GEN 02515

# Sequence and structure of the mouse gene coding for the largest neurofilament subunit

(Intermediate filament; phosphorylation site; neuron-specific gene; CpG-rich island; intron positions; recombinant DNA)

# Jean-Pierre Julien<sup>a</sup>, Francine Côté<sup>a</sup>, Lucille Beaudet<sup>a</sup>, Malak Sidky<sup>b</sup>, David Flavell<sup>c</sup>, Frank Grosveld<sup>c</sup> and Walter Mushynski<sup>b</sup>

<sup>a</sup> Institut du Cancer de Montréal, Centre Hospitalier Notre-Dame, Montreal (Canada H2L 4M1);<sup>b</sup> Department of Biochemistry, McGill University, Montreal (Canada H3G1Y6), Tel. 514-398-7286; <sup>c</sup> National Institute for Medical Research, London NW71AA (U.K.) Tel. 01-959-3660

Received 18 February 1988 Revised 22 April 1988 Accepted 22 April 1988 Received by publisher 10 May 1988

## SUMMARY

We have determined the complete nucleotide sequence of the mouse gene encoding the neurofilament NF-H protein. The C-terminal domain of NF-H is very rich in charged amino acids (aa) and contains a 3-aa sequence, Lys-Ser-Pro, that is repeated 51 times within a stretch of 368 aa. The location of this serine-rich repeat in the phosphorylated domain of NF-H indicates that it represents the major protein kinase recognition site. The nfh gene shares two common intron positions with the nfl and nfm genes, but has an additional intron that occurs at a location equivalent to one of the introns in non-neuronal intermediate filament-coding genes. This additional nfh intron may have been acquired via duplication of a primordial intermediate filament gene.

## INTRODUCTION

Neurofilaments are major cytoskeletal elements of nerve cells that play an important role in the control of axonal caliber. They are formed by the copolymerization of three neuron-specific proteins with apparent  $M_r$ 's of 68 000 (NF-L) 145 000 (NF-M) and 200 000 (NF-H), as determined by SDS-polyacrylamide-gel electrophoresis (Hoffman and Lasek, 1975; Liem et al., 1978; Julien and Mushynski, 1982). In common with other IF proteins, the neurofilament subunits contain a highly conserved  $\alpha$ -helical domain of approx. 40 kDa capable of forming coiled-coil structures (Geisler and Weber, 1982; Geisler et al., 1984; Hanukoglu and Fuchs, 1982; Quax et al., 1983; 1985; Steinert et al., 1983; Julien et al., 1985; 1986). A striking feature of neurofilament proteins is their C-terminal domains, which

Correspondence to: Dr. J.-P. Julien, Institut du Cancer de Montréal, Centre Hospitalier Notre-Dame, 1560 Sherbrooke Est, Montreal (Canada H2L 4M1) Tel. (514)876-5497.

Abbreviations: aa, amino acid(s); bp, base pair(s); IF, intermediate filament; kb, kilobase(s) or 1000 bp; NF-H, the largest neurofilament subunit; *nfh*, gene coding for NF-H; *nfl* and *nfm*, genes coding for the small- and the mid-size neurofilament proteins, respectively; nt, nucleotide(s); SDS, sodium dodecyl sulfate.

308

retain a highly charged character despite their different lengths (Geisler et al., 1984; 1985; 1987; Julien et al., 1986; 1987; Lewis and Cowan, 1986; Robinson et al., 1987; Levy et al., 1987; Myers et al., 1987). In the case of NF-H, the tail domain is highly phosphorylated in axons (Julien and Mushynski, 1982; 1983; Geisler et al., 1985; Carden et al., 1985) and forms cross-links between neurofilaments and their surrounding structures (Julien et al., 1983; Hirokawa et al., 1984).

The *nfl* and *nfm* genes have been sequenced recently (Julien et al., 1986; 1987; Lewis and Cowan, 1986; Levy et al., 1987; Myers et al., 1987; Zopf et al., 1987) and found to be linked in the murine genome (Julien et al., 1986). The exon-intron organizations of the *nfl* and *nfm* genes are similar but the total lack of similarities with those of other IF-coding genes has led to the proposal that the ancestral neurofilament gene originated via an mRNA-mediated transposition event (Lewis and Cowan, 1986; Levy et al., 1987).

We report here the sequence and the exon-intron organization of the mouse n/h gene. The deduced amino acid sequence of NF-H contains in its C-terminal region an unusual serine-rich repeat that probably corresponds to the major phosphorylation site in neurofilament proteins. In addition, the structure of the n/h gene is consistent with the notion that an early duplication event may have separated neurofilament genes from the rest of the IF gene family.

#### MATERIALS AND METHODS

# (a) Cloning

The screening at low stringency of a mouse genomic library in the cosmid vector pLTC with a *nfl* cDNA probe led to the isolation of a cross-hybridizing clone, designated cos3A1, that contained *nfh* exon sequences (Julien et al., 1986).

# (b) Sequencing

For sequence analysis, DNA fragments from the nfh gene were subcloned into the M13mp18 vector or into the Bluescript plasmid (Stratagene, Inc.) to

generate, with exonuclease III and mung-bean nuclease, overlapping deletion mutants. Sequencing was carried out by the dideoxy chain-termination method (Sanger et al., 1980). To confirm the transcription start point, mouse brain RNA ( $20 \mu g$ ) was annealed to a 1.9-kb *NotI-Bam*HI probe labeled at the 5'-end and subjected to S1 nuclease analysis (Maniatis et al., 1982).

#### **RESULTS AND DISCUSSION**

## (a) Sequencing of the *nfh* gene

Screening of a mouse genomic library at reduced stringency with a *nfl* cDNA probe yielded a crosshybridizing cosmid clone, designated cos3A1. The restriction map of the genomic 40-kb insert is shown in Fig. 1. A 1.2-kb *XhoI-BglII* fragment that crosshybridized with the *nfl* probe was found previously to include a small exon sequence corresponding to a highly conserved region of IF proteins and was used to detect on Northern blots a brain-specific mRNA of approx. 4 kb (Julien et al., 1986). This exon sequence is now identified as being in exon 3 of the *nfh* gene.

Mapping of the cos3A1 clone revealed the presence of a NotI site, 4 kb upstream from exon 3 (Fig. 1). The site recognized by NotI occurs infrequently and can indicate the position of a CpGrich island surrounding the transcription start site of a vertebrate gene (Lindsay and Bird, 1987). Fragments flanking the NotI site were subcloned and sequenced by the dideoxy method of Sanger et al. (1980). In agreement with the prediction this region was found to contain a TATAAA box and to encode amino acid sequences corresponding to the N-terminal portion of NF-H (Geisler et al., 1985). The 5' region of the gene was subsequently confirmed by S1 nuclease protection experiments that positioned the cap site at 20 nt downstream from the TATAAA sequence (data not shown).

The 3.9-kb XhoI-XhoI fragment, which contains exon 3, as well as the adjoining 3.0-kb XhoI-EcoRV fragment and the 5-kb EcoRV-SalI fragment (Fig. 1) were subcloned into the Bluescript plasmid vector to produce overlapping deletion mutants that were used for dideoxynucleotide sequencing. The exon-intron

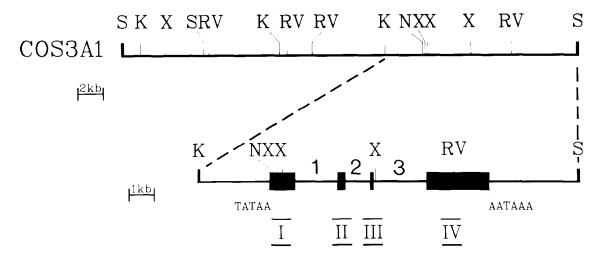


Fig. 1. Restriction cleavage map of the clone cos3A1 and exon-intron organization of the mouse neurofilament *nfh* gene. The clone cos3A1 was found to contain the *nfh* gene (Julien et al., 1986). Abbreviations of restriction enzyme names are: K, KpnI; N, NotI; RV, EcoRV; S, Sal1; X, XhoI. The lowest line is a schematic representation of the *nfh* gene with exons as blackened boxes (marked I-IV). Introns are designated as 1-3 (see also Fig. 2).

organization of the nfh gene was elucidated by comparing the deduced amino acid sequences of open reading frames with partial NF-H sequences reported previously (Geisler et al., 1985; Robinson et al., 1986). The nfh gene has only three introns of 1.7, 1.1 and 2.0 kb, which are illustrated in the schematic presentation in Fig. 1. Approximately 50% of nfh intron sequences have been determined, and the exon-intron boundaries are shown in Fig. 2. All of the junctions follow the typical 5'GT-AG3'rule (Breathnach et al., 1978). The intron positions in the nfh nucleotide sequence are also indicated by arrows in the sequence given in Fig. 3 and their significance will be discussed below.

# (b) The multiple phosphorylation sites of NF-H: a serine-rich sequence repeated in tandem

The *nfh* gene encodes a protein of 1087 as with a calculated  $M_r$  of 116000. This is much smaller than the  $M_r$  200000 estimated on SDS gels. The high content of charged amino acids and of phosphate

moieties in the C-terminal domain appears to be responsible for the anomalous gel-electrophoretic mobility of NF-H (Julien and Mushynski, 1982; Kaufmann et al., 1984; Georges and Mushynski, 1987). Indeed the central portion of the C-terminal domain has a very unusual protein structure. The amino acid sequence Lys-Ser-Pro is repeated 51 times. The repeat units, which are underlined in Fig. 3, are generally flanked by three amino acids (Ala, Gly, Ile)-Glu-(Ala, Val) or Glu-Lys-Ala. The sequences Val-Lys-Glu-Gly-Ala, Val-Lys-Glu-Asp-Ile and Val-Lys-Glu-Glu-Ala separate Lys-Ser-Pro units at the border of the repeated domain. The Lys-Ser-Pro repeats have also been found in the NF-H proteins of other species, including rat (Robinson et al., 1986), pig (Geisler et al., 1987), rabbit (Mack et al., 1988) and human (Lees, J.F., Shneidman, P.S., Skuntz, S.F., Carden, M.J. and Lazzarini, R.A., unpublished). A smaller number of copies of the repeats occur also in the NF-M protein (Myers et al., 1987; Levy et al., 1987). Interestingly, there is a correlation between the number of repeat

INTRON 1 TGGTTCCGAGgtgcgcggggggggggggggggggggggggggggggg
INTRON 2 CTCCTACCAGgtcgagcagaggggggggggg(1.1 kb)tctgatctgtcttcccccagGACGCTATTC
INTRON 3 CCGCTTACAGgtaagatgacccaggagete(2.0 kb)cccattatecactccacagAAAGCTCCTGG

Fig. 2. Exon-intron junctions of the mouse nfh gene. Upper-case letters are exon sequences. For positions of introns 1-3 see Fig. 1.

10 GCGGA <u>TATAAAA</u> GA	20 IGCCGGAGTCC	30 CAGAGCTOC	40 COCAGTGCTGC	50 CTOCCCOTC	60 CCAGCCCGG	70	50 XCGCT0GCGG	90 CGCACCTGCT	100 CCGGCCATGJ	110 TGAGCTTCOG	
130 CGATGCGCTGCTGG	140 IGCGCCCCGTT	150 COCOCCOCTO	160 CACGGAGGCG	170 GCAGCCTGCA	180 CTACTCGCTG	190 MGCCGCAAGO	200 CAGGCCCGGG	210 CCGCACCCC	220	230	240 CCCCTT
DALL 250	260	270	280	290	300	310	320	330	340	350	360
CCACTCGTGGGCGC H S W A	R T S V	ssv	S A S	PSRF	RGA	λSS	T D S I	AGACACCCTA DTL	AGCAACGGCO S N G	PEGC	CGTGGT V V
370 GGCGGCGGTGGCGG A A V A											
490 GGCGGCGCTGCGGC A A L R											
610 GCAGGAGCACCTGC	620	630	640	650	660	670	680	690	700	710	720
Q E H L 730											
GGCGCGCGTGGAGC A R V E			<b>CTGCAGGAGG</b>		CCTGCGGCGC	CACCACCAGO	AGGAGGTGGG	CGAGCTGCTC	GGTCAGATCO	AGGCTGCGG	GGCCGC
850 GCAGGCGCAGGCTC Q A Q A			880 CTCAAGTGCG L K C								
970 GTGGTTCCGAGTGA	980	990 ACTCTCAGAG	1000 GCAGCCAAAG	1010 TGAACACAGA	1020 TGCTATGCGC	1030 AGCGCCCAAG	1040 Baggagataac	1050 TGAGTACCGG	1060 CGGCAGCTG	1070 CAAGCCAGGAC	1080 CACAGA
W F R V 1090	1100	1110	1120	1130	1140	1150	1160	1170 Intron	2 1180	1190	1200
GTTGGAGGCCCTGA L E A L								х ор			
1210 GAGAAACACCAAGT R N T K	1220 GGGAGATGGC W E N A	1230 TGCACAGCTC A O L	1240 CCGAGAGTACC R E Y	1250 AGGACCTGCT Q D L L	1260 CAACGTCAAG N V K	1270 ATGGCCCTGG H A L	1280 GACATTGAGAT D I E J	1290 TGCCGCTTAC	AGAAAGCTCO	1310 TTGGAAGGCGA L E G E	1320 AGAGTG E C
1330 TCGGATTGGCTTTG	1340 GTCCGAGTCC	1350 CTTCTCTCTT	1360 ACTGAAGGAC	1370 TCCCAAAAAT	1380 TCCCTCCATA	1390 TCCACGCACJ	1400 TAAAAGTCAJ	1410 	1420 ATGATAAAG	1430 Гтадтасасаа	1440 ATCCGA
R I) G F 1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560
GAAGGAAACTGTGA K E T V	IVEG	Q T E	EIR	VTEG	VTE	EED	кеас	) G Q E	GEE	A E E G	1680
1570 AAAAGAAGAAGAGG K E E E	1580 AACTAGCAGC E L A A	AGCTACATCI A T S	1600 CCCCCTGCAG P P A	1610 AAGAGGCTGC E E A A	1620 ATCTCCAGAA S P E	1630 АЛАБЛАЛССИ К Е Т	1640 AGTCTCGTGT K S R V	1650 "GAAAGAAGAG ' К Е Е	1660 GCCAAGTCCC A K S	CAGGTGAGGC	CAAGTC
1690 CCCAGGTGAGGCCA P G E A	1700 AGTCCCCAGC	1710 TGAGGCCAAC	1720 TCCCCAGGTG	1730 AGGCCAAGTC E A K S	1740 CCCAGGTGAG P G E	1750 RECCANGTOCO	1760 CAGGTGAGGO PGE	1770 CAAGTCTCCA	1780 GCTGAGCCCI	1790 AGTCTCCAGO KSPA	1800 TGAOCC E P
1810 CAAGTCTCCAGCTG	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920
	1940	P A E	P K S	1970	K S P	G E A	K S P S	2010	S P A 2020	E A <u>K S</u> 2030	2040
TGAGGCCAAATCTC E A K S	C1/00701000	CARACTOCI	OCTOR ACCOUNTS	AGTCACCAGC	TGAAGCCAAG	TCACCAGCTO	AAGCCAAATO	TCCAGCTACA	V K S	CAOGTGAGOC PGEM	K S
2050 ACCATCTGAGGCCAL PSEA	2060 AATCTCCAGCT K S P A	2070 Гдалссала Е а к	2080 TCTCCAGCTGI S P A I	2090 AggCCAAATC E A K S	2100 TCCAGCTGAG P A E	2110 GCCAAATCTC A K S	2120 CAGCTGAGGT P A E \	2130 CAAGTCACCA K S P	2140 AGGTGAGGCC G E A	2150 AAGTCTCCAG K S P	2160 CTGAGCC A E P
2170 CAAGTCACCAGCTG	2180 AGGCCAAATC	2190 TCCAGCTGAG	2200 GTCAAGTCAC	2210 CAGCTGAGGC	2220 CAAATCTCCA	2230 GCTGAGGTCA	2240	2250	2260	2270 GCAGTGAAGT	2280 CACCAGC
<u>K S P</u> A I 2290	2300	2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TGAGGCCAAGTCTCC	P A A V	<u>K 5 P</u>	GEAI	K S P G	Е А <u>к</u>	<u>5 P A</u>	EAKS	5 P A E	A <u>K 5</u>	<u> </u>	<u> </u>
$\begin{array}{c} 2410 \\ TCCAGAGAAGGCCAR \\ \underline{P}  E  K  A  P \end{array}$											
2530 TCCTGAGAAGGCCAI PEKAI											
2650 Agacatcagaccccc	2660 CTGAGCAGGTO	2670 Балалстсст	2680 GCCAAGGAGA	2690 AGGCCAAGTCO	2700 CCCTGAGAAG	2710 GAAGAAGCCA	2720 AGACTTCTG	2730 AAAGGTGGCT	2740 CCCAAGAAG	2750 Gaagagetga	2760 AGTCACC
D I R P I 2770 TGTGAAGGAGGAGG	2780	2790	2800	2810	2820	2830	2840	2850	2860	2870	2880
VKEE											

(Fig. 3)

2890		910 2920	2930 2940	2950 2960	2970 2980 Балдадаадааладссстс	2990 3000 GCCTCAGAGGAGGAGAC
крку	ЕЕККЕ	TPTEKE			<b>G E K K K A V</b>	ASEEET
3010		030 3040	3050 3060	3070 3080	3090 3100 SCCAAAGAACCTAGCAAACCC.	3110 3120 ACAGAGACGGAAAAGCC
PAKL						тетекр
			3170 3180 AGGAGGAGAAGACCACAGA	3190 3200 GTCCAGGAAGCCTGAGGAG	3210 3220 AAGCCCAAAATGGAGGCCAAG	3230 3240 GTCAAGGAGGATGACAA
	МРААР				КРКНЕАК	
3250		270 3280	3290 3300	3310 3320	3330 3340 CCAGAGAAGACCACAGAGGAC	3350 3360 AAGGCCACCAAGGGAGA
	E P S K P		K S S S T I	QKESQP	РЕКТТЕД	KATKGE
3370 Балсталсасался К		390 3400 Заатассаладаласто	3410 3420 CAGGACGGTCCCAGTACTC	3430 3440 Aggggtcggcgtaataaat	3450 3460 TTATTCTTCCTTCCCTCC	3470 3480 Бталдалдаласастос
3490 TTAGATGGTGGGCC		510 3520 Сладляттстаттала	3530 3540 Тталсттасалсасалса	3550 3560 TAACCCTGAGCCTTGTCCCC	3570 3580 CCACGCCGAAAACCCTCCCCA	3590 3600 GGTGATGGACAATTATG
3610 ATAGCTTCTTGTAG		530 3640 ATGCTGAACGCTACGCG1	3650 3660 Алалсасосотсталала	3670 3680 CTGCCCCCCCCTTTCCAAGI	3690 3700 AAGTGCATTTATTTCCTGTA1	3710 3720 IGTCCAACTGACAGATG
3730 Ассяслатаатаа		750 3760 Сосаттатосттолаато	3770 3780 ПТСТААССТАТТССТСАА	3790 3800 TGCCTTCTTGTTTTCCAAAC	3810 3820 GAGTGGTCAGGCCCTTGCCC/	3830 3840 Адтасасстсстддаа
3850 Gagetgeageage		970 3880 Годссастваассасос	3890 3900 Agggtgtactctccactg	3910 3920 AAGTCCATTTCAATTGCTTC	3930 3940 Catgc <u>aataaaac</u> caagtgct	3950 3960 ГТСТБА <u>ЛАТААЛА</u> СТ <del>Б</del> Т
3970		990 4000	4010 4020	4030 4040	4050 4060 Segcatetteateagateace	4070

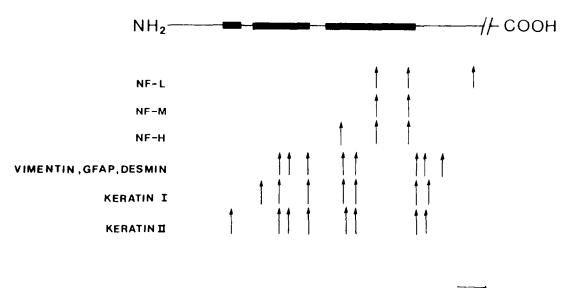
Fig. 3. Nucleotide sequence of the mouse *nfh* gene. The deduced amino acid sequence is shown under the nucleotide sequence. The arrowheads indicate the intron positions (1-3). The rod domain (see blackened boxes in Fig. 4) is delineated by parentheses (nt 395–1327). The TATAA box, the putative polyadenylation sites, and the Lys-Ser-Pro repeats of NF-H are underlined.

units and the extent of in vivo phosphorylation in NF-H and NF-M. The position of Lys-Ser-Pro repeats in the middle of the C-terminal domain corresponds to the location of multiple phosphorylation sites as defined previously by peptide mapping studies (Julien and Mushynski, 1983). Based on the phosphate content of NF-H (Julien et al., 1982; Georges et al., 1986) it can be estimated that about 25% of these sites are in a phosphorylated state in the axon in vivo.

## (c) Structure and evolutionary origin of the nfh gene

The three introns in the *nfh* gene occur in the last  $\alpha$ -helical segment of the protein and they do not delineate obvious functional subdomains (Fig. 4). Two *nfh* introns are found at the same locations in *nfl* and *nfm* genes indicating that the three genes were derived from a common ancestral gene. All introns in the *nfl* and *nfm* genes occur at locations non-homologous to intron locations in other members of the intermediate filament gene family. This led to the proposal that the primordial neurofilament gene evolved via an RNA-mediated transposition event (Lewis and Cowan, 1986; Levy et al., 1987).

The nfh gene contains an additional intron (intron 1) that occurs at an equivalent position in other members of the IF gene family (Fig. 4). Intron 1 of nfh is located only 5 nt upstream from the corresponding intron in the vimentin gene. It may have been integrated following the putative RNA-mediated transposition event. Alternatively, intron 1 of nfh may have been acquired via duplication of an ancestral IF gene with subsequent sliding of the exon-intron junction. The latter phenomenon has also occurred in keratin genes (Steinert et al., 1985). Thus, it is possible that an early gene duplication event separated neurofilament genes from the rest of the IF gene family. Accordingly, members of each IF branch would have gained and/or lost introns at different positions before divergence to give rise to all the different IF genes. Divergence of the ancestral neurofilament gene may have taken place in the earliest metazoa more than 700 million years ago as IF proteins are present in neuronal and nonneuronal cells of several invertebrates (Lasek et al., 1985; Bartnik et al., 1985; 1986). Following the latter evolutionary scheme, the neurofilament branch would have evolved with a first duplication of the ancestral nfh gene to give rise to the precursor of nfm



50a.a.

Fig. 4. Intron positions of intermediate filament genes in the structural regions of the proteins. The conserved  $\alpha$ -helical regions are represented by blackened boxes. Arrows indicate the intron locations in each gene within their corresponding protein sequences. The *nfh* gene shares two intron positions with the *nfl* and *nfm* genes, but in addition its first intron occurs at a position equivalent in other IF genes, i.e., only 1 aa upstream from the corresponding vimentin intron. A more important intron shift of 4 aa occurred also at this position in the human type II keratin (Steinert et al., 1985).

and *nfl* genes. However, intron 1 of this precursor was lost prior to a subsequent gene duplication to yield the *nfm* and *nfl* genes. While the  $\alpha$ -helical regions have been stringently conserved in the three neurofilament proteins, the presence of tandem repeat sequences in the C-terminal domains suggests that this region evolved with several recombination and amplification events. Hence the longer existence of NF-H would account for the more extensive tail region and the larger number of repeat sequences.

# (d) Conclusions

The deduced amino acid sequence of the NF-H protein is remarkable. As expected, the protein shares with other IF proteins a homologous  $\alpha$ -helical domain but the long C-terminal region in NF-H is very rich in charged amino acids and contains a repeated sequence, Lys-Ser-Pro, that represents the major protein kinase recognition site in neuro-filament proteins. This concentration of charged amino acids in a domain of neurofilament subunits that forms a lateral projection may induce charge repulsions that increase the spacing between neurofilaments and thus determine axonal volume (Hoffman et al., 1987).

The actual structure of the nfh gene supports the common origin of neurofilament genes. However, in contrast to the nfl and nfm genes, the nfh gene contains an additional intron at a location homologous to one of the introns in non-neuronal IF genes. This additional nfh intron may reflect an early divergence of the neurofilament gene family via a duplication event.

## ACKNOWLEDGEMENTS

We thank Carolle St-Aubin for secretarial aid and Roger Duclos for photography. This work was supported by the Medical Research Council of Canada (grants MA-9865 to J.-P.J. and MT-5159 to W.E.M.). L.B. is the recipient of an MRC studentship. J.-P.J. is the recipient of a scholarship from Le Fonds de la Recherche en Santé du Québec.

## REFERENCES

Bartnik, E., Osborn, M. and Weber, K.: Intermediate filaments in nonneuronal cells of invertebrates: isolation and biochemical characterization of intermediate filaments from the eso-

- Bartnik, E., Osborn, M. and Weber, K. Intermediate filaments in muscle and epithelial cells of nematodes. J. Cell Biol. 102 (1986) 2033-2041.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P.: Ovalbumin genc: evidence for a leader scquence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. USA 75 (1978) 4853-4857.
- Carden, M.J., Schlaepfer, W.W. and Lee, V.M.-Y.: The structure, biochemical properties, and immunogenicity of neurofilament peripheral regions are determined by phosphorylation state, J. Biol. Chem. 260 (1985) 9805–9817.
- 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., Fisher, S., Vandekerckhove, J., Van Damme, J., Plessmann, U. and Weber, K.: Protein-chemical characterization of NF-H, the largest mammalian neurofilament component; intermediate filament-type sequences followed by a unique carboxy-terminal extension, EMBO J. 4 (1985) 57-63.
- Geisler, N., Kaufmann, E. and Weber, K.: Antiparallel orientation of the two double-stranded coiled-coils in the tetrameric protofilament unit of intermediate filaments. J. Mol. Biol. 182 (1985) 173-177.
- Geisler, N., Vandekerckhove, J. and Weber, K.: Location and sequence characterization of the major phosphorylation sites of the high molecular mass neurofilament proteins M and H. FEBS Lett. 221 (1987) 403-407.
- Georges, E., Lefevre, S. and Mushynski, W.E.: Dephosphorylation of neurofilaments by exogenous phosphatases has no effect on reassembly of subunits. J. Neurochem. 47 (1986) 477-483.
- Georges, E. and Mushynski, W.E.: Chemical modification of charged amino acid moieties alters the electrophoretic mobilities of neurofilament subunits on SDS/polyacrylamide gels. Eur. J. Biochem. 165 (1987) 281–287.
- 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 (243-252).
- 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.
- 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.
- Hoffman, P.N., Cleveland, D.W., Griffin, J.W., Landes, P.W., Cowan, N.J. and Price, D.L.: Neurofilament gene expression: a major determinant of axon caliber. Proc. Natl. Acad. Sci. USA 84 (1987) 3472-3476.

- Julien, J.-P. and Mushynski, W.E.: Multiple phosphorylation sites in mammalian neurofilament polypeptides. J. Biol. Chem. 257 (1982) 10467-10470.
- 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.
- Julien, J.-P., Ramachandran, K. and Grosveld, F.: Cloning of cDNA encoding the smallest neurofilament protein from the rat. Biochim. Biophys. Acta 825 (1985) 398-404.
- Julien, J.-P., Meyer, D., Hurst, J. and Grosveld, F.: Cloning and developmental expression of the murine neurofilament gene family. Mol. Brain Res. 1 (1986) 243-250.
- Julien, J.-P., Grosveld, F., Yazdanbaksh, K., Flavell, D., Meijer, D. and Mushynski, W.E.: The structure of a human neurofilament gene (NF-L): a unique exon-intron organization in the intermediate filament gene family. Biochim. Biophys. Acta 909 (1987) 10-20.
- Kauffman, E., Geisler, N. and Weber, K.: SDS-PAGE strongly overestimates the molecular masses of neurofilament proteins. FEBS Lett. 170 (1984) 81-84.
- Lindsay, S. and Bird, A.: Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 327 (1987) 336-338.
- Lasek, R.J., Phillips, L., Katz, M.J. and Autilio-Gambetti, L.: Function and evolution of neurofilament proteins. Ann. NY Acad. Sci. 455 (1985) 462–478.
- Levy, E., Liem, R.K.H., D'Eustachio, P. and Cowan, N.: Structure and evolutionary origin of the gene encoding NF-M, the middle-molecular-mass neurofilament protein. Eur. J. Biochem. 166 (1987) 71-77.
- Lewis, S.A. and Cowan, N.J.: Anomalous placement of introns in a member of the intermediate filament multigene family: an evolutionary conundrum. Mol. Cell. Biol. 6 (1986) 1529–1534.
- Liem, R.K.H., Yen, S.-H., Salomon, G.D. and Shelanski, M.L.: Intermediate filaments in nervous tissues. J. Cell Biol. 79 (1978) 637-645.
- Mack, K., Currie, J.R. and Soifer, D.: A cDNA coding for the tail region of the high molecular weight rabbit neurofilament protein, NF-H. J. Neurosci. Res. (1988) In press.
- Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
- Myers, M.N., Lazzarini, R.A., Lee, V.M.-Y., Schlaepfer, W.W. and Nelson, D.L.: The human mid-size neurofilament subunit: a repeated protein sequence and the relationship of its gene to the intermediate filament gene family. EMBO J. 6 (1987) 1617-1626.
- Quax, W., Egberts, W.V., Hendricks, W., Quax-Jeuken, Y. and Bloemendal, H.: The structure of the vimentin gene. Cell 35 (1983) 215-223.
- Quax, W., Van den Broek, L., Egberts, W.V., Ramaekers, F. and Bloemendal, H.: Characterization of the hamster desmin gene: expression and formation of desmin filaments in nonmuscle cells after gene transfer. Cell 43 (1985) 327-338.
- Robinson, P.A., Wion, D. and Anderton, A.: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H). FEBS Lett. 209 (1986) 203-205.

- Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H. and Roe, B.A.: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143 (1980) 161–178.
- Steinert, P.M., Rice, R.J., 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 302 (1983) 794-800.
- Steinert, P.M., Steven, A.C. and Roop, D.R.: The molecular biology of intermediate filaments. Cell 42 (1985) 411-419.
- Zopf, D., Hermans-Borgmeyer, I., Gundelfinger, E.D. and Betz, H.: Identification of gene products expressed in the developing chick visual system: characterization of a middlemolecular-weight neurofilament cDNA. Genes Develop. 1 (1987) 699-708.

Communicated by D.T. Denhardt.