

BRM 70019

## Cloning and developmental expression of the murine neurofilament gene family

Jean-Pierre Julien\*, Dies Meyer\*\*, David Flavell, Jacky Hurst and Frank Grosveld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, London (U.K.)

(Accepted 12 August 1986)

*Key words:* Neurofilament; Intermediate filament; Development; Gene expression

DNA clones encoding the 3 mouse neurofilament (NF) genes have been isolated by cross-hybridization with a previously described NF-L cDNA probe from the rat. Screening of a  $\lambda$ gt10 cDNA library prepared from mouse brain RNA led to the cloning of an NF-L cDNA of 2.0 kb that spans the entire coding region of 541 amino acids and of an NF-M cDNA that covers 219 amino acids from the internal  $\alpha$ -helical region and the carboxy-terminal domains of the protein. These cDNA clones were used as probes to screen mouse genomic libraries, and cosmid clones containing both NF-L and NF-M sequences were isolated as well as overlapping cosmids containing the NF-H gene. This strongly suggests that the 3 neurofilament genes are organised in a cluster and derived by gene duplication of a common ancestral gene. RNA blot analyses using specific DNA probes for each of the genes indicate that NF mRNAs are differentially expressed during brain development. The NF-L and NF-M mRNAs are detected early in the embryonal brain, with a progressive increase in their levels during development, while the NF-H mRNA is barely detectable at embryonal stages and accumulates later in the postnatal brain.

### INTRODUCTION

Intermediate filament (IF) genes are characterized by a cell-type specific expression during development. There are 5 classes of IFs; vimentin filaments in cells of mesenchymal origin, keratin filaments in epithelial cells, desmin in muscle cells, glial filaments in astrocytes and, finally, neurofilaments (NFs) in neurons (for review, see refs. 17, 27, 35).

The different types of IF proteins share a common structural feature; a highly conserved domain rich in  $\alpha$ -helices, flanked by amino and carboxy-terminal domains which are hypervariable in sequence and size<sup>4,5,9,10,28,29,30,34</sup>. Protein and DNA sequence data provide evidence that IF genes are derived from a common ancestral gene. In addition, comparison between the organization of 3 different types of IF genes, namely, vimentin, keratin and the glial filament genes, reveals a remarkable conservation of in-

tron positions<sup>1,13,18,23,24,31</sup>.

The 3 NF proteins with apparent molecular weights of 68,000 (NF-L), 145,000 (NF-M) and 200,000 (NF-H)<sup>12</sup> on SDS gel electrophoresis, share a homologous  $\alpha$ -helical domain with other types of IF proteins, but contain a long extension at their carboxy-terminals responsible for their difference in size<sup>6,7,15</sup>.

The isolation of partial cDNA clones encoding the NF-L protein has been reported recently<sup>16,20</sup>. Using cross-hybridization between the different genes, we have isolated either genomic and/or cDNA clones for all the neurofilament genes.

We describe here the isolation and characterization of DNA clones encoding the 3 NF proteins. Specific NF DNA probes for each gene were used to detect NF-L, NF-M and NF-H mRNAs at different stages of the brain development.

\* Present address: Institut du Cancer de Montreal, Centre Hospitalier Notre-Dame, 1560 est Sherbrooke, Montreal, Canada H2L 4M1.

\*\* Present address: Erasmus University, Postbox 1738, 3000 DR Rotterdam, The Netherlands.

Correspondence: F. Grosveld, Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

## MATERIALS AND METHODS

*Cloning and sequencing of NF DNA clones*

Double-stranded cDNA was synthesized from poly(A)<sup>+</sup> mRNA extracted from mouse brain as described<sup>16</sup> and cloned into bacteriophage  $\lambda$ gt10 using

EcoRI linkers. The resulting cDNA library (about 5  $\times$  10<sup>5</sup> plaques) was screened with an XhoI/BglII cDNA probe encoding the rat NF-L protein<sup>16</sup>. The cDNA inserts from hybridizing clones were sub-cloned into either the bacteriophage M13mp8 or into plasmid pUC8 for further analysis. Sequencing was

ACCACTGCTCCTGCCCACTGCAGTCACC - 29

S S F G Y D P Y F S T S Y K R R Y V E T P R V H I S S V R S G Y - 32  
 ATGAGTTCGGTTCGGCTACGATCCGTACTTTTCGACCTCCTACAAGCGCGCTATGTGGAGACGCCCCGGTGCACATCTCCAGCGTGCAGCGGGCTAC - 128

S T A R S A Y S S Y S A P V S S S L S V R R S Y S S S S G S L M P - 65  
 AGCACGGCGCCTCCGGTACTCCAGTACTCCGCGCGGCTCTCTCTCGCTGTCGGTGGCGCGCAGCTACTCGTCCAGCTCTGGCTCTTTGATGCC - 227

S L E N L D L S Q V A A I S N D L K S I R T Q E K A Q L Q D L N D - 98  
 AGCTTGAGAAATCTCGATCTGAGCCAGGTAGCCGCATCAGCAACGACCTCAAGTCTATCCGCACACAAGAGAAGGCACAGCTGCAGGACCTCAACGAT - 326

R F A S F I E R V H E L E Q Q N K V L E A E L L V L R Q K H S E P - 131  
 CGCTTCGCCAGCTTCTCGAGCGGTGCAGGCTGGAGCAGCAGAACAAGTCTTTGGAAGCCGAGCTGTGGTGCTGCGCCAGAACACTCTGAGCCT - 425

S R F R A L Y E Q E I R D L R L A A E D A T N E K Q A L Q G E R E - 164  
 TCCCGCTCCGCGCCTGTACGAGCAGGAGATCCGCGATCTCGCTTGGCAGCGGAAGACGCCACTAACGAGAAGCAAGCGCTGCAGGGCGAGCGCGAG - 524

G L E E T L R N L Q A R Y E E E V L S R E D A E G R L M E A R K G - 197  
 GGGTGGAGGAGACTCTGCGCAACCTCGAGCTCGCTATGAGGAAGAAGTGTGAGCCGCGAGGACGCCGAGGGCCGGCTGATGGAACGCCGCAAGGT - 623

A D E A L A R A E L E K R I D S L M D E I A F L K K V H E E E I A - 230  
 CCGGATGAGGCCCTCCGCCGCGGAGCTGGAGAAGCGCATCGACAGCCTGATGGACGAGATAGCTTCTGAGAAGTGCACGAGGAAGAGATCGCC - 722

E L Q A Q I Q Y A Q I S V E M D V S S K P D L S A A L K D I R A Q - 263  
 GAGCTGCAGGCTCAGATCCAGTATGCTCAGATCTCGTGGAGATGGACGTCTCTCCAAGCCGACCTCTCCGCGCTCTCAAGGACATCCGCGCTCAG - 821

Y E K L A A K N M Q N A E E W F K S R F T V L T E S A A K N T D A - 296  
 TACGAGAAGTGGCCGCCAAGAACATGCAGAACGCCGAAGAGTGTTCAGAGCCGCTTCCAGTGTGTAACCGAGAGCGCCCAAGAACACCGAGCCT - 920

V R A A K D E V S E S R R L L K A K T L E I E A C R G M N E A L E - 329  
 GTGCGCTGCCAAGGACGAGGTGTCGAAAGCCGCCGCTGTCAAGGTAAGACCTGGAGATCGAAGCCTGCCGGGTATGAACGAAGCTCTGGAG - 1019

K Q L Q E L E D K Q N A D I S A M Q D T I N K L E N E L R S T K S - 362  
 AAGCAGTGCAGGAGTAGAGACAAGCAGAAATGCAGACATTAGCCGATGCAGGACACAATCAACAACCTGGAGAATGAGTGAAGACACGAAGAGC - 1118

E M A R Y L K E Y Q D L L N V K M A L D I E I A A Y R K L L E G E - 395  
 GAGATGGCCAGTACCTGAAGGAGTACCAGGACCTCCTCAATGTCAAGATGGCCTTGGACATCGAGATTGCAGCTTACAGAAAACCTTTGGAAGGGCAA - 1217

E T R L S F T S V G S I T S G Y S Q S S Q V F G R S A Y S G L Q S - 428  
 GAGACCAGGCTCAGTTTACCACGCTGGGTAGCATAACCAGCGGCTACTCTCAGAGCTCGCAGTCTTCGGCGTTCTGCTTACAGTGGCTTCAGAGC - 1316

S S Y L M S A R S F P A Y Y T S H V Q E E Q T E V E E T I E A T K - 461  
 AGCTCCTACTTGATGCTGCTCGCTCTTCCAGCCTACTATACCAGCCAGTCCAGGAAGAGCAGACAGAGTTCGAGGAGACCATTTGAGGCTACGAAA - 1415

A E E A K D E P P S E G E A E E E K E K E E G E E E E G A E E E - 494  
 GCTGAGGAGGCAAGGATGAGCCCCCTCTGAAGGAGAAGCAGAAGGAGGAGGAAGGAGAAGAGGAGGAGGAAGGAGAAGCGCTGAGGAGGAA - 1514

E A A K D E S E D T K E E E E G G E G E E E D T K E S E E E E K K - 527  
 GAAGTGC AAGGATGAGTCTGAAGACAAAAAGAAGAAGAAGTGGTGGAGGTGAGGAGGAACACCAAAAGAACTGTAAGAGGAAAGAGAAGAAA - 1613

E E S A G E E Q V A K K K D - 541  
 GAGGAGTGTCTGGAGAGGAGCAGGTGGCTAAGAAGAAAGATTGAGCCCTATTCCTCAACTATCCAGGAAAAGTCTCCCAATCAGTCAACCTCATC - 1712

ACCAACCAACCAGTTGAGTCCAGATCCTATACAAATTAAAGTCAATACATGTATAATTCTGAGATGACTTAGTTGGACTTTCAATGTGTGCTATG - 1811

AATTCCTCCTTACGAGAGTATCTGTTTCTGCTGTCAGAGTGGCTTTCTGGCTTGTGCCAGCCTGTGCATGGTCCATGTTATGAGTTCAGGATCTATG - 1910

GCAATGTGAATCACACAGATGTTTACAATAATAATAAAAAAAAAACCACACACACACCAATAAATGAAT - 1932

Fig. 1. Nucleotide sequence of an NF-L cDNA clone. The predicted amino acid sequence of the open reading frame is shown above the DNA sequence. The demarcation of the  $\alpha$ -helical regions is indicated by a line above the amino acid sequences. Putative polyadenylation sites are underlined.

carried out by the dideoxy chain termination method<sup>32</sup> and/or according to Maxam and Gilbert<sup>25</sup>. The NF-L, NF-M and NF-H genes were isolated from a mouse genomic cosmid library in the vector pLTC, a variant of pTCF in which the HpaI site has been replaced by an XhoI site<sup>8</sup>.

#### RNA blot analysis

Total RNA was prepared as described<sup>22</sup> from mouse brain at different stages of embryonal and postnatal development. The RNA samples were fractionated by electrophoresis on 1% agarose gels in the presence of formaldehyde, blotted and hybridized with the appropriate <sup>32</sup>P-labelled DNA probes as described<sup>16</sup>.

## RESULTS

#### Cloning of a cDNA encoding the mouse NF-L protein

A cDNA library of approximately  $5 \times 10^5$  recombinants in  $\lambda$ gt10 was prepared from poly(A)<sup>+</sup> RNA of the mouse brain. Screening of this library with a cDNA probe encoding the rat NF-L protein<sup>16</sup> yielded 46 positively hybridizing plaques. The largest NF-L cDNA insert of 2.0 kb in size was further analyzed and sequenced. As shown in Fig. 1, the cDNA spans the entire coding region of the mouse NF-L protein with 541 amino acid residues, predicting a molecular weight of approximately 61,000. The  $\alpha$ -helical domain deduced from the heptamer repeats of hydrophobic residues is typical of intermediate filament

proteins. A comparison of the mouse NF-L protein with the porcine sequences<sup>7</sup> indicates a high degree of conservation of NF-L protein with only 4% and 2% divergence in the amino terminal and helical regions, respectively. Indeed, the interspecies differences occur primarily in the carboxy-terminal regions which are characterized by a very high content of glutamic acid residues. In Fig. 1 two potential polyadenylation sites at 277 and 299 nucleotides downstream from the translational stop codon TGA are underlined. The identity of the 29 nucleotides upstream of the ATG initiating codon have been confirmed by S1 nuclease mapping experiments as part of the NF-L mRNA (data not shown), although we do not yet know the actual position of the 5' end of the mRNA. The NF-L cDNA hybridizes to two mRNA species as shown by Northern blot hybridization<sup>16,20</sup> (Fig. 3). The third hybridizing band (2 kb, Fig. 3) is a breakdown product of the 3.5 kb mRNA (data not shown).

#### Cloning of a cDNA encoding the NF-M protein

Screening of the cDNA library at reduced stringency (final wash in 0.3 SSC 55 °C) with the NF-L cDNA probe yielded cross-hybridizing clones which encode the NF-M protein. The sequence of an NF-M cDNA insert of 650 bp is shown in Fig. 2. The cDNA spans the  $\alpha$ -helical region as well as part of the carboxy-terminal tail domain of the NF-M protein. The deduced amino acid sequence of the mouse NF-M shows an extensive homology (99%) with the reported porcine sequences<sup>5</sup>. Southern blot analyses show a single



Fig. 2. Nucleotide sequence of an NF-M cDNA clone with the deduced amino acid sequence. The  $\alpha$ -helical region is indicated by a line above the amino acid sequences.

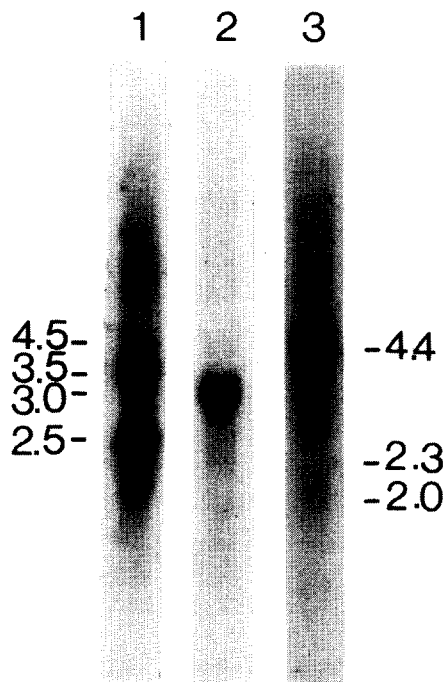


Fig. 3. RNA blot hybridization with neurofilament probes. NF-L (lane 1), NF-M (lane 2) and NF-H (lane 3) DNA probes of 300 bp, 650 bp and 1200 bp respectively, were hybridized to 10  $\mu$ g of total RNA from adult mouse brain. The marker sizes were obtained from  $\lambda$ XHindIII/EcoRI digest (kb). The blots were washed to a stringency of  $0.3 \times$  SSC at 65  $^{\circ}$ C. The additional NF-L 2 kb band (lane 1) is due to a breakdown product of the 3.5 kb mRNA, while the additional signal in NF-H is caused by a breakdown of the 4.4 kb mRNA (lane 3).

strongly hybridizing fragment in several different restriction enzyme digests, which suggests the presence of a single NF-M gene (data not shown). In contrast to the NF-L cDNA, which detects a 2.5 and a 3.5 kb mRNA<sup>16,20</sup>, the NF-M probe detects a single mRNA of 3.0 kb in size on Northern blot analyses (Fig. 3).

#### *Isolation of genomic fragments encoding the NF-L, NF-M and NF-H genes*

Using the NF-L and NF-M cDNA clones as probes, we screened two mouse genomic libraries cloned in the cosmid vector pLTC, a variant of pTCF<sup>8</sup>. Two cosmids from independent libraries hybridized strongly to the NF-L and NF-M cDNA's on Southern blots. Restriction mapping (see Fig. 4, top) showed that the entire NF-L gene was present adjacent to vector sequences at the 5' extremity of the cosmids. The NF-M gene is situated approximately 30 kb 3' from the NF-L gene at the 3' extremity of the

cosmids, although part of the NF-M gene is not present since several restriction sites detected by Southern blots of genomic DNA are not present on the 3' end of the cosmid (not shown). The NF-M cDNA was subsequently used to screen a mouse genomic library cloned in the phage vector EMBL-3 and yielded a single clone corresponding to the NF-M gene. This recombinant overlaps with the cosmids at the 3' extremity, it has all the restriction sites (HindIII EcoRI) and presumably contains the entire NF-M gene.

Screening the genomic cosmid libraries at reduced stringency yielded 3 overlapping clones which generated restriction fragments that did not correspond to the fragments of the NF-L or the NF-M gene on Southern blots of total genomic DNA (not shown). Fig. 4, bottom, shows a restriction map of these cosmid clones that were subsequently identified as containing the NF-H gene. The XhoI/BglII fragment of 1.2 kb, illustrated by the heavy line, has been used as a hybridising probe on Northern blots and detects a brain specific mRNA of 4.5 kb (Fig. 3). This RNA is therefore larger than any known NF of other IF mRNA encoding keratin, vimentin, desmin or the glial fibrillary acidic protein<sup>10,19,30,34,36</sup>.

The NF-L cDNA cross-hybridized with a 1.2 kb XhoI/BglII fragment corresponding to a highly conserved region of IF proteins found in the end of the  $\alpha$ -helical domains. In Fig. 5 this exon sequence of the putative NF-H gene has been aligned with the corresponding amino acid sequences of the NF-L and NF-M proteins. As expected, the 3 NF proteins share a strong homology in this region with 86% sequence identity at the amino acid level. Also identified from the limited sequence data are the positions of two introns which are in different positions in relation to the conserved  $\alpha$ -helical region of the protein, to those of all other IF proteins (for review, see ref. 35), which are mostly found in identical positions. However, the intron positions of the cosmid agree perfectly with the intron positions of the human NF-L gene (Julien et al., submitted), which together with the large mRNA gene strongly argue that this clone represents the murine NF-H gene.

Any attempts to link the NF-H cosmids and the NF-L or M cosmids by chromosomal walking, did not result in the isolation of any new clones, outside the already mapped area. This result, together with the isolation of nearly identical NF-L or M clones from

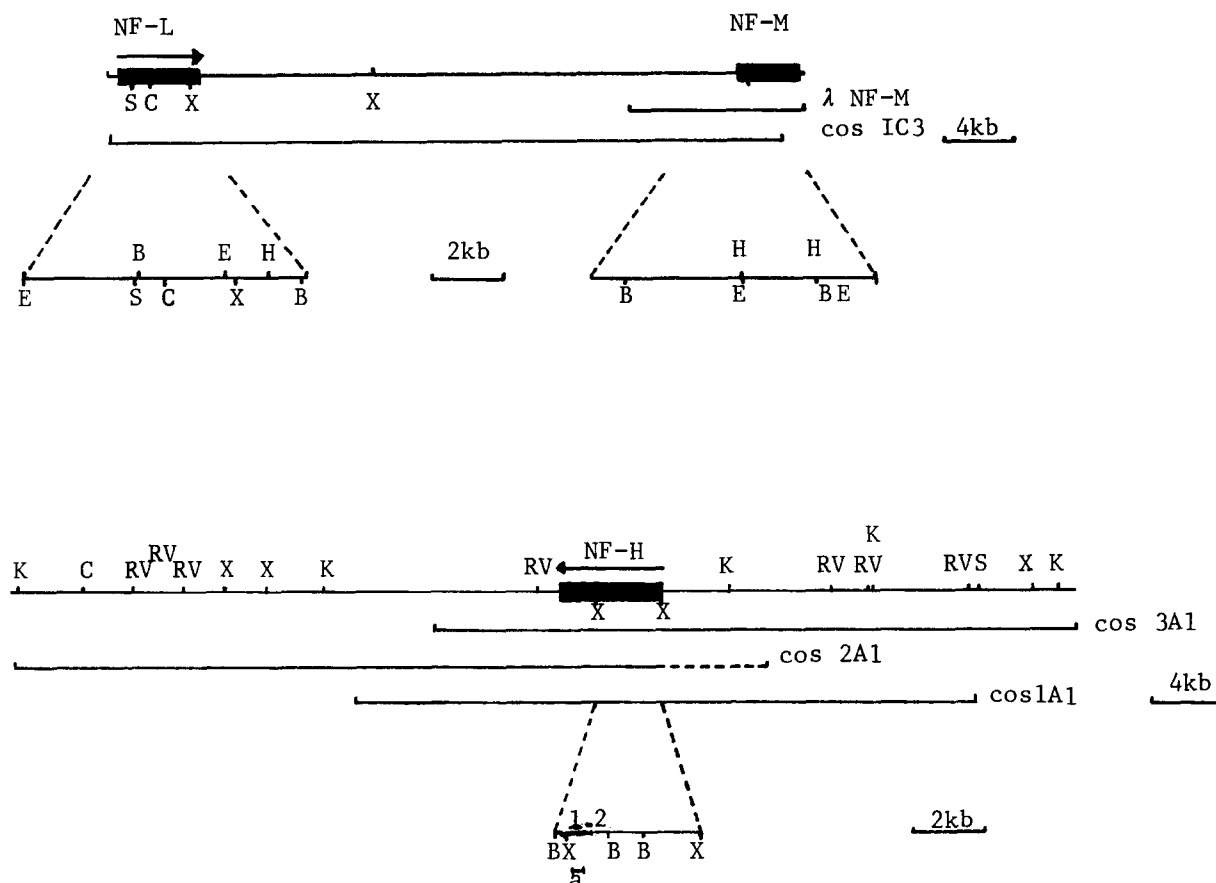


Fig. 4. Top: restriction cleavage map of a genomic clone containing the NF-L and NF-M genes. Arrows indicate direction of transcription, S, SalI; C, ClaI; X, XhoI; B, BglII; E, EcoRI; H, HinIII; K, KpnI; RV, EcoRV. Bottom: restriction cleavage map of genomic clones containing the NF-H gene.

separate libraries, indicates that the further flanking regions might be difficult to obtain as recombinant cosmid clones.

#### Developmental expression of neurofilament mRNAs

Northern blots of RNA obtained at different de-

velopmental stages of the embryonal and postnatal mouse brain were hybridized with specific NF DNA probes that were  $^{32}\text{P}$ -labelled at approximately the same specific activity. For hybridization we have used NF-L and NF-M cDNA probes of 300 bp and 660 bp, respectively, and a BglII/XhoI NF-H probe

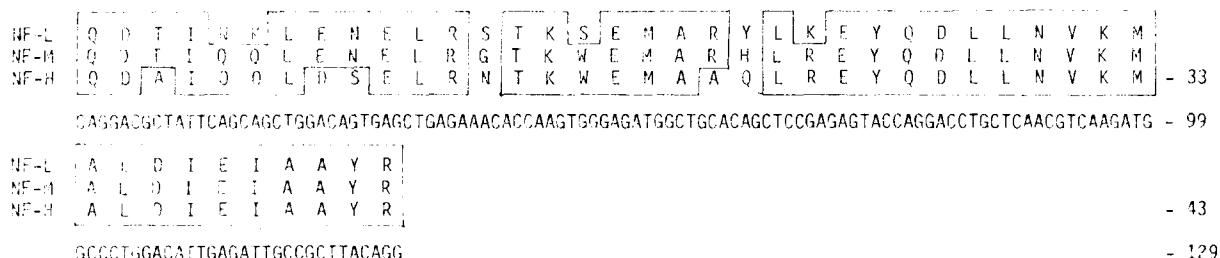


Fig. 5. Exon sequence of the NF-H gene corresponding to the highly conserved region at neurofilament proteins. The nucleotide sequence of the NF-H gene (fragment a in Fig. 4) is shown with the predicted amino acid sequences. The corresponding amino acid sequences of the NF-L and NF-M proteins have been aligned. Boxed residues represent sequence identity.

of 1.2 kb containing at least 130 bp of exon sequences (Fig. 4). The autoradiographs in Fig. 6 reveal that the NF-L and NF-M transcripts are already detected in the brain of the embryo at 11 days of gestation. We did not investigate the expression of NF mRNAs at earlier stages of development because of technical difficulties in obtaining sufficient brain tissues for Northern blot analyses. There is a progressive increase of NF-L and NF-M mRNA levels during embryonal and early postnatal neuronal development, up to a level higher than that found in the adult ani-

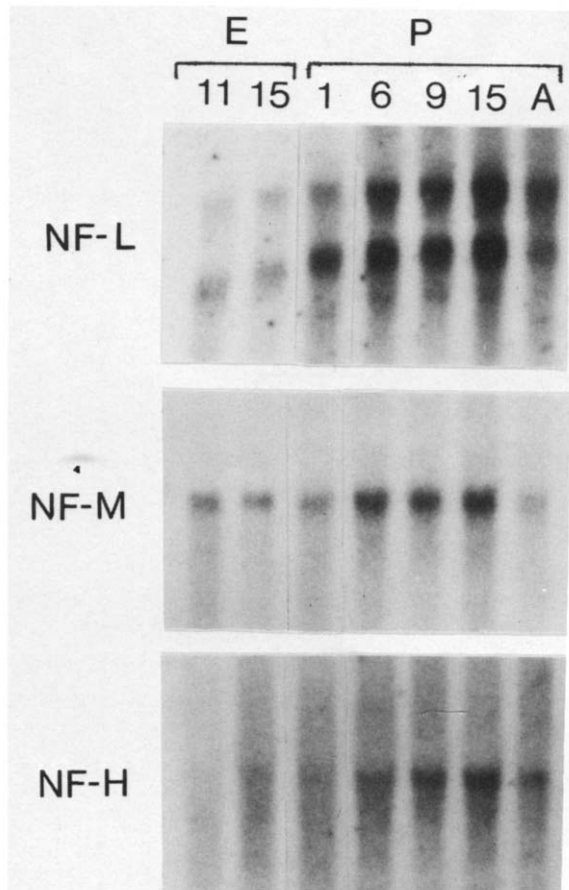


Fig. 6. Developmental expression of neurofilament mRNAs. Total RNA from mouse brain at different stages (days) of embryonal (E) and postnatal (P) development was resolved on 1% agarose gel and transferred to nitrocellulose. Each lane contains 10  $\mu$ g of total RNA with the exception of the adult (A) sample which contains 7  $\mu$ g of RNA. The blots were hybridized to NF-L, NF-M and NF-H DNA probes of 300 bp, 650 bp and 1200 bp which were  $^{32}$ P-labelled at similar specific activities. The NF-L, NF-M and NF-H autoradiographs were exposed for 4 days, 6 h and 4 days, respectively, with an intensifying screen at  $-70^{\circ}\text{C}$ .

mal. In contrast to NF-L and NF-M, barely detectable amounts of the 4.5 kb NF-H mRNA are observed before birth. The NF-H mRNAs accumulate to a significant extent only in the postnatal brain. It is clear that all the NF mRNAs are expressed at lower levels in the adult brain when compared to the young postnatal brain, although it should be noted that the difference is exaggerated in Fig. 6 (7  $\mu$ g of mRNA in the adult RNA lane vs 10  $\mu$ g in the others).

#### DISCUSSION

We report here the isolation of DNA clones encoding the 3 mouse NF proteins. The cloning of the NF-M and NF-H genes was based on cross-hybridization with an NF-L cDNA probe of the rat<sup>16</sup>. The cDNA clones that were isolated for the NF-L and NF-M genes are not full length cDNAs, probably caused by incomplete first strand synthesis and/or cleavage by EcoRI prior to cloning in the  $\lambda$ gt10 vector. This might also explain why we did not pick up any NF-H clones from the library; the NF-H mRNA will contain a large 3' tail region downstream from the  $\alpha$ -helical region and consequently will not be picked up by the probe. To screen the genomic library we used an Xho/BglII restriction fragment that does not cover the region containing the long purine tracts (nucleotides 1445-1655 in Fig. 1) to prevent hybridization to a repetitive element in the mouse genome<sup>20</sup>. A comparison at the nucleic acid sequence level between NF-L, NF-M and NF-H genes indicates that the nucleotides 1067-1232 in Fig. 1 encode the conserved consensus sequence of IF proteins that must be largely responsible for the observed cross-hybridizations. For instance, the consensus amino acid sequence region shown in Fig. 5 has (between NF-L, NF-M and NF-H) over 80% nucleotide sequence identity.

In general, homologies between IF proteins are entirely restricted to the  $\alpha$ -helical regions, although type II cytokeratin proteins display certain common features of domain arrangement in their carboxytermini<sup>14</sup>. A comparison of the sequences in Figs. 1 and 2 reveals that the homologies between NF-L and NF-M proteins are not restricted only to the  $\alpha$ -helical region, but extend also to the so-called hypervariable tail region. In both NF proteins, the tail region is characterized by repeats of glutamic acid residues encoded by long purine tracts, suggesting that the NF

genes arose by duplication of a common ancestral NF gene. The highly charged tail region of NF proteins might represent functionally important domains related to cell-type specific functions. There is evidence that these regions form side-arms on the NF structure which interact with other neuronal components<sup>6,11,15</sup>. A comparison of the NF-H genomic clone with the NF-L and NF-M cDNA sequence shows that at least two intron sequences in the NF-H gene occur at different positions to those in the other (non NF) intermediate filament genes<sup>35</sup>, though in the same positions as those in the human NF-L gene (Julien et al., submitted), suggesting that the neurofilament gene family arose by the reverse transcription of an IF mRNA followed by homologous recombination into the genome and the regaining of introns in different positions to those of other IF genes. Gene duplication would then have created 3 neurofilament genes with identical intron positions.

Expression of IF genes during embryonic development often correlates with the emergence of definitive cell types<sup>17,27</sup>. The development blots in Fig. 5 show that the expression of NF genes is an early event in brain differentiation as NF-L and NF-M mRNA species are detected in the early embryonic brain at 11 days of gestation. The early appearance of NF mRNAs is in accordance with immunofluorescence data, using antibodies specific to NF proteins<sup>2,3</sup>. These studies have shown that the initial appearance of NFs is concomitant with the initiation of axon extension in the postmitotic neurons<sup>3</sup>, although

the NF-M protein was reported to be transiently expressed in a subpopulation of replicating neuroepithelial cells in the chick<sup>2</sup>. Interestingly, NF mRNAs are not coordinately expressed during brain development. The NF-H mRNAs accumulate to a significant extent only in the postnatal brain in contrast to NF-L and NF-M mRNAs (Fig. 6). These results are in accordance with the postnatal detection of the NF-H protein in the brain<sup>33</sup>. However, there is a possibility of embryonic expression of NF-H mRNAs in some neurons since NF-H immunoreactivity has been reported at the embryonic stages<sup>3</sup>. Whether the early and late mRNA accumulation is regulated by transcriptional controls, or, alternatively by preferential mRNA stabilization is not yet known.

In situ hybridization using a specific NF-L DNA probe has recently been used for distinct and rapid localization of NF-L mRNAs in neuronal cell bodies<sup>21</sup>. The availability of specific NF DNA clones encoding the three NF proteins will allow these new approaches to be used to investigate NF gene expression of the developing brain.

#### ACKNOWLEDGEMENTS

We thank Karina Yazdanbakhsh for technical assistance and Cora O'Carroll for the preparation of the manuscript. J.-P. J. was supported by a Fellowship from the Medical Research Council of Canada.

#### REFERENCES

- Balcarek, J.M. and Cowan, N.J., Structure of the mouse glial fibrillary acidic protein gene: implications for the evolution of the intermediate filament multigene family, *Nucl. Acids Res.*, 13 (1985) 5527-5543.
- Bennett, G. and DiLullo, C., Transient expression of a neurofilament protein by replicating neuroepithelia cells of the embryonic chick brain, *Dev. Biol.*, 107 (1985) 107-127.
- Cochard, P. and Paulin, D., Initial expression of neurofilaments and vimentin in the control and peripheral nervous system of the mouse embryo in vivo, *J. Neurosci.*, 4 (1984) 2080-2094.
- Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H.H.M. and Flavell, R.A., The construction of cosmid libraries which can be used to transform eukaryotic cells, *Nucl. Acids Res.*, 10 (1982) 6715-6732.
- Geisler, N. and Weber, K., The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins, *EMBO J.*, 1 (1982) 1649-1656.
- Geisler, N., Fisher, S., Vandekerckhove, J., Plessmann, U. and Weber, K., Hybrid character of a large neurofilament protein (NF-M): intermediate filament type sequences followed by a long acidic carboxy-terminal extension, *EMBO J.*, 3 (1984) 2701-2706.
- Geisler, N., Kaufman, E., Fischer, S., Plessmann, U. and Weber, K., Neurofilament architecture combines structural principles of intermediate filaments with carboxy-terminal extensions increasing in size between triplet proteins, *EMBO J.*, 2 (1983) 1295-1302.
- Geisler, N., Plessmann, U. and Weber, K., The complete amino acid sequence of the major mammalian neurofilament protein (NF-L), *FEBS Lett.*, 182 (1985) 475-478.
- Hanukoglu, I. and Fuchs, E., The cDNA sequence of a human epidermal keratin: divergence of sequence but conservation of structure among intermediate filament proteins, *Cell*, 31 (1982) 243-252.

- 10 Hanukoglu, I. and Fuchs, E., The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins, *Cell*, 33 (1983) 915–924.
- 11 Hirokawa, N., Glicksman, M.A. and Willard, M.B., Organization of mammalian neurofilament polypeptides within the neuronal cytoskeleton, *J. Cell Biol.*, 98 (1984) 1523–1536.
- 12 Hoffman, P.N. and Lasek, R.J., The slow component of axonal transport. Identification of major structural polypeptides of the axon and their generality among mammalian neurons, *J. Cell Biol.*, 66 (1975) 351–366.
- 13 Johnson, L.D., Ioler, W.N., Zhou, X.-M., Roop, D.R. and Steinert, P.M., Structure of a gene for the human epidermal 67-kDa keratin, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 1896–1900.
- 14 Jorcano, J.L., Franz, J.K. and Franke, W.W., Amino acid sequence diversity between bovine epidermal cytokeratin polypeptides of the basic (type II) subfamily as determined from cDNA clones, *Differentiation*, 28 (1984) 155–163.
- 15 Julien, J.-P. and Mushynski, W.E., The distribution of phosphorylation sites among identified proteolytic fragments of mammalian neurofilaments, *J. Biol. Chem.*, 258 (1983) 4019–4025.
- 16 Julien, J.-P., Ramachandran, K. and Grosveld, F.G., Cloning of a cDNA encoding the smallest neurofilament protein, *Biochim. Biophys. Acta*, 825 (1985) 398–404.
- 17 Lazarides, E., Intermediate filaments: a chemically heterogeneous, developmentally regulated class of proteins, *Annu. Rev. Biochem.*, 51 (1982) 219–250.
- 18 Lehnert, M.E., Jorcano, J.L., Zentgraf, H., Blessing, M., Franz, J., Franz, J.K. and Franke, W.W., Characterization of bovine keratin genes: similarities of axon patterns in genes coding for different keratins, *EMBO J.*, 3 (1984) 3279–3287.
- 19 Lewis, S.A., Balcerek, J.M., Krek, V., Shelanski, M. and Cowan, N.J., Sequence of a cDNA clone encoding mouse glial fibrillary acidic protein: structural conservation of intermediate filaments, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 2743–2746.
- 20 Lewis, S.A. and Cowan, N.J., Genetics, evolution and expression of the 68,000-mol-wt neurofilament protein: isolation of a cloned cDNA probe, *J. Cell. Biol.*, 100 (1985) 843–850.
- 21 Liesi, P., Julien, J.-P., Vilja, P., Grosveld, F. and Reichardt, L., Specific detection of neuronal cell bodies: in situ hybridization with a biotin-labelled neurofilament cDNA probe, *J. Histochem. Cytochem.*, in press.
- 22 Maniatis, T., Fritsch, E.F. and Sambrook, J., *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY, 1982.
- 23 Marchuk, D., McCrohon, S. and Fuchs, E., Remarkable conservation of structure among intermediate filament genes, *Cell*, 39 (1984) 491–498.
- 24 Marchuk, D., McCrohon, S. and Fuchs, E., Complete sequence of a gene encoding a human type I keratin: sequences homologous to enhancer elements in the regulatory region of the gene, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 1609–1613.
- 25 Maxam, A. and Gilbert, W., In L. Grossman and K. Moldave (Eds.), *Methods in Enzymology*, Vol. 65, Academic Press, New York, 1980, pp. 499–560.
- 26 Nabeshima, Y., Fukii-Kuriyama, Y., Muramatsu, M. and Ogata, K., Alternative transcription and two modes of splicing result in two myosin light chains from one gene, *Nature (London)*, 308 (1984) 333.
- 27 Osborn, M. and Weber, K., Biology of disease. Tumor diagnosis by intermediate filament typing: a novel tool for surgical pathology, *Lab. Invest.*, 48 (1983) 372–394.
- 28 Quax-Jeuken, Y., Quax, W.J. and Bloemendal, H., Primary and secondary structure of hamster vimentin predicted from the nucleotide sequence, *Proc. Natl. Acad. Sci. U.S.A.*, 80 (1983) 3548–3552.
- 29 Quax, W., Egberts, W.V., Hendricks, W., Quax-Jeuken, Y. and Bloemendal, H., The structure of the vimentin gene, *Cell*, 35 (1983) 215–223.
- 30 Quax, W., Van Den Heuvel, R., Vree Egberts, W., Quax-Jeuken, Y. and Bloemendal, H., Intermediate filament cDNA's from BHK-21 cells: demonstration of distinct genes for desmin and vimentin in all vertebrate classes, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 5970–5974.
- 31 Rieger, M., Jorcano, J.L. and Franke, W.W., Complete sequence of a bovine type I cytokeratin gene: conserved and variable intron positions in genes of polypeptides of the same cytokeratin subfamily, *EMBO J.*, 4 (1985) 2261–2267.
- 32 Sanger, F., Coulsen, A.R., Barrell, A., Smith, J.H. and Roe, B., Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing, *J. Mol. Biol.*, 143 (1980) 161–178.
- 33 Shaw, G. and Weber, K., Differential expression of neurofilament triplet proteins in brain development, *Nature (London)*, 298 (1982) 277–279.
- 34 Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.C., Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments, *Nature (London)*, 302 (1983) 794–800.
- 35 Steinert, P.M., Steven, A.C. and Roop, D.R., The molecular biology of intermediate filaments, *Cell*, 42 (1985) 411–419.
- 36 Zehner, Z.E. and Paterson, B.M., Characterization of the chicken vimentin gene: single copy gene producing multiple mRNAs, *Proc. Natl. Acad. Sci. U.S.A.*, 80 (1983) 911–915.