The EMBO Journal Vol.1 No.10 pp.1155 – 1160, 1982

The structural relation between intermediate filament proteins in living cells and the α-keratins of sheep wool

Klaus Weber* and Norbert Geisler

Max Planck Institute for Biophysical Chemistry, D-3400 Göttingen, FRG

Communicated by K. Weber
Received on 13 July 1982

Although not complete, the available sequence data on smooth muscle desmin, a prototype of 10 nm filaments present in living vertebrate cells, and two wool α-keratin components indicate a common structural motif. A similarly sized rod-like middle domain based mainly on α-helices probably able to form coiled-coils is flanked by differently sized terminal domains of non-α-helical nature. Within the middle domain there seem to be at least two regions where wool keratins and 10 nm filament proteins show a noticeable degree of sequence homology. In general, however, the proteins have diverged to an astonishing degree. Although the analysis seems to support, in general terms, a separation of the rod into two nearly equally long coiled-coils it raises doubts about additional aspects of current models of 10 nm filament organization. We propose that the terminal domains are directly involved in filament assembly making this process permanent in wool α-keratins because of the many disulfide bonds present in these regions. The 10 nm filaments of most living cells seem to avoid this frozen state and lack a similar wealth of cysteine residues.

Key words: desmin/glial acidic fibrillary protein/neurofilaments/vimentin/epithelial keratins

Introduction

Intermediate or 10 nm filaments form the third cytoplasmic fibrous system in most living vertebrate cells. Immunological, biochemical, and protein-chemical results have been used to define a multigene family of related proteins expressed in a histologically meaningful cell- and tissue-specific manner (for recent reviews, see Lazarides, 1980; Weber and Osborn, 1981; Osborn *et al.*, 1982). Isolated or reconstituted 10 nm filaments provide the typical α-type X-ray diffraction pattern (Day and Gilbert, 1972; Steinert *et al.*, 1978; Renner *et al.*, 1981; Zackroff *et al.*, 1981) indicative of coiled-coil proteins (Crick, 1953; McLachlan and Stewart, 1975) raising the question of a structural relation between 10 nm filaments and the α-keratins present in the dead epidermal appendages such as hair and wool. Such a relation may also be indicated by the similar mol. wts. of the structural proteins (for α-keratins, see Jones, 1976; Crewther *et al.*, 1979; Ahmadi *et al.*, 1979) and a related fiber morphology, although wool microfibrils contain several and peculiar additional non-α-helical proteins (for review, see Fraser and McRae, 1972). Furthermore, an antibody raised against helical portions of wool α-keratin decorated the 10 nm filaments of various non-epidermal epithelial cells in culture (Weber *et al.*, 1980). However, the detailed multiple domain model proposed for 10 nm filament organization (Steinert *et al.*, 1980; Zackroff *et al.*, 1981) was difficult to relate to the emerging sequence data on wool α-keratins (Crewther *et al.*, 1979, 1980; Sparrow and Inglis, 1980) beyond possible similarly sized α-helical proteolytic derivatives (Skerrow *et al.*, 1973; Steinert, 1978; Steinert *et al.*, 1980), which could be expected even from non-related coiled-coil proteins. This situation changed to some extent when the carboxy-terminal 140 residues of smooth muscle desmin and mesenchymal vimentin were established in our laboratory and raised some doubts about the proper allocation of α-helical elements in the 10 nm filament model (Geisler and Weber, 1981). The recent extension of sequence data together with the ability to dissect the desmin molecule into three structurally distinct domains (Geisler and Weber, 1982a, 1982b) have now encouraged a direct comparison. Using the currently available sequences, we propose that wool α-keratins and the 10 nm filaments of living cells show a common structural motif. In addition, their indicated rod-like middle domain reveals at least two regions of sequence homology. These findings raise several important questions about the model of 10 nm filaments thought to be firmly established (Steinert *et al.*, 1980; Zackroff *et al.*, 1981).

Results and Discussion

Search for homologous segments and coiled-coil elements in a related rod-like middle region

Desmin protofilaments can be described as containing three structurally distinct domains (Geisler *et al.*, 1982a). Limited chymotryptic digestion of desmin (apparent mol. wt. 50 000) reveals a middle domain (apparent mol. wt. 38 000) of a high α-helical content seen as a thin rod-like particle in electron micrographs. The mass of α-helix (apparent mol. wt. α-helix 32 000) can be envisioned to account for two similarly sized coiled-coil regions separated by a spacer so as to stay within the proposed models of 10 nm filament organization (Skerrow *et al.*, 1973; Steinert *et al.*, 1980). Two non-α-helical regions form the two terminal domains. This information is summarized in Figure 1, which also indicates a similar structural motif for those wool α-keratin components, for which pertinent sequence information is available (Crewther *et al.*, 1980; Sparrow and Inglis, 1980), i.e., components 8c-1 (apparent mol. wt. 50 000) and 7c (apparent mol. wt. 54 000 – 56 000). Alignment has been made to separate the non-α-helical regions occupying the two ends from a proposed common-sized highly α-helical middle region (~330 residues). Attempts at more accurate alignment of the middle region have been helped by comparing available sequence information on the different proteins (Figures 2 and 3). It is proposed (Figure 2) that one of the rare or even sole tryptophan residue(s) situated in the approximate center of the rod can be aligned in all three proteins (homology segment *b*), and that a similar alignment may be possible in the amino-terminal region of the proposed rod (homology segment *a*) (Figure 3). Thus Figure 1 summarizes the previously developed partial sequence information and tries to translate it into a potential common structural motif, which is further analyzed below and in Figures 2 and 3. For additional information the established or tentative position of the proline, cysteine, and tryptophan residues is indicated and so is the

*To whom reprint requests should be sent.

 0261-4189/82/0110-1155$2.00/0.

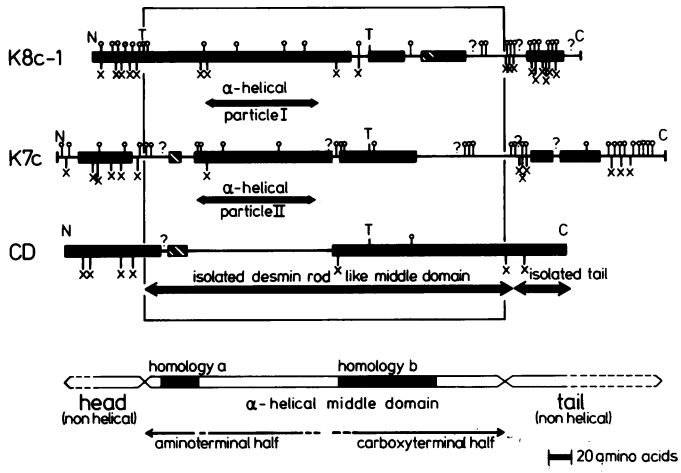

Fig. 1. Schematic presentation of the structure and sequences known for chicken gizzard desmin (CD) and two of the major merino sheep wool α-keratins, components K8c-1 and K7c. Established sequences are provided by boxes; hatched areas indicate that positioning of partial sequences is based on homology (for details see text). Relative arrangement has been made to emphasize the similarity to desmin. A long highly α-helical middle region is flanked by the non-α-helical head and tail domains, which seem to have different lengths. Approximately in the center of the middle region direct alignment is possible at the tryptophan residue (marked T), giving rise to the high homology segment *b*, which is displayed in Figure 2. A similar homology segment, labelled *a*, is shown early in the middle domain and is analyzed in Figure 3. Note the relative wealth of proline residues, labelled X, in the end domains, particularly the headpieces. Cysteine residues are indicated by circles. Question marks indicate uncertainties of length estimate or positioning in the particular segment. N and C label the amino- and carboxy-terminal ends. The small bar corresponds to 20 amino acid residues.

positioning of proteolytically resistant particles (Crewther *et al.*, 1978; Geisler *et al.*, 1982a). In the case of desmin the amino-terminal 88 residues and the carboxy-terminal 209 residues have been reported, leaving a gap of some 150 residues which are assumed to be mainly in α-helical conformation. In order to avoid confusion, the numbering system of the carboxy-terminal 209 residues given already has been retained. It does not refer to residues counted from the amino terminus but rather from the methionine given as residue 1 in Figure 6 of our previous report (Geisler *et al.*, 1982a).

As shown in Figure 2, the 68 residue segment containing the sole tryptophan residue of wool component 7c seems homologous to the corresponding region in chicken desmin. Not only do the two tryptophan residues used as markers in Figure 1 seem to be aligned, but starting with the second horizontal line we recognize 22 identically positioned residues (35%). Thirteen additional highly conservative amino acid exchanges (for details see legend) bring the homology level in these 63 residues to 56%. This value is a lower estimate because in three identically placed positions marked Glx it is not yet known if a glutamic or glutamine residue is present. In addition, we have not counted conventional exchanges such as for instance Ala to Thr, which although allowed by the genetic code as single base changes, could interfere with the structure of a coiled-coil conformation. The close relation indicated between chicken desmin and component 7c is even more apparent when comparison is made to another 10 nm filament protein, the porcine neurofilament component NF68 (Geisler *et al.*, 1982b), which shares a 50% identity with chicken desmin in this region. Using the additional known sequence data on porcine desmin and vimentin (Geisler and Weber, 1981; Geisler *et al.*, 1982b) we estimate the homology

CD							WK 7 c							WK 8 c-1							NF 68						
A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G
		Pro	Asp	Leu	Thr	Ala	Ala			Met	Cys	Cys	Ile	Val													
Leu	Arg	Asp	Ile	Arg	Ala	Gln	Ala	Glu	Glu	Ile	Arg	Ala	Gln														
Tyr	Glu	Ser	Ile	Ala	Ala	Lys	Tyr	Asp	Asp	Ile	Ala	Ser	Arg														
		Asn	Ile	Ala	Glu	Ala			Ser	Arg	Ala	Glu	Ala														
Glu	Glu	Trp	Tyr	Lys	Ser	Lys	Glu	Ser	Trp	Tyr	Arg	Ser	Lys				Trp	Tyr	Ile	Arg	Gln			Trp	Phe	Lys	Ser	Arg
Val	Ser	Asp	Leu	Thr	Gln	Ala	Cys	Glu	Glu	Ile	Lys	Ala	Thr	Thr	Glu	Glu	Leu	Asn	Lys	Gln		Phe	Thr	Val	Leu	Thr	Glu	Ser
Ala	Asn	Lys	Asn	Asn	Asp	Ala	Val	Ile	Arg	His	Gly	Glu	Thr	Val	Val	Ser	Ser	Arg	Glu	Gln		Ala	Ala	Lys	Asn	Thr	Asp	Ala
Leu	Arg	Gln	Ala	Lys	Glx	Glx	Leu	Arg	Arg	Thr	Lys	Glu	Glu	Leu	Gln	Cys	Asn	Gln	Glu	Glu		Val	Arg	Ala	Ala	Lys	Asp	Glu
Met	Leu	Glu	Tyr	Arg	His	Gln	Ile	Asn	Glu	Leu	Asn	Arg	Val	Ile	Ile	Glu	Leu	-	-	-		Val	Ser	Glu	Ser	Arg	Arg	Leu
Ile	Gln	Ser	Tyr	Thr	Cys	Glu	Ile	Gln	Arg	Arg	Thr	Ala	Glx	-	-	-	-	-	-	-		Leu	Lys	Ala	Lys	Thr	Leu	Glu
Ile	Asp	Ala	Leu	Lys	Gly	Thr	Val	Glx						-	-	-	-	-	-	-		Ile	Glu	Ala	Cys			
Asn	Asp	Ser	Leu	Met	Arg	Gln								-	Asp	Ser	Leu	Glu	Asn	Thr								
Met	Arg	Glu	Met	Glu	Glu	Arg								Leu	Thr	Glu	Thr	Glu	Ala	Arg								

Fig. 2. Sequence comparison in homology segment *b*. Comparison of the partial sequences available for wool keratin components 7c and 8c-1 (WK 7c and WK 8c-1) with those of the two 10 nm filament proteins chicken desmin (CD) and porcine neurofilament 68 K protein (NF68) with the tryptophan residue aligned as indicated in Figure 1 (homology region *b*). In the columns for WK 7c and CD, identical residues are underlined by a thick line. Thin lines under corresponding residues indicate highly conservative exchanges (Asp/Glu, Lys/Arg, Met/Ile, Ile/Val, Ala/Val, Ala/Leu). Glx or Asx indicate a residue for which the amide or acid character has not yet been determined. Arrangement is made in the heptade form since a coiled-coil has been proposed to extend from approximately the third line in the desmin presentation for some 130 or 140 residues towards the carboxyl end (for details see Geisler *et al.*, 1982a). Note the high degree of homology (see text). In the third column, comparison is made for wool keratin 8c-1. Symbols are the same as before but indicate relation to either CD or 7c. The fourth column illustrates, for comparative purposes, the sequence divergence for the desmin-related 10 nm filament protein porcine NF68 in this region. Residues marked indicate additional positions where WK 7c is related to 10 nm filament proteins. The extension of the column for 8c-1 shows a tentative alignment of a 13-residue tryptic peptide whose precise positioning is not yet known (Crewther *et al.*, 1980; see Figure 1 and text). Note the degree of homology to the corresponding desmin heptades. Dashes indicate unknown residues, open spaces a possible disturbance of the coiled-coil.

level between wool keratin 7c and 10 nm filament proteins in this region to be at ~60%.

We have previously proposed that a coiled-coil domain starting some 18 residues prior to the sole tryptophan of desmin could extend for ~135 residues towards the carboxy-terminal end before it stops close to proline residue 157 (Geisler *et al.*, 1982a; see Figure 1). A similar extended analysis is not possible for component 7c. The available amino acid composition data (Sparrow and Inglis, 1980) show, however, that within the 99 residues past those given in Figure 2 there are only three proline residues. If they were clustered toward the carboxyl end as observed in a similarly located fragment of component 8c-1 (Crewther *et al.*, 1980; see also Figure 1) we could expect a carboxy-terminal coiled-coil of similar length in keratin 7c and the 10 nm filament proteins. It was previously thought (Dowling *et al.*, 1979) that components 8c-1 and 7c are poorly related outside their amino-terminal α-helical particles where a 30% sequence identity was indicated (Crewther *et al.*, 1978). The available correlation between desmin and 7c seems, however, extendable to 8c-1 assuming a similar location (Figure 1) of the tryptophan residues. Figure 2 shows the established sequence of 30 residues of 8c-1 within this region. When comparison is made to both component 7c and desmin, 40% identity seems possible. The presence of coiled-coil elements in the joining region of 8c-1 towards the carboxyl end is indicated by several observations: the existence of two large chymotryptic fragments obtained by native digestion; in addition the partial sequences so far obtained, but not connected (Crewther *et al.*, 1980), show indications for the 7-residue repeat pattern (see Figure 2) and the prediction that the first proline residue past the tryptophan cannot occur within the next 108 residues (Crewther *et al.*, 1980). In desmin the corresponding number is 123. The difference may be even lower than 15, since the next segment in 8c-1 is ~19 residues but contains three proline residues. We have, therefore, tentatively extended Figure 2 by arranging the sequence of a 13-residue tryptic peptide in a proposed alignment indicated by sequence homology to desmin (Geisler *et al.*, 1981, 1982a), although the precise position of this peptide is not yet established (Crewther *et al.*, 1980). We also note that the 26 known amino-terminal residues of the following fragment (Ch-3) could be aligned parallel to the desmin sequence although the potential homology level would be lower, ~35–40% (not shown; for sequence see Crewther *et al.*, 1980). There is some, but much less sequence information for wool α-keratin component 5, which we have not incorporated into Figures 1–3, since it is already known that this component seems very similar to component 7c (Dowling *et al.*, 1979), although its lower α-helix content in solution remains a puzzle. The combined data indicate that a similarly sized coiled-coil can be tentatively assigned to the carboxy-terminal side of the rod-like middle region in both 10 nm filament proteins and the three wool keratin components 8c-1, 7c, and 5.

In the amino-terminal region of the middle region sequence comparison has so far been limited to the α-helical particles arising from components 8c-1 and 7c and their relatives in the complex mixture of wool low sulfur proteins. After appropriate alignment, a 30% sequence identity was noticed (see for instance Crewther *et al.*, 1978). Possible extension to 10 nm filament proteins is currently frustrated by a virtual lack of information, since our comparative studies used the more easily available carboxy-terminal half (Geisler and Weber, 1981; Geisler *et al.*, 1982a, 1982b). However, among the currently available fragments is the 17-residue segment given in Figure 3. Although it accounts for only some 10% of the gap in the desmin sequence (Figure 1) it is interesting for two reasons. It relates, at the sequence level, chicken desmin and bovine glial fibrillary acidic protein (GFA) and may allow an extension to wool keratins. The 28 residues given in Figure 3 are the only partial sequence data currently available for GFA (Hong and Davison, 1981). They were obtained by Edman degradation on a large cyanogen bromide fragment with an amino acid composition typical of α-helical fragments. Because of its composition (lack of cysteine, proline, and tryptophan), its similarity with the shorter desmin peptide (Figure 3), and its size (132 residues) we tentatively assign the fragment to the proposed amino-terminal coiled-coil region of 10 nm filaments, since GFA and desmin are not only closely related in proteolysis patterns (Rueger *et al.*, 1981), but also show the same relative positioning of their cysteine and tryptophan residues (Geisler *et al.*, 1982b). As indicated in Figure 3, the GFA sequence can be tentatively aligned with residues 62–89 of wool keratin component 8c-1 giving rise to the high homology segment *a* with identities at nine positions (32%) and four very conservative exchanges leading to a homology level of 47%. Within the shorter segment of 17 residues common to all three proteins the homology is even higher, approaching 70%. We can also extend the homology segment *a* to wool α-keratins 7c and 5. A 9-residue segment described earlier for 7c (called no. 3 in Table II of Dowling *et al.*, 1979) has so far not been placed

WK 8 c-1							GFA							CD							WK 7 c						
A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G
MET	GLN	PHE	LEU	ASN	ASP	ARG				(MET)	LEU	ASN	GLU	GLU													
LEU	ALA	SER	TYR	LEU	GLU	LYS	PHE	ALA	ARG	TYR	ILE	GLU	ARG	PHE	ALA	ASX	TYR	ILE	GLU	LYS	PHE	ALA	ALA	PHE	ILE	ASP	LYS
VAL	ARG	GLN	LEU	GLU	ARG	GLU	VAL	VAL	PHE	LEU	GLU	GLU	GLN	VAL	ARG	PHE	LEU	GLU	GLX	GLX	VAL	ARG					
ASN	ALA	GLU	LEU	GLU	SER	ARG	LYS	ARG	ALA	ARG	ALA	ALA	LEU	ASX	ALA	LEU											
ILE	LEU	GLU	ARG	SER	GLN	GLN	LEU	ASP	GLU																		

Fig. 3. Sequence comparison in homology segment *a*. Details of presentation are as in Figure 2 but column two gives a fragment from bovine GFA. The sequence of wool keratin 8c-1 covers residues 59–93 in the known sequence (Crewther *et al.*, 1980; see also Figure 1). The GFA sequence is from Hong and Davison (1981); given the amino acid composition, at most one of the arginine residues shown could also be a histidine residue because of the experimental approach used. Residues labelled in the desmin column are those related to GFA. Residues marked in the 8c-1 sequence are those which relate to either GFA or desmin. The final column provides a short segment of keratin 7c and labelled residues relate to at least one of the other three proteins. Note the potential of a short coiled-coil segment in 8c-1 between residues 59 and 87 (see text). Note that the position of the desmin and GFA sequences are not yet established and that assignment has been made by homology only.

within the 223 amino-terminal residues (Sparrow and Inglis, 1980) and should occur in the yet undetermined region between residues 65 and 123. As suggested in Figure 3 it could be accommodated relative to residues 65–74 of component 8c-1, although this component has been thought to be quite distinct from component 7c outside the 30% identity seen in their α-helical particles, which start within a few residues towards the carboxyl side of segment *a*. Thus, a common sequence relationship between three wool keratins and 10 nm filaments does not seem to be restricted to the carboxy-terminal half of the middle region but, provided future sequence data support our assignment, seems also to emerge in the amino-terminal part (Figure 1).

The analysis of the amino-terminal region in the form of a distinct separate coiled-coil as envisioned from the proteolysis data on 10 nm filaments (Steinert *et al.*, 1980) is currently difficult. This is not only due to some lack of sequence data on component 7c but also because of some peculiarity in proline distribution in the case of the two wool α-keratins (Figure 1) which cannot be easily evaluated without further information on chicken desmin or a related protein. As pointed out by Parry (1979) in a theoretical analysis of the three-stranded rope of fibrinogen, precise delineation of coiled-coil elements is sometimes difficult. Not only is it possible that proper positioning of apolar residues may override some low α-helix-favouring potential of residues found elsewhere in the heptades, but the latter structure may be broken by stutters and an occasionally encountered proline residue. Thus, although the α-helical particles of wool components 8c-1 and 7c seem aligned (Crewther *et al.*, 1978; see also Figure 1), a proline residue present early on in the 7c helical particle is disturbing. In the case of 8c-1 we note the possibility that some 30–40 residues, which seem to relate to the homology region *a* and are possibly of α-helical potential (Figure 2), precede the isolated α-helical particle I, although two closely spaced proline residues (positions 96 and 100; see Crewther *et al.*, 1980) signal a break between the two regions (Figure 1). Thus, a precise estimate of the amino-terminal coiled-coil and a direct relation in size to the postulated carboxy-terminal coiled-coil seems premature and must await further sequence data. For this reason we have also postponed an analysis of the postulated globular spacer proposed to separate two nearly equally sized coiled-coils in 10 nm filaments (Steinert *et al.*, 1980). Although the presence of proline residue(s) in the approximate position of wool α-keratin 8c-1 and desmin may argue for such an interruption, a similar break is not yet indicated by the available data on the wool α-keratin component 7c. Whether such a region exists as a separate entity and can be described as being globular remains to be seen. According to the combined current sequence data a putative interrupting sequence could account for some 20–30 residues and may also be described as a region of low α-helical potential.

The non-α-helical terminal domains

Desmin used as a 10 nm filament prototype, and the two wool components K8c-1 and 7c, reveal non-α-helical amino-terminal domains. These headpieces are difficult to compare directly at the sequence level because of the peculiar wealth of cysteine residues typical for this region in wool α-keratins (Figure 1). Nevertheless, the available sequences and compositions together with multiple proline residues present, clearly point to a non-α-helical structure in which secondary structure prediction rules indicate β-sheet and β-turns for desmin (Geisler *et al.*, 1982a). In the first 40 residues all three headpieces lack acidic residues and their basicity seems due to arginine rather than lysine residues. The precise length of the headpieces seems to vary. Chicken desmin and wool keratin 8c-1 seem to show values close to 65 or 50 residues, respectively, but in the case of component 7c a value >60 residues is envisioned with a more precise evaluation requiring further sequence data between residues 65 and 123 (see also above for the potential length of the amino-terminal helix of the middle region).

The carboxy-terminal non-α-helical regions are difficult to analyze because of the lack of extended sequence information in the case of the wool keratins (Figure 1). As mentioned above, desmin reveals a structural domain of 48 residues, which can be isolated by chymotryptic digestion (Geisler *et al.*, 1982a). Adjacent sequences indicate that the desmin tail sequence following the proposed end of the coiled-coil in the middle region could be of the order of 60 residues. A rather similar size can be expected for the wool component 8c-1 although its wealth of cysteine residues and general composition differs from that of the desmin tail piece. A final comparison with component 7c is difficult because of limited sequence data. What is, however, indicated by all current information is a probably much larger domain – perhaps of the order of 120 residues – whose expected sequences indicated by composition data do not invite α-helical potentials.

Conclusions

The comparison of various wool keratins and 10 nm filament proteins, although still limited because complete amino acid sequences are lacking, leads to several conclusions. The α-keratins of wool seem to share not only a similar structural motif with the 10 nm filaments of living cells but this relation is further indicated by the description of some regions with related amino acid sequences. Two distinct and aligned segments, possibly of special structural importance, reveal homologies at the 70–80% level. Therefore, the α-keratins of wool and hair provide further subgroup(s) of 10 nm filament expression along known pathways of differentiation and cell type specificity. Using the alignment of the sequences given in Figures 1–3 we can also detect a closer relation between two wool keratins, previously thought to be rather different (Dowling *et al.*, 1979), i.e., components 8c-1 and 7c. Furthermore, we can suggest that wool keratins contain not only the previously isolated amino-terminal coiled-coil but may display a similar coil in the carboxy-terminal position where it is already indicated for 10 nm filament proteins (Geisler *et al.*, 1982a). We do note, however, that the various proteins are, even in their α-helical elements, astonishingly distinct.

Throughout our analysis we have placed special emphasis on the possible presence of two coiled-coils suggested by various proteolytic studies dating back to 1973 (Skerrow *et al.*, 1973; Matoltsy, 1975; Steinert, 1978; Steinert *et al.*, 1980). Precise delineation of these domains will require not only more sequence data but must also be combined with future structural data. It is tempting to speculate that the X-ray repeat of 198 Å and the periodicity of 210 Å visualized by electron microscopy in wool microfibrils (Fraser and McRae, 1972; Fraser *et al.*, 1976) and intermediate filaments (Henderson *et al.*, 1982), respectively, are directly related to the presence of two nearly equally long coiled-coils of some 130–140 residues. Although the helical contents of the molecules seem to allow for this possibility, some of the pro-

posed coiled-coils, especially in the case of wool keratins, may become limited in length or locally interrupted by a rare proline residue or a segment of low helical potential. Nevertheless, as long as we remember that this model is only a working hypothesis, the presence of two coiled-coil regions seems currently an appropriate topographical description. None of the discussed or proposed coiled-coils shows the nearly perfect hydrophobic repeat pattern seen in the heptade presentation of the sequence of rabbit muscle tropomyosin (Sodek *et al.*, 1972; McLachlan and Stewart, 1975). It has been speculated (Parry *et al.*, 1977; Parry, 1979) that this may be related to the necessity for allowing a regular, helical repeat pattern of basic and acidic residues on the outer surface of the coil, even if some hydrophobic repeats in *a* and *d* positions have to be dispensed with. The resulting imperfections in the coil structure may be the reason for past difficulty in isolating the proposed carboxy-terminal coiled-coil by proteolysis of α-keratins. We do, however, note that recent, much milder digestion conditions have allowed the isolation of some subfragments in the case of component 8c-1 (Crewther *et al.*, 1980).

The relative alignment of two wool keratins with desmin raises some important questions about several predictions of a general model for 10 nm filaments previously proposed (Steinert *et al.*, 1980). As already stated for desmin (Geisler *et al.*, 1982a), the non-α-helical head domain is not $\sim$1 K but rather 6–7 K. Similarly, the non-α-helical tail domain seems not to be some 24 K but only 7 K for desmin. It is $\sim$7 K for component 8c-1 and $\sim$12 K for component 7c (see Figure 1). In addition, the sequence data on desmin do not allow for either the number or the positions of the tryptophan residues used in proposing the model (see Figure 1 *versus* Steinert *et al.*, 1980). Furthermore, the α-helical spacer envisioned to be able to vary in size between 5 and 17 K, and thought to be the major determinant for mol. wt. variability in the different proteins, seems so far rather small, perhaps in the range of 2–4 K (Figure 1; see above). In addition, we have noticed that the basis for apparent mol. wt. difference in the neurofilament 68-K protein is localized in the carboxy-terminal region, most likely the tailpiece, and not in the spacer proper (Geisler *et al.*, 1982b). Thus, although we have no information of our own on epidermal keratins, which have been used extensively by Steinert (1978, 1980), we feel that in general, size-variability may be mainly reflected by the terminal domains rather than the spacer and we cannot yet even exclude the possibility that even the coiled-coil areas can show some length variations in different proteins. In agreement, we notice that the wool component 7c in line with its higher mol. wt. of 54–56 K (Crewther *et al.*, 1979; Sparrow and Inglis, 1980) has a longer tailpiece than desmin (50 K) and component 8c-1 (50 K) (see Figure 1). It seems unlikely that the coiled-coils of wool keratins and 10 nm filaments differ in rope number. Whereas 10 nm filament structural research has been dominated since 1973 by a three-stranded model (Skerrow *et al.*, 1973; Matoltsy, 1975; Steinert, 1978; Steinert *et al.*, 1980; Zackroff *et al.*, 1981), wool keratin studies have been ambiguous about two- or three-stranded models as already summarized several years ago (Fraser and McRae, 1972; Fraser *et al.*, 1976). However, all data accumulated since 1978 point to a two- rather than a three-stranded coiled-coil (Ahmadi and Speakman, 1978; Ahmadi *et al.*, 1979; Woods and Gruen, 1981). This is further emphasized by model building (McLachlan, 1978) of the sequences of the two α-helical particles of wool keratins. Whereas structural fit was readily obtained for two-stranded models, only instability could be seen for three-roped structures.

Little attention has been drawn so far to the peculiar sequences found in the terminal non-α-helical domains and the distinct differences in cysteine content between wool keratins and 10 nm filament proteins (see Figure 1). For desmin we know that the rod-sized middle domain seems unable to form 10 nm filaments by itself (Geisler *et al.*, 1982a) and therefore one or both of the terminal domains seem to be involved in the transition of protofilaments into the full structure. For the headpiece of desmin, implicated (Geisler *et al.*, 1982a) both in high susceptibility to a recently discovered calcium-dependent protease and a peculiar RNA binding (Nelson and Traub, 1981a, 1981b) there are two indications that its structure may change upon transition to 10 nm filaments. It seems to become less susceptible to proteolysis and also less prone to RNA binding once salt is added, i.e., filament formation is induced. Thus, we tentatively propose that at least the headpiece is involved in filament formation in agreement with experimental data to be published elsewhere (Geisler and Weber, in preparation). It seems, therefore, that this process is made irreversible in the case of wool keratins because of their numerous cysteine residues forming disulfide bonds (Fraser and McRae, 1972). In contrast, the so far characterized 10 nm filament proteins in living cells seem to rely on a different chemistry. Their lack of cysteine residues in these segments indicates a less frozen state. Inspection of the sequence of the desmin headpiece draws attention to a peculiar wealth of arginine residues (Geisler *et al.*, 1982a) possibly involved in salt bridges and electrostatic interactions with the acidic rod portion of neighbouring coiled-coils. Further sequence data, whether from protein or cloned DNA, will surely improve our knowledge of 10 nm filaments but final understanding of these filaments will require a three-dimensional structural model. Towards that aim it may now be possible to correlate the emerging electron microscopical information on paracrystals of epidermal keratins (Aebi *et al.*, 1981) with the X-ray diffraction data on wool (Fraser *et al.*, 1976).

Note added in proof

The now completed amino acid sequence of chicken gizzard desmin confirms the proposed location of the homology *a* segment. In addition, a further strong sequence relation is noted in the previously missing region. This new and very long segment corresponds both in relative location and structure to the α-helical coiled-coil particles of the two wool keratins (Figure 1). Thus the proposed common structure of the large middle domain in hard α-keratins and 10 nm filament proteins is now experimentally verified (Geisler and Weber, in preparation).

References

Aebi,U., Rew,P., and Sun,T.T. (1981) *J. Cell Biol.*, **91**, 234a.

Ahmadi,B., and Speakman,P.T. (1978) *FEBS Lett.*, **94**, 365-367.

Ahmadi,B., Boston,N.M., Dobb,M.G., and Speakman,P.T. (1979) in Parry, D.A.D., and Cramer,L.K. (eds.), *Fibrous Proteins*, Vol. 2, Academic Press, pp. 161-166.

Crewther,W.G., Dowling,L.M., Gough,K.H., Marshall,R.C., and Sparrow, L.G. (1979) in Parry,D.A.D., and Cramer,L.K. (eds.), *Fibrous Proteins*, Vol. 2, Academic Press, pp. 151-175.

Crewther,W.G., Dowling,L.M., and Inglis,A.S. (1981) *Proceedings of the 6th Quinquennial International Wool Textile Research Conference*, Pretoria, Vol. 2, pp. 79-91.

Crewther,W.G., Inglis,A.S., and McKern,N.M. (1978) *Biochem. J.*, **173**, 365-371.

Crick,F.H.C. (1953) *Acta Crystallogr.*, **6**, 689-697.

Day,W.A., and Gilbert,D.S. (1972) *Biochim. Biophys. Acta,* **285**, 503-506.
Dowling,L.M., Gough,K.H., Inglis,A.S., and Sparrow,L.G. (1979) *Aust. J. Biol. Sci.,* **32**, 437-442.
Fraser,R.D.B., and McRae,T.P. (1972) in Kugelmass,I.N. (ed.), *Keratins,* Charles C. Thomas Publications, Springfield, IL, USA.
Fraser,R.D.B., McRae,T.P., and Suzuki,E. (1976) *J. Mol. Biol.,* **108**, 435-452.
Geisler,N., Kaufmann,E., and Weber,K. (1982a) *Cell,* **30**, 277-286.
Geisler,N., Plessmann,U., and Weber,K. (1982b) *Nature,* **296**, 448-450.
Geisler,N., and Weber,K. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 4120-4123.
Henderson,D., Geisler,N., and Weber,K. (1982) *J. Mol. Biol.,* **155**, 173-176.
Hong,B., and Davison,P.F. (1981) *Biochim. Biophys. Acta,* **670**, 139-145.
Jones,L.N. (1976) *Biochim. Biophys. Acta,* **446**, 515-524.
Lazarides,E. (1980) *Nature,* **283**, 249-256.
Matoltsy,A.G. (1975) *J. Invest. Dermatol.,* **65**, 127-142.
McLachlan,A.D. (1978) *J. Mol. Biol.,* **124**, 297-304.
McLachlan,A.D., and Stewart,M. (1975) *J. Mol. Biol.,* **98**, 293-304.
Nelson,W.J., and Traub,P. (1981a) *J. Cell Biol.,* **91**, 232a.
Nelson,W.J., and Traub,P. (1981b) *Eur. J. Biochem.,* **116**,51-57.
Osborn,M., Geisler,N., Shaw,G., Sharp,G., and Weber,K. (1982) *Cold Spring Harbor Symp. Quant. Biol.,* **46**, 413-429.
Parry,D.A.D. (1979) in Parry,D.A.D., and Cramer,L.K. (eds.), *Fibrous Proteins,* Vol. **1**, Academic Press, pp. 393-426.
Parry,D.A.D., Crewther,W.G., Fraser,R.D.B., and McRae,T.P. (1977) *J. Mol. Biol.,* **113**, 449-454.
Renner,W., Franke,W.W., Schmid,E., Geisler,N., Weber,K., and Mandelkow,E. (1981) *J. Mol. Biol.,* **149**, 285-306.
Rueger,D.C., Gardner,E.E., Simonian,H.D., Dahl,D., and Bignami,A. (1981) *J. Biol. Chem.,* **256**, 10606-10612.
Skerrow,D., Matoltsy,A.G., and Matoltsy,M.N. (1973) *J. Biol. Chem.,* **248**, 4820-4826.
Sodek,J., Hodges,R.S., Smillie,L.B., and Jurasek,L. (1972) *Proc. Natl. Acad. Sci. USA,* **69**, 3800-3804.
Sparrow,L.G., and Inglis,A.S. (1980) *Proceedings of the 6th Quinquennial International Wool Textile Research Conference,* Pretoria, Vol. **2**, pp. 237-246.
Steinert,P.M. (1978) *J. Mol. Biol.,* **123**, 49-70.
Steinert,P.M., Zimmermann,S.B., Starger,J.M., and Goldman,R.D. (1978) *Proc. Natl. Acad. Sci. USA,* **75**, 6098-6101.
Steinert,P.M., Idler,W.W., and Goldman,R.D. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 4534-4538.
Weber,K., Osborn,M., and Franke,W.W. (1980) *Eur. J. Cell Biol.,* **23**, 110-114.
Weber,K., and Osborn,M. (1981) *Cell Surf. Rev.,* **7**, 1-53.
Woods,E.F., and Gruen,L.C. (1981) *Aust. J. Biol. Sci.,* **34**, 515-526.
Zackroff,R.V., Steinert,P., Whitman,M.A., and Goldman,R.D. (1981) *Cell Surf. Rev.,* **7**, 56-97.