Journal of Cell Science 108, 3581-3590 (1995)
Printed in Great Britain © The Company of Biologists Limited 1995
JCS4010

Characterisation of neurofilament protein NF70 mRNA from the gastropod *Helix aspersa* reveals that neuronal and non-neuronal intermediate filament proteins of cerebral ganglia arise from separate lamin-related genes

James Adjaye, Uwe Plessmann, Klaus Weber* and Huub Dodemont†

Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, PO Box 2841, D-37018 Goettingen, Germany

*Author for correspondence
†Present address: Department of Molecular Biology, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands

SUMMARY

The neuronal cytoplasmic intermediate filament (IF) protein HeNF70 of the gastropod *Helix aspersa* is identified by sequence analysis of the corresponding 4,600 bp cDNA isolated from a cerebral ganglion cDNA library. HeNF70 shares 60% sequence identity with the neuronal LoNF70 protein of the cephalopod *Loligo pealei* and only 36% identity with the *Helix* non-neuronal IF-A protein. All three molluscan IF proteins display the lamin-type extended coil 1b subdomain harbouring six additional heptads and all have long C-terminal sequences with substantial homology to the tail domains of nuclear lamins. The lamin-like tail domains of the two neurofilament proteins share a unique motif comprising 13 residues, which is absent from *Helix* IF-A and all other known non-neuronal IF proteins. HeNF70 is encoded by a 9.5 kb RNA transcript. The very long 7.2 kb 3′-untranslated sequence contains a unique 26 nucleotide repeat extending over 0.5 kb in its 5′-region. The HeNF70 mRNA is expressed at low abundancy in cerebral ganglia but not in any of the 13 non-neuronal tissues tested. In contrast, all tissues express at fairly high levels the same 4.6 and 4.2 kb mRNAs encoding the *Helix* non-neuronal IF-A/B proteins. Blot hybridisation studies on genomic DNA and ganglion mRNA with subprobes from the cloned HeNF70 cDNA, as well as sequence analysis of an RT-PCR generated partial cDNA encoding a putative HeNF60 protein, indicate at least two different neuronal IF genes in *Helix*.

Key words: gene complexity, intermediate filament, invertebrate, mollusc, neurofilament, nuclear lamin

INTRODUCTION

Intermediate filament (IF) proteins show a typical tripartite structural organisation. A central α-helical rod domain capable of forming double-stranded coiled-coils is flanked by non helical terminal regions, the head and tail domains (for reviews see Fuchs and Weber, 1994; Steinert and Roop, 1988). The neurofilament triplet proteins NF-L, NF-M and NF-H are the adult neuron specific IF proteins of vertebrates. Together with α-internexin, which occurs in embryonic neurons (Fliegner et al., 1990), they form the type IV vertebrate IF proteins. Molecular knowledge of neurofilament proteins from invertebrates is restricted to those from the cephalopod *Loligo pealei*. The three squid polypeptides NF60, NF70 and NF220 share common head and rod domains but have different tail domains, which are generated from a single gene by alternative RNA processing and splicing (Szaro et al., 1991; Way et al., 1992). In contrast, individual vertebrate neurofilament proteins are encoded by separate genes (for review see Eagles et al., 1990).

Like all protostomic non-neuronal IF proteins characterised to date, the *Loligo* neurofilament proteins contain 42 extra residues within the coil 1b region of their rod domains (Bovenschulte et al., 1995; Dodemont et al., 1990, 1994; Riemer et al., 1991; Szaro et al., 1991; Tomarev et al., 1993; Way et al., 1992; Weber et al., 1988, 1989). These additional six heptads, first identified in nuclear lamins (Fisher et al., 1986; McKeon et al., 1986), are absent from all vertebrate cytoplasmic IF proteins (Fuchs and Weber, 1994; Steinert and Roop, 1988). In spite of the differences between vertebrate and squid neurofilament proteins, there are common properties which distinguish neuronal from non-neuronal IF proteins. One shared feature of squid NF220 and mammalian NF-H is their ability to be phosphorylated at repeated KSP (Lys-Ser-Pro) residues within the tail domains. Mammalian NF-H contains 43-52 such motifs while in squid NF220, 21 of these occur together with two additional repeated motifs RSP (Arg-Ser-Pro) and SEK (Ser-Glu-Lys) represented 11 and 32 times, respectively (Julien et al., 1988; Lees et al., 1988; Way et al., 1992). Phosphorylation of mammalian NF-H in the C-terminal tail domain has been implicated in the regulation of axonal calibre and the interactions of neurofilaments with microtubules (Carden et al., 1987).

Here, we describe the full-length and partial cDNA sequence for the neurofilament proteins NF70 and NF60 of the gastropod

Helix aspersa, which arise from IF genes distinct from the gene encoding the major non-neuronal IF proteins (Dodemont et al., 1990).

MATERIALS AND METHODS

Nucleic acid preparation

The following tissues were dissected from *Helix aspersa*: cerebral ganglion, oesophagus, digestive gland, albumen gland, mucus gland, foot sole epidermis, heart, lung, kidney, oviduct, dart sac, receptaculum seminis, vas deferens and penis. From each of these, poly(A)-containing polysomal RNA was isolated as described (Dodemont et al., 1990). Genomic DNA was extracted from digestive gland starting with crude nuclei preparations and purified by standard procedures (Sambrook et al., 1989).

Construction and screening of the plasmid cerebral ganglion cDNA library

A large plasmid library (400,000 primary transformants) representing total cerebral ganglion mRNA was established using the *Escherichia coli* XL-1 Blue/pUC18 host/vector system as described (Dodemont et al., 1990, 1994). Colony hybridisation was performed at reduced stringency (Dodemont et al., 1994) with a cocktail of five ^{32}P-labeled, nick-translated DNA fragments, each at a concentration of 1 ng/ml. Probes were a 174 bp *Pvu*II-*Acc*I fragment comprising the coil 1a region from the *Helix* non-neuronal IF gene (Dodemont et al., 1990) and the corresponding fragments derived from four cloned full-length *Ascaris lumbricoides* IF cDNAs (Dodemont et al., 1994). Following hybridisation, filters were initially washed twice for 15 minutes in 5× SSC/0.1% SDS at 55°C and autoradiographed to Kodak X-Omat AR film using intensifying screens. Filters were then washed to high stringency by two consecutive washes for 15 minutes each in 0.5× SSC/0.1% SDS at 55°C and 0.1× SSC/0.1% SDS at 60°C and re-exposed. Colonies for which the initial hybridisation did not persist through the last washes were selected for rescreening at high density. This yielded only a single novel IF cDNA candidate which was purified by one additional round of screening. Isolation and restriction enzyme analysis of plasmid DNA followed established protocols (Sambrook et al., 1989). Both strands of the cDNA insert were completely sequenced (Dodemont et al., 1994). This led to the identification of the HeNF70 cDNA encoding a neurofilament protein.

Northern analysis

Poly(A)-containing polysomal RNA (2 μg) from oesophagus and cerebral ganglion was treated with glyoxal, electrophoresed in a 1.5% agarose gel and transferred to a nitrocellulose membrane. RNA size standards were rat 18 S and 28 S rRNA (1.9 and 4.7 kb) mixed with *E. coli* 16 S and 23 S rRNA (1.6 and 2.9 kb) and the Gibco-BRL RNA ladder (9.5, 7.5, 4.4, 2.4 and 1.4 kb). The RNA blot filter was subjected to four consecutive hybridisations under conditions of high stringency (Dodemont et al., 1994). Final washes were in 0.1× SSC/0.1% SDS at 60°C for 15 minutes and autoradiography was as above. Prior to rehybridisation, previously annealed DNA probes were effectively removed at 42°C for 30 minutes in a 10 mM boric acid/1 mM disodium tetraborate buffer, pH 8.2, containing 10 mM methylmercury hydroxide. This was followed by a wash in 10 mM Tris base/0.5 M ammonium acetate for 15 minutes at 42°C. DNA probes were prepared from the fragments listed below; nucleotide (nt) designations refer to the HeNF70 cDNA sequence presented in Fig. 1. Probe 1: 1,719 bp *Xba*I-*Pst*I fragment (nt 1-1,706, preceded by the 13 nt *Xba*I-*Sma*I polylinker region from the pUC18 vector) covering the entire 5′-part of the cDNA up to the C-terminal end of the rod domain; probe 2: 96 bp *Dde*I-*Fok*I fragment (nt 1,793-1,888) representing the N-terminal part of the tail domain; probe 3: 327 bp *Pst*I-*Rsa*I fragment (nt 2,103-2,429) spanning the C-terminal half of the tail domain and 89 nt of the 3′-untranslated sequence; probe 4: 174 bp *Pvu*II-*Acc*I fragment comprising the coil 1a region from the *Helix* non-neuronal IF gene (see above) which corresponds to nt 682-855 of the HeNF70 cDNA. Probe 1 was also used on a second RNA blot filter containing mRNA (2 μg) of all 14 tissues.

Genomic DNA analysis

Genomic Southern blots (Sambrook et al., 1989) were hybridised under conditions of high or low stringency (Dodemont et al., 1994). Probes were derived from restriction fragments of the cloned HeNF70 cDNA (see above) or from a subcloned 215 bp PCR fragment generated with the cDNA template and the following primers: (sense, nt 1,567-1,591) 5′-GCCCAGTTGGAGAAACAGTACCAAG-3′ and (antisense, nt 1,781-1,758) 5′-GTTTCCTCTCCCTCCAGCAGCTTC-3′. The 5′-part of the NF70 gene was isolated as two non-overlapping fragments (see main text) cloned from size fractionated subgenomic libraries (Dodemont et al., 1990).

PCR-based isolation of a HeNF60 cDNA

Cerebral ganglion mRNA was converted to cDNA (Dodemont et al., 1990) except that random hexamers were used at a concentration of 50 μg/ml to prime first strand synthesis. The final blunt-end double-stranded cDNA (average size 2,500 bp) was ligated into the *Sma*I site of pUC18. An aliquot of the ligation mixture corresponding to 50 ng of input cDNA template was subjected to the polymerase chain reaction, PCR (Saiki et al., 1988), using conditions given elsewhere (Dodemont et al., 1994). The sense primer covered nt 1,707-1,726 of the HeNF70 cDNA sequence: 5′-GATACTGATGGATGCCAAGC-3′. The antisense primer matched the sequence directly adjacent to the anchored cDNA inserts, comprising the *Pst*I-*Bam*HI region of the pUC18 multiple cloning site: 5′-CTGCAGGTCGACTCTAGAG-GATCC-3′. Amplified DNAs were processed for subcloning in pUC18 (Dodemont et al., 1994) to generate a new plasmid cDNA library. For subsequent screening, the same strategy was followed as outlined above for the library representing total mRNA. The hybridisation probe was prepared from a 396 bp *Pst*I fragment (nt 1,707-2,102 of the HeNF70 cDNA sequence) covering the C-terminal end of the rod domain and most of the tail domain. Clones for which hybridisation signals strongly decreased or were lost completely upon the last wash steps at high stringency were purified. Sequence analysis identified the novel HeNF60 cDNA.

RESULTS

Isolation of a putative neuronal IF cDNA

A large plasmid cDNA library representing total poly(A)-containing polysomal RNA from *Helix aspersa* cerebral ganglion tissue was established. Screening was performed by hybridisation at reduced stringency employing the cloned coil 1a region of the *Helix* non-neuronal IF gene (Dodemont et al., 1990) and the corresponding sequences of four distinct IF cDNAs of the nematode *Ascaris lumbricoides* (Dodemont et al., 1994). Hybridisation of 400,000 clones followed by an initial low stringency wash yielded nearly 200 signals emerging at different intensities above a relatively high background of non specific hybridisation. About 98% of the detected clones showed no obvious loss of hybridisation after the last high stringency wash. Indeed, analysis of several of these clones revealed that they all harboured cDNA inserts derived from the previously characterised non-neuronal IF mRNAs (Dodemont et al., 1990) which are expressed in the glial cells that constitute the vast majority of cells in cerebral ganglia. Hence, the

initially hybridising clones, the signals of which did not persist through the final stages, were considered as putative candidates containing IF cDNA inserts of neuronal origin. Subsequent high density rescreening of these recovered only a single candidate that displayed similar hybridisation properties to those seen during the first round of screening. This was further assessed by Southern blot hybridisation on the *Eco*RI/*Hind*III digest of the still heterogeneous plasmid DNA preparation. A 4.2 kb DNA fragment hybridised at low stringency with the probe cocktail described above but not with the *Helix* coil 1a sequence alone at high stringency. Final purification of the candidate *Helix* neuronal IF cDNA clone revealed that the 4.2 kb fragment originated from a 4.6 kb cDNA insert containing an internal *Eco*RI site.

Sequence characteristics: identification of the novel *Helix* neuronal IF protein HeNF70

The 4,600 bp cDNA sequence shows a single large open reading frame of 1,938 bp flanked by 5′- and 3′-non-translated regions of 402 and 2,260 bp, respectively (Fig. 1). Eight in-frame stop codons occur upstream of the presumptive start codon, which resides within a sequence context that largely fulfils the criteria for efficient translational initiation as originally assessed for vertebrate mRNAs (Kozak, 1991). The 3′-untranslated sequence displays several salient features, some of which have been documented for other molluscan IF specific mRNAs (Dodemont et al., 1990; Riemer et al., 1991; Tomarev et al., 1993). These include a high overall A/T content of 68% (versus 58% and 49% for the 5′-untranslated and coding sequences), the presence of large TA repeat regions and a high number (50 copies) of TATT/ATTT$(T)_n$ sequence motifs representing putative mRNA instability determinants (Jackson, 1993; Shaw and Kamen, 1986; Wilson and Treisman, 1988). Additional features of the 3′-untranslated sequence are the unprecedented repeat region extending over 507 bp in the 5′-part and the presence of three canonical AATAAA polyadenylation signals (Birnstiel et al., 1985). Obviously none of these are recruited upon transcription termination since the corresponding transcripts were not detected in the poly (A)-containing RNA population by northern analysis (see below).

The open reading frame translates into a polypeptide comprising 646 residues corresponding to a calculated molecular mass of 73,907 and a theoretical isoelectric point of 5.79. The predicted protein sequence fully complies with the IF specific tripartite structural organisation based on a central α-helical rod domain with coiled-coil forming ability and the flanking non-helical head and tail domains (Fig. 2; for reviews see Fuchs and Weber, 1994; Steinert and Roop, 1988). Since the tail domain does not contain the hallmark nuclear localisation signal and C-terminal CaaX box of nuclear lamins (Fisher et al., 1986; McKeon et al., 1986), the sequence represents a true cytoplasmic IF protein. The rod domain shows the characteristic segmentation into coils 1a, 1b and 2, separated by short non-helical linkers. The coil 1b subdomain harbours six additional heptads (42 residues), first identified in nuclear lamins (Fisher et al., 1986; McKeon et al., 1986), which are also found in all protostomic cytoplasmic IF proteins known to date (for references see Introduction). The novel *Helix* IF sequence clearly does not represent the gastropod counterpart of the squid neurofilament-like protein LoSNLK, which has a coil 1b subdomain that seems to be extended by only two additional heptads (Adjaye et al., 1993).

Sequence comparisons were made for the rod and tail domains of the novel IF protein and the corresponding regions from a wide variety of both neuronal and non-neuronal IF proteins including all known invertebrate cytoplasmic IF sequences. Table 1 documents that by far the highest sequence identity is shared with the medium molecular mass neurofilament NF70 protein from the squid *Loligo pealei* (Szaro et al., 1991). The scores of 65 and 52% for their rod and tail domains represent a sharp rise from the values of only 38 and 32% identity between the novel *Helix* IF specific sequence and the non-neuronal IF-A protein from the same source. Thus the newly characterised gastropod IF sequence has been identified as a neurofilament protein which, by analogy to its cephalopod counterpart, was designated HeNF70 although its molecular mass is slightly higher (HeNF70: ~74 kDa; LoNF70: ~71 kDa).

Comparison of HeNF70, LoNF70 and HeIF-A: documentation of molluscan NF70-specific motifs in the head and the lamin-like tail domains

The alignments in Fig. 2 highlight the far better sequence conservation among the two NF70 proteins derived from different mollusc classes than the homology shared by the neuronal and non-neuronal IF proteins from the same mollusc species. The close relation of HeNF70 and LoNF70 is particularly well displayed in their identically sized rod domains comprising 361 residues. It is most obvious for the entire coil 1a and second spacer region where sequence homology approaches 100%. The first linker, the N-terminal part of the coil 1b subdomain and the entire coil 2 subdomain represent additional regions contributing to the enhanced sequence identities of both neurofilament proteins.

Beyond the rod regions the NF70 proteins again show better overall sequence conservation for their respective non-helical domains. For the head domains, sequence identity between HeNF70 and LoNF70 is 54% but only 40% for HeNF70 and HeIF-A. Although quite distinct in size and sequence all three head domains are basic and show an equally large preponderance (~35%) of serine and threonine residues. The head domains of both NF70 proteins display several characteristic sequence motifs which constitute putative phosphorylation sites for different serine and threonine protein kinases. As these elements are less prominently present in the head domain of HeIF-A, this may indicate functional differences between IF proteins of neuronal and non-neuronal origin. All but one of the possible phosphorylation sites in the HeNF70 head domain are contained within a centrally located cluster adjacent to the four consecutive aspartic acid residues. This sequence S*PMYYRSS*IT*PRQTS*ITRTS*LPTS*P (residues 45-69, asterisks mark possible phosphoacceptor residues) contains three recognition sites (SP, TP, SP) for proline dependent protein kinase, two sites (RSS, RTS) for cAMP dependent protein kinase and one site (RQTS) for the multifunctional calmodulin dependent protein kinase II. The RQTS sequence is part of the larger motif RQTSITR which represents a possible target sequence for protein kinase C (for review see Kemp and Pearson, 1990). An additional site (RSS) for cAMP dependent protein kinase occurs outside the cluster within the N-terminal part of the HeNF70 head domain. The distribution and nature of putative phosphorylation sequence motifs (RPSS,

RTS, RSS, SP, RSS) in the head domain of LoNF70 are similar to HeNF70, except that there is no obvious protein kinase C site and that clustering in the central region is less tight. In contrast, the head domain of the non-neuronal HeIF-A protein

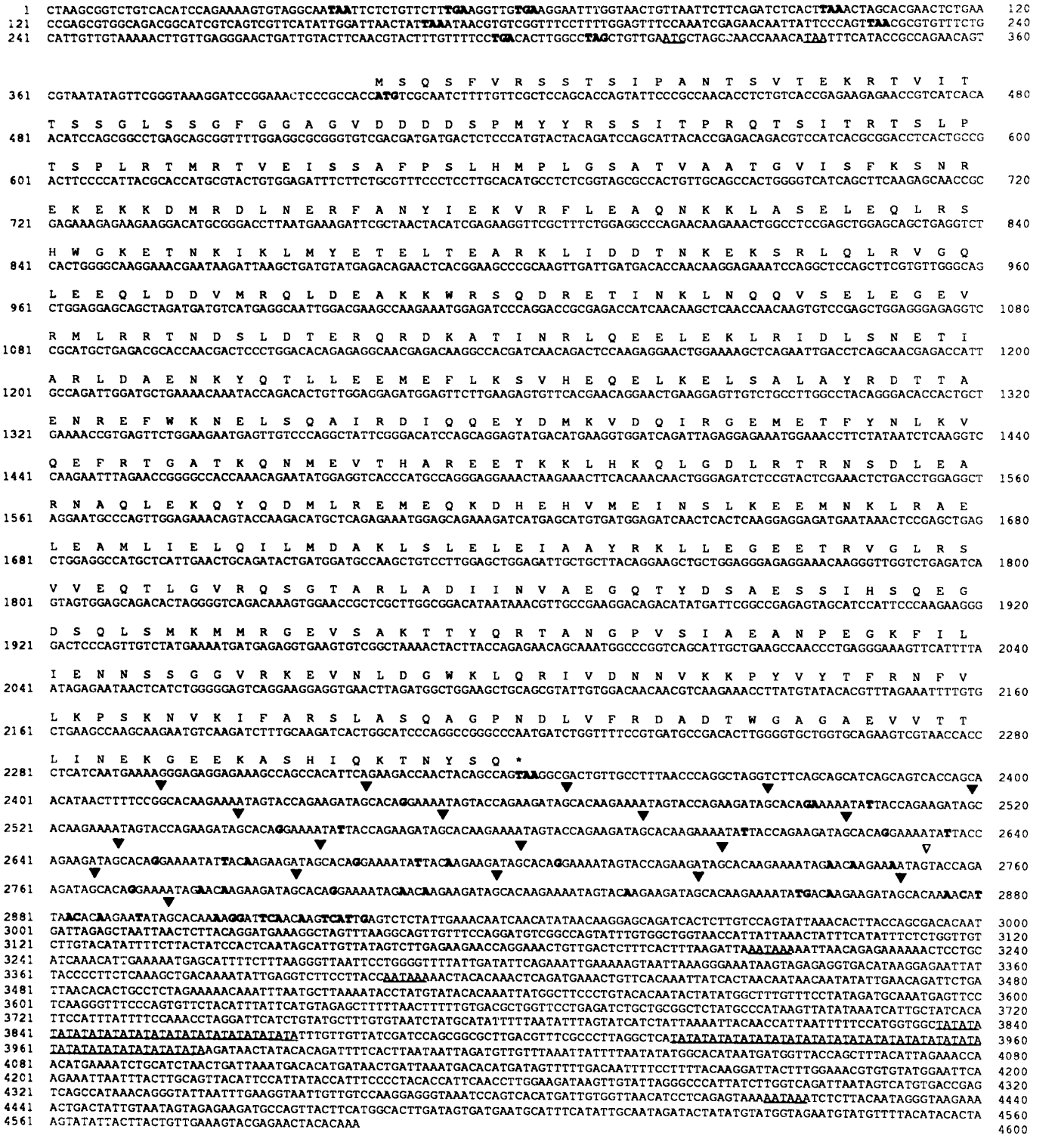

Fig. 1. Nucleotide sequence and conceptual translation of the HeNF70 cDNA. Predicted amino acid residues are given in the single letter code above the first nucleotide of each codon. The translational initiation and termination codons as well as the eight in-frame stop codons within the 5′-untranslated region are shown in bold. The only other ATG triplet within the leader sequence and the nearby in-phase TAA triplet are underlined. Arrowheads in the 3′-non coding region mark the starts of a 26 bp sequence repeated 19 times in full (filled arrowheads) and once as half a copy comprising the 3′ 13 bp (open arrowhead). Nucleotides depicted here in bold differ from those at the corresponding positions in the first repeat. The three canonical AATAAA polyadenylation signals and the two 19- and 30-fold TA repeat stretches are underlined.

contains only two types of sequence motifs (three copies of an RQS sequence and one TP motif), none of which are located within the central region.

As reported for LoNF70 (Szaro et al., 1991), the tail domain of HeNF70 is entirely distinct from the corresponding sequences of mammalian NF proteins. The C-terminal sequences of both mollusc NF70 proteins lack a high content of glutamic acid and lysine, and are completely devoid of short stretches solely consisting of acidic residues. They also lack the repeated Lys-Ser-Pro motifs which are hallmark phosphorylation target sites of vertebrate NF proteins (for review see Eagles et al., 1990). Instead, the last ~120 residues of their tail domains display substantial homology to the corresponding sequences of nuclear lamins. Lamin-like tail domains have also been documented for most invertebrate IF proteins of non-neuronal origin (for references see Introduction). However, Table 2 shows that the tail domains of the *Helix* and *Loligo* NF70 proteins reveal a closer relation to nuclear lamins than their non-neuronal counterparts from either the same or other molluscan species.

Over the lamin-homology regions of their tail domains, HeNF70 and LoNF70 share 54% sequence identity against only 28% between HeNF70 and HeIF-A. Sequence identity over the entire tail domains of the latter two IF proteins rises to 32% due to enhanced homology among their N-terminal tail sequences. Curiously these regions cover the only extended area relative to the whole polypeptides where sequence conservation between HeNF70 and HeIF-A is significantly higher than between HeNF70 and LoNF70. Located between this area and the onset of the long lamin-homology segment, the snail and squid NF70 proteins share at 100% similarity a 13 amino acid sequence with only a single conservative substitution (Fig. 2). This sequence is not present in either HeIF-A or in all other invertebrate non-neuronal IF proteins known to date and therefore may be a molluscan NF70 protein specific insert.

Fig. 2. Sequence comparison of molluscan neuronal and non-neuronal IF proteins. The predicted sequence for the neurofilament NF70 protein of the snail *Helix aspersa* (middle lines; see Fig. 1) is aligned with the sequences of the corresponding NF70 protein of the squid *Loligo pealei* (top lines; Szaro et al., 1991) and the *Helix* non-neuronal IF-A protein (bottom lines; Dodemont et al., 1990). Dashes mark positions where gaps were introduced for optimal alignment. The three structural domains (head, rod, tail) are indicated and the boundaries of the α-helical subdomains (coils 1a, 1b and 2) of the rods are delineated by arrowheads. The bar above the central part of the coil 1b subdomain sequence spans the additional 42 residues (six heptads) present in all nuclear lamins and in all currently known cytoplasmic IF proteins of invertebrates. *Lam hom* marks the onset of the large region (~ 120 residues) within all three tail domains with pronounced homology to the corresponding sequences of nuclear lamins (Dodemont et al., 1990; Weber et al., 1989). Total residue numbers are given at the ends of the sequences. Identical residues are marked by vertical lines whereas homologous amino acid substitutions are indicated by colons. Note that *inter*species sequence conservation for both neurofilament proteins is considerably higher than the *intra*species homology between the neuronal and non-neuronal IF proteins. The mollusc NF70 proteins share a unique 13 amino acid sequence (bracketed) which occurs neither in *Helix* IF-A nor in any other known metazoan non-neuronal or neuronal IF protein.

Table 1. Comparison of HeNF70 with other cytoplasmic IF proteins

Source	IF protein designation	Reference	Sequence identity	
			Rod	Tail
Squid *Loligo pealei*	Neuronal IF: Lo NF70	a	65	52
Squid *Loligo forbesi*	Neuronal IF-like: Lo SNLK	b	12	NA
Squid *Ommastrephes sloanei*	Non-neuronal IF: Om nnIF-1	c	40	28
Octopus *Octopus dofleini*	Non-neuronal IF: Oc nnIF-1	c	41	30
Snail *Helix aspersa*	Non-neuronal IF: He nnIF-A	d	38	32
Snail *Aplysia californica*	Non-neuronal IF: Ap nnIF-A	e	37	30
Nematode *Ascaris lumbricoides*	Non-neuronal IF: Asc IF-A_1	f	47	40
Nematode *Ascaris lumbricoides*	Non-neuronal IF: Asc IF-B_1	f	43	44
Nematode *Caenorhabditis elegans*	IF: Cel IF a_1 to a_4	g	41-49	39-43
Nematode *Caenorhabditis elegans*	IF: Cel IF b_1 and b_2	g	30-42	26-43
Nematode *Caenorhabditis elegans*	IF: Cel IF c_1 and c_2	g	23	NA

Sequence identity values (%) are given for the entire rod and tail domains of the neuronal NF70 protein from *Helix aspersa* versus the corresponding regions of other invertebrate cytoplasmic IF proteins designated by name and origin. Because of their entirely different sequences the tail domains of the LoSNLK and Cel IF c1/c2 proteins have not been considered (NA, not applicable). References a to g are: Szaro et al., 1991 (a); Adjaye et al., 1993 (b); Tomarev et al., 1993 (c); Dodemont et al., 1990 (d); Riemer et al., 1991 (e); Weber et al., 1989, H. Dodemont, D. Riemer, N. Ledger, K. Weber, unpublished (f); and Dodemont et al., 1994 (g).

Co-expression of neuronal and non-neuronal IF mRNAs in cerebral ganglia: multiple rare NF-specific transcripts and two abundant non neuronal species

Neuronal and non-neuronal IF mRNAs in cerebral ganglion tissue differ in size, abundance and type specific isoform complexity. On northern blots of cerebral ganglion mRNA, DNA probes derived from non-overlapping regions of the cloned HeNF70 cDNA (probes 1-3, see Materials and Methods) together detect multiple RNA transcripts of which three appear as the main bands corresponding to mRNAs of 9.5, 8 and 7 kb (Fig. 3, panels 1-3, lanes g). Several lines of evidence strongly support the identification of these transcripts as authentic neuronal IF mRNAs. First, their expression is strictly limited to cerebral ganglion tissue. Significant hybridisation with probes 1-3 was observed neither in oesophagus epithelium mRNA (panels 1-3, lanes o) nor in mRNA preparations from 12 other non-neuronal tissues (not shown, see Materials and Methods). Second, under the high stringency conditions applied, these probes do not cross-hybridise (compare panels 1-3 with panel 4) to the previously characterised non-neuronal IF mRNAs which are expressed at high levels in the glial cells of cerebral ganglia (Dodemont et al., 1990). Furthermore, since for the northern analyses polysomal RNA was used instead of total cellular RNA, detection of unspliced precursor RNAs is precluded and thus all hybridising transcripts represent genuine mRNAs.

The neuronal IF mRNAs occur at very low expression levels inferred from the rather long exposure times required for their detection. This is consistent with the finding that only a single NF70 specific cDNA was retrieved from as many as 400,000 primary clones (see above). Of all neuronal IF transcripts detected, only the 9.5 kb mRNA hybridises to each of the three probes tested and thus represents the NF70 mRNA from which the 4.6 kb cDNA originates. Obviously the 2,260 bp 3′-trailer sequence in the cDNA (Fig. 1) covers only the 5′-part of a very long 3′-non-translated region of some 7.2 kb including the poly(A) tail. Strikingly all other hybridising transcripts, including the 8 and 7 kb mRNAs, specifically anneal to either probe 2 or 3 but not to probe 1. The hybridisation data imply the presence of at least two subpopulations of *Helix* neuronal IF mRNAs which diverge substantially over their entire 5′-sequences (covered by probe 1) up to the C-terminal ends of the rod domains, while discrete regions of tail domain sequences (covered by probes 2 and 3) are either shared or highly conserved among different transcripts.

The *Helix* neuronal IF mRNA characteristics contrast sharply with the features documented for *Helix* non-neuronal IF mRNAs which have identical 5′-sequences but distinct 3′-ends (Dodemont et al., 1990). Of the two abundant transcripts detected in cerebral ganglion mRNA by the non-neuronal IF specific DNA probe 4 (see Materials and Methods), the less prominent 4.6 kb mRNA encodes HeIF-A which has a unique tail domain extension versus the C-terminally truncated but otherwise identical HeIF-B protein defined by the major 4.2 kb mRNA (panel 4, lane g). The same non-neuronal IF transcripts are also present in oesophagus epithelium (panel 4, lane o) and in all other tissues of non-neuronal origin tested (not shown).

Helix neuronal IF proteins are encoded by more than one gene: divergence of HeNF60 and HeNF70

The non-neuronal IF mRNAs of *Helix* arise from a single gene through different splicing pathways due to alternative utilisa-

Table 2. Sequence identity between the tail domains of molluscan cytoplasmic IF proteins and various nuclear lamins

IF protein	Invertebrate lamins	Vertebrate lamins	
		Type A	Type B
He NF70	22-30	29-30	27-33
Lo NF70	18-26	26-28	26-32
He nnIF-A	14-16	17-19	19-23
Ap nnIF-A	17-18	20-21	21-25
Om nnIF-1	17-20	22-26	23-28
Oc nnIF-1	17-21	22-23	19-25

The tail domains of all neuronal and non neuronal IF proteins from molluscs (gastropods and cephalopods) known to have lamin-like sequences, have been compared with the corresponding regions of nuclear lamins. For the lamins considered here see Stick (1995) and references cited therein. Sequence identity values (%) are restricted to the lamin-like homology region covering the ~120 C-terminal residues. For IF protein designations see Table 1.

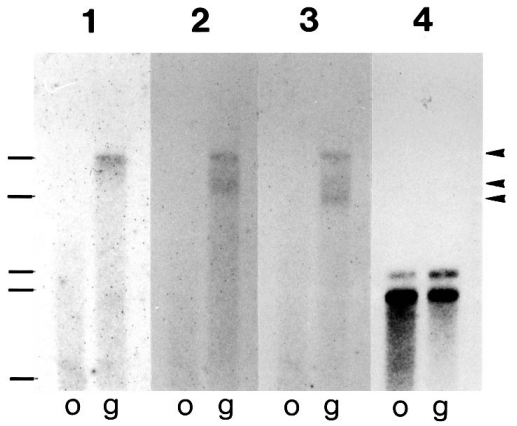

Fig. 3. Northern analysis of *Helix* IF mRNAs. Glyoxylated poly(A)-containing polysomal RNA (2 μg) from oesophagus (o) and cerebral ganglion (g) was electrophoresed in a 1.5% agarose gel and blotted onto a nitrocellulose membrane. The RNA blot filter was challenged in four successive rounds of hybridisation at high stringency, stripping and reprobing with the IF specific DNA probes designated 1 to 4 (see Materials and Methods). In each case the final wash was in 0.1× SSC/0.1% SDS at 60 °C. Autoradiography was five days for probes 1 to 3 but only five hours for probe 4. Bars mark the positions of RNA size standards: 9.5, 7.5, 4.7, 4.4 and 2.9 kb. Arrowheads denote the three NF specific RNA transcripts detected by probes 1 to 3.

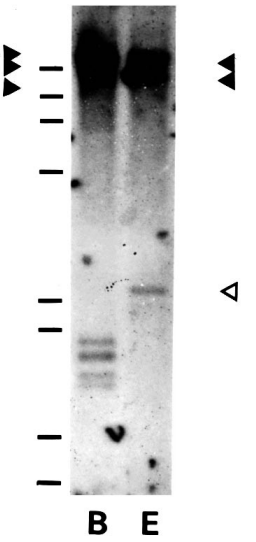

Fig. 4. Genomic representation of putative NF specific sequences in *Helix.* Genomic DNA digests (5 μg) generated with various restriction endonucleases, were electrophoresed in an 0.8% agarose gel and transferred onto a nitrocellulose membrane. DNA fragments were hydridised at low stringency to a 215 bp PCR fragment covering the C-terminal part of the rod domain of the HeNF70 cDNA sequence (see Materials and Methods). After a preliminary wash in 2× SSC/0.1% SDS at 50°C, the Southern blot was overexposed (five days of autoradiography) to enable detection of weakly cross-hybridising fragments. As many as eight bands can be discerned in the *Bam*HI (B) digest, while at least four bands are revealed in the *Eco*RI (E) digest. Note the faintly hybridised 2.5 kb *Eco*RI fragment (open arrowhead) derived from the *Helix* single non-neuronal IF gene (cf. Fig. 4B in Dodemont et al., 1990). The corresponding *Bam*HI fragment (>50 kb) is eclipsed by the large and prominent hybridisation bands, which originate from other genes. Filled arrowheads denote hybridising bands which persist through to the final high stringency wash. Positions of DNA size standards (λ/*Hind*III and ΦX174/*Hae*III digests): 23.1, 9.4, 6.6, 4.4, 2.3, 2.0, 1.4 and 1.1 kb are indicated.

tion of two distinct polyadenylation sites upon transcription termination (Dodemont et al., 1990). In contrast, both the multiplicity and the complex hybridisation characteristics of the *Helix* neuronal IF mRNAs seem difficult to reconcile with the possibility that all of these represent different RNA splicing variants generated from a unique *Helix* neurofilament gene. The 5′-end of the gene corresponding to the HeNF70 cDNA has been cloned. An 8 kb *Bam*HI-*Hind*III and a 10 kb *Hind*III fragment, detected by cDNA probe 1 (see above) on Southern blots, were isolated from size fractionated genomic DNA libraries (not shown). The two non-overlapping DNA fragments span the cDNA sequence from nucleotides 382 to 1,784 (Fig. 1) and thus encompass the coding sequences of the head and the entire rod domain. The cloned portion of the gene contains six introns which occupy precisely the same positions as the first six introns of the *Helix* non-neuronal IF gene and the *Loligo* neuronal IF gene (Dodemont et al., 1990; Way et al., 1992).

To assess whether additional neuronal IF genes besides the NF70 gene occur in *Helix*, Southern analysis was performed with a 215 bp hybridisation probe that covers the sixth exon of the gene and defines the C-terminal sequences of the NF70 rod domain. These sequences seem sufficiently conserved among different IF genes (e.g. see Fig. 2; for reviews see Fuchs and Weber, 1994; Steinert and Roop, 1988) to allow for cross-hybridisation at reduced stringency. Despite the fact that the probe sequences are entirely contained within a single exon and are not cut by the restriction enzymes used for the genomic DNA digests, a complex multiband pattern is obtained (Fig. 4). At least seven *Bam*HI and three *Eco*RI fragments are detected, that cannot be attributed to the *Helix* NF70 gene. One of these is a 2.5 kb *Eco*RI fragment which originates from the single non neuronal IF gene (Dodemont et al., 1990) and contains the sixth exon that shares 52% overall sequence identity with the HeNF70 hybridisation probe. While all weak bands including the 2.5 kb *Eco*RI band are no longer visible after washing at moderate stringency (0.5× SSC, 55°C), two to three strong bands in each digest still persist through to the final wash at high stringency (0.1× SSC, 60°C) (not shown). Since only one band can be ascribed to the *Helix* NF70 gene, the additional one or two hybridising fragments represent closely related sequences probably derived from neurofilament specific genes.

To test whether these genes are actually expressed, 3′-anchored RT-PCR was performed on cerebral ganglion mRNA using a sense primer nested within the 3′-end of the 215 bp probe used for the Southern analysis (see Materials and Methods). PCR-generated cDNAs were subcloned and the resulting plasmid library was screened by colony hybridisation at reduced stringency with a probe covering most of the tail domain sequences of the HeNF70 cDNA. Clones of which the hybridisation signals substantially diminished under conditions of increasing stringency, were selected for further characterisation. Thus, the strategies followed for the PCR and the subsequent screening were aimed at the amplification and isolation of cDNAs with similar if not identical 3′ rod domain sequences as the HeNF70 cDNA but containing putatively different tail

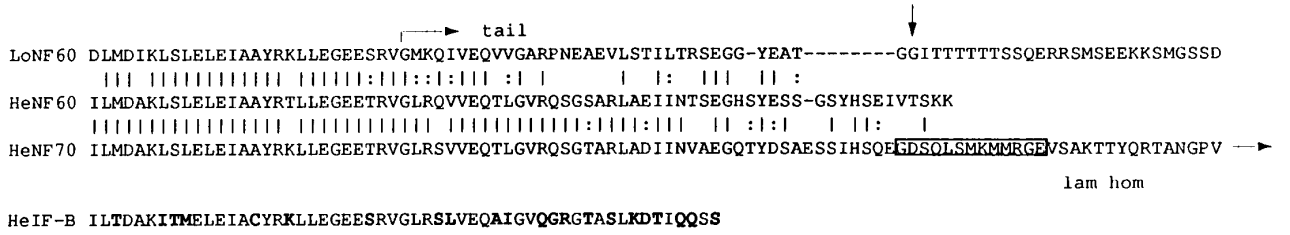

Fig. 5. Sequence divergence of the gastropod HeNF60 and HeNF70 neurofilament proteins. The amino acid sequence covering the C-terminal end of the rod and the entire tail domain of the newly identified neurofilament NF60 protein from *Helix aspersa* (HeNF60) is aligned with the corresponding sequences of the neurofilament proteins NF70 from *Helix* (HeNF70; Figs 1, 2) and NF60 from *Loligo pealei* (LoNF60; Szaro et al., 1991). The HeNF60 sequence was deduced from partial cDNA clones which were isolated by RT-PCR; the first six amino acid residues, ILMDAK, are defined by the sense primer that was originally derived from the HeNF70 cDNA sequence (see Materials and Methods). Identical and homologous residues are marked as in Fig. 2. The start of the tail domains is indicated. Complete tail sequences comprising 47 and 62 residues, respectively, are shown for HeNF60 and LoNF60; the tail sequence of HeNF70 (184 residues) continues beyond the horizontal arrow (see Figs 1, 2). Note that HeNF60 (and also LoNF60, see Fig. 2) is devoid of the NF70 specific lamin-like tail extension and its preceding 13 residue sequence motif. The vertical arrow denotes the position where the LoNF60/70 sequences diverge (see also Fig. 2); the C-terminal 27 residues of LoNF60 are encoded by an alternatively spliced exon (Way et al., 1992). Below the three molluscan neurofilament protein sequences, the complete C-terminal region is shown of the *Helix* non-neuronal IF-B protein (HeIF-B) which is the C-terminally truncated version of HeIF-A (Dodemont et al., 1990; see Fig. 2). Bold residues in the HeIF-B sequence mark differences with respect to HeNF60. The HeNF60 and HeIF-B sequences shown share 60% identity which is substantially lower than the 76% between the displayed HeNF60 and HeNF70 sequences. Note that HeNF60 and HeNF70 have highly similar rod domains with only one amino acid exchange but gradually diverge over their tail domains, and thus arise from separate genes.

domain sequences. This led to the identification of several cDNAs that represented a single novel IF sequence, HeNF60, which is the *Helix* homologue of the *Loligo* small neurofilament protein NF60 (Szaro et al., 1991). Sequences encoding a putative HeNF220 protein as the counterpart of the large neurofilament protein from *Loligo* (Way et al., 1992) have not yet been found.

The longest HeNF60 cDNA sequence comprises 298 bp (not shown) of which the 5′-end coincides with nucleotide 1,707 of the HeNF70 cDNA sequence (Fig. 1). This marks the onset of an open reading frame of 223 bp which predicts the 27 C-terminal amino acid residues of the rod domain and a tail domain of only 47 residues (Fig. 5). Thus, like LoNF60, HeNF60 neither contains the lamin-like tail extension nor the immediately preceding 13 amino acid sequence motif, both present in the larger NF70 protein counterparts (Figs 2, 5). The partial HeNF60 protein sequence shows 76% overall identity with the corresponding region of HeNF70. For their rod and tail domain sequences taken separately, these values are 96 and 64%, respectively. Fig. 5 shows that sequence divergence of the HeNF60/70 proteins proceeds gradually from the rod domains through their tail domains. The number of amino acid exchanges increases from one (Thr vs Lys) among the last 27 residues of the rod domains to 17 substitutions in the 47 residues tail domain of HeNF60 versus the corresponding region of HeNF70. At the nucleotide level these regions show 11 and 53 substitutions, respectively, while the HeNF60 coding sequence is flanked by a 75 nucleotide 3′-untranslated region which totally diverges from the HeNF70 sequence. These data document that the differences between the HeNF60 and HeNF70 sequences are not confined to delimited regions due to alternatively spliced exons as is the case for the LoNF60/70 proteins (Szaro et al., 1991; see Figs 2, 5). Thus the HeNF60/70 sequences characterised here are encoded by separate genes, in contrast to the *Loligo* triplet NF proteins which arise from a single gene (Szaro et al., 1991; Way et al., 1992).

DISCUSSION

The major aim of this study was to make comparisons of both neuronal and non-neuronal IF proteins from the same mollusc and to determine whether these proteins arise from the same or different genes. Screening of a large cDNA library constructed with cerebral ganglion mRNA from the gastropod *Helix aspersa* yielded a single cDNA clone of 4,600 bp, which encodes a novel neuronal IF protein referred to as HeNF70. Northern blots show that HeNF70 is encoded by a 9.5 kb transcript (Fig. 3). No signal was found in the 13 non neuronal tissues tested. These express the previously characterised 4.6 and 4.2 kb mRNAs encoding the *Helix* non-neuronal IF-A and IF-B proteins, respectively (Dodemont et al., 1990). The low abundance of the HeNF70 transcript in cerebral ganglion is in line with the assumption that it is neuron specific as these ganglions contain primarily glial cells. Direct support for this view comes from the high sequence identity between HeNF70 and the established neurofilament protein LoNF70 from the cephalopod *Loligo pealei.* The two proteins derived from different mollusc classes share over their rod and tail domains 65 and 52% sequence identity. These values are strikingly higher than the 38 and 32% sequence identities shown by the corresponding regions of HeNF70 and the non-neuronal HeIF-A protein from the same mollusc species (Table 1, Fig. 2). The close relation between HeNF70 and LoNF70 is also emphasised by the unique motif of 13 amino acid residues, GDSQI(L)SMKMMRGE, which precedes the lamin homology segment of the tail domains in both proteins (Fig. 2). This motif is absent from all other invertebrate IF sequences currently known and thus could be a distinctive neurofilament specific sequence in molluscs. Although formal proof of the neuronal nature of HeNF70 by immunocytochemistry or in situ hybridisation is not yet available, the combined data argue strongly that HeNF70 is the snail homologue of the squid LoNF70 protein.

In situ hybridisation on squid stellate ganglion showed location of the NF220 probe throughout the cytoplasm, while the NF70 probe hybridised most strongly to the axon hillock region. This raises the question whether the long 3′-untranslated region of the squid NF70 mRNA, which was not sequenced, influences cellular location and might contain a targeting signal (Way et al. 1992). Interestingly, the 9.5 kb NF70 mRNA of *Helix aspersa* shows in its early 3′-untranslated region a unique 26 nucleotide sequence which is repeated 19 times (Fig. 1 and Results). Future experiments using cultured gastropod neurons (see for instance Davis et al., 1992) can explore in transfection experiments with suitable cDNA derivatives whether the repeat sequences or another sequence in the 3′-untranslated region of the NF70 mRNA contain putative targeting signals. Cultured neurons may also be a suitable system to ascertain the finding on frozen squid tissue sections that mRNA encoding NF220 and NF70 can be found in the axons (Way et al. 1992). Such cultures should also help to identify other neuronal IF proteins and their mRNAs and allow some insight into neurofilament assembly.

Non-neuronal cytoplasmic IF proteins of molluscs, nematodes and annelids have the extended coil 1b typical of nuclear lamins, while all vertebrate IF proteins display a coil 1b subdomain shortened by 6 heptads i.e. 42 residues. Many but not all of these protostomic IF proteins also share with the lamins a long lamin homology segment covering a major part of the tail domain (Bovenschulte et al., 1995; Dodemont et al., 1990, 1994; Weber et al., 1989). Thus it is thought that the archetypal IF protein arose from a mutated nuclear lamin by the loss of two topogenic sequences related to lamin functionality, the nuclear localisation signal and the C-terminal CaaX box (Dodemont et al., 1990; Döring and Stick, 1990). Since the long coil 1b version is also found in the neurofilament proteins of molluscs (Szaro et al., 1991; Way et al., 1992; this study) we speculate that because of evolutionary parsimony the (protostomic) neurofilament gene arose from a duplicated ancestral non-neuronal IF gene. Of the currently known molluscan neurofilament proteins only HeNF70 and LoNF70 have retained the lamin homology segment in the tail domain while LoNF220 has acquired an entirely different tail domain (Way et al., 1992). This situation is not without precedent. Three of the 12 IF proteins of the nematode *Caenorhabditis elegans* have also lost the lamin tail homology segment and acquired entirely different sequences (Dodemont et al., 1994).

It now seems well established that both neuronal and non-neuronal IF proteins of molluscs, annelids and nematodes carry particular lamin-like features absent from chordate IF proteins. Are these sequence features just evolutionary remnants of the postulated lamin progenitor (see above), or do they have direct structural and functional significance? A possible direct influence on IF structure can be analysed. Thus, the regular beading pattern of metal shadowed filaments changes from 22 nm in the case of vertebrate IF to 26 nm for the lamins so far tested. Current IF models attribute this change to the six extra heptads (see Fuchs and Weber, 1994; Heins and Aebi, 1994). It should be possible to verify this view with those protostomic IF proteins which are known to provide in vitro self assembly into IF (Weber et al., 1988, 1989). Since these filaments show control of the diameter at 10 nm, while nuclear lamins assemble in vitro beyond filaments in the form of bundles and paracrystalline arrays (Heins and Aebi, 1994), the sequence differences between lamins and protostomic IF proteins may lead to a future understanding of their different self assembly properties. It seems firmly established that the tail domains protrude from the filament wall. This raises the question of whether the lamin-like tail domains, when present, provide different contacts from other tail domains in interactions with neighbouring cellular components. Alternatively, IF binding proteins recognised by the tail domains may show compensating sequence changes. This problem, however, can only be approached once the field of IF binding proteins is better defined.

The pioneering work on squid neurofilaments described three cDNAs encoding LoNF60, LoNF70 and LoNF220, and indicated that the corresponding mRNAs arise from a single not yet fully characterised gene via different splicing pathways (Way et al., 1992). We have now shown, that in the gastropod *Helix aspersa* non-neuronal and neuronal IF proteins are encoded by distinct genes. The gene encoding the non-neuronal IF proteins has been previously documented (Dodemont et al., 1990). Of the *Helix* neurofilament proteins, we have so far characterised a complete cDNA for HeNF70 and a partial cDNA for HeNF60. Their sequence differences argue against a derivation from a single neurofilament gene in the way suggested for the squid. This view is supported by Southern and northern analyses, which indicate that HeNF60 and HeNF70 are encoded by different genes. A clearer appreciation of the differences in neurofilament gene number and organisation in the two molluscs will require the full characterisation of the corresponding genes in both *Helix* and squid.

We thank Tomma Eisbein for skilful technical assistance, Irene Gajewski for photography and Gillian Paterson for typing the manuscript. J. A. was supported by a postdoctoral fellowship from the Max Planck Society. The sequences reported above are available from the EMBL/GenBank under accession numbers X86347 (HeNF70) and X86348 (HeNF60).

REFERENCES

Adjaye, J., Marsh, P. J. and Eagles, P. A. M. (1993). Immunological properties and cDNA sequence analysis of an intermediate-filament-like protein from squid neuronal tissue. *J. Cell Sci.* **106**, 1283-1290.

Birnstiel, M. L., Busslinger, M. and Strub, K. (1985). Transcription termination and 3′ processing: the end is in site! *Cell* **41**, 349-359.

Bovenschulte, M., Riemer, D. and Weber, K. (1995). The sequence of a cytoplasmic intermediate filament (IF) protein from the annelid *Lumbricus terrestris* emphasizes a distinctive feature of protostomic IF proteins. *FEBS Lett.* **360**, 223-226.

Carden, M. J., Trojanowski, J. Q., Schlaepfer, W. W. and Lee, V. M. -Y. (1987). Two stage expression of the neurofilament polypeptides during rat neurogenesis with early establishment of adult phosphorylation patterns. *J. Neurosci.* **7**, 3489-3504.

Davis, L., Dou, P., DeWit, M. and Kater, S. B. (1992). Protein synthesis within neuronal growth cones. *J. Neurosci.* **12**, 4867-4877.

Dodemont, H., Riemer, D. and Weber, K. (1990). Structure of an invertebrate gene encoding cytoplasmic intermediate filament (IF) proteins: implications for the origin and the diversification of IF proteins. *EMBO J.* **9**, 4083-4094.

Dodemont, H., Riemer, D., Ledger, N. and Weber, K. (1994). Eight genes and alternative RNA processing pathways generate an unexpectedly large diversity of cytoplasmic intermediate filament proteins in the nematode *Caenorhabditis elegans*. *EMBO J.* **13**, 2625-2638.

Döring, V. and Stick, R. (1990). Gene structure of nuclear lamin L_{III} of *Xenopus laevis;* a model for the evolution of IF proteins from a lamin-like ancestor. *EMBO J.* **9**, 4073-4081.

Eagles, P. A. M., Pant, H. C. and Gainer, H. (1990). Neurofilaments. In

Cellular and Molecular Biology of Intermediate Filaments (ed. R. D. Goldman and P. M. Steinert), pp. 37-94. Plenum Publishing Corporation, New York.

Fisher, D. Z., Chaudhary, N. and Blobel, G. (1986). cDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins. *Proc. Nat. Acad. Sci. USA* **83**, 6450-6454.

Fliegner, K. H., Ching, G. Y. and Liem, R. K. H. (1990). The predicted amino acid sequence of α-internexin is that of a novel neuronal intermediate filament protein. *EMBO J.* **9**, 749-755.

Fuchs, E. and Weber, K. (1994). Intermediate filaments: Structure, dynamics, function and disease. *Annu. Rev. Biochem.* **63**, 345-382.

Heins, S. and Aebi, U. (1994). Making head and tails of intermediate filament assembly, dynamics and networks. *Curr. Opin. Cell Biol.* **6**, 25-33.

Jackson, R. J. (1993). Cytoplasmic regulation of mRNA function: The importance of the 3′ untranslated region. *Cell* **74**, 9-14.

Julien, J. -P., Côté, F., Beaudet, L., Sidky, M., Flavell, D., Grosveld, F. and Mushynski, W. (1988). Sequence and structure of the mouse gene coding for the largest neurofilament subunit. *Gene* **68**, 307-314.

Kemp, B. E. and Pearson, R. B. (1990). Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* **15**, 342-346.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell Biol.* **115**, 887-903.

Lees, J. F., Shneidman, P. S., Skuntz, S. F., Carden, M. J. and Lazzarini, R. A. (1988). The structure and organization of the human heavy neurofilament subunit (NF-H) and the gene encoding it. *EMBO J.* **7**, 1947-1955.

McKeon, F. D., Kirschner, M. W. and Caput, D. (1986). Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. *Nature* **319**, 463-468.

Riemer, D., Dodemont, H. and Weber, K. (1991). Cloning of the non-neuronal intermediate filament protein of the gastropod Aplysia californica; identification of an amino acid residue essential for the IFA epitope. *Eur. J. Cell Biol.* **56**, 351-357.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning : A Laboratory Manual, 2nd edition.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Shaw, G. and Kamen, R. (1986). A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659-667.

Steinert, P. M. and Roop, D. R. (1988). Molecular and cellular biology of intermediate filaments. *Annu. Rev. Biochem.* **57**, 593-625.

Stick, R. (1995). Nuclear lamins and the nucleoskeleton. In *The Cytoskeleton* vol. I: *Structure and Assembly* (ed. J. E. Hesketh and I. Pryme), pp. 257-296. Jai Press Incorporated, Greenwich, Connecticut, USA.

Szaro, B. G., Pant, H. C., Way, J. and Battey, J. (1991). Squid low molecular weight neurofilament proteins are a novel class of neurofilament protein. *J. Biol. Chem.* **266**, 15035-15041.

Tomarev, S. I., Zinovieva, R. D. and Piatigorsky, J. (1993). Primary structure and lens-specific expression of genes for an intermediate filament protein and a β-tubulin in cephalopods. *Biochim. Biophys. Acta* **1216**, 245-254.

Way, J., Hellmich, M. R., Jaffe, H., Szaro, B., Pant, H. C., Gainer, H. and Battey, J. (1992). A high-molecular-weight squid neurofilament protein contains a lamin-like rod domain and a tail domain with Lys-Ser-Pro repeats. *Proc. Nat. Acad. Sci. USA* **89**, 6963-6967.

Weber, K., Plessmann, U., Dodemont, H. and Kossmagk-Stephan, K. (1988). Amino acid sequences and homopolymer-forming ability of the intermediate filament proteins from an invertebrate epithelium. *EMBO J.* **7**, 2995-3001.

Weber, K., Plessmann, U. and Ulrich, W. (1989). Cytoplasmic intermediate filament proteins of invertebrates are closer to nuclear lamins than are vertebrate intermediate filament proteins; sequence characterization of two muscle proteins of a nematode. *EMBO J.* **8**, 3221-3227.

Wilson, T. and Treisman, R. (1988). Removal of poly(A) and consequent degradation of c-*fos* mRNA facilitated by 3′ AU-rich sequences. *Nature* **336**, 396-399.

(Received 21 June 1995 - Accepted 11 August 1995)