The EMBO Journal Vol.2 No.8 pp.1295 – 1302, 1983

Neurofilament architecture combines structural principles of intermediate filaments with carboxy-terminal extensions increasing in size between triplet proteins

Norbert Geisler, Eckhard Kaufmann, Stefan Fischer, Uwe Plessmann and Klaus Weber*

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

Communicated by K. Weber
Received on 22 April 1983

Mammalian neurofilament triplet proteins (68 K, 160 K and 200 K) have been correlated by a biochemical, immunological and protein chemical study. The 160 K and 200 K triplet proteins are intermediate filament proteins in their own right, since they reveal the α-helical coiled-coil rod domain analyzed in detail for the 68 K protein. Triplet proteins display two distinct arrays. Their amino-terminal region built analogously to non-neuronal intermediate filament proteins should allow a co-polymerization process *via* the interaction of coiled-coil domains. The extra mass of all triplet proteins is allocated to carboxy-terminally located extensions of increasing size and unique amino acid sequences. These may provide highly charged scaffolds suitable for interactions with other neuronal components. Such a domain of 68 K reveals, in sequence analysis, 47 glutamic acids within 106 residues. The epitope recognized by a monoclonal antibody reacting probably with all intermediate filament proteins has been mapped. It is located within the last 20 residues of the rods, where six distinct intermediate filament proteins point to a consensus sequence.

Key words: intermediate filaments/neurofilaments/desmin/α-keratin/coiled-coils

Introduction

Slow axonal transport studies (Hoffman and Lasek, 1975; Micko and Schlaepfer, 1978) and biochemical purification schemes (Liem *et al.*, 1978; Delacourte *et al.*, 1980; Geisler and Weber, 1981a) have shown that mammalian neurofilaments in general contain three major polypeptides, the triplet proteins. These have apparent mol. wts. of 68 K, 160 K and 200 K. Isolated 68 K protein has been related in sequence to non-neuronal intermediate filament proteins by an analysis of a short 43 residue fragment (Geisler *et al.*, 1982a), and by its ability to form homopolymeric smooth 10 nm filaments (Geisler and Weber, 1981a; Liem and Hutchinson, 1982) revealing the 21 nm lateral periodicity typical of other intermediate filaments (Henderson *et al.*, 1982; Milam and Erickson, 1982). Very little is known about the two high mol. wt. proteins. So far, they seem unable to form filaments on their own but are readily incorporated when mixed with 68 K protein (Geisler and Weber, 1981a; Liem and Hutchinson, 1982). Antibodies have given a rather complex view. Some are specific for an individual triplet component, some show cross-reaction between two or more polypeptides (see, for instance, Willard and Simon, 1981; Shaw and Weber, 1981; Debus *et al.*, 1982; Lee *et al.*, 1982; Anderton *et al.*, 1982), and there is even a monoclonal antibody which can react with probably all intermediate filament proteins including the three triplet polypeptides (Pruss *et al.*, 1981). Thus, both the relation between the individual triplet proteins and their mutual interaction is still unclear. The two high mol. wt. proteins have been considered either as peripherally bound associated proteins (Willard and Simon, 1981; Sharp *et al.*, 1982; Geisler and Weber, 1981a) or to repeat tandemly the α-helical coiled-coil arrays present in the 68 K protein and the non-neuronal intermediate filament proteins (Anderton, 1981).

We have used a combined biochemical, protein chemical and immunological approach to study the triplet proteins. Mild *in vitro* proteolysis with chymotrypsin shows that all three proteins reveal the domain structure typical of several non-neuronal intermediate filament proteins, i.e., a basic amino-terminal non-α-helical headpiece followed by an α-helical rod of some 330 residues, which is able to form interpolypeptide coiled-coils, and then by a carboxy-terminal tailpiece (Geisler *et al.*, 1982b; Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982). Tailpiece extensions of unique amino acid sequences account both for the extra mass of all triplet proteins and their 'neurofilament specificity'. We propose that the two high mol. wt. triplet proteins have two major and distinct regions. Their amino-terminal part resembles other intermediate filament proteins and allows their firm incorporation into the filament backbone by a co-polymerization process with 68 K. Their tailpieces providing a scaffold for interactions with other components.

Results

Mild chymotryptic digestion provides stable subdomains

Figure 1 shows the gel patterns after a short incubation with chymotrypsin. All three triplet proteins gave rise to components of apparent mol. wt. 40 K. Whereas these products were well-defined in the case of 68 K and 160 K, the 200 K protein yielded several closely spaced components, which had much lower protease stability since they were seen only in brief digests. The 40 K components are shown below to reflect an α-helical rod-like domain resembling the muscle desmin rod fragment known to contain interpolypeptide coiled-coil arrays (Geisler *et al.*, 1982b; Geisler and Weber, 1982). The second degradation product, corresponding to a major part of the non-coiled-coil carboxy-terminal tailpieces, increases in size from ~12 K (68 K protein) through 100 K (160 K protein) to the 150 K component derived from the 200 K protein. The latter two mol. wts. are apparent values indicated by SDS-polyacrylamide gel electrophoresis. Smaller chymotryptic fragments too short for gel analysis were further characterized by fingerprinting on paper using a native digest of 68 K protein. They account for the amino-terminal basic headpiece and the remainder (domain *a*) of the tailpiece (see below).

Rod-like middle domains and extensions of the amino acid sequence of 68 K

The three purified 40 K components revealed a c.d. pattern typical of α-helical proteins with an α-helix content in the range of 70–85% (Figure 2). Amino acid compositions (Table I) resembling those of the desmin rod (Geisler *et al.*, 1982b) are in agreement with this property. As expected for

*To whom reprint requests should be sent.

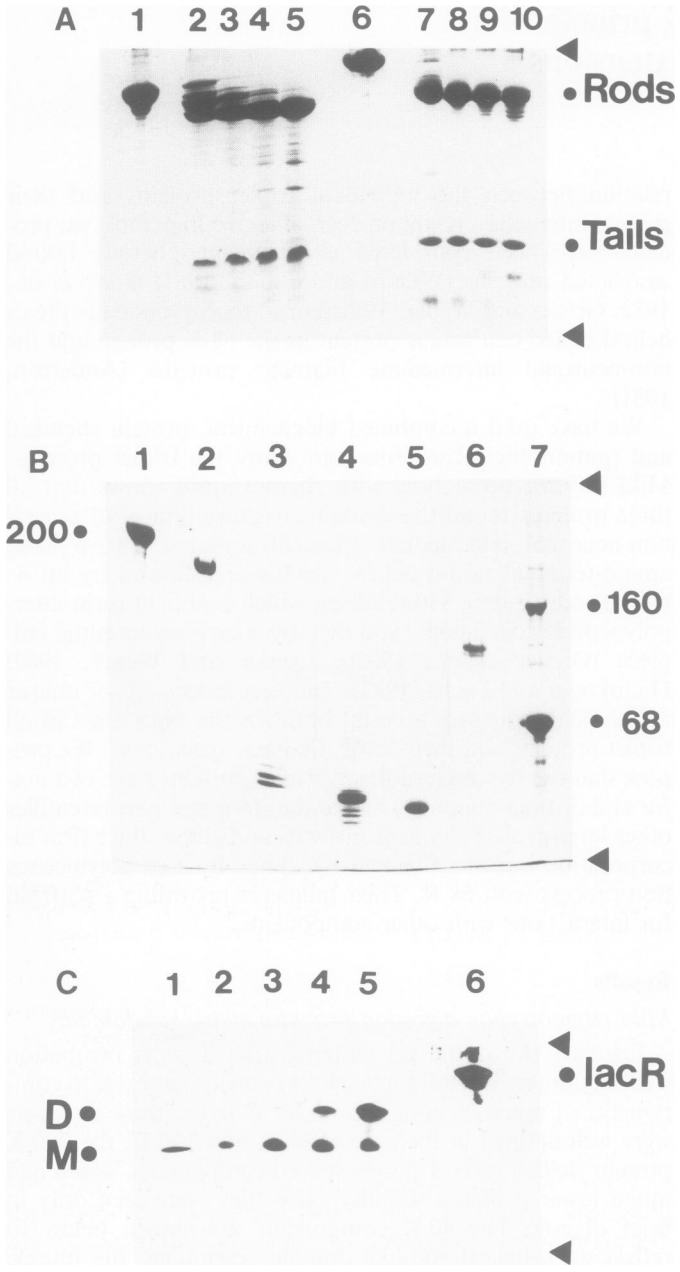

Fig. 1. Time course of a chymotryptic digestion of 68 K protein and purity of the different domains of the triplet proteins in gel electrophoresis. **(A)** Chymotryptic digestion of 68 K protein (**slots 6** and **7–10**) compared with digestion of chicken desmin (**slots 1** and **2–5**). The native proteins (**1** and **6**) are rapidly converted into the rod and tail domains (digestion times 5, 15, 30 and 60 min). For the 68 K protein the resistant tail corresponds to domain *b* '(see text). Digestion of headpieces and domain *a* (68 K protein) leads to small mol. wt. peptides. **(B)** Fragments derived by chymotryptic digestion of triplet proteins. **Slots 2** and **3**, the tailpiece and the rod derived from 200 K protein. **Slot 4**, the rod derived from 68 K protein (for tailpiece see **C, slot 1**). **Slots 5** and **6**, the rod and the tail derived from 160 K protein. Note that the rod in 200 K protein is rather sensitive to further digestion (see text). **Slot 1**, pure 200 K protein and **slot 7**, a mixture of the other triplet proteins. **(C)** Cross-linking of the tail domain *b* of the 68 K protein with dimethyl suberimidate. **Slot 1**, unreacted fragment (mol. wt. 12 000); **slots 2–5**, the increase in dimer formation (D; apparent mol. wt. 25 000) by increasing concentration of cross-linker (see Materials and methods). **Slot 6**, *lac* repressor (mol. wt. 37 000) used as marker.

α-helices arranged as coiled-coils, the compositional data reveal a very low level of proline and high amounts of glutamic acid, leucine, arginine and alanine. In addition, as

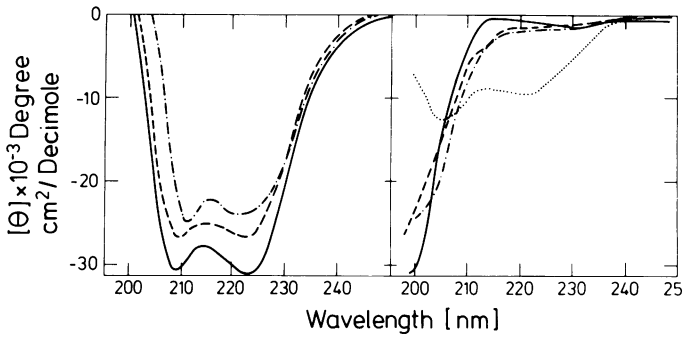

Fig. 2. C.d. spectra of rod **(left panel)** and tail **(right panel)** domains of the three triplet proteins. —·—·—·(200 K-derived), — — — (160 K-derived) and ——— (68 K-derived). Note that at neutral pH the rod domains are mainly α-helical while the tail domains show spectra related to random coils. The *b* domain of the 68 K tailpiece can be introduced to form some secondary structure with α-helical elements at pH 4.0 (see text). Ellipticities were calculated using amino acid compositions (Table I). Protein concentrations were determined by amino acid analysis.

shown below, all rod domains are related to each other and the desmin rod by the presence of the epitope recognized by a mouse monoclonal antibody in most if not all intermediate filament proteins (Pruss *et al.*, 1981).

The 43 residue fragment of 68 K previously aligned in parallel to residues 286–329 of the desmin polypeptide lies within the rod domain (Geisler *et al.*, 1982a; Geisler and Weber, 1982). We have extended the comparison by a protein chemical characterization of the fragment spanning the remaining carboxy-terminal part in the 68 K protein. It was obtained by cleavage at the single cysteine residue with 2-nitro-5-thiocyanobenzoic acid (Geisler *et al.*, 1982a; see also Figure 3), and was subjected to either enzymatic cleavage using trypsin, thermolysin, lysine protease, protease V8 or to fragmentation with CNBr. Peptides separated by two-dimensional paper electrophoresis and h.p.l.c. were analyzed by amino acid composition, stepwise Edman degradation and in one case by automatic Edman degradation (see Materials and methods). The combined results allow us to propose the amino acid sequence shown in Figure 3.

The extreme wealth of glutamic acid residues present at the carboxyl side of the large fragment has made sequence analysis sometimes difficult. We therefore consider the proposal a working hypothesis (Figure 3). Nevertheless, the accumulated data allow the following conclusions. First, starting with the uniquely positioned tryptophan residue (desmin 286 position), the 68 K triplet protein and desmin are aligned for 122 residues. This alignment is very good. No introduction of gaps or addition of residues is necessary to detect the consecutive heptades thought to be responsible for the formation of interpolypeptide coiled-coils. Thus, nearly the complete helix II of the desmin model (Geisler and Weber, 1982) has its determined counterpart in the rod domain of 68 K protein and a sequence identity of 54% is easily detected. Second, the sequence data divide the following tailpiece into two arrays. The first 43 residues (domain *a*) are very low in charged residues revealing only two arginines. They are rich in hydroxy amino acids including tyrosine. Whereas domain *a* is highly susceptible to chymotrypsin, the following domain *b* (106 residues) is resistant and shows a totally different sequence. It lacks the following eight amino acids: asparagine, leucine, tyrosine, phenylalanine, arginine, cysteine, methionine and tryptophan. Instead it is very rich in glycine (10%), alanine (15%), lysine (12%) and particularly in

Table I. Amino acid composition of the three domains in the triplet proteins

	200 K tail	160 K tail	68 K tail (*b*)	200 K rod	160 K rod	68 K rod	D rod	160 K head	68 K head	D head
Asx	4.3	3.2	3	10.5	8.0	8.4	10.7	4.5	4	–
Thr	2.6	1.6	4	4.1	3.3	3.5	3.5	5.6	3	10
Ser	7.0	5.5	3	4.6	9.3	7.0	2.9	21.6	25	18
Glx	23.5	30.3	52	21.6	22.7	25.4	23.2	5.9	4	3
Pro	11.2	7.4	2	2.4	1.6	1.4	0.7	4.2	4	4
Gly	4.1	8.6	9	4.5	4.8	4.0	2.6	8.4	2	7
Ala	11.6	11.8	16	11.0	9.7	12.9	10.7	3.0	5	3
Cys	n.d.	n.d.	–	0.8	n.d.	0.3	0.3	–	–	–
Val	6.9	8.3	2	5.5	4.3	3.8	4.1	4.1	6	7
Met	0.5	0.5	–	1.6	1.3	2.3	3.2	–	1	–
Ile	0.8	1.4	2	2.6	2.7	2.9	5.2	–	2	–
Leu	2.6	2.9	–	11.0	8.0	10.6	11.6	4.3	6	2
Tyr	0.8	0.4	–	2.7	2.9	2.1	2.9	3.9	9	1
Phe	0.8	0.2	–	2.4	2.3	2.2	2.3	5.0	1	4
Lys	20.5	16.7	12	5.6	5.8	6.8	5.2	–	2	–
His	0.7	0.3	1	1.4	3.3	0.9	1.7	–	1	–
Arg	1.7	1.2	–	7.7	5.3	5.8	8.7	9.4	7	10
Trp	n.d.	n.d.	–	0.4	n.d.	0.3	0.3	–	–	–
	~100	~100	106	~100	~100	~100	~100	~80	82	73

Values determined by standard amino acid analysis are compared with the established domains in muscle desmin (Geisler *et al.*, 1982b). Values are expressed as mol% when the sum is given as ~100. Otherwise the sum is taken from sequence data or from the estimates of domain size (headpiece of 160 K protein). Composition of 68 K tail covers only domain *b*. For details see text. Boxes indicate differences in the compositions of the tailpieces. n.d. not determined; dash indicates absence of residue.

glutamic acid (44%). The final residue of domain *b* is aspartic acid as previously shown by end group analysis of intact 68 K protein (Geisler *et al.*, 1982a). Because of the extreme sequence divergence beyond helix II no attempt was made to align the short tailpiece of desmin with the much longer tailpiece of the 68 K protein in Figure 3.

Although secondary structural rules applied to domain *b* indicate several long α-helical stretches, c.d. data revealed only random coil in samples prepared at pH 7.8 (Figure 2). This is in line with the known inability of poly-L-glutamic acid to form α-helix at neutral pH (Doty, 1959; Chou and Fasman, 1978). As known for this model system, partial protonation of the carboxyl groups induces α-helix and at pH 4 we have observed some ordered structure for domain *b*. That this domain must have, even at physiological pH, some secondary structure is shown by chemical cross-linking studies using dimethyl suberimidate. The *b* domain is easily recognized as a dimer by this procedure (Figure 1C).

The carboxy-terminal 11 residues of α-tubulin form a rather short but separate domain, which is very acidic due to several residues of glutamic acid (Postingl *et al.*, 1979). Given this sequence peculiarity, we note a certain similarity with residues 44–53 of the *b* domain.

Mapping the epitope recognized by a monoclonal antibody in diverse intermediate filament proteins

The antibody was used on various fragments derived both from chicken desmin and from the three porcine triplet proteins. Table II confirms its reaction with these proteins (Pruss *et al.*, 1981) and shows that the antigenic site is retained in the 40 K rod domain of all proteins. Thus it must lie between residues 69 and 415 of the desmin polypeptide (Geisler *et al.*, 1982b; Geisler and Weber, 1982). Use of desmin fragments obtained by cleavage at the sole cysteine or tryptophan restricts the location to residues 325–415. A refined analysis was possible using the CNBr fragments of fragment II of the 68 K triplet protein. A large fragment reaching from the carboxyl end into the last part of helix II was aligned there with desmin residues following position 387 (Figure 3; see Materials and methods). As shown in Table II, this fragment retained the antigenic site. Since the few residues following helix II in the rod domain show strong sequence variability among the different proteins (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982; see also Figure 3), the combined results assign the epitope to the region of residues 388–408 of the desmin molecule. Inspection of the available sequences shows that these 20 residues are among the most conserved arrays along the rod domains of six known intermediate filament proteins (see Geisler and Weber, 1982, and also Figure 3).

The carboxy-terminal tailpieces are acidic, highly charged non-coiled-coil fragments of increasing size

The second major fragments of 160 K and 200 K obtained by mild chymotryptic digestion have apparent mol. wts. of 100 K and 150 K, respectively (see above). C.d. studies of the purified fragments after removal of the urea (see Materials and methods) showed no α-helical elements and pointed instead to random coils. Although we cannot exclude the possibility that these domains have not renatured properly,

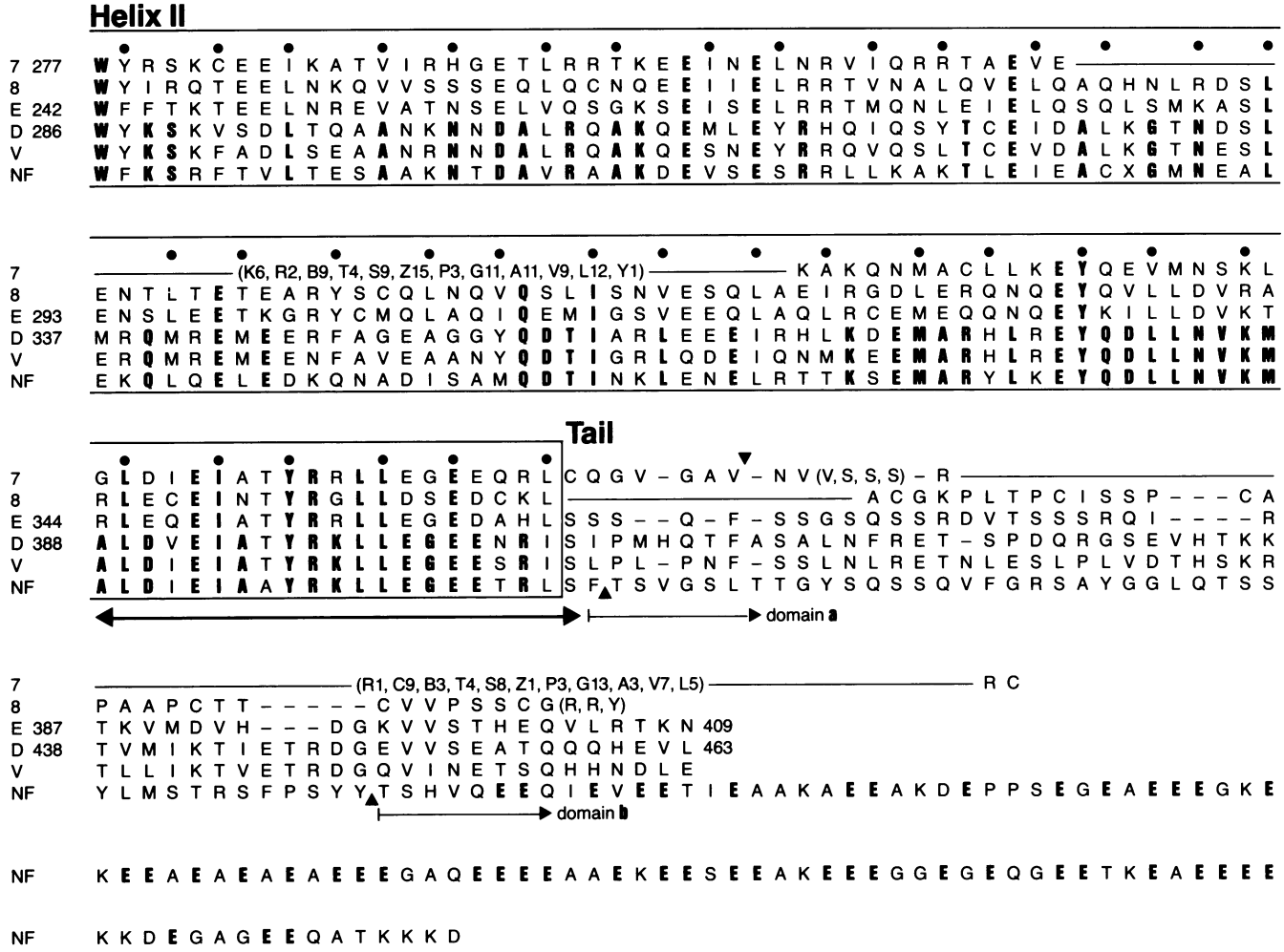

Fig. 3. Sequence comparison in the carboxy-terminal of six different intermediate filament proteins. The sequences cover the full tail-piece region and nearly the full helix II of the topographical model (Geisler and Weber, 1982). Primary information and nomenclature is 7 for sheep wool α-keratin 7 (Sparrow and Inglis, 1980), 8 for sheep wool α-keratin 8c-1 (Crewther *et al.*, 1980), E for a 50 K human epidermal keratin (Hanukoglu and Fuchs, 1982), D for chicken desmin (Geisler and Weber, 1982; and references therein), V for porcine vimentin (Geisler and Weber, 1981b; Geisler *et al.*, 1982a) and NF for the 68 K neurofilament triplet protein (see text; for earlier references Geisler *et al.*, 1982a). α-Keratin sequences have recently been extended and a few corrections have been incorporated (Dowling *et al.*, 1983). In the α-helix supposed to form inter-polypeptide coiled-coils the preferentially hydrophobic *a* and *d* positions of the consecutive heptades are indicated by dots. Arrowhead pointing down marks the end of the biochemically isolated desmin rod. Identical residues in the three non-epithelial proteins D, V and NF are given by bold letters. When all proteins show an identical residue in the helical domain this is marked by a bold letter. Note the increase in sequence identity towards the carboxyl side of helix II where the approximate position of the epitope recognized by a monoclonal antibody against intermediate filament proteins is indicated by an arrow (for details see text). Horizontal lines indicate not yet established sequences, which in some cases are indicated by compositional data provided in parenthesis (for documentation see original references). Dashes used in the tail regions of E, D, V provide for a better sequence alignment. They have not been used to align the hypervariable regions of the α-keratins or the 68 K triplet protein. Residue X following the single cysteine in 68 K protein is most likely either lysine or arginine. Note the subdivision of the tailpieces of 68 K into domains *a* and *b* due to their distinctly different sequences (see text). Note the very much longer tailpiece of 68 K and the sequence peculiarity of domain *b*, i.e., the presence of many glutamic acid residues (bold letters) and long runs of charged residues.

their peculiar amino acid compositions (Table I) argue strongly against α-helical arrays able to form coiled-coils. They show very low values for the large hydrophobic residues, particularly leucine, known to be involved in α-helically arranged coiled-coils of intermediate filaments. Instead a wealth of glutamic acid, alanine and lysine is observed, while the arginine content is very low. Although these features are reminiscent of the sequence of the *b* domain in 68 K, the compositional data (Table I see boxes) argue for differences. For instance, with increasing mol. wt. of the chymotryptic tailpieces there is a decrease in relative glutamic acid content and an increase in lysine and proline. Thus, if the two high mol. wt. tailpieces contained a domain resembling the *b* domain of 68 K there should be further domains of rather peculiar composition.

The secondary structure of the two large tailpieces is not yet known. They were stable against further chymotryptic proteolysis and in sucrose gradients showed sedimentation values of 2–3. Considering their apparent mol. wts. on gels, some extended asymmetric organization is indicated. None of the triplet tailpieces reacted with the monoclonal antibody recognizing the rod portions (Table II). Similar experiments also showed that the mouse monoclonal antibody NF_1 previously found to be specific for the 200 K protein (Debus *et al.*, 1982) recognized the large tailpiece and not the rod portion of this protein (Table II).

Basic non-α-helical amino-terminal headpieces

Following the work on desmin (Geisler *et al.* 1982b), the headpiece region of 68 K was isolated either as a basic CNBr

Table II. Location of the epitopes of monoclonal antibodies

	A	B
Segment tested		
Desmin	+	−
Triplet 68 K	+	−
160 K	+	−
200 K	+	+
Mixtures of wool α-keratins	+	n.d.
68 K rod	+	−
160 K rod	+	−
200 K rod	+	−
Desmin rod	+	n.d.
68 K tail	−	−
160 K tail	−	−
200 K tail	−	+
Desmin reisdues 1 − 286	−	n.d.
residues 287 − 463	+	n.d.
residues 1 − 323	−	n.d.
residues 324 − 463	+	n.d.
CNBr fragment of 68 K protein aligned amino-terminally with desmin residues 388 and so on	+	n.d.
Proposed location		
A: desmin residues 388 − 408		
B: tailpiece 200 K		

Mouse monoclonal antibodies were tested on purified proteins or fragments bound to nitrocellulose. Symbols for positive and negative reactions are + and − respectively; n.d. stands for not done. A is the general intermediate filament antibody described by Pruss *et al.* (1981), B is the NF_1 antibody specific for the 200 K triplet protein (Debus *et al.*, 1982).

fragment retained by CM-cellulose chromatography or as a larger fragment resulting from a digest with a lysine-specific protease. The sequence proposal for the 82 residues is given in Figure 4. Within the first 66 residues we note an abundance of hydroxyamino acids including tyrosine, seven arginines, four prolines and only one acidic residue. The compositional (Table I) as well as the sequence data (Figure 4) indicate a rather general structural similarity with the desmin headpiece. As in that case (Geisler *et al.*, 1982b), the first 60 residues of the 68 K headpiece reveal several regions of very strong β-turns when analyzed for secondary structure. This explains the high protease sensitivity of this domain. Again, as noted previously, α-helical potential is first expressed starting around residue 66 and we expect that extension of the sequence by only some 10 residues would lead into the beginning of the first interpolypeptide coiled-coil, called helix Ia in the desmin model (Geisler and Weber, 1982).

The headpiece of the 160 K triplet protein was isolated by double chemical cleavage (Geisler *et al.*, 1982a). Fragmentation with ^{14}C-labelled 2-nitro-5-thiocyanobenzoic acid led to an unlabelled and therefore amino-terminally located 37 K derivative, which upon further treatment with BNPS-skatole was converted to a 30 K and a 7 K fragment. The 30 K species had a composition similar to coiled-coil α-helical fragments (not shown) whereas the 7 K band was related to

```
NF   1   S S F Y S Q P Y Y S T S Y K R R Y V E T - - - - - - P R
D    1   S Q S Y S S S Q R V S S Y - R R T F G G G T S P V F P R

NF  23   V H I S S V R - S G Y S - T A R S A - Y S S Y S A - P -
D   28   A S F G S - R G S G S S V T S R V Y Q V S R T S A V P T

NF  46   V S S S L - S V R R S - Y - S S S S G L M P S L E N L
D   55   L S T F - R T T R V T P L R T Y Q S A Y Q G A G E L L

NF  70   D L S Q V A A I S N D L K    82
D   81   D F S L A D A M N Q E F L    93
```

Fig. 4. Amino acid sequence of the headpiece of 68 K protein (NF) and its relation to the same domain in desmin (D). Dashes are introduced to indicate a better degree in hypothetical sequence relation. Data on desmin have been given (summarized in Geisler and Weber, 1982). Positions with identical residues are underlined below the desmin sequence. Note that both domains are very basic due to numerous arginine residues. With the exception of positon 19 in NF, there are no acidic residues prior to either residue 66 (NF) or 78 (D). In the desmin sequence the first coiled-coil region (helix Ia) starts nine residues downstream from the last residue given here. A similar-sized 7 K headpiece of 160 K protein (Table I) digested by chymotrypsin also provides mainly basic peptides, some of which have been characterized by composition and end groups: Arg (Thr_2, Gln, Val); Arg (Thr, Ser_2, Gln, Gly, Val, Phe, Arg); Arg (Ser_2, Gln); Ser (Ser_4, Pro, Gly_2, Val, Phe, Arg); Thr (Asp, Asn, Ser_2, Pro, Gly, Ala, Leu_2, Tyr); Ser-Phe.

known headpieces by amino acid composition (Table I). Note the wealth of hydroxyamino acids and arginine as well as the presence of several prolines. Chymotryptic digestion led to several peptides some of which were characterized by amino acid composition and by their amino-terminal residue (legend to Figure 4). As expected, the headpiece of 160 K triplet protein is very basic and the available chemical information excludes identity with the corresponding headpiece of 68 K protein.

Discussion

Earlier descriptions of intermediate or 10 nm filament proteins, in particular the mammalian epidermal keratins, proposed the sequence of two long α-helical segments and made various size assignments and positionings for the non-α-helical domains (see, for instance, Skerrow *et al.*, 1973; Steinert *et al.*, 1980). Whereas the first aspect has been upheld by later sequence information, most of the other aspects could either not be supported or had to be corrected (Geisler *et al.*, 1982b; Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982; Weber and Geisler, 1982). The latter studies also incorporated the wool hard α-keratins and identified them as a special set of epithelial keratins. Current sequence data covering, to varying extents, seven distinct intermediate filament proteins support the topographical model proposed for muscle desmin, which is the only protein whose sequence has been fully determined (Figure 5; for a review, see Geisler and Weber, 1982b). The dominant structural motif is a rod-like middle domain of some 330 residues displaying, at a first approximation, two long α-helices able to form interpolypeptide coiled-coil structures. Thus, the sequence data were directly connected with the X-ray diffraction patterns (Geisler *et al.*, 1982b). Within these rod domains, sequence identity values ranged from ~30 to 80% when different proteins are compared (Geisler and Weber, 1982b; Hanukoglu

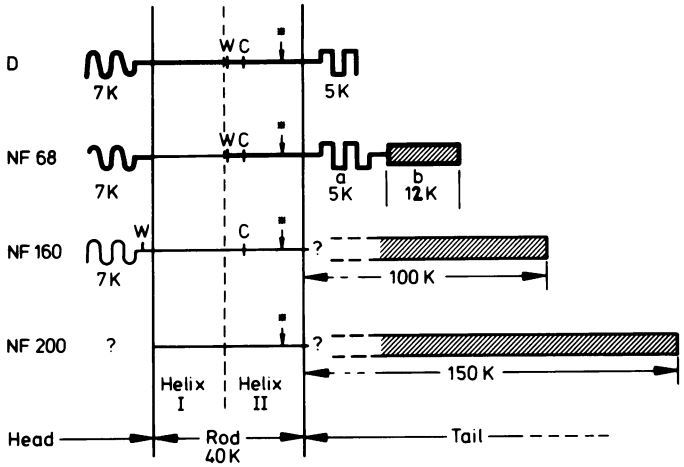

Fig. 5. A schematic presentation of the three domains of the triplet proteins in comparison with desmin (Geisler *et al.*, 1982b; Geisler and Weber, 1982). A similar sized α-helical coiled-coil rod domain is present in all proteins. Rods are preceded at the amino-terminal by a headpiece domain, which is non-α-helical and very sensitive to protease. Note that a headpiece has not yet been documented for 200 K protein. The rod is followed by non-coiled-coil tailpieces, which increase in length with increasing mol. wt. of the protein. In the tailpiece of 68 K protein, subdivision into domains *a* and *b* (see text and Figure 3) indicates structurally distinct arrays. Whether subdivision also occurs in 160 K and 200 K is not yet known. The thickened lines represent established sequences. W and C mark the position of rare tryptophan and cysteine residues used to provide large fragments. The star shows the binding site of the rod-specific monoclonal antibody used in Figure 3 and Table II.

and Fuchs, 1982). The rod is flanked by the non-α-helical terminal domains. The amino-terminal headpiece and the carboxy-terminal tailpiece reveal themselves as hypervariable domains of an otherwise rather conservative structure. They vary both in sequence and in length. We have now analyzed the three components of the mammalian neurofilaments and have shown that they follow the same rules (Figure 5). We have also established that neurofilament 'specificity' lies in the hypervariable tailpieces increasing in size from the 68 K protein to the 200 K protein. Sequence analysis of this region in the 68 K protein documents a uniquely charged subdomain in which glutamic acid accounts for 44% of the residues. Such sequences are unknown for any other intermediate filament protein (Geisler and Weber, 1982b) and invite future studies to detect if they could act as highly charged scaffolds able to bind other components of the neurone.

All three triplet proteins contain a coiled-coil rod domain of similar amino acid composition, size (40 K) and α-helical content (70–80%), which carries the epitope for a monoclonal antibody reported to recognize probably all intermediate filament proteins (Pruss *et al.*, 1981). This epitope has now been mapped within some 20 amino acids at the carboxyl side of the desmin and 68 K rod and lies in an array which belongs to most conserved regions when six different intermediate filament protein sequences are compared (Figure 3; Geisler and Weber, 1982). The amino acid sequence information available for 68 K (Figure 3) now covers nearly the complete helix II in the topographical model developed for desmin. Within this array an overall level of sequence identity of ~54% is observed in a comparision of desmin and 68 K. This value is noticeably lower than the 86% found for desmin and vimentin (Geisler and Weber, 1981b, 1982), but is still much higher than the 34% value typical of a comparison between non-epithelial intermediate filament proteins and an epidermal keratin or the two α-keratin prototypes of sheep wool (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982; Dowling *et al.*, 1983). Future experiments have to determine the precise degree of sequence relationship between the 68 K rod and its counterpart in the two higher mol. wt. triplet proteins.

Parallel to the desmin model (Geisler *et al.*, 1982a; Geisler and Weber, 1982), we have shown that at least two triplet proteins have a very basic non-α-helical headpiece as amino-terminal domain. Whereas the sequence of the 68 K headpiece has been completed (Figure 4), the corresponding fragment of the 160 K protein is so far only characterized by amino acid composition and by chymotryptic peptides described by composition and end group determination (Figure 4). The combined results are in agreement with the hypothesis that headpieces are a hypervariable structure of intermediate filament proteins (Geisler and Weber, 1982; Weber and Geisler, 1982). Thus, the proposed alignment in Figure 4 for these domains of desmin and 68 K necessitates the introduction of gaps to imply only a rather limited sequence relatedness. This sharply contrasts with the situation along the interpolypeptide coiled-coils of the rod domains where smooth alignment without any gaps is easily seen (Figure 3). The limited data on the headpiece of 160 K already argue that it is distinct from the 68 K headpiece. All headpieces known so far (desmin, 68 K, 160 K) are highly basic due to the presence of several arginine residues and contain few if any acidic amino acids. Their high content of hydroxyamino acids and glycine, as well as the presence of several proline residues, points to a non-α-helical organization.

The second hypervariable domain is the carboxy-terminal tailpiece which follows the rod (Weber and Geisler, 1982; Geisler and Weber, 1982). This is very clearly seen in the case of all neurofilament triplet proteins and documented in detail by amino acid sequence data for the 68 K protein (Figures 3, 5). These show that the increase in mol. wt. of 68 K protein in comparison with desmin (53 K) proposed to occur towards the carboxyl side of the molecule (Geisler *et al.*, 1982a) is due to an extended tailpiece. The sequence data also show that the tailpiece reveals two structurally distinct domains. The first 43 residues (domain *a*) are highly sensitive to chymotrypsin, devoid of negatively charged residues and possibly related to headpiece sequences by their wealth of hydroxyamino acids and the presence of several β-turns. Although a similar relationship between headpiece and tailpiece is not immediately obvious for desmin, vimentin and a human epidermal 50 K protein (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982), it seems possible in at least two other protein types. Thus, wool α-keratin 8c-1 has proline- and cysteine-rich terminal domains (Crewther *et al.*, 1980; see also Weber and Geisler, 1982) and long regions of tandem repeats of several glycine residues have been reported in abstract form for both domains in two higher mol. wt. mouse epidermal keratins (Roop and Steinert, 1982). The functional importance of these regions has to be explored in the future, since potentially they could be involved in a head-to-tail type assembly. More important seems the large domain *b* of the 68 K triplet protein, which, contrary to domain *a*, is resistant to chymotrypsin and distinctly different in chemistry. This sequence of 106 residues is noteworthy for two reasons. Charged residues account for as much as 60% and these are provided by three aspartic acids, 12 lysines and 47 glutamic acid

residues, making this 12 K fragment a uniquely glutamic acid-rich structure. The much longer tailpieces of the two higher mol. wt. triplet proteins show a certain similarity to domain *b* in the following aspects: they are resistant to chymotryptic digestion in spite of their size of 100 K and 150 K, respectively; they reveal, at neutral pH, a c.d. spectrum typical of random coils and have an amino acid composition dominated by glutamic acid, alanine and lysine with only low levels of the large hydrophobic residues and arginine. Because of the large mol. wt. differences involved we do not yet know if the larger tailpieces contain, as subdomain, a region corresponding to domain *b* of 68 K protein. If they do, then the remainder of their sequences will have rather peculiar amino acid compositions. The c.d. spectra indicating random coil conformation for the three tail domains do not rule out the possibility that some ordered structure may be induced either by ions or other proteins. The potential of at least the 68 K *b* domain to show some ordered structure is indicated by the change in the spectrum at pH 4 (Figure 2) and the presence of presumed dimers at pH 8 as shown by cross-linking (Figure 1).

Bodian's silver stain executed on polypeptides separated by gel electrophoresis is specific for all three triplet proteins (Gambetti *et al.*, 1981) and does not detect other intermediate filament proteins. If this is true, our sequence data suggest that the target of the stain lies outside the head and rod domains of 68 K protein. Within the remaining tailpiece, region domain *b*, with its unique glutamic- and lysine-rich sequence, seems a possible target. In this case also the two higher mol. wt. tailpieces of 160 K and 200 K protein might contain sequences related to domain *b*.

Our results show that not only 68 K but also the two high mol. wt. triplet proteins follow the structural principles developed for non-neuronal intermediate filament proteins (for references see above). Thus, the 160 K and 200 K protein are not simply associated proteins attached to a filament backbone built from 68 K protein (Willard and Simon, 1981; Geisler and Weber, 1981a; Sharp *et al.*, 1982) but rather are intermediate filament proteins in their own right. Their amino-terminal part comprising the headpiece and the rod should allow, by interaction of appropriate coiled-coils, a copolymerization process with 68 K protein, the most abundant triplet protein which is also known to form on its own 10 nm filaments (Geisler and Weber, 1981a; Henderson *et al.*, 1982; Liem and Hutchinson, 1982). The large and extended tailpiece can be envisioned as 'extra-filament' material protruding from the backbone or enwrapping it. The established properties of the long tailpieces rule out the possibility that the higher mol. wt. triplet proteins have tandem repeats of those coiled-coils present already in 68 K protein (Anderton, 1981) and provide a new view of the 200 K protein. This component has been thought to be located peripherally to the filament backbone when immunoelectron microscopy was performed on neurofilaments (Willard and Simon, 1981; Sharp *et al.*, 1982). Since we have mapped the epitope of such a 200 K-specific antibody (Debus *et al.*, 1982), the monoclonal antibody NF_1, in the tailpiece (Table II) there is no discrepancy between the electron micrographs and our description of the 200 K protein. It seems, therefore, that the 200 K tailpiece protrudes from the filament wall and is very likely responsible for the cross-bridges seen between neighbouring neurofilaments. Currently, we do not know if the putative cross-bridge structure is solely due to the tailpiece of the 200 K protein or if, when present, also involves contributions from the tailpieces of the other two triplet proteins. Given the extreme sequence peculiarity of the 68 K tailpiece and a possible structural relation among the three triplet tailpieces, further sequence data on the higher mol. wt. proteins are desirable.

The model developed for the mammalian triplet proteins is easily extended to neurofilaments from the invertebrate *Myxicola*. Here filaments are built from two polypeptides of apparent mol. wts. of 165 K and 172 K. Limited proteolysis mainly with papain resulted in progressive degradation most likely occurring mainly from one end only (Eagles *et al.*, 1981a, 1981b) and the final stable product of ~35 K probably corresponds to the 40 K rod domain in our gel system. The long proteolyzable extensions should reflect our tailpieces, which also are not resistant to papain. The existence of a small and basic headpiece, which we would expect (see above), could have gone undetected because no amino acid compositional data on the various fragments have been reported for *Myxicola* neurofilaments.

That studies on the rather complex neurofilament triplet proteins can enlarge our knowledge of intermediate filaments with simpler composition is indicated by two results. First, the epitope recognized by a general intermediate filament monoclonal antibody (Pruss *et al.*, 1981) has now been mapped within the rod domain to a region which belongs to the most conserved sequences of the various proteins known (Geisler and Weber, 1982b). Thus, this antibody should provide a valuable tool for characterizing associated proteins. For instance, we can now propose that a 66 K protein recognized by this antibody as a minor component in many different filament preparations (Pruss *et al.*, 1981) is probably a component co-polymerized into the filament backbone rather than attached to it. Second, it has so far remained unclear if different intermediate filament proteins, which are under strict rules of gene expression parallel to known embryonic differentiation programs (for a review, see Osborn and Weber, 1982), could perform differential functions (Lazarides, 1982). The triplet proteins of neurofilaments may offer the first answer. Not only are they extremely rich in components of very high mol. wt. but also all components reveal unique tailpieces. If the established domain *b* of the 68 K protein is really typical of all three tailpieces, we are encountering long arrays of extreme charge density particularly rich in glutamic acid. These could provide scaffolds suitable for interaction with other cellular components, either ions or certain cytoplasmic proteins. Concentration on the chemistry of these tail-domains should provide a key to understanding the interactions between neurofilaments and the cytoplasm that are indicated by morphological studies.

Materials and methods

Purification of triplet proteins

Triplet proteins from porcine spinal cord were separated by DEAE-cellulose in urea buffer (Geisler *et al.*, 1981a). When necessary, individual proteins were further purified by gel filtration through Sepharose 6B-C1 equilibrated in 6 M urea buffer. Proteins were recovered for chemical studies by ethanol precipitation followed by dialysis against water and lyophilization. For native digests, triplet proteins in urea buffer were dialyzed into 10 mM Tris-HCl pH 7.8 in the presence of 0.5 mM dithiothreitol.

Amino acid sequence studies

The fragment covering the carboxy-terminal 230 residues was obtained by cleavage at the single cysteine residue of 68 K protein using 2-nitro-5-thiocyanobenzoic acid. The fragment was isolated by preparative SDS-gel electrophoresis and freed of SDS as described (Geisler *et al.*, 1982a). The fragment was processed using digestion with trypsin, clostripain, thermolysin, chymotrypsin and V8 protease. Peptides isolated by two-dimensional separa-

tion on paper were characterized by amino acid composition and the stepwise Edman procedure either by the dansyl or the modified technique (Chang *et al.*, 1978). To establish the full sequence in domain *b* of the tailpiece, the domain was isolated as the chymotrypsin-resistant tailpiece present in a native chymotryptic digest of 68 K protein (see below). The digest was separated on DEAE-cellulose in 6 M urea and domain *b* recovered as expected as the most acidic fragment. Domain *b* was cleaved with lysine-specific protease (Boehringer, FRG) and the resulting peptides separated by reversed phase h.p.l.c. using a Zorbax C8 column and a trifluoroacetic acid-acetonitrile system. When necessary, peptides were further digested with V8 protease and fragments purified by the same approach. Overlapping fragments between the individual peptides were present both in the lysine-specific protease digest and the V8 protease digest. All peptides were manually sequenced as above. In addition, a liquid phase sequencer run performed on the first 33 residues of domain *b* gave very useful information. We thank J Friedrich and N. Hilschmann for their kind support.

The amino acid sequence of the headpiece region of 68 K protein was obtained from two fragments. First, CNBr-fragmented 68 K protein was passed through DEAE-cellulose in 6 M urea. The flowthrough fraction was recovered and chromatographed on CM-cellulose as described (Geisler *et al.*, 1982b). Residues 1 – 63, being the most basic fragment, eluted as the final peak. The DEAE column was developed with a salt gradient to isolate an overlapping CNBr fragment, which extends from the carboxyl end of the 68 K protein just into helix II of the rod domain. This fragment was also used to locate the epitope recognized by the general intermediate filament antibody. Second, 68 K protein was digested in 6 M urea buffer by lysine-specific protease (Boehringer, FRG). Subsequent chromatography on CM-cellulose in the presence of urea provided residues 1 – 82 as the most basic domain. Both headpiece fragments were characterized by the methods given above using chymotrypsin, trypsin, thermolysin, V8 protease and a proline-specific protease (Miles-Yeda, Israel). Separation was either by paper methods or a Zorbax C8 column. Final assignment of the first three residues following the blocking group, which in analogy to other related proteins could be the acetyl group (Steinert *et al.*, 1980), was provided by a chymotryptic tripeptide containing two serines per one phenylalanine. Given the specificity of the enzyme, we propose the sequence shown (Figure 4).

Triplet protein 160 K was subjected to chemical cleavage at cysteine using ^{14}C-labelled 2-nitro-5-thiocyanobenzoic acid. Preparative SDS-gel electrophoresis provided a single unlabelled fragment. This 37 K fragment was cleaved with the tryptophan reagent BNPS-skatole (for general methods see Geisler *et al.*, 1982a). SDS-gel electrophoresis yielded two fractions, a 30 K and a 7 K fragment. The 7 K fragment was digested with chymotrypsin and peptides separated by paper methods were further characterized.

Native digests and domain isolation

Urea-freed triplet proteins (see above) were digested at room temperature with chymotrypsin at a ratio of 1:400 (w/w). Time-dependent digestion was monitored by gel electrophoresis to obtain optimal conditions. In preparative digests, enzymatic activity was stopped by passing the solution through a Sepharose column containing anti-chymotrypsin IgGs. Separation of rod and tail domain was achieved by gel chromatography on Sepharose 6B-C1 in urea buffer for the digests of 68 K and 200 K protein. The *b* domain of the 68 K tailpiece was obtained as the most acidic fragment when chromatography on DEAE-cellulose was used. In the case of 160 K, protein chymotryptic domains were separated on DEAE-cellulose equilibrated in 10 mM Tris-HCl pH 7.6 without resorting to urea buffer.

Purity of domains was monitored by gel electrophoresis. Isolated domains were dialyzed into 10 mM Tris-HCl pH 7.8 and characterized by their c.d. spectra recorded by a Iasco-Polarimeter. Some experiments made use of 50 mM sodium phosphate buffer pH 4. Mean residue weights for calculation of ellipticities and α-helical content were provided by amino acid analyses using internal standards as given (Geisler *et al.*, 1982b).

Miscellaneous procedures

Cross-linking of the isolated *b* domain of 68 K protein was performed by adding to 40 μl of fragment dialyzed into 0.1 M Na-phosphate buffer pH 8.0 (0.5 mg/ml) 5, 10, 20 and 50 μl of a 10 mg/ml freshly prepared solution of dimethyl suberimidate in the same buffer. After 2 h at room temperature, the reaction was stopped by addition of 20 μl of 1 M Tris-acetate pH 8. After 10 min, protein was precipitated by trichloroacetic acid. Pellets were extracted with acetone, dissolved in sample buffer and subjected to gel electrophoresis.

Monoclonal antibodies were assayed on various proteins and fragments using the peroxidase technique on electro-blotted material (Western blots) or the spotting test on pure fragments applied to nitrocellulose paper (Towbin *et al.*, 1979; Debus *et al.*,, 1982). The mouse monoclonal antibodies used were NF_1, specific for the 200 K triplet protein (Debus *et al.*, 1982), and an antibody raised against glial fibrillary acidic protein (Pruss *et al.*, 1981), which seems to recognize all intermediate filament proteins.

Acknowledgements

We thank Drs. E. Debus and M. Osborn for samples of mouse monoclonal antibodies. Dr. M. Raff kindly provided his hybridoma line. Dr. G. Flügge contributed some of the earlier data used in Table II. S. Gaspar helped in protein purification and W. Koch in spotting experiments. We thank Dr. M. Osborn for constructive help in assembling this manuscript. We are very grateful to Dr. J.Friedrich and N. Hilschmann for a run on their sequenator. This work was supported by the Max Planck Society and a grant from the Deutsche Forschungsgemeinschaft. E.K. and S.F. acknowledge Max Planck predoctoral fellowships.

References

Anderton,B.H. (1981) *J. Muscle Res. Cell Motil.*, **2**, 141-166.

Anderton,B.H., Breinburg,D., Downes,M.J., Green,P.J., Tomlinson,B.E., Ulrich,J., Wood,J.M. and Kahn,J. (1982) *Nature*, **298**, 84-86.

Chang,Y., Bauer,D. and Wittmann-Liebold,B. (1978) *FEBS Lett.*, **93**, 205-214.

Chou,P.Y. and Fasman,G.D. (1978) *Annu. Rev. Biochem.*, **47**, 251-276.

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.

Debus,E., Fluegge,G., Weber,K. and Osborn,M. (1982) *EMBO J.*, **1**, 41-45.

Delacourte,A., Filliatreau,G., Boutteau,F., Biserte,G. and Schrevel,J. (1980) *J. Biochem.*, **191**, 543-546.

Doty,P. (1959) in Oncley,J.E. (ed.), *Biphysical Sciences*, Wiley, NY.

Dowling,L.M., Parry,D.A.D. and Sparrow,L.G. (1983) *Biosci. Rep.*, **3** 73-78.

Eagles,P.A.M., Gilbert,D.S. and Maggas,A. (1981a) *Biochem. J.*, **199**, 101-111.

Eagles,P.A.M., Gilbert,D.S. and Maggs,A. (1981b) *Biochem. J.*, **199**, 89-100.

Gambetti,P., Auitilio-Gambetti,L. and Papasozomenos,S.C. (1981) *Science (Wash.)*, **213**, 1521-1522.

Geisler,N. and Weber,K. (1981a) *J. Mol. Biol.*, **151**, 565-571.

Geisler,N. and Weber,K. (1981b) *Proc. Natl. Acad. Sci. USA*, **78**, 4120-4123.

Geisler,N. and Weber,K. (1982) *EMBO J.*, **1**, 1649-1656.

Geisler,N., Plessmann,U. and Weber,K. (1982a) *Nature*, **296**, 448-450.

Geisler,N., Kaufmann,E. and Weber,K. (1982b) *Cell*, **30**, 277-286.

Hanukoglu,I. and Fuchs,E. (1982) *Cell*, **31**, 243-252.

Henderson,D., Geisler,N. and Weber,K. (1982) *J. Mol. Biol.*, **155**, 173-176.

Hoffman,P.N. and Lasek,R.J. (1975) *J. Cell Biol.*, **66**, 351-366.

Lazarides,E. (1982) *Annu. Rev. Biochem.*, **51**, 219-250.

Lee,V., Wu,H.L. and Schlaepfer,W.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6089-6092.

Liem,R.K.H. and Hutchinson,S.B. (1982) *Biochemsitry (Wash.)*, **21**, 3221-3226.

Liem,R.K.H., Yen,S., Salomon,G.D. and Shelanski,M.L. (1978) *J. Cell Biol.*, **79**, 637-645.

Micko,S. and Schlaepfer,W.W. (1978) *J. Neurochem.*, **30**, 1041-1049.

Milam,L. and Erickson,H.P. (1982) *J. Cell Biol.*, **94**, 592-596.

Osborn,M. and Weber,K. (1982) *Cell*, **31**, 303-306.

Postingl,H., Little,M., Krauhs,E. and Kempf,T. (1979) *Nature*, **282**, 423-424.

Pruss,R.M., Mirsky,R., Raff,M.C., Thorpe,R., Dowding,A.J. and Anderton,B.H. (1981) *Cell*, **27**, 419-428.

Roop,D.R. and Steinert,P.M. (1982) *J. Cell Biol.*, **95**, 228a.

Sharp,G., Shaw,G. and Weber,K. (1982) *Exp. Cell Res.*, **137**, 403-413.

Shaw,G. and Weber,K. (1981) *Exp. Cell Res.*, **136**, 119-125.

Skerrow,D., Matoltsy,A.G. and Matoltsy,M.N. (1973) *J. Biol. Chem.*, **248**, 4820-4826.

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., Idler,W.W. and Goldman,R.D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4534-4538.

Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350-4354.

Weber,K. and Geisler,N. (1982) *EMBO J.*, **1**, 1155-1160.

Willard,M. and Simon,C. (1981) *J. Cell Biol.*, **89**, 198-205.

Note added in proof

Using mildly proteolyzed neurofilaments in gel electrophoresis followed by immune blotting experiments with polyclonal antibodies, Julien and Mushynski (*J. Biol. Chem.*, **258**, 4019-4025, 1983) have recently mapped the phosphorylation sites of the two large triplet proteins into fragments resembling the tailpieces described above. The combined studies indicate that the tailpieces seem to contain at least two distinct subdomains.