

# Amino acid sequence data on glial fibrillary acidic protein (GFA); implications for the subdivision of intermediate filaments into epithelial and non-epithelial members

Norbert Geisler and Klaus Weber\*

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

Communicated by K. Weber

Received on 22 July 1983; revised on 8 August 1983

**Determination of 50% of the sequence of the astrocyte-specific intermediate filament (IF) protein documents the hypervariable regions as well as parts of the coiled-coil array of glial fibrillary acidic protein (GFA). The results show that the four non-epithelial IF proteins (myogenic desmin, mesenchymal vimentin, GFA and neurofilament 68 K protein) known to form homopolymers are much more closely related than the epithelial keratins, which seem to form heteropolymers only. Of the four non-epithelial proteins, desmin and vimentin are the most closely related, since GFA has a shorter non- $\alpha$ -helical array at the amino terminus. We discuss the possibility that the non- $\alpha$ -helical terminal arrays, because of their sequence and length variability, are responsible for differences of distinct IF with respect to physical-chemical properties such as the low ionic strength-induced depolymerization into protofilaments.**

**Key words:** intermediate filaments/GFA/neurofilaments/desmin/keratin

## Introduction

Immunological data have been instrumental in subdividing the complex multigene family of intermediate filaments (IF) in a histologically meaningful manner (neurofilaments, myogenic desmin, glial-specific GFA, mesenchymal vimentin, epithelial cytokeratins) (for a recent review, see Osborn and Weber, 1983). Amino acid sequence studies not only confirmed this distinction but also allowed the development of a common structural model for IF proteins (Geisler *et al.*, 1982a). Here, a rod-like middle domain displaying extensive  $\alpha$ -helical arrays able to form interpeptide coiled-coil structure accounts for the  $\alpha$ -type X-ray diffraction patterns known for all IFs. Whereas the rod of some 310 residues is highly related in all IF proteins, the two flanking terminal domains reveal themselves as hypervariable regions differing both in sequence and length (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982). Although this topographical description was based originally only on sequence data for desmin (Geisler *et al.*, 1982a; Geisler and Weber, 1982), vimentin (Geisler and Weber, 1981a; Geisler *et al.*, 1982b), a human 50 K epidermal keratin (Hanukoglu and Fuchs, 1982) and the alignment of partial sequence data for two hard  $\alpha$ -keratins from sheep wool (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982; Weber and Geisler, 1982; Dowling *et al.*, 1983), subsequent studies on a 59 K epidermal keratin (Steinert *et al.*, 1983), the neurofilament 68 K protein (Geisler *et al.*, 1983) and extension of the vimentin sequence (Quax-Jeuken *et al.*, 1983) have fully supported the model. In particular, it has become obvious (Geisler and Weber, 1982) that contrary to previous assumptions (Steinert *et al.*, 1980), even the short regions

separating the coiled-coil arrays of the rods are highly preserved both in sequence and in length.

Here we have characterized the last member of the four major non-epithelial IF proteins, the glial-specific protein GFA (glial fibrillary acidic protein). Determination of the amino-terminal 82 and carboxy-terminal 136 residues fully covers the hypervariable regions and provides additional information on the rod region. In line with several common biochemical properties and their histological expression patterns we show that the four non-epithelial IF proteins form a highly related subgroup, with vimentin and desmin being the closest members.

## Results and Discussion

### Determination of porcine GFA sequences

In order to fully cover the hypervariable regions and still obtain sequence information on the rod domain (Geisler and Weber, 1982) we aimed at two particular fragments. Following our earlier fragmentation studies, GFA was cleaved at its sole cysteine residue (Geisler *et al.*, 1982b) and the carboxy-terminal 15 K fragment was isolated and subjected to sequence analysis. We also probed the molecule for an arginine-rich headpiece by digesting the protein with lysine-specific protease in the presence of 6 M urea followed by chromatography on CM cellulose, a procedure already found useful in the isolation of the corresponding region of the neurofilament 68 K protein (Geisler *et al.*, 1983). This approach provided a 10 K fragment retaining the blocked amino terminus of the whole molecule. Digests obtained by various proteases were separated by paper methods and h.p.l.c. Peptides were recovered and characterized by amino acid composition and stepwise Edman degradation. The combined information characterized the 10 K and 15 K fragments as shown in Figure 1. In each case the sequence gave an amino acid composition in good agreement with the value obtained by acid hydrolysis. In the case of the 15 K fragment we note: (i) although the nature of the three residues following the blocked cysteine are known, their order based only on indirect arguments is still tentative; (ii) the two residues following the cysteine at positions 43 and 59, which are also underlined in Figure 1, have been assigned by compositional data only; (iii) we lack a smaller overlap peptide covering residues 9 and 10 from the carboxyl end, although a larger fragment supports the sequence as given.

The sequence information on the amino-terminal 82 and carboxy-terminal 136 residues of the GFA molecule is added to our previous alignment of chicken gizzard desmin, a 50 K human epidermal keratin, the two major components of sheep wool hard  $\alpha$ -keratin and the sequences known for porcine vimentin (Geisler and Weber, 1982). The alignment for vimentin can now be greatly extended towards the amino-terminal side by the sequence deduced for hamster vimentin using DNA data (Quax-Jeuken *et al.*, 1983). Figure 1 also incorporates a recent report on a murine 59 K epidermal keratin (Steinert *et al.*, 1983) and the partial sequence of porcine neurofilament 68 K protein (Geisler *et al.*, 1983). All sequences emphasize the structural motif proposed for IF pro-

\*To whom reprint requests should be sent.

teins; a highly related coiled-coil domain flanked by hyper-variable terminal domains.

*Relation between GFA and the other three non-epithelial IF proteins*

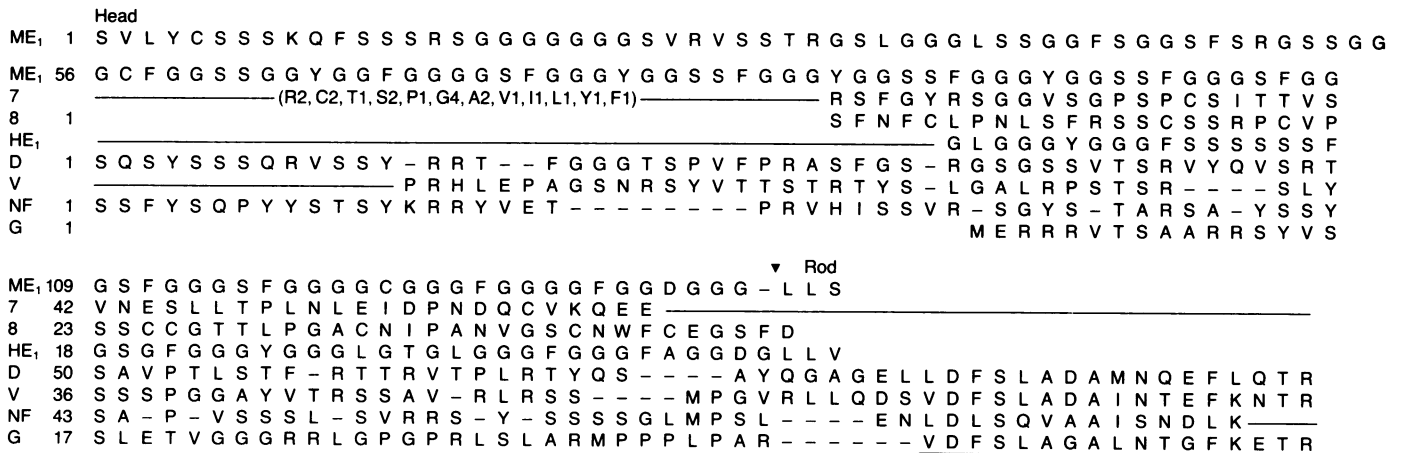
As seen in Figure 1, the presumptive rod domain of GFA is well delineated at its two ends since the data already emphasize the common part of IF proteins corresponding to desmin residues 97–407 (Geisler and Weber, 1982). We note that the two ends of the coiled-coil array, i.e., the start of coil Ia and the end of coil II each approach as predicted a consensus sequence. The common sequence in the latter region explains a monoclonal antibody elicited with human GFA, which seems to recognize all IF proteins (Pruss *et al.*, 1981) and whose epitope lies within the last 20 residues of coil II (Geisler *et al.*, 1983). That the end of coil II may be some three residues earlier than previously thought (Geisler *et al.*, 1982a, 1983) is indicated by the sequence of a mouse epidermal keratin (Steinert *et al.*, 1983) whose non-helical tailpiece sequence starts somewhat closer to the consensus sequence than expected from the sequences of the other proteins. We expect from the sequence data a second type of general IF antibody whose epitope could lie in the high homology segment of coil Ia (i.e., desmin residues 103–121). The formation of mixed dimers upon Cu<sup>2+</sup>-phenanthroline oxidation of certain heteropolymeric IF (Quinlan and Franke, 1982, 1983) taken together with the identical location along coil II of the sole cysteine residues of desmin, vimentin (Geisler and Weber, 1981a) and now also of GFA (Figure 1) may put several restrictions on future detailed models of IF structure.

Our data on porcine GFA are somewhat at variance with an earlier proposed short sequence of 28 residues for the antigenic 15 K cyanogen bromide fragment of bovine GFA (Hong and Davison, 1981), which we have previously aligned by homology into coil Ia (Weber and Geisler, 1982). Some of

the few discrepancies may be species-specific residues although this phenomenon is rare along the rod domain of other IF proteins (Geisler and Weber, 1981a; Quax-Jeuken *et al.*, 1983). The combined results show, however, that the major antigenic site(s) of rabbit anti-GFA antibodies are located on the 15 K CNBr fragment which spans most of coil Ia and Ib (Figure 1). It remains to be seen if the site(s) are on one of the two coils (Ia, Ib) or in the short non- $\alpha$ -helical spacer I located between them.

Whereas residues 46–82 of GFA can be readily aligned with residues 80–116 of desmin, the preceding 45 amino acids of GFA are difficult to accommodate in a straightforward manner (Figure 1). The hypervariability of the headpiece domains has already necessitated the introduction of several gaps in order to align the headpieces of vimentin, desmin, and neurofilament 68 K protein, and even after this is done only a rather remote sequence relation is achieved. Thus, for instance, vimentin and desmin show 70% sequence identity in the rod region (desmin residues 97–407) but only a rare identical tripeptide in their headpieces (Geisler and Weber, 1981a; Quax-Jeuken *et al.*, 1983). The GFA headpiece does not allow a further easy alignment. It is characterized by a noticeably shorter length (46 residues *versus* 70–80) and a clustering of arginine and particularly of proline residues. Nevertheless, the headpieces of all non-epithelial IF proteins are clearly non- $\alpha$ -helical, reveal several  $\beta$ -turns in line with their high proteolytic susceptibility and are rich in arginine, proline, hydroxyamino acids and glycine. They rarely display an acidic residue or a basic amino acid other than arginine (Geisler *et al.*, 1982a, 1983; Quax-Jeuken *et al.*, 1983).

The carboxy-terminal 50 residues of GFA form the variable and non- $\alpha$ -helical tailpiece. This region can be easily aligned with the corresponding arrays in desmin and in vimentin, if allowance for some deletions and amino acid ex-



**Fig. 1.** Sequence relationship between GFA and other IF proteins. Alignment is based on previous arguments (Geisler and Weber, 1982). For primary sequence data see the following references (Geisler *et al.*, 1982a, 1983; Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982; Steinert *et al.*, 1983; Quax-Jeuken *et al.*, 1983; Dowling *et al.*, 1983). Abbreviations for individual proteins are ME<sub>1</sub> (mouse epidermal keratin 59 K), HE 1 (human epidermal keratin 50 K), 8 and 7 (sheep wool hard  $\alpha$ -keratin components 8c-1 and 7c), D (chicken desmin), V (hamster vimentin), G (porcine GFA) and NF (porcine 68 K neurofilament protein). Horizontal lines indicate as yet unestablished sequences. X is an arginine or lysine residue in NF. In the carboxy-terminal part of GFA a few minor ambiguities are underlined (see text). All proteins seem to have a blocking group, most probably the acetyl group, at their amino-terminal residue. The three structural domains are indicated as are the hydrophobic *a* and *d* positions (dots) in the consecutive heptades of the presumptive coiled-coils. Note some general irregularity early in coil II and the *a* to *d* reversal around desmin residue 342 also common to all proteins. Arrowheads mark the location of the isolated desmin rod (Geisler *et al.*, 1982a). A leader sequence (underlined) occurring early in the rod is typical only of non-epithelial IF proteins. Identical residues in all four non-epithelial IF proteins or all proteins are given in bold letters, when the highly related sequence arrays of the rods (desmin residues 97–407) are concerned. Deletions (dashes) allow for better alignment of the short spacer regions and to some extent of the hypervariable non- $\alpha$ -helical terminal domains. In the latter case there is only a rather remote relation of D, V and NF through the entire headpiece region but a convincing homology is seen in the tailpieces of D, V and G, which can be extended with a lower degree of homology to HE1. Due to a mistake in sequence storage and transfer which was not recognised previously, the following residues in the desmin sequence must be corrected (see original data in Geisler and Weber, 1982): 268 I, 270 A, 275 I, 280 I and 281 A.

	Coil Ia	Spacer I
ME, 140	G N G R V T M R N L N D R L A S Y M D K V R A L E E S N Y E L E G K I K E V	V R E A R Q L K P R E P R D Y S K
7	R F A A F I D K V R	R Q - C C E - S N L E P L - -
8 53	G N E K E T M Q F L N D R L A S Y L E K V R Q L E R E N A E L E S R I L E R	S Q Q Q E P L V C P N Y Q S - - -
HE, 50	G S E K V T M Q N L N D R L A S Y L D K V R A L E E A N A D L E V K I R D W	Y Q R Q R P A E I K D Y S P - - -
D 97	T N E K V E L Q E L N D R F A N Y I E K V R F L E Q Q N A L M V A E V N R L	R G K Q - P T R V - - A E - M - -
V 83	T N E K V E L Q E L N D R F A D Y I D K V R F L E Q Q N K I L L A E L E Q L	K G - Q G K S R L - - G D - L - -
G 63	A S E R A E M M E L N D R F A S Y I E K	

	Coil Ib
ME, 195	Y Y K T I E D L K G Q I L T L T T D N A N V L L Q I D N A R L A A D D F R L K Y E N E V T L R Q S V E
7 134	F E G Y I E T L R R E A E C V E A D S G R L S S E L N H V Q E V L E G Y K K K Y E E E V A L R A T A E
8 105	Y F R T I E E L Q Q K I L C A K S E N S R L V I E I D N A K L A S D D F R T K Y E S E R S L R Q L V E
HE, 102	Y F K T I E D L R N K I L T A T V D N A N V L L Q I D N A R L A A D D F R T K Y E T E L N L R M S V E
D 146	Y E E E L R E L R R Q V D A L T G Q R A R V E V E R D N L L D N L Q K L K Q K L Q E E I Q L K Q E A E
V 132	Y E E E M R E L R R Q V D Q L T N D K A R V E V E R D N L A E D I I R L R E K L Q E E M L Q R E E A E

	Spacer II
ME, 246	A D I N G L R R V L D E L T L S Q S V L E L Q I E S L N E E L A Y L K K N L E E E M R D L Q N V S T G
7 185	N E F V A L K K D V D C A Y V R K S D L E A N V E A L I Q E I D F L R R L Y E E E I R V L Q A N I S D
8 156	S D I N S L R R I L D E L T L C K S N L E A E V E S L K E E L L C L K Q N H E E E V N T L R S Q L G D
HE, 153	A D I N G L R R V L D E L T L A R A D L E M Q I E S L K E E L A Y L K K N H E E E M N A L R G Q V G G
D 197	N N L A A F R A D V D A A T L A R I D L E R R I E S L Q E E I A F L K K V H E E E I R E L Q A Q L Q E
V 183	S T L Q S F R Q D V D N A S L A R L D L E R K V E S L Q E E I A F L K K L H D E E I Q E L Q A Q I Q E

	Coil II
ME, 297	- D V N V - E M N N A - P G V D L T - - - Q L L N N M R N Q Y E Q L A E K N R K D A E E
7 236	T S V I V S K M D N S R - - - D L N M D C I V A E E I R A Q Y D D I A S R S R A E A E S
8 207	- R L N V - E V D A A - P T V D L N - - - V R L N E T R A Q Y E A L V E T N R R D V E E
HE, 204	- D V N V - E M D A A - P G V D L S - - - R I L N E M R D Q Y E K M A E K N R K D A E E
D 248	Q H I Q V - E M D I S K P - - D L T - - - A A L R D V R Q Q Y E S V A A K N L Q E A E E
V 234	Q H V Q I - D V D V S K P - - D L T - - - A A L R D V R Q Q Y E S V A A K N L Q E A E E

ME, 335	W F N Q K S K E L T T E I D S N I A Q M S S H K S E I T E L R R T V Q G L E I E L Q S Q L A L K Q S L
7 277	W Y R S K C E E I K A T V I R H G E T L R R T K E E I N E L N R V I Q R R T A E V E
8 245	W Y I R Q T E E L N K Q V V S S S E Q L Q C N Q E E I I E L R R T V N A L Q V E L Q A Q H N L R D S L
HE, 242	W F F T K T E E L N R E V A T N S E L V Q S G K S E I S E L R R T M Q N L E I E L Q S Q L S M K A S L
D 286	W Y K S K V S D L T Q A A N K N D A L R Q A K Q E M L E Y R H Q I Q S Y T C E I D A L K G T N D S L
V 272	W Y K S K F A D L S E A A N R N N D A L R Q A K Q E S N E Y R R Q V Q S L T C E V D A L K G T N E S L
NF	W F K S R F T V L T E S A A K N T D A V R A A K D E V S E S R R L L K A K T L E I E A C X G M N E A L
G	C E V D S L R G T N E S L

ME, 386	E A S L A E T V E S L L R Q L S Q I Q S Q I S A L E E Q L Q Q I R A E T E C Q N A E Y Q Q L L D I K T
7	K A K Q N M A C L L K E Y Q E V M N S K L
8 296	E N T L T E T E A R Y S C Q L N Q V Q S L I S N V E S Q L A E I R G D L E R Q N Q E Y Q V L L D V R A
HE, 293	E N S L E E T K G R Y C M Q L A Q I Q E M I G S V E E Q L A Q L R C E M E Q N Q E Y K I L L D V K T
D 337	M R Q M R E M E R F A G E A G G Y Q D T I A R L E E E I R H L K D E M A R H L R E Y Q D L L N V K M
V 323	E R Q M R E M E E N F A L E A A N Y Q D T I G R L Q D E I Q N M K E E M A R H L R E Y Q D L L N V K M
NF	E K Q L Q E L E D K Q N A D I S A M Q D T I N K L E N E L R T T K S E M A R Y L K E Y Q D L L N V K M
G	E R Q M R E Q E E R H A R E A A S Y Q E A L T R L E E E G Q S L K D E M A R H L Q E Y Q E L L N V K M

	Tail ▼
ME, 437	R L E N E I Q T Y R S L L E G E G S S S G G G G R R R G G S G G G S Y G G S S G G G S Y G G S S G G G
7	G L D I E I A T Y R R L L E G E E Q R L C Q G V - G A V - N V (V, S, S) - R
8 347	R L E C E I N T Y R G L L D S E D C K L A C G K P L T P C I S S P - - - C A
HE, 344	R L E Q E I A T Y R R L L E G E D A H L S S S - - Q - F - S S G S Q S S R D V T S S S R Q I - - - R
D 388	A L D V E I A T Y R K L L E G E E N R I S I P M H Q T F A S A L N F R E T S P D Q R - G S E V H T K K
V 374	A L D I E I A T Y R K L L E G E E S R I S L P L - P N F - S S L N L R E T N L E S L P L V D T H S K R
NF	A L D I E I A A Y R K L L E G E E T R L S F T S V G S L T T G Y S Q S S Q V F G R S A Y G G L Q T S S
G	A L D I E I A T Y R K L L E G E E N R I T I P V - Q T F - S N L Q I R E T S L D T K S V S E G H L K R

ME, 488	G S Y G G S S G G G S Y G G G S S G C G G R G G G S G G G Y G G G S S S G G A G G R G G G S G G G Y
7	(R1, C9, B3, T4, S8, Z1, P3, G13, A3, V7, L5) R C
8	P A A P C T T - - - - C V V P S S C G (R, R, Y)
HE, 387	T K V M D V H - - - D G K V V S T H E Q V L R T K N 409
D 438	T V M I K T I E T R D G E V V S E A T Q Q Q H E V L 463
V 423	T L L I K T V E T R D G Q V I N E T S Q H H D D L E
NF	Y L M S T R T S F P S Y T S H V Q E E Q I E V E E T I E A A K A E E A K D E P P S E G E A E E E G K E
G	N V - - K T S V E M R D G E V I K E S K Q E H K D V

ME, 539	G G G G S S G R R G G S G G F S G T S G G G D Q S S K G P R Y 569
NF	K E E A E A E A E A E E E G A Q E E E E A A E K E E S E E A K E E E G G E G E Q G E E T K E A E E E E

NF K K D E G A G E E Q A T K K K D

changes is made as suggested in Figure 1. Thus, of the four non-epithelial IF proteins, desmin, vimentin and GFA are the closest related members when only their coiled-coil arrays are compared. In the currently established rod regions the sequence identity levels lie around 70%, a value noticeably higher than the 55% seen for the neurofilament 68 K protein. The striking relation between these proteins is still sometimes considered as to indicate only some homology (Quinlan and Franke, 1983). It should therefore be stressed that the sequence identity levels observed in the rods approach those known for more distant vertebrate haemoglobins. In the headpiece region GFA diverges noticeably from the three other non-epithelial IF proteins, whereas in the tailpiece domain the neurofilament protein is the most remote. Thus, in an overall view, desmin and vimentin are clearly the closest pair of the non-epithelial IF proteins.

The suggestion from gel electrophoresis that GFA has most likely a shorter polypeptide length than desmin (Rueger *et al.*, 1981) is now confirmed. Since the common rod domain shows very little change in length even in the spacer regions of the more remote IF proteins (Figure 1), we would predict from the noticeably shorter headpiece that porcine GFA is some 4 K smaller than chicken desmin where the chemically determined mol. wt. is 53 K (Geisler and Weber, 1982). The good alignment at the carboxyl end and the presence of a blocking group at the amino-terminal methionine indicate that GFA purified from spinal cord is not a degraded molecule. The data on the highly protease-sensitive headpiece (for enzymatic digestion studies see Rueger *et al.*, 1981) and its sequence organization also explain several studies on GFA from different glial material reporting multiple components with different solubilities, mol. wts. and isoelectric points (see for instance, Dahl and Bignami, 1975; Newcomb *et al.*, 1982; Bigbee *et al.*, 1983). As discussed before for desmin and vimentin (Geisler *et al.*, 1982a) a zipper-type proteolytic removal of fragments from the basic headpiece must lead to a more pronounced influence of the acidic rod portion and therefore gives rise, in two-dimensional gels, to a 'staircase pattern' of fragments with decreasing mol. wt. and isoelectric point, as also observed for GFA (Bigbee *et al.*, 1983). In addition, it has been shown that the 40 K rod domain of desmin, which can be obtained by chymotryptic digestion *in vitro*, behaves as a soluble protein at physiological conditions where normal desmin filaments are insoluble (Geisler *et al.*, 1982a). Thus the lower mol. wt. derivatives of GFA, which probably arise due to the action of a  $\text{Ca}^{2+}$ -activated protease, should also become more soluble. In order to avoid this enzyme activity our preparation (see Materials and methods) includes EGTA as do a few previous procedures (Rueger *et al.*, 1979, 1981).

#### *Non-epithelial IF proteins form a closely related subgroup*

Whereas all non-epithelial IF proteins are able to form homopolymeric 100 Å filaments *in vitro* (Rueger *et al.*, 1979; Geisler and Weber, 1981b; Steinert *et al.*, 1981) similar studies on epidermal keratins indicate the necessity for obligatory heteropolymers (Steinert *et al.*, 1976, 1981), and no epithelial cell type displaying a single cytokeratin is known (Moll *et al.*, 1982). Current sequence data offer only a partial answer for this difference between epithelial and non-epithelial IF. Although at least parts of the terminal domains are necessary for filament integrity (Geisler *et al.*, 1982a) the framework of the structure itself seems to rely on the interactions between coiled-coils (Fraser *et al.*, 1976). With the

molecular characterization of GFA (see above) and the neurofilament 68 K protein (Geisler *et al.*, 1983) the previous speculation that the rod sequences of keratins are more remote (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982) is now consolidated. Among non-epithelial proteins, sequence identity of at least 55% is seen whereas transition to the keratins provides a much lower value of ~30%. In addition, the two hard  $\alpha$ -keratins (8c-1 and 7c) known to interact *in vitro* when their coil Ib fragments are mixed (Gruen and Woods, 1983) share only some 30% sequence identity. Thus it seems that keratin structure requires at least two distinct rod domains, either as  $\alpha$ -helices or as coiled-coils, which by acting in a complementary manner, lead to a self assembly-competent component. In non-epithelial IF the necessary information seems already contained within a single protein. Since the two epidermal keratin sequences that are currently known clearly reflect component 8c-1 it is to be expected that there are other epidermal keratins more closely related to 7c and that the two prototypes interact to allow self assembly.

Rod sequences of various IF proteins reveal a pronounced 28 residue repeat with a regular pattern of positive and negative charges thought to be responsible for highly specific electrostatic interactions necessary to pack parallelly or anti-parallelly oriented coiled-coils (Parry *et al.*, 1977; McLachlan and Stewart, 1982). It remains to be seen if comparative Fourier analysis can locate the differences between non-epithelial IF proteins and keratins. In this context, we note that sequence homology between the two subgroups is not randomly distributed (Geisler and Weber, 1982). The start of coil Ia, the end of coil Ib and both ends of coil II are particularly well preserved among all IF proteins and these consensus sequences should be of general structural importance. In addition, we have observed that all non-epithelial IF proteins reveal a leader sequence with the potential for  $\alpha$ -helix formation but with poor coiled-coil forming ability. This sequence of 17 residues precedes the highly related part of the rod (Figure 1, underlined). Its function, if any, remains unknown but it has not been detected in the keratin sequences so far established.

#### *Divergence of the terminal hypervariable domains*

Since previous descriptions of these regions were restricted to a rather small sample of proteins (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982; Steinert *et al.*, 1983) one could not foresee how general the proposed hypervariability would be. As summarized in Figure 1 we now recognize several distinct motifs. The long tandem repeats of several glycine residues flanked by large hydrophobic residues can occur with epidermal keratins both in the headpieces and in the tailpieces (Hanukoglu and Fuchs, 1982; Steinert *et al.*, 1983). That such sequences are not a necessary requirement for keratin structure is shown by the hard  $\alpha$ -keratins of sheep wool where proline- and cysteine-rich regions reveal a different chemistry probably leading to extensive disulfide cross-linking (Crewther *et al.*, 1980; Sparrow and Inglis, 1980; Weber and Geisler, 1982). The headpieces of all non-epithelial proteins are arginine- and proline-rich (Geisler *et al.*, 1982a, 1983; Quax-Jeuken *et al.*, 1983, and Results), and curiously a preference for arginine can also be seen in some of the emerging head- and tailpieces of epidermal keratins (Steinert *et al.*, 1983). Finally, the neurofilament triplet proteins have extremely complicated and very long tailpieces, where one subdomain has been shown to contain 46 glutamic acids within 106 residues (Geisler *et al.*, 1983; Figure 1).

Although the coiled-coils and their interaction must be the key to IF structure, the non- $\alpha$ -helical terminal domains have an independent and necessary contribution to filament stability. The rod domain of desmin (residues 73–415) isolated by chymotryptic trimming is at physiological salt and pH a soluble protein, although intact desmin forms extensive filament arrays under the same conditions (Geisler *et al.*, 1982a; Geisler and Weber, 1982). However, it is difficult to decide if differences in physical-chemical properties of distinct IF such as low ionic strength-induced depolymerization into protofilaments are only due to changes in the terminal domains or also to differences in the organization of higher order coiled-coil interactions or both. Thus further biochemical and more sequence data are required to pinpoint the higher tendency of GFA versus desmin/vimentin to depolymerize in low ionic strength buffers (Rueger *et al.*, 1979; Steinert *et al.*, 1981). Nevertheless, it is very likely that the extreme insolubility of epidermal keratins reported previously (Steinert, 1978) should be related to their unique 'polyglycine' tracts, which can occur in either one or both terminal domains and possibly form a special structure. Although all epidermal keratins characterized by amino acid composition seem rich in glycine (Zackroff *et al.*, 1981; Hanukoglu and Fuchs, 1982) this is not the case for certain cytokeratins from interior epithelia. We have recently found that a higher mol. wt. cytokeratin of porcine intestinal epithelium has a nearly normal glycine content and that the cytokeratin mixture of these cells shows a solubility more related to desmin than to epidermal keratins (our unpublished results). Thus a further possibility is indicated to account for the broad spectrum of molecularly distinct keratins.

## Materials and methods

GFA was isolated from porcine spinal cord by a modification of our general procedure to isolate IF proteins. After extraction with a physiological buffer [0.1 M MES, pH 6.5; 0.1 M NaCl, 5 mM EGTA, 1 mM phenylmethylsulphonyl fluoride (PMSF), 0.5 mM dithiothreitol (DTT)] followed by buffer with 1% Triton X-100 and again by buffer, the insoluble material was dissolved in 6 M urea buffer in the presence of EGTA and chromatographed on DEAE-cellulose using a shallow salt gradient. GFA eluted well-separated from the later emerging neurofilament triplet proteins (Geisler and Weber, 1981b). GFA purified in this manner can be isolated within a few days in 50 mg amounts. It has a polypeptide chain of apparent mol. wt. 51 000 and is obtained in 95% purity. The material is specifically detected in Western blots by two GFA antisera which include a rabbit serum kindly provided by A. Bignami. Removal of urea under the conditions described by Rueger *et al.* (1979) leads to extensive formation of 10 nm filaments.

Protein was dialyzed against water and lyophilized. The amino-terminal 10 K fragment was obtained by digestion with lysine-specific protease (Boehringer Mannheim, FRG) in 6 M urea followed by chromatography on CM-cellulose as described for neurofilament 68 K protein (Geisler *et al.*, 1983). The carboxy-terminal 15 K fragment resulted from treatment with 2-nitro-5-thiocyanobenzoic acid leading to cleavage at the single cysteine residue (Geisler *et al.*, 1982b). The fragment was isolated by gel filtration through Sepharose 6B in the presence of urea. Digestion with trypsin, chymotrypsin, thermolysin and V8 protease as well as separation by two-dimensional paper methods and h.p.l.c. was as before. Peptides were recovered and characterized by amino acid composition and stepwise degradation using either the normal dansyl-Edman or a modified technique (Geisler and Weber, 1982a; Geisler *et al.*, 1983). Information from the Offord plot was used to check the acid and amide assignments. Polyacrylamide gel electrophoresis in the presence of SDS was used to monitor purity of GFA and its fragments.

## Acknowledgements

We thank W. Koch for preparing GFA and greatly appreciate the protein-chemical expertise of U. Plessmann. Dr. A. Bignami kindly supplied a sample of rabbit anti-GFA serum. This work was supported by the Max Planck Society and a grant from the Deutsche Forschungsgemeinschaft to N.G. and K.W.

## References

- Bigbee, J.W., Bigner, D.D., Pegram, C. and Eng, L.F. (1983) *J. Neurochem.*, **40**, 460-467.
- Crewther, W.G., Dowling, L.M. and Inglis, A.S. (1980) *Proceedings of the 6th Quinquennial International Wool Textile Research Conference*, Pretoria, Vol. 2, pp. 79-91.
- Dahl, D. and Bignami, A. (1975) *Biochim. Biophys. Acta*, **386**, 41-51.
- Dowling, L.M., Parry, D.A.D. and Sparrow, L.G. (1983) *Biosci. Rep.*, **3**, 73-78.
- Fraser, R.D.B., MacRae, T.P. and Suzuki, E. (1976) *J. Mol. Biol.*, **108**, 435-452.
- Geisler, N. and Weber, K. (1981a) *Proc. Natl. Acad. Sci. USA*, **79**, 4120-4123.
- Geisler, N. and Weber, K. (1981b) *J. Mol. Biol.*, **151**, 565-571.
- Geisler, N. and Weber, K. (1982) *EMBO J.*, **1**, 1649-1656.
- 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., Kaufmann, E., Fischer, S., Plessmann, U. and Weber, K. (1983) *EMBO J.*, **2**, 1295-1302.
- Gruen, L.C. and Woods, E.F. (1983) *Biochem. J.*, **209**, 587-595.
- Hanukoglu, I. and Fuchs, E. (1982) *Cell*, **31**, 253-252.
- Hong, B. and Davison, P.F. (1981) *Biochim. Biophys. Acta*, **670**, 139-145.
- McLachlan, A.D. and Stewart, M. (1982) *J. Mol. Biol.*, **162**, 693-698.
- Moll, R., Franke, W.W., Schiller, D.L., Geiger, B. and Krepler, R. (1982) *Cell*, **31**, 11-24.
- Newcomb, J., Glynn, P. and Cuzner, M.L. (1982) *J. Neurochem.*, **38**, 267-274.
- Osborn, M. and Weber, K. (1983) *Lab. Invest.*, **48**, 372-394.
- Parry, D.A.D., Crewther, W.G., Fraser, R.D.B. and MacRae, T.P. (1977) *J. Mol. Biol.*, **113**, 449-454.
- Pruss, R.M., Mirsky, R., Raff, M.C., Thorpe, R., Dowling, A.J. and Anderton, B.H. (1981) *Cell*, **27**, 419-428.
- Quax-Jeuken, Y.E.F.M., Quax, W.J. and Bloemendal, H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3548-3552.
- Quinlan, R.A. and Franke, W.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3452-3456.
- Quinlan, R.A. and Franke, W.W. (1983) *Eur. J. Biochem.*, **132**, 477-484.
- Rueger, D.C., Huston, J.S., Dahl, D. and Bignami, A. (1979) *J. Mol. Biol.*, **135**, 53-68.
- Rueger, D.C., Gardner, E.E., Simonian, H.D., Dahl, D. and Bignami, A. (1981) *J. Biol. Chem.*, **256**, 10606-10612.
- 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., Idler, W.W. and Zimmerman, S.B. (1976) *J. Mol. Biol.*, **108**, 547-567.
- Steinert, P.M., Idler, W.W. and Goldman, R.D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4534-4538.
- Steinert, P.M., Idler, W.W., Cabral, F., Gottesman, M.M. and Goldman, R.D. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3692-3696.
- Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.D. (1983) *Nature*, **302**, 794-800.
- Weber, K. and Geisler, N. (1982) *EMBO J.*, **1**, 1155-1160.
- Zackroff, R.V., Steinert, P.M., Whitman, M.A. and Goldman, R.D. (1981) *Cell Surf. Rev.*, **7**, 56-97.

## Note added in proof

A recent DNA sequence study on a 56-kd human epidermal keratin shows a very close sequence relationship with wool  $\alpha$ -keratin 7c (I. Hanukoglu and E. Fuchs, *Cell*, **33**, 915-924, 1983). Thus, as predicted from  $\alpha$ -keratins, there are only two prototype rod sequences present in epidermal keratins.