Biochem. J. (1986) 236, 705–712 (Printed in Great Britain)

Secondary structure of component 8c-1 of α-keratin

An analysis of the amino acid sequence

Lyndsay M. DOWLING,*‡ W. Gordon CREWTHER* and David A. D. PARRY†
*C.S.I.R.O. Division of Protein Chemistry, 343 Royal Parade, Parkville, Vic. 3052, Australia, and †Department of Physics and Biophysics, Massey University, Palmerston North, New Zealand

The amino acid sequence of component 8c-1 from α-keratin was analysed by using secondary-structure prediction techniques, homology search methods, fast Fourier-transform techniques to detect regularities in the linear disposition of amino acids, interaction counts to assess possible modes of chain aggregation and assessment of hydrophilicity distribution. The analyses show the following. (i) The molecule has two lengths of coiled-coil structure, each about 20 nm long, one from residues 56–202 with a discontinuity from about residue 91 to residue 101, and the other from residues 219–366 with discontinuities from about residue 238 to residue 245 and at about residue 306. (ii) The acidic and basic residues in the coiled-coil segment between residues 102 and 202 show a 9,4-residue structural period in their linear disposition, whereas between residues 246 and 366 a period of 9.9 residues is observed in the positioning of ionic residues. Acidic and basic residues are out of phase by 180°. Similar repeats occur in corresponding regions of other intermediate-filament proteins. The overall mean values for the repeats are 9.55 residues in the *N*-terminal region and 9.85 residues in the *C*-terminal region. (iii) The regions at each end of the protein chain (residues 1–55 and 367–412) are not α-helical and contain many potential β-bends. (iv) The regions specified in (iii) have a significant degree of homology mainly due to a semi-regular disposition of proline and half-cystine residues on a three-residue grid; this is especially apparent in the *C*-terminal segment, in which short $(Pro\text{-}Cys\text{-}Xaa)_n$ regions occur. (v) The coiled-coil segments of component 8c-1 bear a striking similarity to corresponding segments of other intermediate-filament proteins as regards sequence homology, structural periodicity of ionic residues and secondary/tertiary-structure predictions. (vi) The assessments of the probabilities that these homologies occurred by chance indicate that there are two populations of keratin filament proteins. (vii) The non-coiled-coil regions at each end of the chain are less hydrophilic than the coiled-coil regions. (viii) Ionic interactions between the heptad regions of components 8c-1 and 7c from the microfibrils of α-keratin are optimized when a coiled-coil structure is formed with the heptad regions of the constituent chains both parallel and in register.

INTRODUCTION

The preceding paper (Dowling *et al.*, 1986) describes the determination of the amino acid sequence of component 8c-1, one of the eight proteins constituting the microfibril in cortical cells of the Merino wool fibre. The microfibrillar proteins, isolated as *S*-carboxymethyl derivatives, comprise two families of proteins, the component 7 family and the component 8 family, both of which have four homologous members (Crewther *et al.*, 1980*a*). Studies on solutions of these proteins and of the two types (I and II) of helical segment excised from them by chymotryptic digestion (Crewther & Dowling, 1971) indicate that members of the component 7 family associate specifically with members of the component 8 family to form complexes stabilized by interactions between α-helices (Crewther *et al.*, 1968; Crewther *et al.*, 1980*b*; Gruen & Woods, 1983).

X-ray-diffraction analysis of porcupine quill, which gives a diffraction pattern characteristic of hard α-keratins, indicates a high degree of regularity in the microfibrils (see Fraser *et al.* 1972, 1976), and helical models for the structure of the microfibril have been proposed on the basis of these data (Fraser *et al.*, 1972, 1976; Fraser & MacRae, 1983; Crewther *et al.*, 1983). In the present paper, predictive method are applied to the amino acid sequence of component 8c-1 to indicate the most probable structure and location of the native chain in the microfibril.

EXPERIMENTAL

Methods of sequence analysis

(i) Secondary structure. The Robson (Garnier *et al.*, 1978) and Chou–Fasman (Chou & Fasman, 1974*a*,*b*) methods were used to predict secondary structure in component 8c-1 on the basis of the amino acid sequence.

(ii) Sequence homology. The MATCH program, which incorporates the methods and statistical tests described by Gibbs & McIntyre (1970), was used extensively to search for both intra-chain and inter-chain sequence homologies. Sections of amino acid sequence up to 240 residues in length were compared and coincidences of residues tabulated. The number of coincidences were

‡ To whom reprint requests should be addressed.

compared with those arising from a random sequence and a measure of the significance of any homology discovered was calculated by using (primarily) the χ^2 statistical test.

(iii) Periodic features. Regularities in the linear disposition of apolar, acidic and/or basic residues in the sequences of fibrous proteins (Fraser & MacRae, 1976; Parry, 1975, 1979, 1981*a*,*b*; McLachlan & Stewart, 1976; Parry *et al.*, 1977) were determined by using the fast Fourier-transform technique (Fraser & MacRae, 1976; Parry, 1975, 1979; McLachlan & Stewart, 1976).

(iv) Inter-chain interactions. In α-helical regions heptapeptide pseudo-repeats ($(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$ with *a* and *d* apolar) indicate the presence of a coiled-coil rope structure (Crick, 1953). In addition to apolar interactions, inter-chain ionic interactions of the type 2*e*′–1*g*, 1*g*′-2*e*, 2*a*′–1*g*, 1*g*′–2*a*, 1*e*′–1*d* and 1*d*′-1*e* are possible (see McLachlan & Stewart, 1975; Parry *et al.*, 1977). Such interactions were calculated as a function of heptad stagger, chain combination and relative chain direction (McLachlan & Stewart, 1975; Parry *et al.*, 1977). Where appropriate, sequence data for component 8c-1 were compared with similar data for microfibrillar protein, component 7c (Sparrow & Inglis, 1980) and desmin (Geisler & Weber, 1982) or other proteins from intermediate filaments (Crewther *et al.*, 1983; Dowling *et al.*, 1983).

(v) Hydrophilicity. Variations in the hydrophilicity of short segments of chain in relation to their position in the chain were determined by using hydrophilicity values assigned to individual amino acid residues by Hopp & Woods (1981). Mean values for six consecutive residues along the chain were plotted against the position of the centre of the hexapeptide in the chain.

RESULTS AND DISCUSSION

Structural features of component 8c-1 from keratin

For ease of discussion, segments of the sequence showing common features are described separately. Although residue numbers have been assigned to these segments, the boundaries of each segment cannot be precisely identified.

(*a*) Residues 1–55 (*N*-terminal segment). The Chou–Fasman and the Robson methods both predict that this region is relatively rich in β-turns but is unlikely to contain any α-helix (Table 1); thus this segment may adopt a globular and compact structure. The sequence is rich in cystine and proline residues and in the uncharged polar residues serine, threonine and asparagine, often in consecutive runs. In globular proteins comprising several subunits, uncharged polar residues frequently occur at the interface between the subunits, where they appear to mediate the interactions leading to aggregation. Hence this segment of component 8c-1 may facilitate the lateral organization of component 8c-1, or structural units containing component 8c-1, in the microfibril. Homology of this segment with the *C*-terminal segment is discussed in (*i*) below.

(*b*) Residues 56–90 (segment 1A). Both prediction methods indicate that this 35-residue region is totally α-helical (Table 1). More importantly, it comprises five pseudo-repeating heptapeptides of the form (*a*-*b*-*c*-*d*-*e*-*f*-*g*) (Stewart & McLachlan, 1975), where *a* and *d* are usually apolar residues. This provides strong evidence that this segment interacts with another to form a coiled-coil rope of length about 5.2 nm. No significant homology exists between residues 56–90 and other segments of the component 8c-1 chain. Also, excluding the regular disposition of apolar residues, fast Fourier-transform techniques have not revealed any significant periodicities in the linear distribution of other amino acids of the type found in segments 1B and 2B (Table 2). Segment 1A is probably too short for such periodicities to be apparent.

(*c*) Residues 91–101 (linker L1). Although the secondary structure of this segment cannot be predicted with confidence, several features of the sequence indicate that the coiled-coil is disrupted over this region: many of the residues in the segment do not favour the formation of an α-helix, and there is no regular disposition of apolar residues in heptads (Table 1); the sequence from residue 99 to residue 102 strongly indicates the presence of a β-turn (Table 1); furthermore the ready cleavage of the chain by chymotrypsin between residues 97 and 98 (Crewther & Dowling, 1971; Gough *et al.*, 1978) in associated complexes of components 7 and 8 indicates that there is little regular structure of the chain around these residues.

Since both proline and cystine can occur in the first turn of an α-helix (Chou & Fasman, 1974*a*,*b*), account was taken of the possibility that the sequence Val-Cys-Pro-Asn (residues 98–101) may form the *N*-terminal turn of an α-helix. The apolar residue valine is in the appropriate position to continue the band of apolar residues in the adjacent coiled-coil region. Examination of the sequences of proteins from other intermediate filaments, however, indicated that the residue in the equivalent position is not usually apolar in character. Hence it seems probable that the adjacent coiled-coil region commences at residue 102. Residues 91–101 may protrude into the matrix or into the centre of the microfibril such that the terminal residues 91 and 101 lie close together, possibly within 1 nm of one another. The 11-residue segment contains five uncharged polar residues (serine, asparagine, glutamine) and, as for residues 1–55, may have an important role in determining modes of chain and/or molecular aggregation.

(*d*) Residues 102–202 (segment 1B). This region, from which the type I helical segment (Crewther & Dowling, 1971) is derived, is predicted to be largely α-helical (Table 1). It also shows a regular disposition of apolar residues in positions *a* and *d* of pseudo-repeating heptads over its entire length (Table 1), though six of the 29 residues in these positions are polar. Since, however, other α-helical proteins known to be totally in the form of coiled-coils, such as tropomyosin, also have some polar residues in positions *a* and *d* of the heptads, it seems probable that the segment takes the form of a coiled-coil with a length of 15 nm. Provided that the limit to the axial dimensions of the structure suggested for residues 91–101 is correct, residues 56–90 and 102–202 would together form an effectively continuous piece of coiled-coil with a length of about 21 nm. When associated with component 7 the segment between residues 98 and 200 is resistant to digestion with either chymotrypsin (Crewther & Dowling, 1971; Gough *et al.*, 1978) or trypsin (Woods & Inglis,

Table 1. Prediction of secondary structure in component 8c-1 based on primary structure

Segments, indicated by residue numbers, with structure predicted to be:						
					β-Turn (Chou–Fasman)	
α-Helix (Chou–Fasman)	α-Helix (Robson)	Coiled-coil (heptad repeat of apolar residues)	β-Strand (Chou–Fasman)	β-Turn (Robson)	Of $P > \times 10^4$	Of $P = 1 \times 10^{-4}$–2×10^4
			1–5	1–5	6–9	
				7–28	15–18	
					23–26	
				31–34	30–33	
				38–50	42–45	47–50
						52–55
56–95	54–90	59–87				63–66
				91–94		
				100–105	99–102	
			104–108		(10.7*)	
109–125	108–203	102–189				
			126–130			
131–157						
						157–160
161–199						171–174
208–217						
225–255	209–258	220–234	218–224			238–241
263–266	264–289	246–302	256–259			
						259–262
						265–268
269–274						
			275–280			
281–290						291–294
295–305	296–307					
		306–362	307–319			306–309
320–337	320–355					335–338
			338–343			
344–351						
			352–356			353–356
						361–364
				363–368		
					369–372	
					(3.3*)	
				375–386	375–378	
				383–386	387–390	
				391–395		393–396
				400–409		399–402
					405–408	
					(3.3*)	

* Values in parentheses indicate actual probabilities ($\times 10^{-4}$).

Table 2. Significant periodicities in the disposition of apolar, acidic and basic residues in the 1B and 2B segments of component 8c-1

Residues	Type	Period (residues)	Scaled Fourier intensity (I)	Probability [$\exp(-I)$]
1B (102–202)	Apolar*	3.49	10.73	2.2×10^{-5}
		2.33	5.48	4.2×10^{-3}
	Acidic	9.29	9.55	7.1×10^{-5}
	Basic	9.49	6.45	1.6×10^{-3}
2B (246–366)	Apolar*	3.52	11.62	9.0×10^{-6}
		2.31	5.23	5.4×10^{-3}
	Acidic	9.93	5.53	4.0×10^{-3}
		2.79	7.86	3.9×10^{-4}
	Basic	9.97	4.11	1.6×10^{-2}
		2.30	5.41	4.5×10^{-3}

* Periodicities of approx 3.5 ($\frac{7}{2}$) and 2.3 ($\frac{7}{3}$) result from the heptapeptide repeat in apolar residues.

1984). This supports the view that the coiled-coil is continuous at least beween residues 102 (tyrosine) and 189 (leucine) inclusive. There are no sites readily cleaved by chymotrypsin or trypsin between residues 190 and 200 and the cleavage of the chain by chymotrypsin adjacent to residue 200 (leucine) may mean that the coiled-coil structure does not extend to this residue.

Fast Fourier-transform techniques applied to the linear disposition of the acidic residues (aspartic acid, glutamic acid) and the basic residues (arginine, lysine) have revealed a highly significant periodicity of about 9.4 residues in both acidic and basic residues. The phases of the periods for acidic and basic residues differ by about 180°. Similar periods were reported by Parry *et al.* (1977) for the type I (residues 98–200, component 8) helical segment obtained from wool keratin by partial proteolysis (Crewther & Dowling, 1971). The mean period for this region of all known intermediate-filament protein sequences is 9.55 residues.

***(e)* Residues 203–218 (linker L1,2).** The segment from residue 201 to residue 218 is almost certainly non-coiled-coil since (i) the chain when in association with component 7, is readily cleaved by chymotrypsin at residue 200 (leucine) (Crewther & Dowling, 1971; Gough *et al.*, 1978), (ii) proline located at residue 216 would dislocate an α-helix, and (iii) the prediction techniques indicate a low probability for an α-helix from residue 204 to residue 208 (Table 1). The considerably greater size of linker L1,2 as compared with linker L1 or L2 (see below) suggests that this segment plays an important role in the formation of the four-chain microunit that is arranged helically in the microfibril (see Crewther *et al.*, 1983).

***(f)* Residues 219–237 (segment 2A).** In the segment from residue 219 to residue 237 four out of five residues corresponding to positions *a* and *d* in three consecutive heptids are hydrophobic; the other residue, threonine at position 227, has some hydrophobic character (Nozaki & Tanford, 1971) and occurs in coiled-coil regions of this and other filamentous proteins. In all known sequences of intermediate-filament proteins the corresponding position is occupied by apolar residues. It seems probable therefore that this segment has a coiled-coil conformation of length approx. 3 nm.

***(g)* Residues 238–245 (linker L2).** The helicity scores for this section of chain (Table 1) indicate that much of it could be α-helical. There is, however, a discontinuity in the heptad pattern of apolar residues between residues 235 and 245. Assuming that the α-helix is continuous, this indicates a disruption in the coiled-coil structure. Geisler & Weber (1983) have not taken this disruption into account and have designated the sequences of other intermediate-filament proteins corresponding to residues 222–366 as a continuous coiled-coil region.

***(h)* Residues 246–366 (segment 2B).** This segment, which is predicted to be largely α-helical (Table 1), contains a few short regions, three to five residues in length, which have a relatively low α-helix score. However, as the repeat of hydrophobic residues is complete except for positions 250, 260, 288, 292, 334, 345 and 362, there can be little doubt that the 121-residue region adopts a coiled-coil structure (length 18 nm). Positioning of hydrophobic residues in successive heptads to give a sequence *a*, *d*, *a*, *d* reverses in the vicinity of residue 306 to give *a*, *d*, *a*, *d*, *d*, *a*, *d*, *a*. Similar reversals have been reported in the sequences of segments from the α-, β- and γ-chains of fibrinogen (Doolittle *et al.*, 1977; Parry, 1978, 1979) and in both skeletal-muscle and nematode myosin (Parry, 1981*b*; McLachlan & Karn, 1982). All intermediate-filament proteins exhibit the same feature at the corresponding position in the sequence. Such a reversal implies a major discontinuity in the structure of the coiled-coil, and would be expected to affect the coiled-coil pitch length locally and hence the mode of interaction of the chain with its nearest neighbours. It may be of significance that the α-helix score around residue 308 is low and that a half-cystine residue is positioned at this point.

Segment 2B, like segment 1B, has a regular disposition of both acidic and basic residues (Table 2), but in this case the periodicity is 9.9 residues (cf. 9.4 residues period for segment 1B). The phases of the period for the acidic and the basic residues again differed by about 180°. The mean period for this region of all known filament proteins is 9.85 residues.

***(i)* Residues 367–412 (*C*-terminal segment).** This region has no predicted α-helix or β-sheet structure, but is expected to contain a number of β-turns (Table 1). On the basis of this evidence alone the segment would be expected to adopt a compact globular structure, which may be located in the matrix or in the central core of the microfibril. Like the *N*-terminal segment, the 46-residue sequence is rich in cystine and proline residues and contains many uncharged polar residues. Hence, like the *N*-terminal segment, this segment may be involved in lateral interactions between chains, molecules or structural units of the microfibril. The proline residues are highly regular in their disposition, all but one being separated from other proline residues by multiples of three residues (see Fig. 1). The half-cystine residues show a similar three-residue regularity, and in two cases triplets of the form (Pro-Cys-Xaa) are adjacent. Residue Xaa is either apolar or an uncharged polar residue. In two other positions proline occurs with the same three-residue regularity, though not adjacent to half-cystine; likewise half-cystine occurs in two additional positions with this pattern, though not adjacent to proline. The probability of this repeat pattern occurring by chance is very small. These observations suggest that the triplet repeats are involved in forming a specific secondary structure with a functional role in the assembly of molecules forming the microfibril. Further, it suggests that this structure is stabilised by disulphide cross-linkages. The most likely structure for such three-residue repeats would be a helix containing three residues per turn and resembling the polyproline II helix ($h = 0.312$ nm, $t = -120°$). Both proline and half-cystine residues would then be aligned along such a helix. It is noteworthy that the five proline residues occurring between positions 7 and 39 in the *N*-terminal segment also exhibit a three-residue period, though of the eight half-cystine residues in this segment only four lie on a corresponding three-residue grid.

Residues 1–55 and 367–412 have a significant degree of homology, as deduced by using the MATCH program, mainly with respect to proline and cystine residues. The correspondence between the sequences is not of a simple 1:1 type, but nonetheless the probability (*P*) that they arose independently is about 6.7×10^{-3}. Simple (linear)

Table 3. Interchain ionic interactions as a function of chain stagger and chain direction for segments from components 8c-1 and 7c of wool keratin and desmin

Segment 1A ≡ residues 56–90 in component 8c-1; unknown in component 7c; residues 100–134 in desmin. Segment 1B ≡ residues 102–202 in component 8c-1; residues 155–255 in component 7c; residues 143–243 in desmin. Segment 2A ≡ Residues 219–237 in component 8c-1; residues 273–291 in component 7c; residues 260–278 in desmin. Segment 2B(1) ≡ residues 246–303 in component 8c-1; residues 300–357 in component 7c; residues 287–344 in desmin. Segment 2B(2) ≡ Residues 305–366 in component 8c-1; residues 359–420 in component 7c; residues 346–407 in desmin. Numbering of residues for component 7c is tentative, as the sequence is incomplete.

Interacting proteins	Stagger (heptads)	Score for parallel interaction of segments						Score for antiparallel interaction of segment				
		1A	1B	2A	2B(1)	2B(2)	Total	1A	1B	2A	2B(1)	2B(2)
8c-1–8c-1	−2	2	1	0	0	−1	2	3	−1	0	0	0
	−1	3	3	0	0	2	8	2	−4	−1	−2	−1
	0	0	2	0	2	0	4	−2	1	0	0	0
	1	3	3	0	0	2	8	0	−2	0	−2	−5
	2	2	1	0	0	−1	2	−4	−8	0	0	0
7c–7c	−2	—	−3	−1	−1	1	−4	—	3	0	2	4
	−1	—	2	1	−2	−1	0	—	−4	−1	3	−4
	0	—	2	0	0	2	4	—	3	2	2	6
	1	—	2	1	−2	−1	0	—	−4	1	−1	−3
	2	—	−3	−1	−1	1	−4	—	−7	0	4	0
Desmin–desmin	−2	−3	0	−1	0	0	−4	4	3	0	0	−2
	−1	2	0	0	0	0	2	1	−6	−1	0	−5
	0	2	6	0	4	2	14	−2	1	0	−2	−2
	1	2	0	0	0	0	2	−4	0	0	−5	−5
	2	−3	0	−1	0	0	−4	2	−5	0	2	−2
8c-1–7c	−2	—	2	−1	1	0	2	—	−2	0	0	2
	−1	—	3	0	−2	1	2	—	−1	−1	0	−2
	0	—	7	0	2	1	10	—	1	0	2	3
	1	—	−1	1	1	1	2	—	−2	1	−3	−4
	2	—	5	0	0	−1	4	—	−6	0	3	0

homologies exist between residues 15–39 and residues 384–408 (36% homology; $P = 7.6 \times 10^{-2}$) with greatest similarity in residues 19–27 and residues 397–405 (78% homology). The levels of significance for these shorter segments, however, are not high.

Ionic interactions between components 8c-1 and 7c

A comparison was made of the potential ionic interactions between coiled-coil regions of components 8c-1 and 7c when arranged in parallel. The ionic interactions of chain 8c-1 with chain 8c-1 (type I–type I), chain 7c with chain 7c (type II–type II) and chain 8c-1 with chain 7c (type I–type II) were assessed with the coiled-coil regions in register and when displaced both by one and by two heptads. A similar assessment was made for an antiparallel arrangement, but scores for the individual segments of coiled-coil cannot be summed since the segments cannot be suitably matched in this arrangement. The antiparallel arrangements of the type I and type II helical regions have low scores (Table 3). The parallel arrangement of these segments gave a higher score than any other pairing of identical or homologous segments, and the complete chains of components 7c and 8c-1 gave an even higher score. The parallel in-register arrangements of two chains of desmin also gave a high score (Table 3); hence a parallel in-register interaction of components 7 and 8 appears probable.

Structural model for the molecular unit in the microfibril

The data suggest that component 8c-1 will tend to form a rod-shaped structure of the form shown in Fig. 1. The positions of interruptions in the coiled-coil regions and the positioning and periodicity of charged groups along the chain suggest that the chain is unlikely to fold back upon itself but rather forms in association with component 7 a two-chain parallel coiled-coil rope with the dimensions shown in Fig. 1. Studies on the interactions between components 7 and 8 and between type I and type II helical segments derived from them strongly favour a 1:1 ratio of these components in the complex (Crewther *et al.*, 1980*b*; Gruen & Woods, 1983). We can therefore envisage a rod-shaped structure consisting of two chains, one of component 8, the other of component 7, with a length somewhat greater than 40 nm and with interruptions or marked changes in pitch at its centre and at three other positions, one towards the *N*-terminus and two towards the *C*-terminus. At each end the rod carries the material that does not have a coiled-coil structure.

Homologies with proteins of intermediate filaments

The obvious homologies between the amino acid sequences of the coiled-coil regions of component 8c-1 and the corresponding regions of component 7c and other proteins derived from intermediate filaments has been discussed elsewhere (Dowling *et al.*, 1983). These

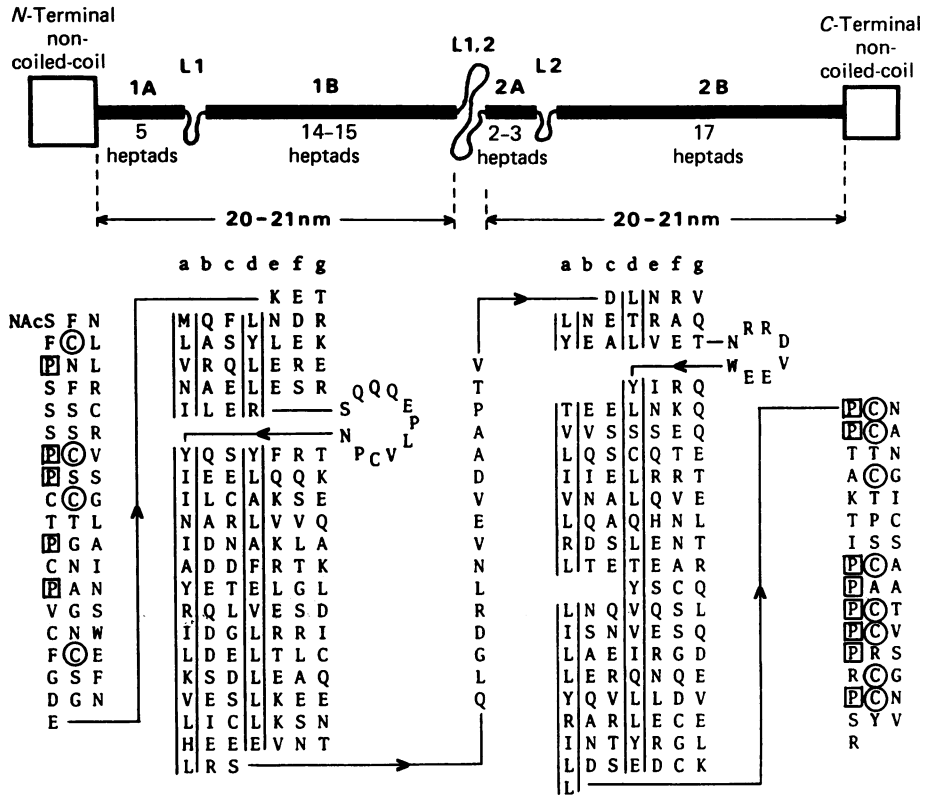

Fig. 1. Primary and predicted secondary structure of component 8c-1

Coiled-coil segments (1A, 1B, 2A, 2B) are indicated by heavy lines, terminal non-α-helical segments rich in cystine and proline are indicated by boxes, and linker segments (L1, L1,2, L2) are indicated by irregular lines. In the terminal non-α-helical segments proline and cystine residues that lie on the three-residue grid are boxed and circled respectively. Residues in the *a* and *d* positions in the coiled-coil segments are outlined.

homologies were used in proposing a general model for the structure of intermediate filaments (Crewther *et al.*, 1983). The significance of these homologies is considered in the preceding paper (Dowling *et al.*, 1986).

Hanukoglu & Fuchs (1982) have classified the proteins obtained from epidermal keratin filaments as of type I or type II according to their degree of homology with the type I or type II helical segments excised from the microfibrillar proteins of wool by chymotryptic digestion (Crewther & Dowling, 1971). These correspond to segment 1B of component 8 (Fig. 1) and component 7 proteins respectively. Desmin, vimentin and glial fibrillary acid protein constitute a third type of intermediate-filament protein (Steinert *et al.*, 1984*a*; Dowling *et al.*, 1986). The type I proteins include the 50000-M_r protein from human epidermis (Hanukoglu & Fuchs, 1982), the 59000-M_r protein from mouse epidermis (Steinert *et al.*, 1983), component 8c-1 and other members of the component 8 family. The type II proteins include the 56000-M_r protein from human epidermis (Hanukoglu & Fuchs, 1983), the 60000-M_r protein from mouse epidermis (Steinert *et al.*, 1984*b*), component 7c and other members of the component 7 family of microfibrillar proteins. A statistical analysis of the probabilities that such sequence homologies occur by chance has confirmed this classification (Table 4; see also Table 1 of Dowling *et al.*, 1983).

When the type I and type II helical segments or the component 8 and component 7 proteins from which they were derived are mixed in solution, they interact specifically in a 1:1 molar ratio to give increased helicity and increased stability of the α-helices (Crewther *et al.*, 1980*b*; Gruen & Woods, 1983). It has been concluded (Crewther *et al.*, 1983) that the molecules constituting the

Table 4. Sequence homology between corresponding segments of component 8c-1 and other intermediate-filament proteins of types I, II and III

The following abbreviations are used. The sources of proteins from intermediate filaments are indicated by ME, mouse epidermis, and HE, human epidermis; individual proteins are indicated by the appropriate M_r values. Comp. 7c indicates the microfibrillar protein, component 7c, from wool keratin, and GFA indicates the glial fibrillary acidic protein.

Type	Intermediate-filament protein	Probability of chance homology		
		Segment 1A	Segment 1B	Segment 2B
I	ME 59000 M_r	6.1×10^{-14}	1.4×10^{-37}	7.8×10^{-30}
	HE 50000 M_r	3.3×10^{-14}	1.0×10^{-39}	8.9×10^{-32}
II	Comp. 7c	—	2.3×10^{-9}	7.9×10^{-4}
	ME 60000 M_r	1.4×10^{-3}	9.1×10^{-7}	3.0×10^{-7}
	HE 56000 M_r	—	3.2×10^{-6}	5.2×10^{-8}
III	Desmin	2.6×10^{-4}	2.5×10^{-7}	1.3×10^{-7}
	Vimentin	6.6×10^{-4}	3.1×10^{-7}	9.4×10^{-5}
	GFA	2.8×10^{-5}	3.3×10^{-6}	1.6×10^{-6}

microfibrils of wool consist of two chains, one from the component 7 family, the other from the component 8 family. It is probable that the fundamental structural units in all keratin filaments are made up of type I and type II chains interacting similarly.

Close homologies between the amino acid sequences of proteins constituting the diverse range of intermediate filaments found in hard keratins, epidermis, muscle cells, fibroblasts, glial cells and nerve cells occur only in or near the sequences showing a heptapeptide repeat. The coiled-coil structures formed from segments showing such a repeat are known from X-ray diffraction studies on α-keratins (Fraser *et al.*, 1980) to be located in the outer shell of the filament (microfibril). We will call this shell 'the annulus'. Since electron micrographs of a wide range of intermediate filaments (Steinert, 1981; Steinert *et al.*, 1981) have shown them also to gave an annular structure, it is highly probable that the homologous segments of chain comprise the basic structural component of the filament.

A few residues adjacent to the *N*-terminus of segment 1A (residues 53–55 in component 8c-1) and two of the non-coiled-coil segments linking the coiled-coil segments (namely linkers L1,2 and L2) show homology in all known sequences of intermediate-filament proteins. Furthermore on the basis of homology these linking segments from keratin proteins can be classified as of type I or type II in parallel with the associated coiled-coil segments. This suggests that these sections of chain also constitute an essential part of the annulus.

Location of the terminal non-coiled-coil segments and the structure of the microfibril in wool

Conceivably the non-coiled-coil segments could be located either in the hollow centre of the annulus or as fringes on its outer surface (Crewther *et al.*, 1968, 1983; Fraser & MacRae, 1983). The generally hydrophobic nature (Fig. 2) of the terminal segments of component 8c-1 and the known presence of material of different staining characteristics at the centre of the wool microfibril (Filshie & Rogers, 1961) leads to the prediction that most of the sequence comprising these segments is located at the centre of the annulus. The hydrophilicity of a peptide chain largely reflects the distribution of ionic residues along the chain, and, whereas the coiled-coil regions of component 8c-1 contain 91 ionic residues in 276 residues (33%) and even the terminal segments of desmin contain 33 ionic residues in 155 residues (21%), the terminal segments of component 8c-1 contain only nine ionic residues in 101 residues (9%). Four of these lie within 12 residues of the ends of the chain, and these terminal portions may form short fringes at the surface of the annulus. The incomplete data for component 7c indicate that it resembles component 8c-1 in this respect. The *C*-terminal non-coiled-coil segments of desmin (30% ionic), the 68000-M_r neurofilament protein (41% ionic) and glial fibrillary acidic protein (36% ionic) on the other hand are likely to be located on the outer surface of the filament unless there is free access of water to the centre of the filament.

The *N*-terminal segments of desmin and neurofilament protein contain several arginine residues, which are distributed along the segments. Conceivably these could be accommodated in an essentially hydrophobic core in the form of salt linkages with the numerous anionic side chains in the coiled-coil segments.

The suggestion that the terminal segments of the chain contain helices with three residues per turn that may be similar to the polyproline II helix is of importance in understanding the structure of the microfibril. Since all but five of the proline and cystine residues in the two terminal segments of component 8c-1 lie on the three-residue grid, it seems probable that a structure of this type is an important feature in these segments. The occurrence of proline and cystine residues at intervals of three residues would then result in a line of each of these residues along the helix. Disulphide-bond formation between half-cystine residues either along the helix or between adjacent helices would stabilize the structures.

Blundell *et al.* (1981) have shown that the single α-helix in a 36-residue peptide isolated from the pancreas of turkeys is stabilized by an adjacent polyproline-like helix containing three proline residues occurring at intervals of three residues. The proline side chains interdigitate with hydrophobic side chains in the α-helix. It is possible therefore that in the keratin microfibril the proline side chains of disulphide-linked 'polyproline' helices interact with hydrophobic residues in the coiled-coil structures of the microfibrillar annulus and so stiffen the microfibril.

It has previously been pointed out (Crewther, 1965) that the stiffness of wool fibres during longitudinal extension is markedly decreased if either the disulphide bonds in the fibre or non-covalent bonds between and within protein constituents of the fibre are destroyed. This

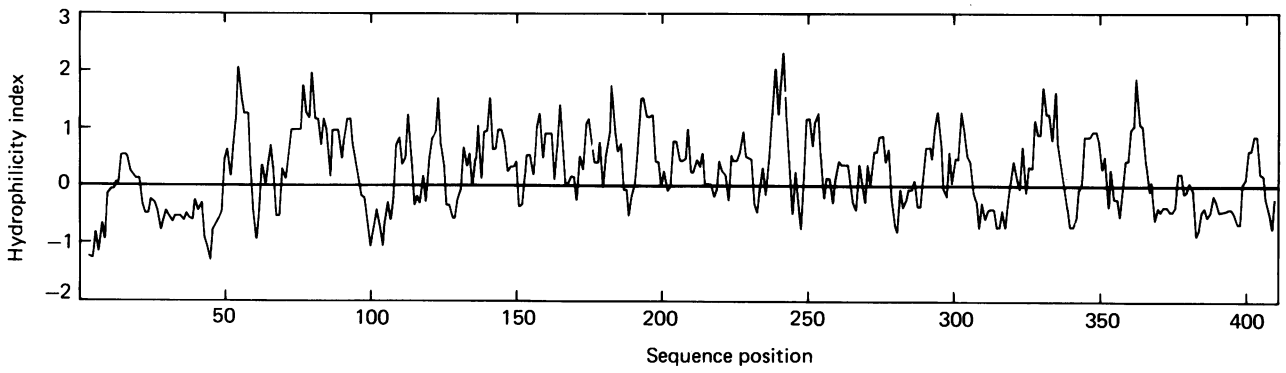

Fig. 2. Hydrophilicity profile of component 8c-1

Hydrophilicity values were repetitively averaged over six-residue spans of sequence, and the average values are plotted against the position in the sequence (Hopp & Woods, 1981).

was interpreted in terms of the stabilization of the α-helices of the microfibrils by disulphide-rich components of the fibre through the mediation of non-covalent interactions. The present proposal that disulphide-linked 'poly-proline' helices stabilize the α-helical coiled-coils is in accordance with that interpretation. Furthermore, the presence of multiple 'poly-proline' helices at the centre of the microfibril could account for the observation by Filshie & Rogers (1961) that some of the material constituting the core of the microfibril, like the annulus, does not stain with osmium.

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

REFERENCES

Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C.-W. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 4175–4179
Chou, P. Y. & Fasman, G. D. (1974*a*) Biochemistry **13**, 211–222
Chou, P. Y. & Fasman, G. D. (1974*b*) Biochemistry **13**, 222–245
Crewther, W. G. (1965) Text. Res. J. **35**, 867–877
Crewther, W. G. & Dowling, L. M. (1971) Appl. Polym. Symp. **18**, 1–20
Crewther, W. G., Dobb, M. G., Dowling, L. M. & Harrap, B. S. (1968) in Symposium on Fibrous Proteins (Crewther, W. G., ed.), pp. 329–340, Butterworths, Sydney
Crewther, W. G., Dowling, L. M., Gough, K. H., Marshall, R. C. & Sparrow, L. G. (1980*a*) in Fibrous Proteins: Scientific, Industrial and Medical Aspects (Parry, D. A. D. & Creamer, L. K., eds.), vol. 2, pp. 151–159, Academic Press, London
Crewther, W. G., Dowling, L. M., Gruen, L. C., Sparrow, L. G. & Woods, E. F. (1980*b*) Proc. Int. Wool. Text. Res. Conf. 6th, 1980, vol. 2, pp. 1–12
Crewther, W. G., Dowling, L. M., Steinert, P. M. & Parry, D. A. D. (1983) Int. J. Biol. Macromol. **5**, 267–274
Crick, F. H. C. (1953) Acta Crystallogr. **6**, 689–697
Doolittle, R. F., Cassman, K. G., Cottrell, B. A., Freizner, S. J. & Takagi, T. (1977) Biochemistry **16**, 1710–1715
Dowling, L. M., Parry, D. A. D. & Sparrow, L. G. (1983) Biosci. Rep. **3**, 73–78
Dowling, L. M., Crewther, W. G. & Inglis, A. S. (1986) Biochem. J. **236**, 695–703
Filshie, B. K. & Rogers, G. E. (1961) J. Mol. Biol. **3**, 784–786
Fraser, R. D. B. & MacRae, T. P. (1976) Proc. Int. Ornithological Congress 16th, Canberra, pp. 443–451, Australian Academy of Science, Canberra
Fraser, R. D. B. & MacRae, T. P. (1983) Biosci. Rep. **3**, 517–525
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
Fraser, R. D. B., MacRae, T. P. & Suzuki, E. (1976) J. Mol. Biol. **108**, 435–452
Fraser, R. D. B., Jones, L. N., MacRae, T. P., Suzuki, E. & Tulloch, P. A. (1980) Proc. Int. Wool Text. Res. Conf. 6th, 1980, vol. 1, pp. 1–33
Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) J. Mol. Biol. **120**, 97–120
Geisler, N. & Weber, K. (1982) EMBO J. **1**, 1649–1656
Geisler, N. & Weber, K. (1983) EMBO J. **2**, 2059–2063
Gibbs, A. J. & McIntyre, G. A. (1970) Eur. J. Biochem. **16**, 1–11
Gough, K. H., Inglis, A. S. & Crewther, W. G. (1978) Biochem. J. **173**, 373–385
Gruen, L. C. & Woods, E. F. (1983) Biochem. J. **209**, 587–595
Hanukoglu, I. & Fuchs, E. (1982) Cell (Cambridge, Mass.) **31**, 243–252
Hanukoglu, I. & Fuchs, E. (1983) Cell (Cambridge, Mass.) **33**, 915–924
Hopp, T. P. & Woods, K. R. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 3824–3828
McLachlan, A. D. & Karn, J. (1982) Nature (London) **299**, 226–231
McLachlan, A. D. & Stewart, M. (1975) J. Mol. Biol. **98**, 293–304
McLachlan, A. D. & Stewart, M. (1976) J. Mol. Biol. **103**, 271–298
Nozaki, Y. & Tanford, C. (1971) J. Biol. Chem. **246**, 2211–2217
Parry, D. A. D. (1975) J. Mol. Biol. **98**, 519–535
Parry, D. A. D. (1978) J. Mol. Biol. **120**, 545–551
Parry, D. A. D. (1979) in Fibrous Proteins: Scientific, Industrial Medical Aspects (Parry, D. A. D. & Creamer, L. K., eds.), vol. 1, pp. 393–427, Academic Press, London
Parry, D. A. D. (1981*a*) J. Mol. Biol. **146**, 259–263
Parry, D. A. D. (1981*b*) J. Mol. Biol. **153**, 459–464
Parry, D. A. D., Crewther, W. G., Fraser, R. D. B. & MacRae, T. P. (1977) J. Mol. Biol. **113**, 449–454
Sparrow, L. G. & Inglis, A. S. (1980) Proc. Int. Wool Text. Res. Conf. 6th. 1980, vol. 2, pp. 237–246
Steinert, P. M. (1981) in Electron Microscopy of Proteins (Harris, J. R., ed.), vol. 1, pp. 125–166, Academic Press, London
Steinert, P. M., Starger, J. M. & Goldmann, R. D. (1981) in Fibrous Proteins: Scientific, Industrial and Medical Aspects (Parry, D. A. D. & Creamer, L. K., eds.), vol. 2, pp. 227–236, Academic Press, London
Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L. & Steven, A. C. (1983) Nature (London) **302**, 794–800
Steinert, P. M., Jones, J. C. R. & Goldman, R. D. (1984*a*) J. Cell. Biol. **99**, 22s–27s
Steinert, P. M., Parry, D. A. D., Racoosin, E. L., Idler, W. W., Steven, A. C., Trus, B. L. & Roop, D. R. (1984*b*) Proc. Natl. Acad. Sci. U.S.A. **81**, 5709–5713
Stewart, M. & McLachlan, A. D. (1975) Nature (London) **257**, 331–333
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