if the synthetic mechanism also showed unique features. Before any investigation of the synthetic mechanisms can be usefully carried out it is important to be able to characterize the end products, i.e. the high-sulphur proteins. This work can be regarded as a further stage in such studies in progress in this laboratory.

II. Materials and Methods

(*a*) *Wool*

Merino wool top from a commercial flock (Wintoc) was cleaned by standard procedures of solvent and water extraction (Gillespie 1962).

(*b*) *Preparation of* S-*Carboxymethyl High-sulphur Proteins* (*SCMKB*)*

The preferential extraction method of Gillespie (1962) was used. Wool (50 g) was extracted for 16 hr at 0°C with 0·8M potassium thioglycollate (1·5 litres) at pH 10·2. The solution was filtered and the pH adjusted to 5 to precipitate the extracted proteins. These were filtered off and redissolved in 0·1M potassium thioglycollate (1 litre) at pH 9·0 in a completely filled and sealed Waring Blendor jar. Iodoacetate solution (40 g iodoacetic acid in 150 ml water adjusted to pH 7) was added and the pH kept between 8 and 9 until a negative nitroprusside test indicated completion of the reaction (approx. 5 min). The solution was then made just nitroprusside-positive with thioglycollate to prevent non-specific carboxymethylation by excess iodoacetate. The solution was dialysed overnight against deionized water and then brought to ionic strength 0·5 and pH 4·1 with sodium acetate and acetic acid to precipitate all the proteins except the high-sulphur fraction. After centrifugation the supernatant was dialysed against deionized water and the SCMKB recovered by lyophilization.

(*c*) *Chromatography of SCMKB*

Chromatography was carried out on a DEAE-cellulose (Eastman Kodak) column at pH 4·5 and 20°C with a linear gradient of sodium acetate–acetic acid, ionic strength 0·2–0·8. SCMKB (1 g in 10 ml dialysed against the starting buffer) was applied to a column (1 by 120 cm) and eluted at a flow rate of 24 ml/hr. After 182 fractions of 7·5 ml each had been collected the buffer ionic strength was increased to 1·0 to complete the elution. The protein concentration in the effluent was monitored continuously at 280 mμ with an Isco model 222 ultraviolet analyser. This procedure is similar to that used by Joubert and Burns (1967).

(*d*) *Amino Acid Analyses*

The samples were hydrolysed for 18 hr *in vacuo* at 110°C with 6N HCl, and analysed with a Spinco model 120C amino acid analyser.

(*e*) *High-voltage Electrophoresis*

This was carried out in a Michl-type apparatus and the position of the peptides was subsequently revealed by the Cl_2–tolidine staining technique.

(*f*) *Enzyme Digestions*

Digestions with α-chymotrypsin (Worthington), trypsin (Mann), and thermolysin (Calbiochem) were carried out at 40°C in 1% (w/v) ammonium carbonate solution (pH 8·3) using 0·1 mg of enzyme for each 10 mg of protein. Chymotrypsin was pre-incubated for 30 min with 20% of its weight of lima bean trypsin inhibitor (Koch-Light) before addition to the protein. After 4 hr the digests were shell-frozen and freeze-dried to remove water and ammonium carbonate.

(*g*) *Dansylation*

This was performed by the procedure outlined by Gray (1967).

*S-Carboxymethylkerateine-B, following the nomenclature of Goddard and Michaelis (1935).

TABLE 1

AMINO ACID COMPOSITION OF FRACTIONS OF HIGH-SULPHUR PROTEINS OF MERINO WOOL SEPARATED ACCORDING TO THE SCHEME OF FIGURE 1

Results are expressed as number of residues per 10^4 g of protein

Amino Acid	Fraction No.																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Lysine	0·8	0·8	0·9	0·8	0·7	0·7	0·5	0·6	0·6	0·4	0·3	0·3	0·3	0·2	0·4	0·5	0·8
Histidine	0·8	1·3	1·7	1·6	1·3	1·3	0·9	0·9	0·6	0·5	0·3	0·3	0·2	0·3	0·5	0·5	0·8
Arginine	6·6	5·2	4·6	5·2	5·1	5·5	6·1	6·2	6·9	5·7	6·8	6·7	5·5	4·2	4·0	4·2	4·6
SCM-cysteine	11·4	12·7	13·7	15·0	15·2	15·2	16·4	16·9	18·4	19·2	19·8	17·3	16·8	19·6	20·3	21·0	19·6
Aspartic acid	2·7	4·3	6·1	5·1	4·8	5·0	3·9	3·5	2·8	2·1	1·9	1·7	1·3	1·2	1·1	1·0	<1·0*
Threonine	8·8	9·3	9·0	9·2	9·3	9·1	9·4	9·3	9·3	9·6	8·8	8·0	8·6	8·4	8·3	8·9	9·7
Serine	13·2	11·3	8·4	8·1	8·1	7·8	8·0	8·1	7·7	8·5	8·8	11·6	12·3	11·5	11·1	10·5	11·0
Glutamic acid	5·8	5·4	4·4	4·9	5·2	4·9	5·4	5·6	5·8	6·3	6·2	7·4	8·8	9·3	9·5	9·1	9·1
Proline	10·2	11·1	12·3	12·1	13·0	12·3	12·7	12·9	12·6	12·7	11·7	11·6	10·4	9·4	9·3	10·1	10·9
Glycine	5·6	4·3	4·2	3·7	3·4	4·3	3·6	3·4	3·6	3·2	4·1	5·3	6·8	6·9	6·5	5·4	5·2
Alanine	2·4	2·3	2·6	2·5	2·5	2·5	2·4	2·3	2·1	2·1	2·1	2·4	3·1	2·9	2·8	2·6	2·7
Valine	7·7	6·9	5·5	5·6	5·0	5·2	5·3	5·3	5·4	5·6	5·5	5·2	4·8	3·9	3·8	3·7	4·2
Isoleucine	2·5	2·9	3·3	2·9	2·7	2·8	2·5	2·4	1·9	2·0	2·0	2·5	3·0	3·0	3·1	2·5	2·0
Leucine	4·6	5·3	6·0	5·5	5·1	5·0	4·2	4·2	3·2	3·2	2·7	2·4	2·3	2·0	1·9	1·8	1·9
Tyrosine	2·9	2·3	1·6	1·5	1·6	1·6	1·4	1·2	0·9	1·1	1·1	1·8	2·1	2·0	1·8	1·8	1·9
Phenylalanine	2·3	2·0	2·4	2·1	2·5	2·3	2·1	2·0	1·9	1·9	1·6	1·1	0·9	1·0	1·0	0·8	0·6

* Because of the high ratio of SCM-cysteine to aspartic acid no accurate value can be given in this case.

III. Results

Figure 1 shows the elution pattern obtained by chromatography of 1 g of the high-sulphur proteins on DEAE-cellulose. Fractions were pooled as shown and freeze-dried after dialysis against deionized water. Yields are given below the figure, and the results of amino acid analysis of each fraction appear in Table 1. Methionine was absent from all fractions, and no tryptophan analyses were carried out. From the results it can be seen that significant differences in composition occur between fractions. A similar separation has been described by Joubert and Burns (1967), who found in addition that each fraction contains several proteins.

Figure 2 shows a high-voltage paper electrophoretogram of tryptic digests of each of the 17 fractions run side-by-side on the same sheet at pH 6·5. The figure immediately suggests the presence of a number of families of proteins with many common sequences within each family. High-voltage electrophoresis of the same digests at pH 3·5 supported this observation. Although Figure 1 shows considerable overlap of components between adjacent fractions, the common features in the electrophoretograms persist over a much larger range of fractions than could reasonably be expected on the basis of cross-contamination alone.

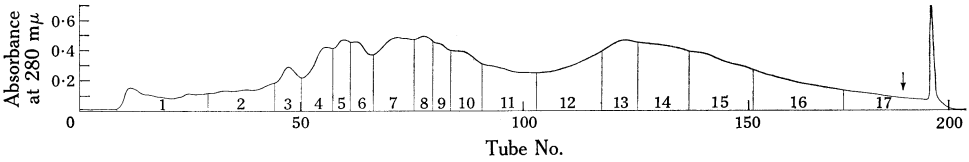
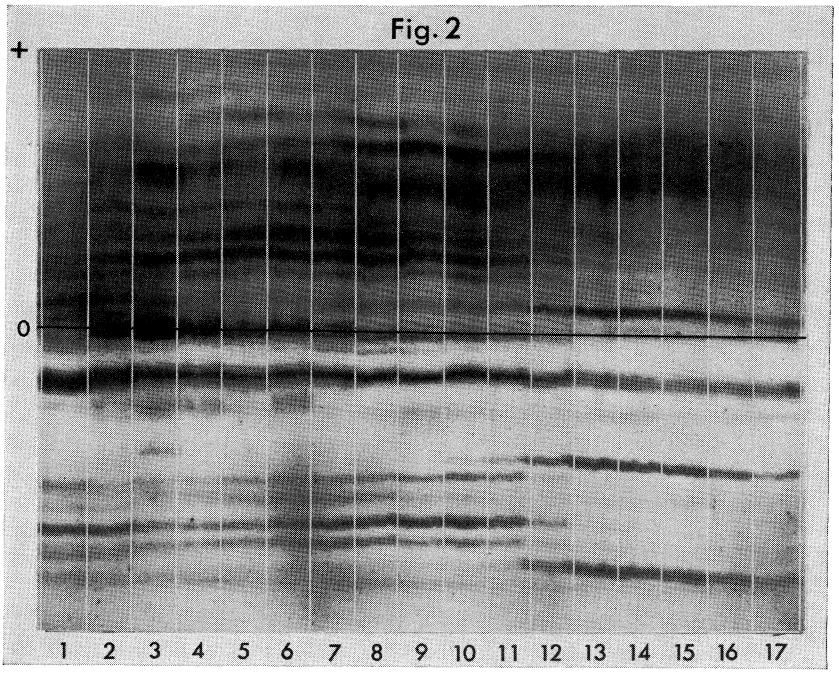


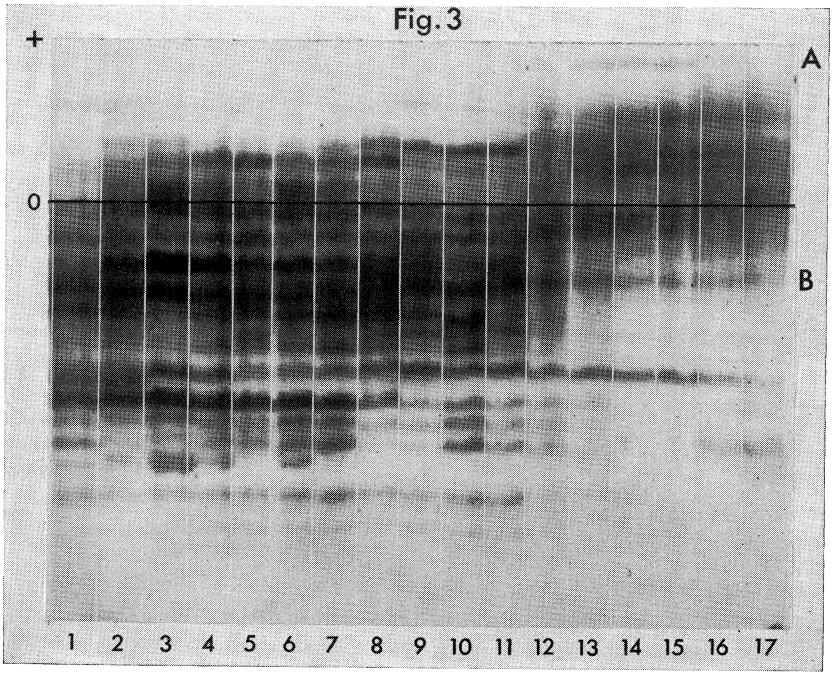
Fig. 1.—Chromatographic separation of 1 g of high-sulphur proteins into 17 fractions on DEAE-cellulose at pH 4·5. Experimental conditions are given in the text. The arrow indicates change to eluent with ionic strength 1·0. Yields of fractions (mg): *1*, 24; *2*, 35; *3*, 38; *4*, 48; *5*, 35; *6*, 46; *7*, 71; *8*, 32; *9*, 29; *10*, 43; *11*, 57; *12*, 78; *13*, 62; *14*, 88; *15*, 88; *16*, 76; *17*, 48. Total recovery, 0·90 g.

Fractions 13–17 show obvious similarities to patterns previously published for the high-sulphur protein fraction SCMKB2 (Gillespie, Haylett, and Lindley 1968) and on this basis some of the peptides can be identified, since it has been found (Gillespie and Darskus, unpublished data) that SCMKB2 is eluted as a broad peak in this region. Tryptic digests run at pH 3·5 show evidence of a very fast anodic component in fractions 3–7. This is almost certainly free *S*-carboxymethylcysteine since it has the correct mobility at pH 3·5, and after dansylation of these tryptic digests free dansyl-*S*-carboxymethylcysteine was detected. This would imply that all these fractions contain proteins having a *C*-terminal sequence of arginyl- (or less probably lysyl-) *S*-carboxymethylcysteine.*

High-voltage electrophoresis at pH 3·5 of the chymotryptic digests of the 17 fractions (Fig. 3) again suggests the possibility of a family of proteins with common sequences, and once more fractions 13–17 show similarities to SCMKB2 (Gillespie, Haylett, and Lindley 1968). The *N*-terminal sequence acetylAla-SCMCys-SCMCys-Ser-Thr-Ser-Phe (band *A*, Fig. 3) is probably common to all, as is the sequence Gly-Gln-Val-Gly-Ser-Gly-Ser-Ala-Ser-Val (band B, Fig. 3). Possibly this latter peptide extends over an even wider range of fractions.

* Lysine is present in only small amounts in these fractions (Table 1).





Figs. 2 and 3.—Electrophoretograms run at pH 6·5 (Fig. 2) and 3·5 (Fig. 3) showing comparative peptide maps of tryptic (Fig. 2) and chymotryptic (Fig. 3) digests of the 17 high-sulphur protein fractions defined in Figure 1. *O*, origin.

The electrophoretogram obtained at pH 6·5 from the thermolysin digests (Fig. 4) seems especially convincing in its demonstration of families of proteins with common sequences. This overall conclusion is further supported by electrophoretograms (not reproduced) of the chymotryptic digest at pH 6·5 and the thermolysin digest at pH 3·5.

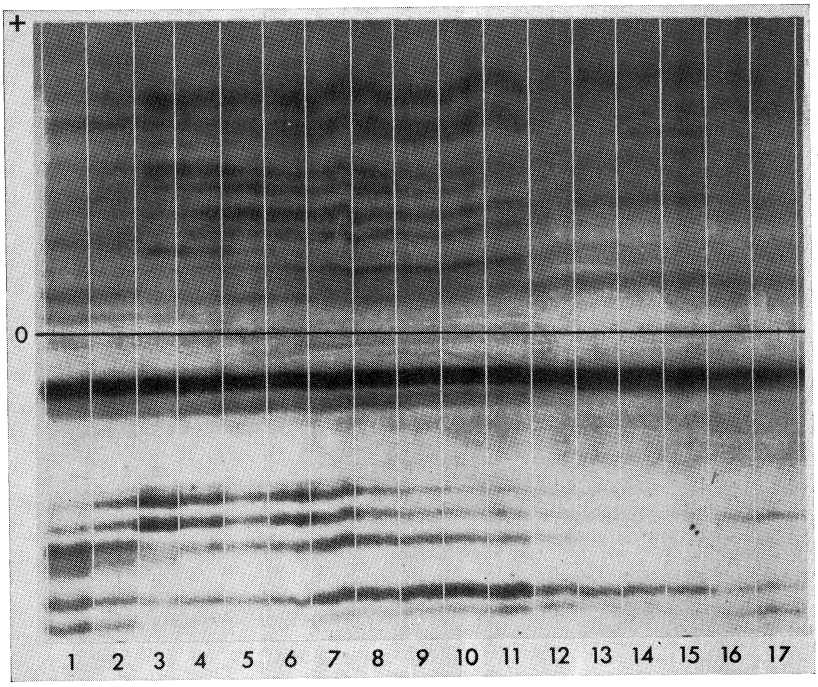


Fig. 4.—Electrophoretogram run at pH 6·5 showing comparative peptide maps of thermolysin digests of the 17 high-sulphur protein fractions defined in Figure 1. *O*, origin.

IV. Discussion

Earlier work (Gillespie 1965) has suggested a positive correlation between sulphur content and molecular weight of high-sulphur protein fractions. A re-examination of the data of Joubert and Burns (1967) by Lindley, Gillespie, and Rowlands (unpublished data) has shown this relationship to be significant at a probability level of greater than 99·9%. Hence there is probably an increase both in charge and molecular weight with increasing fraction number in Figure 1.

In the interpretation of the electrophoretic data reproduced in Figures 2–4 it should be remembered that identity of mobility in a one-dimensional electrophoretogram is not unequivocal proof of the chemical identity of two peptides. Nevertheless the fact that the protein fractions can be arranged in families each of which has its own characteristic peptide pattern for each of their different enzyme digestions suggests very strongly that common structural features persist throughout these

families. The present data would suggest the occurrence of three families comprising fractions 1–2, 3–11, and 12–17, but the family comprising fractions 3–11 can probably be subdivided into two very closely related groups, fractions 3–6 and 7–11. Joubert and Burns (1967) have also grouped similar fractions on the basis of their amino acid compositions. Their results, as well as those in Table 1, show a marked change in the tyrosine:phenylalanine ratio after fractions 2 and 11.

If substantial lengths of amino acid sequence are in fact common to a number of high-sulphur protein fractions, this could be of some importance. Most examples of families of proteins relate to molecules in which single amino acid replacements or deletions have occurred, and perhaps loss of a segment with three or four residues (Eck and Dayhoff 1966). In the case of high-sulphur proteins from wool, however, the previously mentioned correlation between sulphur content and molecular weight suggests that this familial relationship covers an unusually large molecular weight range. The exact nature of this heterogeneity and the way in which it arises could therefore be of considerable interest outside the field of keratin chemistry.

There is a considerable body of evidence against any suggestion that this heterogeneity is an artefact of the method of preparation. Amongst the possibilities which have been suggested, peptide bond fission during extraction at pH 10·2 (Corfield 1962) can be ruled out because all the proteins have acetylated *N*-termini and no evidence has been found for the formation of significant amounts of fission peptides from the *C*-termini during reductive or oxidative extraction procedures (Haylett *et al.* 1963; Gillespie 1964). Non-specific alkylation with iodoacetate as a major contributor of heterogeneity (Gillespie *et al.* 1960) is unlikely because lysine and histidine are present in only small amounts, and no *N*-carboxymethylamino acids have been detected in acid hydrolysates of these proteins. The absence of methionine further limits the possibility of non-specific carboxymethylation.

We have now shown by tritium exchange procedures (in collaboration with S. J. Leach and L. A. Holt) that racemized residues as another possible source of heterogeneity probably are only present to the extent of 1 residue in 2000 and certainly no more frequently than 1 in 100. The positive finding that typical high-sulphur proteins can be extracted from the wool follicle with no more drastic a reagent than 8M urea at pH 7 suggests that to a large extent these proteins are naturally heterogeneous (Downes *et al.* 1966).

V. References

Corfield, M. C. (1962).—*Biochem. J.* **84**, 602.

Crewther, W. G., Fraser, R. D. B., Lennox, F. G., and Lindley, H. (1965).—*Adv. Protein Chem.* **20**, 191.

Downes, A. M., Sharry, L. F., and Rogers, G. E. (1963).—*Nature, Lond.* **199**, 1059.

Downes, A. M., Ferguson, K. A., Gillespie, J. M., and Harrap, B. S. (1966).—*Aust. J. biol. Sci.* **19**, 319.

Eck, R. V., and Dayhoff, M. O. (1966).—"Atlas of Protein Sequence and Structure." (National Biomedical Research Foundation: Silver Springs.)

Gillespie, J. M. (1962).—*Aust. J. biol. Sci.* **15**, 572.

Gillespie, J. M. (1964).—*Aust. J. biol. Sci.* **17**, 282.

GILLESPIE, J. M. (1965).—In "Biology of the Skin and Hair Growth". (Eds. A. G. Lyne and B. F. Short.) p. 377. (Angus & Robertson Ltd.: Sydney.)

GILLESPIE, J. M., HAYLETT, T., and LINDLEY, H. (1968).—*Biochem. J.* **110**, 193.

GILLESPIE, J. M., O'DONNELL, I. J., THOMPSON, E. O. P., and WOODS, E. F. (1960).—*J. Text. Inst.* **51**, T703.

GODDARD, D. R., and MICHAELIS, L. (1935).—*J. biol. Chem.* **112**, 361.

GRAY, W. R. (1967).—In "Methods in Enzymology". (Eds. S. P. Colowick and N. O. Kaplan.) Vol. 11. p. 139. (Academic Press, Inc.: New York.)

HAYLETT, T., JOUBERT, F. J., SWART, L. S., and LOUW, D. F. (1963).—*Text. Res. J.* **33**, 639.

JOUBERT, F. J., and BURNS, M. A. C. (1967).—*Jl S. Afr. chem. Inst.* **20**, 161.

MERCER, E. H. (1961).—"Keratin and Keratinization: an Essay in Molecular Biology." p. 219. (Pergamon Press: Oxford.)