The primary structure of component 8c-1, a subunit protein of intermediate filaments in wool keratin

Relationships with proteins from other intermediate filaments

Lyndsay M. DOWLING,* W. Gordon CREWETHER and Adam S. INGLIS

C.S.I.R.O. Division of Protein Chemistry, 343 Royal Parade, Parkville, Vic. 3052, Australia

Component 8c-1, one of four highly homologous component-8 subunit proteins present in the microfibrils of wool, was isolated as its *S*-carboxymethyl derivative and its amino acid sequence was determined. Large peptides were isolated after cleaving the protein chemically or enzymically and the sequence of each was determined with an automatic Sequenator. The peptides were ordered by sequence overlaps and, in some instances, by homology with known sequences from other component-8 subunits. The *C*-terminal residues were identified by three procedures. Full details of the various procedures used have been deposited as Supplementary Publication SUP 50133 (4 pp.) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in Biochem. J. (1986) 233, 5. The result showed that the protein comprises 412 residues and has an M_r , including the *N*-terminal acetyl group, of 48 300. The sequence of residues 98–200 of component 8c-1 was found to correspond to the partial or complete sequences of four homologous type I helical segments previously isolated from helical fragments recovered from chymotryptic digests of microfibrillar proteins of wool [Crewther & Dowling (1971) Appl. Polym. Symp. 18, 1–20; Crewther, Gough, Inglis & McKern (1978) Text. Res. J. 48, 160–162; Gough, Inglis & Crewther (1978) Biochem. J. 173, 373–385]. Considered in relation to amino acid sequences of other intermediate-filament proteins, the sequence is in accord with the view that keratin filament proteins are of two types [Hanukoglu & Fuchs (1983) Cell (Cambridge, Mass.) 33, 915–924]. Filament proteins from non-keratinous tissues, such as desmin, vimentin, neurofilament proteins and the glial fibrillary acidic protein, which form monocomponent filaments, constitute a third type. It is suggested that as a whole the proteins from intermediate filaments be classed as filamentins, the three types at present identified forming subgroups of this class. The significant homologies between types I, II and III occur almost exclusively in segments of the chain that have been identified as having a coiled-coil structure together with the relatively short sections connecting these segments. The non-coiled-coil segments at the *C*- and *N*-termini show no significant homology between types, nor is homology in these segments apparent in all members of one type. Component 8c-1 does not show homology in its terminal segments with the known sequence of any other filamentin. There is evidence that the portion of the gene coding for the 14 heptads comprising the type I and type II helical segments arose by the duplication of DNA coding for 49 amino acid residues. The portion of the gene coding for seven heptads adjacent to the non-helical *C*-terminal segment of component 8c-1 appears to have originated from a similar segment of DNA. The amino acid sequence of component 8c-1 is subjected to detailed structural analysis in the following paper [Dowling, Parry & Crewther (1986) Biochem. J. 236, 705–712].

INTRODUCTION

Cortical cells from wool and other hard α -keratins contain microfibrils set in a non-filamentous matrix (see Fraser *et al.*, 1972). The many constituent proteins of these structures are highly cross-linked by disulphide bonds. After reduction of these bonds in aqueous denaturing solutions and *S*-carboxymethylation of cysteine residues so formed, up to 85% of the keratin can be extracted. The extracts from wool contain three classes of protein (see Crewther, 1976): high-sulphur and high-tyrosine proteins derived from the matrix, and low-sulphur proteins from the microfibrils (Jones, 1976). The primary structures of 18 high-sulphur proteins (see

Swart *et al.*, 1976) and two high-tyrosine proteins (Dopheide, 1973; Marshall *et al.*, 1980) are known.

The low-sulphur proteins, which range in M_r from 45 000 to 58 000 (Woods, 1979), comprise two families, each consisting of four homologous subunits: components 5, 7a, 7b and 7c, and components 8a, 8b, 8c-1 and 8c-2 (Crewther *et al.*, 1980a). Earlier attempts to obtain amino acid sequence data from a fraction containing low-sulphur proteins (Corfield *et al.*, 1967; Corfield & Fletcher, 1969) and from a fraction enriched in component 8 (Thompson & O'Donnell, 1967; Hosken *et al.*, 1968; O'Donnell, 1969) provided some short segments of sequence and demonstrated heterogeneity in the preparations used. Complete amino acid sequences have been determined for

Abbreviation used: h.v.p.e., high-voltage paper electrophoresis.

* To whom reprint requests should be addressed.

representatives of two types of helical segments [type I, 103 residues (Gough *et al.*, 1978); type II, 109 residues (Crewther *et al.*, 1978b)] isolated from partial chymotryptic digests of the low-sulphur fraction (Crewther & Dowling, 1971). Partial sequence data have also been obtained from the *N*-termini of other homologous helical segments (Crewther *et al.*, 1978a).

The present study reports the complete amino acid sequence of component 8c-1. The amino acid sequence of the *N*-terminal peptide obtained by cleaving the preparation of component 8c-1 used in the present work with CNBr has already been reported by Gough & Inglis (1980), who also showed that the *N*-terminus is acetylated.

EXPERIMENTAL AND RESULTS

Fractionation of proteins and large peptides

Chromatography of proteins or peptides derived from them was performed in urea buffer containing 8 M-urea (to prevent association), 10 mM-Tris, 50 mM-ethanolamine (to remove CNO⁻ ions) and 1 mM-EDTA at either pH 7.4 or pH 8.5. Fractionations were monitored by polyacrylamide-gel electrophoresis in buffers containing 8 M-urea (Sparrow & Crewther, 1972) or SDS (Laemmli, 1970) for gels of uniform concentration and SDS (Weber & Osborn, 1969) for gradient gels.

Preparation of component 8c-1

The low-sulphur protein fraction prepared from an extract of *S*-carboxymethylated Merino wool by the method of Dowling & Crewther (1974) was fractionated on DEAE-cellulose (Whatman DE-52) with a linear KCl gradient (Crewther *et al.*, 1976). Fractions containing component 8c-1 as the sole subunit of *M_r* approx. 50000, as indicated by gel electrophoresis in SDS, were pooled and contaminants of high *M_r* were removed by gel filtration on a column of Sephadex G-200 as described by Thompson & O'Donnell (1965). The preparation was dialysed against deionized water and freeze-dried.

Analytical methods

For the determination of amino acid compositions, samples were hydrolysed *in vacuo* for 24 h at 108 °C in constant-boiling HCl containing 10 mM-phenol, and the hydrolysates were analysed on a modified Beckman analyser. For the determination of tryptophan the hydrolysis was performed *in vacuo* for 48 h at 115 °C in 4 M-methanesulphonic acid containing tryptamine (Simpson *et al.*, 1976). An automatic Sequenator of the Edman-Begg type with a metal reaction cup and a modified extraction procedure was used for amino acid sequence analysis (Inglis *et al.*, 1981; Inglis, 1983).

Amino acid thiazolinones from the Sequenator were converted into phenylthiohydantoin derivatives in 25% (v/v) trifluoroacetic acid at 60 °C for 15 min. This ensured good yields of the phenylthiohydantoin derivative of *S*-carboxymethylcysteine. The phenylthiohydantoin derivatives were identified by h.p.l.c. with a Zorbax ODS column (du Pont) at 40 °C, with elution at 0.9 ml/min with a gradient formed by mixing 20 mM-ammonium acetate, pH 4.8, with acetonitrile. The acetonitrile concentration was increased linearly from 15% to 40% (v/v) in 11 min, held constant for 3 min, increased to 50% (v/v) in 2 min, and then to 99% (v/v) in a further 2 min.

With this procedure the phenylthiohydantoin derivative of *S*-carboxymethylcysteine is clearly resolved from the phenylthiohydantoin derivatives of aspartic acid and glutamic acid.

The strategy for sequence determination involved the isolation of large peptide fragments for analysis with the Sequenator.

Cleavage of component 8c-1 and recovery of peptides

(1) **At methionine.** Component 8c-1 was cleaved by CNBr at the sole methionine residue. The cleavage products were separated by gel-filtration chromatography to give the peptides CNBr-1 and CNBr-3 (see Thompson & O'Donnell, 1967), shown schematically in Fig. 1(a). The intermediate-sized fraction, CNBr-2, observed by the latter authors, and derived from homologous components, was virtually absent. The large peptide (CNBr-1) was found to have no free terminal α -amino group and, as the *N*-terminal amino acid residue is known to be glutamine (Thompson & O'Donnell, 1967), this was presumed to have cyclized to a pyroglutamic acid residue during preparation. Accordingly, peptide CNBr-1 was cleaved to yield smaller peptides suitable for analysis with the Sequenator.

(2) **At tryptophan.** The dimethyl sulphoxide/HBr procedure of Savage & Fontana (1977) was used. The large fragments Trp-1 and Trp-2 were separated from each other and from uncleaved component 8c-1 by gel-filtration chromatography. Their positions and the extent of the sequences obtained from them are shown schematically in Fig. 1(b).

(3) **At hydrophobic residues.** Component 8c-1 was partially digested with chymotrypsin under conditions identical with those used for the production of helical fragments from the unfractionated low-sulphur proteins (Crewther & Dowling, 1971), except in the use of a digestion time of only 10 min. After separation into two fractions on the basis of solubility in 0.5 M-KCl at pH 4.0, the complex mixtures of peptides were subjected to further fractionation by gel-filtration chromatography followed by ion-exchange chromatography. Large peptides were further purified by gel-filtration chromatography, whereas small peptides were recovered by preparative h.v.p.e. The peptides used in elucidating the sequence were labelled Ch-1 to Ch-7 according to their positions along the protein chain (shown schematically in Fig. 1c). Their amino acid compositions are recorded in Table 1.

Cleavage of peptide CNBr-1 and recovery of peptides

(1) **At lysine.** Peptide CNBr-1 was cleaved with the lysine-specific proteinase from *Armillaria mellea* under the conditions described by Doonan *et al.* (1975). After an initial fractionation on the basis of solubility at pH 4.0, the soluble peptides were further fractionated by preparative h.v.p.e. and the insoluble peptides by gel-filtration chromatography. The peptides were labelled AP-1 to AP-8 according to their positions along the protein chain (Fig. 1d), and their amino acid compositions are recorded in Table 2.

(2) **At lysine and arginine.** Peptide CNBr-1 was cleaved with trypsin[1-chloro-4-phenyl-3-tosylamidobutan-2-one

Fig. 1. Location in component 8c-1 of the major peptides used in determining its amino acid sequence

Peptides were obtained by cleavage of component 8c-1 with CNBr (a), dimethyl sulphoxide/HBr (b), chymotrypsin (c) and *A. mellea* proteinase (d), by cleavage of *S*-carboxymethylated wool protein SCMK-A with chymotrypsin at 15 °C (e), and by cleavage of peptide CNBr-1 with *S. aureus* V8 proteinase (f). The positions of the methionine and tryptophan residues are indicated by arrows. Sequences determined on peptides from component 8c-1 are represented by solid black bars. Sequences from a homologous peptide or a preparation containing homologous peptides (see the text) are represented by hatching. Peptides of uncertain length are represented by broken lines.

Table 1. Amino acid compositions of component 8c-1 and peptides obtained from it by partial chymotryptic digestion

The compositions predicted from sequence data are given in parentheses. Abbreviation: Cm, carboxymethyl.

Amino acid	Composition (residues/molecule)												
	Ch-1 Residues 12–45		Ch-2 Residues 62–106		Ch-4 Residues 246–306		Ch-5 Residues 307–355		Ch-6 Residues 356–375		Ch-7 Residues 376–410		8c-1 Residues 1–412
Lys	0.1	(–)	1.0	(1)	1.0	(1)	0.4	(–)	0.9	(1)	1.1	(1)	11.9 (12)
His					0.7	(1)	0.2	(–)					2.5 (2)
Arg	1.8	(2)	4.9	(5)	4.7	(5)	4.4	(4)	1.1	(1)	2.1	(2)	34.2 (32)
Trp	N.D.	(1)	N.D.	(–)	N.D.	(–)	N.D.	(–)	N.D.	(–)	N.D.	(–)	1.6 (2)
CmCys	5.2	(6)	0.9	(1)	0.9	(1)	1.7	(2)	2.8	(3)	6.0	(7)	23.9 (25)
Asp	2.3	(3)	4.0	(4)	5.1	(5)	5.9	(6)	3.8	(4)	1.3	(1)	46.6 (46)
Thr	2.0	(2)	0.3	(–)	4.3	(6)	1.5	(1)	1.8	(2)	2.7	(3)	21.8 (23)
Ser	5.6	(7)	3.2	(4)	4.1	(5)	3.9	(4)	1.0	(1)	3.7	(4)	28.8 (34)
Glu	0.1	(–)	11.6	(12)	14.9	(17)	14.0	(13)	1.3	(1)	0.2	(–)	74.6 (74)
Pro	4.5	(4)	1.9	(2)	< 0.4	(–)	< 0.3	(–)	2.3	(2)	6.3	(7)	16.5 (17)
Gly	2.8	(3)	0.3	(–)	0.5	(–)	0.9	(1)	1.1	(1)	2.1	(2)	15.2 (12)
Ala	2.3	(2)	2.0	(2)	3.2	(3)	2.4	(2)	1.1	(1)	3.6	(4)	23.9 (24)
Val	1.8	(2)	2.1	(2)	3.4	(4)	4.2	(4)	0.2	(–)	1.0	(1)	24.7 (25)
Met													1.2 (1)
Ile	1.1	(1)	1.1	(1)	1.9	(3)	2.8	(3)	0.2	(–)	1.6	(2)	15.2 (16)
Leu	0.9	(1)	6.6	(7)	7.3	(8)	7.1	(7)	2.8	(3)	0.1	(–)	50.7 (49)
Tyr			2.7	(3)	1.7	(2)	2.0	(2)			0.9	(1)	11.5 (10)
Phe			0.6	(1)	0.2	(–)	0.1	(–)					8.7 (8)

(‘TPCK’)-treated] and the digest fractionated by gel-filtration chromatography on four coupled Sephadex columns (two of Sephadex G-50, two of Sephadex G-25) as described previously (Gough & Inglis, 1980). Further fractionation was accomplished by preparative h.v.p.e. Four peptides, numbered CNBr-1,T1 to CNBr-1,T4 according to their positions in the protein chain, were purified and their sequences determined.

(3) **At glutamic acid.** Peptide CNBr-1 was digested with the proteinase of *Staphylococcus aureus* (V8 strain) as described by Drapeau *et al.* (1972). A digestion time of only 1 h was used in order to obtain partial cleavages. The products were subjected to gel-filtration chromatography, and a large peptide, CNBr-1,S, was isolated that required no further purification. The *N*-terminal 23 residues of this peptide were identified (Fig. 1f).

Table 2. Amino acid compositions of peptides from an *A. mellea* proteinase digest of peptide CNBr-1

The compositions predicted from the sequence data are given in parentheses. Abbreviation: Cm, carboxymethyl.

Amino acid	Composition (residues/molecule)													
	AP-1 Residues 72–114		AP-2 Residues 115–119		AP-3 Residues 134–142		AP-4 Residues 143–171		AP-6 Residues 190–254		AP-7 Residues 255–378		AP-8 Residues 379–412	
Lys	1.4	(1)	1.0	(1)	1.0	(1)	1.0	(1)	1.4	(1)	2.6	(2)	0.7	(1)
His									1.0	(1)	1.0	(1)		
Arg	4.6	(5)			1.0	(1)	2.9	(3)	6.4	(7)	8.9	(9)	2.9	(3)
Trp									0.9	(1)	0.1	(–)		
CmCys	1.1	(1)			0.7	(1)	1.0	(1)	0.9	(–)	6.0	(7)	5.2	(6)
Asp	3.0	(2)			2.0	(2)	3.1	(3)	10.4	(11)	13.2	(14)	1.4	(1)
Thr	1.3	(1)			1.0	(1)	1.8	(2)	4.1	(5)	6.5	(8)	2.7	(3)
Ser	2.3	(3)			0.3	(–)	0.9	(1)	2.3	(2)	7.9	(10)	4.2	(4)
Glu	12.6	(15)					4.9	(5)	13.5	(14)	24.3	(28)	0.3	(–)
Pro	1.5	(2)					0.2	(–)	1.8	(1)	3.0	(2)	7.7	(7)
Gly	0.7	(–)	0.3	(–)			1.6	(2)	2.0	(1)	4.0	(3)	1.4	(1)
Ala	1.4	(1)	0.8	(1)	1.8	(2)	0.4	(–)	4.0	(4)	6.9	(7)	3.2	(3)
Val	2.0	(2)					1.3	(1)	6.1	(7)	7.7	(8)	2.0	(2)
Met														
Ile	2.1	(2)	1.0	(1)			1.7	(2)	1.3	(1)	4.5	(5)	1.7	(2)
Leu	6.3	(5)	1.2	(1)	1.1	(1)	6.7	(7)	7.5	(7)	15.9	(17)	0.2	(–)
Tyr	1.8	(2)					1.0	(1)	2.0	(2)	3.2	(3)	1.0	(1)
Phe	0.9	(1)			1.1	(1)	0.1	(–)	0.7	(–)	0.9	(–)	0.2	(–)

Further digestion of Ch- and AP-peptides

In order to complete sequences or to obtain overlapping sequences, several peptides were digested further with trypsin, *S. aureus* proteinase or chymotrypsin.

Peptides isolated from the digests by preparative h.v.p.e. were given the appropriate Ch- or AP-designation followed by T, S or C to indicate the enzyme used, and, if appropriate, were numbered to indicate the position in the parent peptide.

C-Terminal analysis of peptide CNBr-1

(1) **C-Terminal amino acid.** Peptide CNBr-1 was selectively tritiated by the method of Matsuo *et al.* (1966). The tritiated peptide was hydrolysed, and constituent amino acids were separated on paper by two-dimensional electrophoresis/chromatography. The amino acids were located with ninhydrin, then excised, and their radioactivities were measured by liquid-scintillation counting. Arginine was identified as the C-terminal amino acid.

(2) **C-Terminal peptides.** The 'on-diagonal' paper-electrophoresis method of Duggleby & Kaplan (1975) was used to identify C-terminal peptides. The major 'on-diagonal' peptide from both thermolytic and chymotryptic digests of peptide CNBr-1 that had been amidinated with ethanolamine contained valine, arginine and ethanolamine in approximately equal proportions.

(3) **Carboxypeptidase digestion.** Peptide CNBr-1 was digested with a mixture of carboxypeptidases A and B. The amino acids most readily released were arginine, valine and tyrosine. A time-course study of amino acid release indicated a probable C-terminal sequence of -Tyr-Val-Arg.

Supplementary information

Details of procedures used in determining the sequence of component 8c-1 have been deposited as Supplementary Publication SUP 50133 at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K. The information comprises: (1) details of chemical and enzymic methods used for cleavage of component 8c-1, peptide CNBr-1 and other peptides; (2) details of procedures used for the fractionation and purification of peptides from (1); (3) Figures showing elution profiles obtained from gel-filtration and ion-exchange chromatographic fractionations of peptides; (4) details of the methods used to determine the C-terminal sequence of peptide CNBr-1 and a Figure showing the results of the time course of digestion of peptide CNBr-1 with a mixture of carboxypeptidases A and B.

Reconstruction of the complete amino acid sequence

The complete amino acid sequence of component 8c-1 (Fig. 2) was assembled by using data obtained in the present study together with data from earlier studies in this laboratory. The latter data, which are included in the schematic representation of the relative positions of peptides (Fig. 1), include the following.

(i) The sequence of peptide CNBr-3 (residues 1–59) determined by Gough & Inglis (1980). In the present work the sequences of residues 12–29 and 46–59 were confirmed by the N-terminal sequences of peptides Ch-1 and Trp-1 respectively.

(ii) The N-terminal sequence of peptide CNBr-1 (residues 60–89) prepared from a mixture of component-8 subunits (Thompson & O'Donnell, 1967). This sequence confirmed the positioning of peptide Ch-2, as also did data of Corfield & Fletcher (1969), which corresponded to residues 57–76.

(iii) The complete sequence (residues 98–200) of a type

Fig. 2. Complete amino acid sequence of component 8c-1

The designation systems for the peptides are described in the text. Primary peptides formed by cleaving component 8c-1 or peptide CNBr-1 are indicated by solid lines. Secondary peptides formed by cleavage of primary peptides are indicated by dotted lines. These lines are continuous for identified amino acid sequences and broken where sequences were not obtained. The positions of C-termini for peptides Ch-3, AP-5 and CNBr-1,S are uncertain.

I helical segment (Gough *et al.*, 1978) obtained from a partial chymotryptic digest of the unfractionated low-sulphur proteins (Crewther *et al.*, 1978b). By homology with this sequence peptides accounting for all residues between 98 and 200, except for residues 120–125, were

ordered unequivocally. Differences between the sequences of these peptides and that of the type I helical segment (see the Discussion section) indicated that the helical segment was derived from a subunit other than component 8c-1 (see iv).

(iv) The *N*-terminal sequence (residues 98–129) of a second type I helical segment [only the first 27 residues were reported by Crewther *et al.* (1978a), but 32 residues were in fact identified]. Since the sequences identified for residues 98–119 and 126–129 were identical with the corresponding sequences in this helical segment, it was assumed that the segment is derived from component 8c-1. Residues 120–125 were assigned on this assumption.

(v) The *N*-terminal sequence (residues 268–295) of a peptide isolated from a low-temperature (15 °C) partial chymotryptic digest of the unfractionated low-sulphur proteins (L. M. Dowling & W. G. Crewther, unpublished work). This sequence was homologous with residues 268–295 and allowed the unequivocal positioning of peptide AP-7,S2. Substitutions of alanine, leucine, glutamic acid, glutamine, leucine and aspartic acid at positions 269, 275, 282, 290, 292 and 294 respectively indicated that the peptide was derived from a subunit other than component 8c-1. Since *S*-carboxymethyl-cysteine is located at position 267 in the sequence of component 8c-1, chymotrypsin would not be expected to cleave at this point. In at least one component-8 subunit, however, tyrosine occurs at this position (Woods & Inglis, 1984).

It is seen from Fig. 2 that with one exception all the other peptides were aligned by overlaps of sequence of at least three residues. Peptide AP-7,S4 overlapped the identical peptides CNBr-1,T3 and Ch-4,T1 by only two residues (304–305), but the amino acid composition of another peptide (Ch-4,T2) was the same as that of peptide Ch-4,T1 with an additional tyrosine residue corresponding to the *C*-terminus of peptide Ch-4. Evidence from the *C*-terminal analyses described above confirmed the *C*-terminal sequence of component 8c-1 as -Tyr-Val-Arg.

DISCUSSION

General aspects of the analysis

The amino acid composition of component 8c-1 (Table 1) is in good agreement with the composition calculated from the sequence. Also, the value of 48300 for the M_r of component 8c-1, calculated from the amino acid sequence of 412 residues, is almost identical with the value of 48000 obtained by Woods (1979) using ultracentrifugation.

In an earlier study on the sequences of peptides from component 8c-1 (Crewther *et al.*, 1980b), t.l.c. was used for identifying phenylthiohydantoin derivatives of amino acids. In the present study the use of h.p.l.c. in Sequenator analysis and additional checks involving hydrolysis of amino acid phenylthiohydantoin derivatives and amino acid analyses have shown that some of these preliminary sequences require amendment. These changes have been confirmed in most instances by overlapping sequences or sequence analysis of small peptides.

Although 230 residues were successfully identified by isolating large peptides from partial digests with chymotrypsin, this approach is appropriate only if adequate supplies of purified protein are available.

Digestion of the large peptide CNBr-1 with the lysine-specific proteinase from *A. mellea* was particularly useful, in providing eight segments accounting for most of the sequence. Lysine was identified as the *N*-terminal residue in all of these peptides, so confirming the specificity claimed by Walton *et al.* (1972).

Sequences of a total of 21 tryptic and chymotryptic peptides isolated in an early study on a fraction US3 from oxidized wool proteins (Corfield *et al.*, 1967; Corfield & Fletcher, 1969) can be matched with sequences in component 8c-1 totalling 70 residues. Other peptides from fraction US3 were homologous with sequences in component 8c-1. Others again were identical or homologous with sequences of component 7c (Sparrow & Inglis, 1980). The fraction US3 thus contained a mixture of proteins from the low-sulphur class, and the attempt by Corfield *et al.* (1967) to estimate an M_r from the peptide composition was not valid.

The type I helical segments in the component-8 family of microfibrillar proteins that correspond with the 14-heptad coiled-coil segment 1B [see Fig. 1 in Dowling *et al.* (1986)] probably have a major part to play in structuring the four-chain unit on which the microfibril is based (Ahmadi *et al.*, 1980; Crewther *et al.*, 1983; Woods & Inglis, 1984). The completion of the amino acid sequence of component 8c-1 makes possible a comparison of two complete type I sequences, each of 103 residues. The sequence of component 8c-1 was found to differ from that of the type I helical segment in 17 residues. The most important changes appear to be those resulting in the net gain of one anionic group, the gain of two cationic groups and the rearrangement of the charge pattern. Such changes would be expected to influence significantly the affinity between different type I and type II (from component 7) helical segments and as a consequence probably determine the pairing of individual subunits of component 8 with subunits of component 7 in the microfibril.

Structural similarities in component 8c-1 and proteins from intermediate filaments: classification and nomenclature

In the following paper (Dowling *et al.*, 1986) the amino acid sequence of component 8c-1 has been subjected to computational analysis to identify potential structural features. The analysis indicates that the chain, like those from other intermediate filaments, takes the form of a rod consisting of coiled-coil segments comprising approximately five, 14 or 15, two or three and 17 heptads of sequence. These segments, following the general notation of Geisler & Weber (1982), are designated segments 1A, 1B, 2A and 2B respectively [see Fig. 1 in Dowling *et al.* (1986)]. Large non-coiled-coil segments of chain are located at each end of the rod, and smaller such segments between the coiled-coil segments. The homology of the helical regions with the sequence of helical regions in intermediate filaments (Dowling *et al.*, 1983) and the significance of these homologies in increasing our understanding of the structure of all filaments in this class (Crewther *et al.*, 1983) have been discussed elsewhere.

Hanukoglu & Fuchs (1983) have divided the proteins comprising intermediate filaments from keratins into two classes according to their sequence homology with type I and type II helical segments excised from the microfibrillar proteins of wool (Crewther & Dowling, 1971; Gough *et al.*, 1978; Crewther *et al.*, 1978a,b). Undoubtedly, in terms of sequence homology, not only in segment 1B but also in segments 1A, 2A and 2B in the coiled-coil regions, component 8c-1 from wool keratin, the 50000- M_r filament protein from human epidermis (Hanukoglu & Fuchs, 1982) and the 59000- M_r filament protein from mouse epidermis (Steinert *et al.*, 1983) form a subgroup, designated type I. On the same basis,

component 7c from wool keratin (Sparrow & Inglis, 1980) the 56000- M_r filament protein from human epidermis (Hanukoglu & Fuchs, 1983) and the 60000- M_r filament protein from mouse epidermis (Steinert *et al.*, 1984a) form another subgroup, designated type II. It is equally clear from the amino acid sequences of proteins constituting intermediate filaments in non-keratinous tissues that these form a third subgroup, designated type III (see Steinert *et al.*, 1984b). This subgroup includes desmin (Geisler & Weber, 1982), vimentin (Geisler & Weber, 1981; Quax-Jeuken *et al.*, 1983), glial fibrillary acidic protein (Geisler & Weber, 1983; Lewis *et al.*, 1984) and the 68000- M_r neurofilament protein (Geisler *et al.*, 1982, 1983).

Within these subgroups differences in the degree of homology in segment 1B are apparent, which appear to relate chiefly to the tissue of origin of the filament proteins. Thus there is much greater homology between sequences of segment 1B for different epidermal proteins of one type ($> 80\%$ identity) than between corresponding segments of an epidermal protein and a protein of the same type derived from wool (type I-type I, 60–70% identity; type II-type II, slightly less than 50% identity).

The proteins comprising these three subgroups all have three important features in common: they are major constituents of intermediate-sized filaments (7–12 nm diameter), they exhibit significant sequence homology, one with the other, over a section amounting to about 300 residues, and their predicted secondary structures show an identical arrangement of coiled-coil segments. On this basis it is proposed that, as a group, these proteins be classified as filamentins. The use of this general term for proteins from intermediate filaments should obviate the undesirable use of the term 'keratin' for individual proteins derived from keratins. Keratins such as wool, hair, horn, epidermis, feather or reptile scales are complex structures containing a wide range of apparently unrelated subunit proteins quite as characteristic of keratins as those comprising the filamentous structures in the individual cells of α -keratins. It would avoid confusion if the term 'keratin' retained its original meaning and applied only to tissues.

Internal homology in component 8c-1 and other filamentins

A statistical examination of the sequences of the type I and type II helical segments derived from components 8 and 7 respectively has shown (Elleman *et al.*, 1978) a strong 49-residue repeat ($P = 6 \times 10^{-5}$) in the type II sequence. This was significant at the 0.05 level even when account was taken of the heptad repeat in the sequence. Although the homology between the two corresponding portions of the type I segment was not of comparable significance, the homology between the 49 residues at the C-terminus of type I and the N-terminal portion of type II was even more significant. It was concluded that during the evolution of a precursor gene for both type I and type II material in hard α -keratins duplication of a segment coding for 49 amino acid residues may have occurred. Alternatively, a requirement for 49-residue repeating units in the micro-unit structure of the microfibril may have given rise to the repeat sequence. Examination of the type I segment (segment 1B) in component 8c-1 disclosed that residues repeating at an interval of 49 residues were located at positions identical with those showing a similar repeat (18.4% identity, $P = 0.025$) in the first type I sequence to be determined (Gough *et al.*, 1978; Elleman

et al., 1978). This repeat is not significant when account is taken of the heptapeptide repeat.

In order to examine the corresponding sequences of other filamentins for similar repeats, the most commonly occurring residue was identified for each corresponding position in the sequences of a range of filamentins (three of type I, three of type II and three of type III). The normalized amino acid sequence obtained in this way was then examined for a 49-residue repeat. Homology was, in fact, apparent (Fig. 3) between the two halves of segment 1B (41% identity), and similar homology between the C-terminal half of segment 1B and the C-terminal section of segment 2B (29% identity). In assessing homology at positions in the normalized sequence with multiple residues (Fig. 3) identity with any one of the residues was counted. If positions with multiple residues were not counted the homologies in the remaining positions were 28% and 26% respectively. As with the repeat in the type II segment, these data may indicate a doubling of segments in a precursor gene or they may result from a structural requirement for particular residues in structural repeats of 49 residues.

The normalized sequence for coiled-coil regions of all nine filamentins (Fig. 3) is highly homologous with that for the corresponding regions of the type III filamentin desmin (76% identity). As would be expected, the normalized sequences for type III and for all filamentins show similar homology. This suggests that the ancestral gene (see Krieg *et al.*, 1985) from which the filamentins developed coded for a protein with type III coiled-coil regions somewhat similar to those of desmin. This view is strengthened by the greater internal homology between the 49-residue pseudo-repeats in type III as compared with type I or type II filamentins.

Sequence homology in the terminal segments of component 8c-1 and other filamentins

Comparison of the amino acid sequences of the non-coiled-coil regions of component 8c-1 with the corresponding regions of other filamentins gives rise to four unexpected results: (1) the terminal sequences of component 8c-1 show no significant homology with those of other filamentins; (2) The degree of identity between terminal sequences of component 8c-1 and those of other type I filamentins (mouse epidermal 59000- M_r filamentin, human epidermal 50000- M_r filamentin) are no greater than that observed when comparing component 8c-1 with type II filamentins (component 7c from wool, mouse epidermal 60000- M_r filamentin, human epidermal 56000- M_r filamentin); (3) the N-terminal and C-terminal segments of component 8c-1, however, show significant homology (21% identity) one with the other; in this respect component 8c-1 resembles mouse epidermal 59000- M_r filamentin (Steinert *et al.*, 1983); (4) the three-residue pseudo-repeat (Pro-Cys-Xaa) found in the terminal segments of component 8c-1 (Dowling *et al.*, 1986) is totally unrelated to the four-, five-, nine- or ten-residue pseudo-repeats found in other filamentins derived from keratins.

These dissimilarities between terminal regions of component 8c-1 and other type I filamentins are the more extraordinary in view of the development of the wool follicle as an invagination of epidermal cells (Hardy, 1969), which subsequently differentiate to produce proteins characteristic of the wool fibre. They suggest that the portions of the genes coding for the terminal segments

Fig. 3. Comparison of amino acid sequences in 49-residue pseudo-repeats in desmin (Geisler & Weber, 1982) and in a normalized sequence combining data from nine filamentins

The filamentins comprise: from type I, component 8c-1, human epidermal 50000- M_r filamentin (Hanukoglu & Fuchs, 1982) and mouse epidermal 59000- M_r filamentin (Steinert *et al.*, 1983); from type II, component 7c (Sparrow & Inglis, 1980), human epidermal 56000- M_r filamentin (Hanukoglu & Fuchs, 1983) and mouse epidermal 60000- M_r filamentin (Steinert *et al.*, 1984a); from type III, chicken gizzard desmin (Geisler & Weber, 1982), hamster vimentin (Quax-Jeukens *et al.*, 1983) and mouse glial fibrillary acidic protein (Lewis *et al.*, 1984). Boxes have been used to indicate identical residues in corresponding positions of the sequences being compared. Residues underlined in the sequence of desmin are identical with those in corresponding positions in the normalized sequence. (a) Comparison of the first with the second half of segment 1B in desmin and in the normalized sequence. (b) Comparison of the 49-residue sequences at the C-terminus of segment 2B in desmin and in the normalized sequence with the second half of segment 1B in the normalized sequence.

of the filamentins were built up in part by the incorporation in a precursor gene of new genetic material with subsequent evolutionary development.

The homology between the N-terminal and C-terminal segments of component 8c-1 is no longer apparent if proline and cystine residues are discounted (two identical residues in 30), and hence the triplet arrangement of proline and cystine is responsible for the homology. This suggests that the two segments evolved independently following insertion of a gene segment coding for Pro-Cys-Xaa or Pro-Cys-Xaa-Pro-Cys-Yaa with subsequent modification by duplication and mutation. The fact that proline and cystine are the only amino acids showing significant homology in the terminal segments strengthens the view that the Pro-Cys-Xaa repeat has an important structural function in the microfibril (Dowling *et al.*, 1986).

We thank Dr. L. G. Sparrow for providing us with sequence data for component 7c before its publication.

REFERENCES

- Ahmadi, B., Boston, N. M., Dobb, M. G. & Speakman, P. T. (1980) in *Fibrous Proteins: Scientific, Industrial and Medical Aspects* (Parry, D. A. D. & Creamer, L. K., eds.), vol. 2, pp. 161–166, Academic Press, London and New York.
- Corfield, M. C. & Fletcher, J. C. (1969) *Biochem. J.* **115**, 323–334.
- Corfield, M. C., Fletcher, J. C. & Robson, A. (1967) *Biochem. J.* **102**, 801–814.
- Crewther, W. G. (1976) *Proc. Int. Wool Text. Res. Conf.* 5th 1975, vol. 1, pp. 1–101.
- Crewther, W. G. & Dowling, L. M. (1971) *Appl. Polym. Symp.* **18**, 1–20.
- Crewther, W. G., Dowling, L. M., Gough, K. H., Inglis, A. S., McKern, N. M., Sparrow, L. G. & Woods, E. F. (1976) *Proc. Int. Wool Text. Res. Conf.* 5th, 1975, vol. 2, pp. 233–242.
- Crewther, W. G., Gough, K. H., Inglis, A. S. & McKern, N. M. (1978a) *Text. Res. J.* **48**, 160–162.
- Crewther, W. G., Inglis, A. S. & McKern, N. M. (1978b) *Biochem. J.* **173**, 365–371.
- Crewther, W. G., Dowling, L. M., Gough, K. H., Marshall, R. C. & Sparrow, L. G. (1980a) in *Fibrous Proteins: Scientific, Industrial and Medical Aspects* (Parry, D. A. D. & Creamer, L. K., eds.), vol. 2, pp. 151–159, Academic Press, London and New York.
- Crewther, W. G., Dowling, L. M. & Inglis, A. S. (1980b) *Proc. Int. Wool Text. Res. Conf.* 6th, 1980, vol. 2, pp. 79–91.
- Crewther, W. G., Dowling, L. M., Steinert, P. M. & Parry, D. A. D. (1983) *Int. J. Biol. Macromol.* **5**, 267–274.
- Doonan, S., Doonan, H. J., Hanford, R., Vernon, C. A., Walker, J. M., Airolidi, L. P. da S., Bossa, F., Barra, D., Carloni, M., Fasella, P. & Riva, F. (1975) *Biochem. J.* **149**, 497–506.
- Dopheide, T. A. A. (1973) *Eur. J. Biochem.* **34**, 120–124.
- Dowling, L. M. & Crewther, W. G. (1974) *Prep. Biochem.* **4**, 203–226.
- Dowling, L. M., Parry, D. A. D. & Sparrow, L. G. (1983) *Biosci. Rep.* **3**, 73–78.
- Dowling, L. M., Crewther, W. G. & Parry, D. A. D. (1986) *Biochem. J.* **236**, 705–712.
- Drapeau, G. R., Boily, Y. & Houmard, J. (1972) *J. Biol. Chem.* **247**, 6720–6726.
- Duggleby, R. G. & Kaplan, H. (1975) *Anal. Biochem.* **65**, 346–354.
- Elleman, T. C., Crewther, W. G. & Van Der Touw, J. (1978) *Biochem. J.* **173**, 387–391.

- Fraser, R. D. B., MacRae, T. P. & Rogers, G. E. (1972) *Keratins, their Composition, Structure and Biosynthesis* (Kugelmass, I. N., ed.), Charles C. Thomas, Springfield
- Geisler, N. & Weber, K. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4120–4123
- Geisler, N. & Weber, K. (1982) *EMBO J.* **1**, 1649–1656
- Geisler, N. & Weber, K. (1983) *EMBO J.* **2**, 2059–2063
- Geisler, N., Plessman, U. & Weber, K. (1982) *Nature (London)* **296**, 448–450
- Geisler, N., Kaufmann, E., Fischer, S., Plessmann, U. & Weber, K. (1983) *EMBO J.* **2**, 1295–1302
- Gough, K. H. & Inglis, A. S. (1980) *Aust. J. Biol. Sci.* **33**, 295–301
- Gough, K. H., Inglis, A. S. & Crewther, W. G. (1978) *Biochem. J.* **173**, 373–385
- Hanukoglu, I. & Fuchs, E. (1982) *Cell (Cambridge, Mass.)* **31**, 243–252
- Hanukoglu, I. & Fuchs, E. (1983) *Cell (Cambridge, Mass.)* **33**, 915–924
- Hardy, M. H. (1969) *Adv. Biol. Skin* **9**, 35–60
- Hosken, R., Moss, B. A. O'Donnell, I. J. & Thompson, E. O. P. (1968) *Aust. J. Biol. Sci.* **21**, 593–595
- Inglis, A. S. (1983) *Methods Enzymol.* **91**, 443–450
- Inglis, A. S., Rubira, M. R. & Strike, P. McK. (1981) *J. Biochem. Biophys. Methods* **4**, 279–285
- Jones, L. N. (1976) *Biochim. Biophys. Acta* **446**, 515–524
- Krieg, T. M., Schafer, M. P., Cheng, C. K., Filpula, D., Flaherty, P., Steinert, P. M. & Roop, D. R. (1985) *J. Biol. Chem.* **260**, 5867–5870
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lewis, S. A., Balcerek, J. M., Krek, V., Shelanski, M. & Cowan, N. J. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2743–2746
- Marshall, R. C., Gillespie, J. M. & Frenkel, M. J. (1980) *Proc. Int. Wool Text. Res. Conf.* 6th, 1980, vol. 2, pp. 147–158
- Matsuo, H., Fujimoto, Y. & Tatsuno, T. (1966) *Biochem. Biophys. Res. Commun.* **22**, 69–74
- O'Donnell, I. J. (1969) *Aust. J. Biol. Sci.* **22**, 471–488
- Quax-Jeuken, Y. E. F. M., Quax, W. J. & Bloemendal, H. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3548–3552
- Savigne, W. E. & Fontana, A. (1977) *Methods Enzymol.* **47**, 459–469
- Simpson, R. J., Neuberger, M. R. & Liu, T.-Y. (1976) *J. Biol. Chem.* **251**, 1936–1940
- Sparrow, L. G. & Crewther, W. G. (1972) *J. Text. Inst.* **63**, 619–621
- Sparrow, L. G. & Inglis, A. S. (1980) *Proc. Int. Wool Text. Res. Conf.* 6th, 1980, vol. 2, pp. 237–246
- Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L. & Steven, A. C. (1983) *Nature (London)* **302**, 794–800
- Steinert, P. M., Parry, D. A. D., Racoosin, E. L., Idler, W. W., Steven, A. C., Trus, B. L. & Roop, D. R. (1984a) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5709–5713
- Steinert, P. M., Jones, J. C. R. & Goldman, R. D. (1984b) *J. Cell Biol.* **99**, 22s–27s
- Swart, L. S., Joubert, F. J. & Parris, D. (1976) *Proc. Int. Wool Text. Res. Conf.* 5th, 1975, vol. 2, pp. 254–264
- Thompson, E. O. P. & O'Donnell, I. J. (1965) *Aust. J. Biol. Sci.* **17**, 973–989
- Thompson, E. O. P. & O'Donnell, I. J. (1967) *Aust. J. Biol. Sci.* **20**, 1001–1024
- Walton, P. L., Turner, R. W. & Broadbent, D. (1972) *Br. Patent* 1263956
- Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406–4412
- Woods, E. F. (1979) *Aust. J. Biol. Sci.* **32**, 423–435
- Woods, E. F. & Inglis, A. S. (1984) *Int. J. Biol. Macromol.* **6**, 277–283

Received 7 October 1985/24 December 1985; accepted 10 February 1986