The EMBO Journal vol.3 no.11 pp.2701 – 2706, 1984

Hybrid character of a large neurofilament protein (NF-M): intermediate filament type sequence followed by a long and acidic carboxy-terminal extension

Norbert Geisler, Stefan Fischer, Joel Vandekerckhove[1], Uwe Plessmann and Klaus Weber

Max Planck Institute for Biophysical Chemistry, D-3400 Göttingen, FRG, and [1]Rijksuniversiteit Gent, Laboratory of Genetics, B-9000 Gent, Belgium

Communicated by K.Weber

The sequence of the amino-terminal 436 residues of porcine neurofilament component NF-M (apparent mol. wt. in gel electrophoresis 160 kd), one of the two high mol. wt. components of mammalian neurofilaments, reveals the typical structural organization of an intermediate filament (IF) protein of the non-epithelial type. A non-α-helical arginine-rich headpiece with multiple β-turns (residues 1 – 98) precedes a highly α-helical rod domain able to form double-stranded coiled-coils (residues 99 – 412) and a non-α-helical tailpiece array starting at residue 413. All extra mass of NF-M forms, as a carboxy-terminal tailpiece extension of ~500 residues, an autonomous domain of unique composition. Limited sequence data in the amino-terminal region of this domain document a lysine- and particularly glutamic acid-rich array somewhat reminiscent of the much shorter tailpiece extension of NF-L (apparent mol. wt. 68 kd), the major neurofilament protein. NF-M is therefore a true intermediate filament protein co-polymerized with NF-L *via* presumptive coiled-coil type interactions and not a peripherally bound associated protein of a filament backbone built exclusively from NF-L. Along the structurally conserved coiled-coil domains the two neurofilament proteins show only ~65% sequence identity, a value similar to that seen when NF-L and NF-M are compared with mesenchymal vimentin. The highly charged and acidic tailpiece extensions of all triplet proteins particularly rich in glutamic acid seem unique to the neurofilament type of IFs. They could form extra-filamentous scaffolds suitable for interactions with other neuronal components. One such example is the cross-bridge between neighbouring neurofilaments, which in axonal material seems to contain the tailpiece extension of NF-H.

Key words: axons/coiled-coils/intermediate filaments/neurofilaments/keratin

Introduction

Amino acid sequence data on muscle desmin and the biochemical properties of its subdomains obtained by limited proteolysis have led to a general model of intermediate filament (IF) proteins, which accounts for both the common and diverse properties of this complex class of proteins (Geisler *et al.*, 1982; Geisler and Weber, 1982). Based originally on the complete sequence of desmin and partial sequence data of vimentin, the neurofilament component NF-L, two α-keratins and one epidermal keratin, subsequent extension of the sequence bank has strongly supported the model (Hanukoglu and Fuchs, 1982, 1983; Geisler *et al.*, 1983a; Geisler and Weber, 1983; Quax-Jeuken *et al.*, 1983; Crewther *et al.*, 1983; Steinert *et al.*, 1983; Dowling *et al.*, 1983). The α-type X-ray diffraction pattern is due to a centrally located α-helical domain of ~310 residues. This rod-like domain is structurally preserved in its ability to allow exensive coiled-coil rope formation leading to a protofilamentous tetramer built from double-stranded coiled-coils (Geisler and Weber, 1982; Crewther *et al.*, 1983) rather than from one triple strand (Steinert *et al.*, 1980). Whereas the rod domain is well preserved in sequence and in length, the two flanking non-α-helical domains – the amino-terminal head- and carboxy-terminal tailpiece – are hypervariable in both properties. Parallel to their much higher sequence homology (Geisler and Weber, 1983) all four major non-epithelial IF proteins [myogenic desmin, mesenchymal vimentin, glial fibrillary acidic protein (GFAP) and the major neurofilament protein NF-H] can form homopolymeric IF whereas the epithelial keratins and the α-keratins as a special epithelial derivation are obligatory heteropolymers due to a complementarity between two rod prototypes: keratin I and II (reviewed by Hanukoglu and Fuchs, 1983; Franke *et al.*, 1983; Weber and Geisler, 1984).

One of the remaining problems of IF structure concerns the neurofilament proteins, which in mammals reveal three major components with apparent mol. wts. of 68 kd (NF-L), 160 kd (NF-M) and 200 kd (NF-H) in SDS-gel electrophoresis (Geisler and Weber, 1981; Liem and Hutchinson, 1982) (see Discussion). Our sequence data covering 65% of the polypeptide chain identified porcine NF-L as a non-epithelial IF protein carrying a tailpiece extension particularly rich in glutamic acid and lysine. This autonomous domain, not found in other major IF proteins, accounts for the increased mol. wt. of NF-L in comparison with desmin, vimentin and GFA (Geisler *et al.*, 1983a). As expected, NF-L shows self-assembly into smooth IF (Geisler and Weber, 1981; Liem and Hutchinson, 1982) revealing the typical 210 Å lateral periodicity (Henderson *et al.*, 1982; Milam and Erickson, 1982). Much less is known for the two larger mol. wt. components. Proteolytic studies on total neurofilaments (Julien and Mushynski, 1983; Chin *et al.*, 1983) and the isolated components (Geisler *et al.*, 1983a) identify two major domains for H and M. In addition to a very large domain thought to provide an extension from the filament wall, a 40-K domain is observed. Two models were proposed. Emphasizing that NF-M and H could act primarily as peripherally bound associated proteins of a filament made from NF-L, the 40-K domain was considered as anchor (Julien and Mushynski, 1983; Chin *et al.*, 1983). Alternatively, more detailed biochemical and immunological data suggested that the 40-K domain could correspond to the α-helical rod domain already delineated in all the IF proteins of much lower mol. wt. (Geisler *et al.*, 1983a). Thus NF-M and NF-H could be firmly co-polymerized with NF-L *via* the interaction between neighbouring coiled-coils. Since coiled-coil forming ability can be detected in amino acid sequences *via* the heptade repeat pattern, a decision between the two models should be possible. Here we identify NF-M by extended sequence analysis as a hybrid molecule. It carries in its amino-terminal region the structural

information typical of a non-epithelial IF protein. All its extra mass is located to a carboxy-terminal tailpiece extension forming an autonomous domain of unique amino acid composition.

Results

Sequence approach

Figure 1 summarizes our sequence data on the amino-terminal 436 residues of NF-M from porcine spinal cord. The amino-terminal region of 102 residues was obtained as an arginine-rich headpiece domain by digestion with lysine-specific protease followed by chromatography on CM-cellulose. This headpiece domain was fully sequenced (see Materials and methods). Of the seven major fragments obtained by CNBr cleavage fragments 2, 3 and 4 were fully sequenced. CNBr 1 was identified as the first fragment of the headpiece region by amino acid composition and the presence of a blocked N terminus. CNBr 5 was characterized in its amino-terminal 35 residues by a sequenator run and the results were confirmed by peptide data derived from digestion with trypsin. The amino acid composition (Table I) of CNBr 7 identified this large fragment as being closely related with the tailpiece extension of NF-M previously derived from a mild chymotryptic digest of the protein (Geisler *et al.*, 1983a). CNBr 1 (residues 1–67) and 2 (residues 68–285) are directly ordered by the overlapping headpiece region (residues 1–102). The resulting unambiguous alignment of the emerging NF-M sequence (residues 1–285) with the sequences of other non-epithelial IF proteins (Figure 1) previously determined by us was unexpected. It made, however, the further ordering of CNBr fragments 3, 4 and 5 by homology so easy that we did not isolate overlapping methionine peptides. The continuous sequence proposed in Figure 1 completely covers CNBr fragments 1–4 and the subsequent amino-terminal part of 5. Further partial sequences of the remainder of fragment 5 and fragment 7 are summarized below. Fragment 6, tentatively located between 5 and 7 (see below), was not available in sufficiently pure form to allow characterization by sequence data.

The α-helical coiled-coil domain

Residues 99–412 of NF-M display all the characteristic features of the conserved α-helical rod domain characteristic of IF proteins. Amino acid composition shows that this region is covered by the 40-kd domain previously excised from NF-M using mild chymotryptic proteolysis and found to have a high α-helix content by its circular dichroism properties (Geisler *et al.*, 1983a). Along the rod domain of NF-M the typical three α-helical regions (coil 1a, 1b and 2) show the usual heptade repeat pattern indicative of coiled-coil forming activity. The two interspersed non-coiled-coil segments of low α-helical potential show only minimal length variability and lack proline in the case of NF-M. In the rod region NF-M contains two tryptophan residues. The first (position 291) is found in most IF whereas the second (position 375) is so far unique. Coil 2 stops around residue 412 and is followed by the non-α-helical tailpiece (Figure 1). The only sequence ambiguity along the coiled-coil domain is the relative order of glutamine and glutamic acid at positions 304 and 305.

Over the three coiled-coil domains, hamster vimentin and chicken desmin show 74% sequence identity (Quax *et al.*, 1983; Geisler and Weber, 1983) while NF-M shares 50 and 52% identity with desmin and vimentin, respectively. Although these comparisons emphasize a considerable sequence flexibility along the coiled-coils they also delineate the previously recognized segments of highest homology among all IF proteins including the keratins (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982). These consensus-type sequences are located early in coil 1 and late in coil 2. The latter

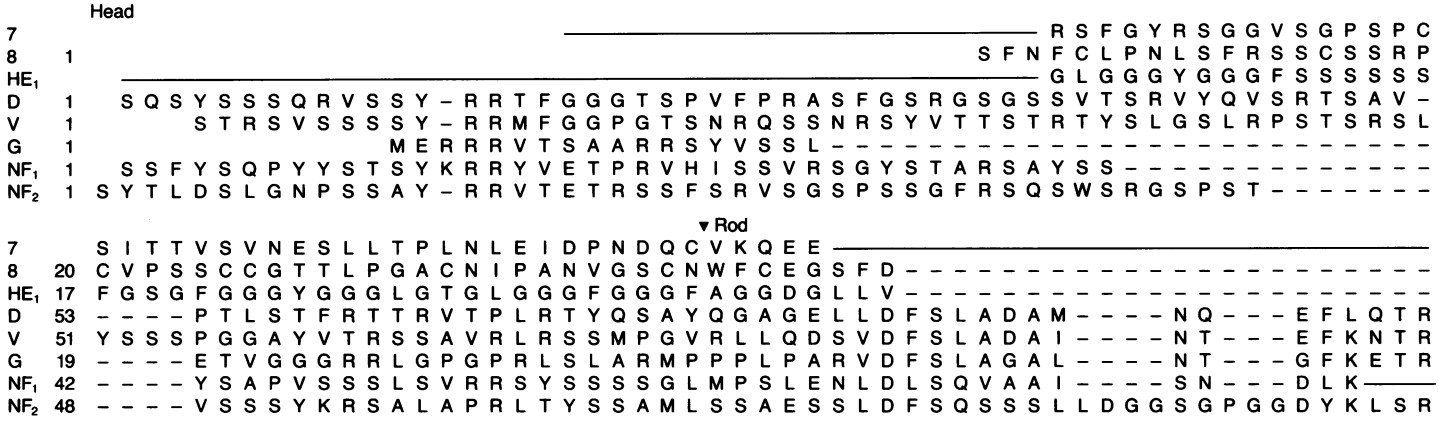

Fig. 1. Sequence relationship between NF-M and other IF proteins. Alignment is based on previous arguments (Geisler and Weber, 1982). For primary sequence data see the following references (Geisler *et al.*, 1982, 1983a; Geisler and Weber, 1982, 1983; Hanukoglu and Fuchs, 1982, 1983; Quax-Jeuken *et al.*, 1983 as modified in Quax *et al.*, 1983). Abbreviations are HE_1 and HE_2 human epidermal keratins 50 kd and 56 kd, respectively; 8 and 7 sheep wool α-keratins 8c-1 and 7c; D chicken desmin; V hamster vimentin; G porcine GFAP, NF_1 porcine neurofilament component L (NF-L); NF_2 porcine neurofilament component M (NF-M). α-Keratin sequences (Crewther *et al.*, 1980; Sparrow and Inglis, 1980) are arranged as previously proposed (Geisler and Weber, 1982; Geisler *et al.*, 1983a; for supporting evidence see Dowling *et al.*, 1983, and Crewther *et al.*, 1983). Horizontal lines indicate as yet unestablished sequences. X is an arginine or lysine residue in NF_1. For the possibility of some minor ambiguities in the sequence of G see original reference (Geisler and Weber, 1983). The three structural domains are indicated as are the hydrophobic **a** and **d** positions (dots) in the consecutive heptades of the presumptive coiled-coils (lines above the sequence blocks). Bold letters along the rod domain indicate identical residues among the different members of each of the three prototype sequences, i.e., non-epithelial IF proteins, keratins I and keratins II. Note some general irregularity early in coil II and the reversal around desmin residue 342 also common to other proteins. Arrowheads mark the location of the isolated desmin rod (Geisler *et al.*, 1982). Deletions (dashes) allow for better alignment of the short spacer regions and to some extent of the hypervariable non-α-helical terminal domains (see text). Note that the amino-terminal sequence of 436 residues for NF-M is well aligned with other non-epithelial IF proteins. An additional 18 residue fragment not yet connected forms the carboxyl end of CNBr 5. The amino-terminal sequence of the large CNBr 7 is given in Figure 2 (see text).

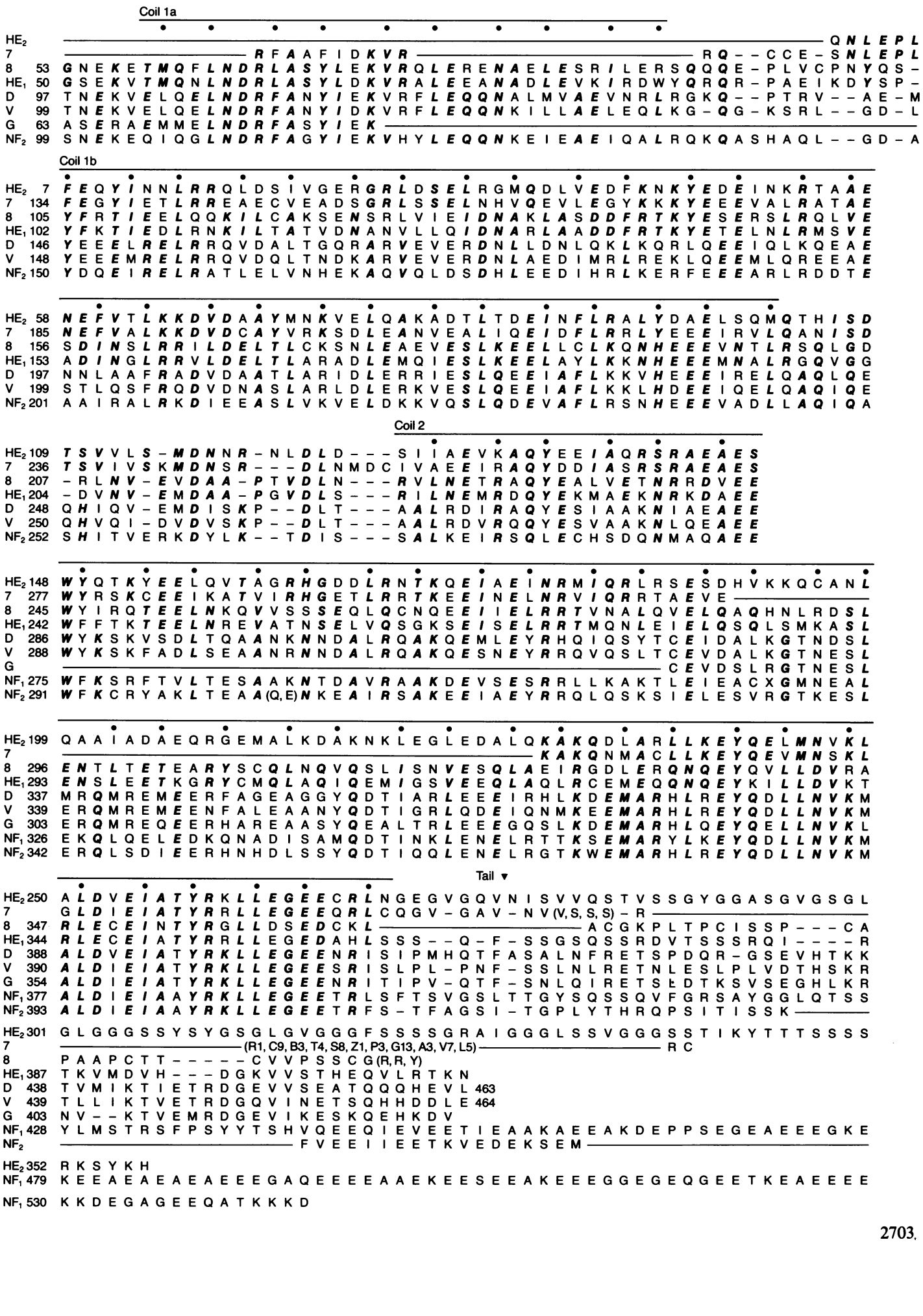
Coil 1a
Coil 1b
Coil 2
Tail ▼
HE2
7
8 53
HE1 50
D 97
V 99
G 63
NF2 99
HE2 7
7 134
8 105
HE1 102
D 146
V 148
NF2 150
HE2 58
7 185
8 156
HE1 153
D 197
V 199
NF2 201
HE2 109
7 236
8 207
HE1 204
D 248
V 250
NF2 252
HE2 148
7 277
8 245
HE1 242
D 286
V 288
G
NF1 275
NF2 291
HE2 199
7
8 296
HE1 293
D 337
V 339
G 303
NF1 326
NF2 342
HE2 250
7
8 347
HE1 344
D 388
V 390
G 354
NF1 377
NF2 393
HE2 301
7
8
HE1 387
D 438
V 439
G 403
NF1 428
NF2
HE2 352
NF1 479
NF1 530
(R1, C9, B3, T4, S8, Z1, P3, G13, A3, V7, L5)
463
464

Table I. Amino acid composition of CNBr fragment 7 (A) and the chymotryptic derivative of the tailpiece extension (B) of NF-M

	A	B
Asx	2.2	3.2
Thr	2.6	1.6
Ser	5.8	5.5
Glx	31.4[a]	30.3
Pro	7.2	7.4
Gly	9.0	8.6
Ala	12.2	11.8
Cys	n.d.	n.d.
Val	6.5	8.3
Met	–	0.5
Ile	1.1	1.4
Leu	2.0	2.9
Try	0.3	0.4
Phe	0.3	0.3
Lys	18.3	16.7
His	–	0.3
Arg	1.0	1.2
Trp	n.d.	n.d.

Values determined by standard amino acid analysis using acid hydrolysis are given in mol %. Values for B are from Geisler *et al.*, 1983a. Dash indicates absence of residue; n.d. not determined. Note the strong similarity between the two analyses and the wealth of Glx and Lys in particular.
[a]More than 90% of the value of Glx are provided by glutamic acid as shown by enzymatic hydrolysis using pronase followed by leucine aminopeptidase.

array, which harbors the epitope of a probably general monoclonal antibody for IF proteins, including NF-M (Geisler *et al.*, 1983a), is preceded in non-epithelial proteins by a further 17 very highly conserved residues. Within this region 29 identical residues (positions 383–411) form by far the longest array of sequence identity when NF-L and NF-M are compared. Since the sequence of NF-L is still incomplete, comparison is restricted to the established part of coil 2. Over this region sequence identity values are 79% for desmin *versus* vimentin, 56% for vimentin *versus* NF-L, 60% for vimentin *versus* NF-M and 63% for NF-L *versus* NF-M. Thus the two neurofilament proteins have distinctly diverged during evolution from a putative common precursor.

The amino-terminal headpiece of NF-M is related to other non-epithelial IF proteins

Given the alignment of residues 99–412 with the coiled-coil forming domain of other IF proteins, the preceding 98 residues cover a non-α-helical headpiece region with several β-turns. Residues 1–74 display a non-epithelial headpiece in their wealth of hydroxyamino acids (35% serine), the very basic character based nearly exclusively on arginine residues and the presence of several prolines (Geisler and Weber, 1982, 1983; Geisler *et al.*, 1982, 1983a; Quax-Jeuken *et al.*, 1983; Quax *et al.*, 1983). In spite of the general structural similarity, a convincing alignment of the five sequences of non-epithelial proteins is very difficult in this region, which also seems to show a pronounced species-specific drift in evolution (Geisler *et al.*, 1983b). The presentation used in Figure 1 stresses two points. First, a consecutive pair of arginines in a similar sequence environment located relatively close to the N-terminal end. Second, a leader-type sequence of ~16 residues, directly preceding the conserved region of 310 residues, showed for desmin, vimentin, GFAP and NF-L α-helical potential but poor coiled-coil forming ability (reviewed in Geisler and Weber, 1983). In NF-M this sequence is not only somewhat longer but also strongly reduced in helix formation due to the presence of a proline and several glycine residues. The structural importance of this array, which is so far not detected in the keratins, is not understood. The five headpiece arrays of non-epithelial IF proteins lack the repeats of ordered oligo-glycine sequences documented for certain epidermal keratins (Hanukoglu and Fuchs, 1982; Steinert *et al.*, 1983) and also the multiple cysteine residues found in wool α-keratins, which must be considered a different epithelial differentiation (Geisler and Weber, 1982; Hanukoglu and Fuchs, 1982, 1983). The amino-terminal blocking group of NF-M is most likely as in other IF proteins the acetyl group (Steinert *et al.*, 1980).

Partial characterization of the large carboxy-terminal domain

The sequence of CNBr 5 is not complete. The amino-terminal 35 residues derived by automated sequence analysis were extended by nine residues using an overlapping chymotryptic peptide. This established part of CNBr 5 covers the consensus type sequence at the end of coil II (positions 393–412) which is immediately followed by a non-α-helical tailpiece. In the first 24 residues of this domain (positions 413–436) are two prolines, two glycines and 10 hydroxyamino acids. This non-α-helical domain continues for probably 15–20 residues before the last 18 residues of CNBr 5 are reached. The latter sequence (Figure 1) was established on a tryptic peptide and its seven glutamic acid residues signal a new structural theme.

A small fragment, CNBr 6, located between fragments 5 and 7 (see below) was not available in pure form and remains poorly characterized. Fragment 6 was, however, present as carboxy-terminal extension of an overlapping fragment with CNBr 5. The amino acid composition and the limited information on tryptic and chymotryptic peptides of this overlap fragment indicate that CNBr 6 is ~30–40 residues long and extends the glutamic acid-rich sequence type already seen in the carboxy-terminal 18 residues of CNBr 5.

The wealth of glutamic acid continues in the amino-terminal 39 residues of CNBr 7 proposed from automated sequence analysis in Figure 2. CNBr 7 has an apparent mol. wt. of ~100 kd in SDS-polyacrylamide gel electrophoresis (see Discussion). Its amino acid composition (Table I) shows that it covers most of the tailpiece extension of NF-M previously isolated as autonomous domain by mild chymotryptic proteolysis (Geisler *et al.*, 1983a). Unique for these fragments is an abundance of lysine and particularly of glutamic acid now directly verified by enzymatic hydrolysis of CNBr 7 (Table I). In a few positions in the amino-terminal end of CNBr 7 we have either been unable to identify the residue (indicated by X) or have observed two different derivatives which we have given (Figure 2). Although the latter feature could indicate sequence heterogeneity it seems equally or even more likely that it results either from insufficient purity or some stuttering due to incomplete reaction during the sequenator run of a particularly curious sequence. Thus the sequence past residue 21 cannot yet be taken as an indication for heterogeneity of NF-M in this region but rather as a documentation for the continuation of the lysine- and particularly glutamic acid-rich sequences, which started already in CNBr 5. Given the amino acid composition of CNBr 7 (Table I) we expect such curious sequences to cover most of the remaining tailpiece extension of NF-M.

```
1                                   20
K E E V K E E E A E E K E E K Q E A E E

21                                  39
V E A A K K X X V K A E A X E L K G E
  V E           E
```

Fig. 2. The amino-terminal sequence of CNBr 7 from the neurofilament M component obtained by automated sequence analysis. Unidentified residue is indicated by X. Possible interpretations of more than one type of residue occurring past position 21 are given in Results.

Discussion

The continuous amino acid sequence of the amino-terminal 436 residues of porcine NF-M proposed above unambiguously identifies this high mol. wt. neurofilament component as a true intermediate filament protein of the non-epithelial type. It reveals a typical non-α-helical headpiece rich in hydroxyamino acids and arginine which is followed by the ~310 residues displaying the coiled-coil forming α-helical segments assumed to be the structurally dominant feature of all IF proteins. Past this rod domain the non-α-helical tailpiece and its extension have been subjected only to a partial characterization. The combined results fully confirm for NF-M our proposal derived from biochemical and immunological studies that the large mol. wt. neurofilament proteins are IF proteins in their own right co-polymerized within the filament with the major component NF-L, which has a much lower mol. wt. (Geisler *et al.*, 1983a). This interaction most likely occurs *via* the extended coiled-coil forming arrays. Although no sequence data are yet available on NF-H the excellent agreement between the biochemical and sequence data found for NF-M suggests a similar structural organization for the third mammalian neurofilament component.

The second major conclusion from the sequence data concerns the 'hybrid' structure of the large neurofilament proteins proposed from proteolysis data (Chin *et al.*, 1983; Julien and Mushynski, 1983; Geisler *et al.*, 1983). All the mass of NF-M exceeding that of other non-epithelial IF proteins such as desmin, vimentin and GFAP, which have mol. wts. between 51 and 54 kd (Geisler and Weber, 1982, 1983; Quax *et al.*, 1983; Geisler *et al.*, 1983b), is now located at the carboxyl end of the molecule. This array – the tailpiece extension – forms an independent structural domain, which can be clipped off by trypsin or chymotrypsin. It has a low α-helical content and a remarkable amino acid composition characterized by a wealth of glutamic acid and lysine (Table I). We have recently questioned the actual mol. wt. of porcine NF-M. Whereas SDS-polyacrylamide gel electrophoresis indicated a value ~160 kd, a much lower number of ~107 kd was found in 6 M guanidine-HCl by gel filtration and sedimentation equilibrium centrifugation. The aberrant behaviour in electrophoresis seems to be due to the tailpiece extension with its unique amino acid composition (Kaufmann *et al.*, 1984). Using the mol. wt. value of 107 kd, NF-M should contain ~950 residues with the acidic tailpiece extension providing ~500 residues. Of these we have determined only some 10%, a region too small to allow a detailed comparison with the 106 residue long tailpiece extension of the smaller NF-L (Geisler *et al.*, 1983a). Nevertheless, the 50 residues so far established in this region of NF-M reveal a general sequence type homology because of their wealth of lysine and particularly of glutamic acid. This observation raises the possibility that all three neurofilament proteins may share some sequence homology in their acidic tailpiece extensions. We do, however, note that amino acid compositional data on these regions (Geisler *et al.*, 1983a) already indicate strong individual differences and this is reinforced by recent data on further fragments in the case of NF-H (our unpublished results).

What is the function of the tailpiece extensions in neurofilament architecture? The results of limited proteolytic degradation of individual triplet components or intact neurofilaments (Chin *et al.*, 1983; Geisler *et al.*, 1983a; Julien and Mushynski, 1983) are now well connected with the emerging sequence data. A location of the extensions towards the outside of the filaments is for instance indicated by the finding that short tryptic digestion removes the large tailpiece regions of NF-M and NF-H without interfering with filament integrity (Chin *et al.*, 1983). In addition, these arrays seem highly charged due to a preponderance of lysine and glutamic acid (Geisler *et al.*, 1983a, and Results) and therefore most likely surface-exposed. The phosphorylated hydroxyamino acids noticeably found for NF-M and NF-H seem also predominantly located in the extensions (Julien and Mushynski, 1983). That they are not detected as extra material in negatively stained filaments could be related to thin and very extended structures possibly of high flexibility. Neighbouring neurofilaments particularly in axonal material are extensively connected *via* cross-bridges (for review, see Hirokawa *et al.*, 1984). A peripheral and discontinuous deposition of NF-H-specific polyclonal antibodies sometimes even cross-linking neighbouring filaments indicated that this protein could be involved in cross-bridge formation (Willard and Simon, 1981; Sharp *et al.*, 1982). This interpretation has been consolidated using the quick-freeze deep-etch technique which gives better structural preservation (Hirokawa *et al.*, 1984). Since a monoclonal antibody to NF-H whose epitope is located in the tailpiece extension (Geisler *et al.*, 1983a) provided the same peripheral decoration pattern (see Figure 3 in Debus *et al.*, 1982) as originally observed (Willard and Simon, 1981; Sharp *et al.*, 1982), this special domain is a very likely candidate for the cross-bridge structure. That in this region NF-H differs distinctly from NF-M is seen by compositional data (Geisler *et al.*, 1983a) and further emphasized by a study of smaller fragments performed on NF-H (our unpublished data). The high charge density of the extensions could also provide scaffolds for the interaction with other components of the neuronal cytoplasm. Extensions by non-α-helical material are also known for certain epidermal keratin polypeptides (Hanukoglu and Fuchs, 1982, 1983; Steinert *et al.*, 1983). They reveal, however, a quite distinct chemistry. Being often based on ordered arrays of oligoglycine they are not highly charged. In addition they can occur on the amino- as well as the carboxy-terminal end and are noticeably shorter than the tailpiece extensions of the neurofilament components NF-M and NF-H. It is, however, interesting that vimentin IF reveal an additional rather minor component of apparent mol. wt. 230 000 (synemin) (Granger and Lazarides, 1982), which although present only at very low concentration could be a non-neuronal counterpart to NF-M and NF-H.

Our sequence data necessitate a reevaluation of previous self-assembly studies using individual triplet proteins separated in the presence of urea. Removal of the denaturant by dialysis led to filament formation only in the case of NF-L

(Geisler and Weber, 1981; Liem and Hutchinson, 1982). It now seems very likely that the experimental conditions previously used were insufficient for filament assembly or that the large tailpiece extension interfered in the renaturation process of NF-M. After our sequence data were essentially complete, Bignami and co-workers showed that a quite different choice of pH and salt conditions led indeed to self-assembly of bovine NF-M (Gardner *et al.*, 1984). *In vitro* self-assembly ability of NF-M and possibly also NF-H observed under their experimental conditions raises certain questions as to a regulation of neurofilament assembly *in vivo*. How does the cell suppress segregation of triplet proteins into different filaments given the immunological data pointing to a copolymerization process for axonal or neurite filaments (Hirokawa *et al.*, 1984; Willard and Simon, 1981; Sharp *et al.*, 1982; Shaw and Weber, 1982)? How can NF-H be nearly exclusively located to axons and be nearly absent in dendrites (Hirokawa *et al.*, 1984; Shaw *et al.*, 1981; Shaw and Weber, 1982)? The sequences so far obtained for NF-L and NF-M argue against a derivation of the three triplet proteins from a common transcript differentially processed by RNA splicing. The three neurofilament proteins also cannot result from translational read-through mechanisms from the same transcript and are therefore expected to be the products of independent genes. Current sequence data covering NF-L and NF-M to 65% and 54%, respectively, have not proven sequence heterogeneity. Thus we suggest that these two proteins do not arise from multiple genes and that the charge heterogeneity seen in two-dimensional gels most likely results from post-translational modifications in line with the presence of multiple phosphate groups (Julien and Mushynski, 1983).

Materials and methods

Neurofilament triplet proteins from porcine spinal cord were isolated and separated using DEAE-cellulose and gel filtration in the presence of urea as described (Geisler and Weber, 1981; Geisler *et al.*, 1983a). NF-M was digested with lysine-specific protease in the presence of urea and an arginine-rich headpiece fragment was obtained by chromatography on CM-cellulose in 6 M urea as described for NF-L (Geisler *et al.*, 1983a). Six major CNBr fragments of NF-M were isolated by a combination of DEAE-chromatography and preparative gel electrophoresis in the presence of SDS (CNBr 2, 3, 5 and 7), gel filtration on G-100 in 0.1 M NH_4HCO_3 (CNBr 4) and chromatography on CM-cellulose in urea buffer (CNBr 1). Individual sequences were determined from tryptic, thermolytic and V8-protease peptides separated by two-dimensional fingerprint methods on paper as well as h.p.l.c. chromatography, and characterized by amino acid composition and stepwise Edman degradation using the modified technique (Chang *et al.*, 1978). [For further details see References (Geisler and Weber, 1982; Geisler *et al.*, 1983a)] In addition information was derived from chymotryptic peptides (headpiece, CNBr 2 and 5) and peptides obtained by clostripain and lysine-specific protease (CNBr 2). Using a gas phase sequenator the following regions were additionally determined: 68–88, 236–253, 275–285, 352–371 and 393–427. The carboxy-terminal region of CNBr 5 and the amino-terminal region of CNBr 7 are based on the same approach. A further sequenator run on residues 61–90 as part of a large tryptic peptide isolated from the headpiece was kindly performed by Drs Friedrich and Hilschmann.

Acknowledgements

We thank S.Gaspar for expert technical assistance and J.Van Damme and C. Ampe for runs on the gas phase sequenator. This work was supported by the Max Planck Society and a grant from the Deutsche Forschungsgemeinschaft to N.G. and K.W. J.V. is *Bevoegdverklaard Navorser* at the Belgian National Fund for Scientific Research (N.F.W.O.). S.F. was supported by a predoctoral fellowship from the Max Planck Society.

References

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

Chin,T.K., Eagles,P.A.M. and Maggs,A. (1983) *Biochem. J.*, **215**, 239-252.

Crewther,W.G., Dowling,L.M. and Inglis,A.S. (1980) *Proc. 6th Quinquennial Int. Wool Textile Res. Conf.*, Vol. **2**, Pretoria, pp. 79-91.

Crewther,W.G., Dowling,L.M., Steinert,P.M. and Parry,D.A.D. (1983) *Int. J. Biol. Macromol.*, **5**, 267-274.

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

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

Franke,W.W., Schiller,D.L., Hatzfeld,M. and Winter,S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7113-7117.

Gardner,E.E., Dahl,D. and Bignami,A. (1984) *J. Neurosci.*, **11**, 145-155.

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

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

Geisler,N. and Weber,K. (1983) *EMBO J.*, **2**, 2059-2063.

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

Geisler,N., Kaufmann,E., Fischer,S., Plessmann,U. and Weber,K. (1983a) *EMBO J.*, **2**, 1295-1302.

Geisler,N., Plessmann,U. and Weber,K. (1983b) *FEBS Lett.*, **163**, 22-24.

Granger,B.L. and Lazarides,E. (1982) *Cell*, **30**, 263-275.

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

Hanukoglu,I. and Fuchs,E. (1983) *Cell*, **33**, 915-924.

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

Hirokawa,N., Glicksman,M.A. and Willard,M.B. (1984) *J. Cell Biol.*, **98**, 1523-1536.

Julien,J.P. and Mushynski,W.E. (1983) *J. Biol. Chem.*, **258**, 4019-4025.

Kaufmann,E., Geisler,N. and Weber,K. (1984) *FEBS Lett.*, **170**, 81-84.

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

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

Quax,W.J., Egbert,W.V., Hendriks,W., Quax-Jeuken,Y. and Bloemendal, H. (1983) *Cell*, **35**, 215-223.

Quax-Jeuken,Y.E.F.M., Quax,W.J. and Bloemendal,H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3548-3552.

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

Shaw,G., Osborn,M. and Weber,K. (1981) *Eur. J. Cell Biol.*, **26**, 68-82.

Shaw,G. and Weber,K. (1982) *Nature*, **298**, 277-279.

Sparrow,L.G. and Inglis,A.S. (1980) *Proc. 6th Quinquennial Int. Wool Textile Res. Conf.*, Vol. **2**, Pretoria, pp. 237-246.

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

Steinert,P.M., Rice,R.H., Roop,D.R., Trus,B.L. and Steven,A.D. (1983) *Nature*, **302**, 794-800.

Weber,K., Shaw,G., Osborn,M., Debus,E. and Geisler,N. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **48**, 717-729.

Weber,K. and Geisler,N. (1984) in *Cancer Cells Vol. 1 The Transformed Phenotype*, Cold Spring Harbor Laboratory Press, NY, pp. 153-159.

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

Received on 5 July 1984

Note added in proof

We have meanwhile extended the protein chemical analysis to NF-H by a partial characterization of the non-α-helical headpiece located at the N-terminal end. In addition we have aligned a 17-kd fragment with the corresponding sequences of the rod domains of other non-epithelial IF proteins (Geisler *et al.*, in preparation). Our model on neurofilament structure (Geisler *et al.*, 1983a) is therefore essentially verified, although more extended sequence data are desirable.