Eur. J. Biochem. *166*, 71 – 77 (1987)

Structure and evolutionary origin of the gene encoding mouse NF-M, the middle-molecular-mass neurofilament protein

Efrat LEVY[1], Ronald K. H. LIEM[2], Peter D'EUSTACHIO[1] and Nicholas J. COWAN[1]

Departments of [1] Biochemistry and [2] Pharmacology, New York University Medical Center

(Received November 26, 1986/February 13, 1987) – EJB 86 1260

We describe the complete sequence of the gene encoding mouse NF-M, the middle-molecular-mass neurofilament protein. The coding sequence is interrupted by two intervening sequences which align perfectly with the first two intervening sequences in the gene encoding NF-L (the low-molecular-mass neurofilament protein); there is no intron in the gene encoding NF-M corresponding to the third intron in NF-L. Therefore, both the number of introns and their arrangement in the genes coding NF-L and NF-M contrast sharply with the number and arrangement of introns in the genes of known sequence, encoding other members of the intermediate filament multigene family (desmin, vimentin, glial fibrillary acidic protein and the acidic and basic keratins); with the exception of a single truncated keratin gene that lacks an encoded tailpiece, these genes all contain eight introns, of which at least six are placed at homologous locations. Assuming the existence of a primordial intermediate filament gene containing most (if not all) the introns found in contemporary non-neurofilament intermediate filament genes, it seems likely that an RNA-mediated transposition event was involved in the generation of an ancestral gene encoding the NF polypeptides. A combination of insertional transposition and gene-duplication events could then explain the anomalous number and placement of introns within these genes. Consistent with this notion, we show that the genes encoding NF-M and NF-L are linked.

The eucaryotic cytoskeleton is assembled from three distinct classes of structural protein: actins, which are the subunits of microfilaments; tubulins, which polymerize to form microtubules; and the intermediate filament (IF) proteins, which assemble to form characteristic 10-nm filaments. The actins, tubulins and IF subunits are each encoded by a multigene family recognizable by homologies at the protein and nucleic acid sequence level [3, 5, 6]. In the case of the multigene family encoding IFs, the relationship between individual members is relatively distant, reflecting their early evolutionary divergence from a presumptive common ancestor and a subsequent functional specialization in terms of structure and characteristic patterns of tissue-specific expression [6–9, 16, 17]. Nonetheless, all IF proteins share a conserved central α-helical region that is responsible for their assembly into characteristic coiled coils [7–9, 40].

The neurofilament polypeptides are members of the IF multigene family that are expressed exclusively in cells of neural origin. In mammals there are three subunits, with apparent molecular masses of 68 kDa, 150 kDa and 200 kDa [9, 12, 23], hereafter referred to as NF-L, NF-M and NF-H respectively. Each is present predominantly in axons [11, 35] where they are thought to be determinants of axonal calibre [13, 14] as well as structurally involved in the maintenance of neural cell shape [14, 28]. Inspection of sequence data, obtained either directly or via cloned cDNAs, shows that, with the exception of the more divergent keratins, the members of the IF multigene family have a similar degree of homology within the coiled regions that form the backbone of the filaments [8, 19, 20, 24, 31, 32]. It therefore came as a complete surprise to discover that, while the currently characterized genes encoding non-neurofilament IF proteins (i.e. desmin, vimentin, glial filament proteins and keratins) share identical or near identical structures with respect to the placement of their introns, the gene encoding NF-L is anomalous, containing only three introns at entirely non-homologous locations [22]. The fact that the three mammalian neurofilament polypeptides are restricted in their expression to neuronal cells, and that each possesses a highly charged carboxy-terminal domain, suggests that they share a more recent common ancestor within the IF multigene family; thus, it was to be expected that the gene encoding NF-M would have a similar structure to that encoding NF-L. Here we report that, while the two genes do indeed have a similar structure, they differ in that the third intron is absent in the gene encoding NF-M. We also show that the genes encoding the NF-L and NF-M polypeptides are linked. The implications of these observations are discussed in terms of the evolution of the IF multigene family.

Correspondence to N. J. Cowan, Department of Biochemistry, New York University Medical Center, 550 First Avenue, New York City, New York, USA-10016

EXPERIMENTAL PROCEDURE

Isolation and sequencing of the gene encoding the mouse NF-M polypeptide

A 2.7-kb cDNA clone, containing sequences encoding the entire rat NF-M polypeptide [29], was used to screen [25] a mouse genomic library cloned in bacteriophage λ EMBL4. A single positively hybridizing bacteriophage plaque was purified and amplified. The extent of sequences encoding the mouse NF-M gene was determined by restriction mapping and Southern blotting [38] using the rat NF-M-encoding cDNA labelled with ^{32}P by nick-translation [33] as probe. Two adjacent *Eco*RI fragments, of 3.1 kb and 3.8 kb (Fig. 2)

that hybridized with this probe were subcloned into plasmid pUC8. Exonuclease *Bal*31 was used as described [21] to generate a set of overlapping, deleted fragments from each subclone, which were cloned into bacteriophage M13 and sequenced by the dideoxy-chain-termination method [34].

Chromosomal assignment of the genes encoding the mouse NF-L and NF-M polypeptides

Genomic DNA, extracted from interspecies somatic hybrid cell lines and from mouse livers, was analyzed by Southern blotting [38] as described [4].

RESULTS

A recently isolated full-length cDNA clone encoding the rat NF-M polypeptide [29] was used to determine the copy number of mouse genomic sequences encoding NF-M (Fig. 1). The data show the presence of two hybridizing bands when DNA from two strains of mouse was digested with *Eco*RI, and one hybridizing band in each *Tag*I digest. The same cDNA probe was used to screen a mouse genomic library cloned in bacteriophage λ. A single positively hybridizing plaque was picked, purified and amplified for further study. The region within the recombinant fragment that contained the NF-M gene was determined by restriction mapping and Southern blot analysis using the cloned NF-M rat cDNA as probe. The data thus obtained are shown in Fig. 2. The entire hybridizing region is contained within two *Eco*RI restriction fragments of 3.1 kb and 3.8 kb consistent with a single-copy gene (see Fig. 1). These two fragments were subcloned into pUC, for sequence analysis. The region sequenced is indicated in Fig. 2; the sequence data are shown in Fig. 3 together with the encoded amino acid sequence. The mouse NF-M gene encodes a polypeptide of 849 amino acids that is 96.5% homologous to the sequence encoded by the corresponding rat cDNA. The gene contains two intervening sequences that interrupt the coding region at amino acids 359 and 401. Alignment of the genes encoding NF-L and NF-M for maximum amino acid homology shows that the two introns present in the NF-M gene correspond precisely in location to the first two introns in the NF-L gene. Both these introns occur in regions where there is the most extensive homology between the NF-L and the NF-M polypeptides, and indeed among all the intermediate filament subunits, including the lamins of the nuclear envelope [27]. However, there is no interruption of the coding sequence at the location corresponding to the third intron in the NF-L gene, or at any other position further downstream. The carboxy-terminal tail region is remarkable for its high content of purine residues (>72%) which frequently occur as long uninterrupted tracts and encode an amino acid sequence that is rich (51%) in charged residues. The evolutionary significance of the purine tracts is discussed below.

Because of the overall structural similarity between the genes encoding the NF-L and NF-M polypeptides, it seemed likely that they arose from a common ancestor by a mechanism involving gene duplication. In that event the two genes might be expected to be linked. To determine the chromosomal locations of the DNA sequences in the mouse homologous to the NF-L and NF-M cDNA probes, a panel of eight somatic hybrid cell lines, carrying various combinations of mouse chromosomes, was examined by Southern blotting. The same four hybrids had retained mouse-specific DNA

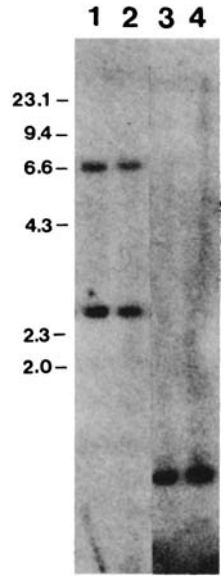

Fig. 1. *A single-copy gene encodes mouse NF-M.* Genomic DNA from two strains of mice (C57 black, tracks 1 and 3; Balb/c, tracks 2 and 4) (10 µgm) was digested with either *Eco*RI (tracks 1 and 2) or *Tag*I (tracks 3 and 4). Fragments were resolved on a 0.8% agarose gel, transferred to nitrocellulose [38] and hybridized with the cDNA probe encoding MF-M (see text) labelled with ^{32}P by nick translation [33]. The hybridization and wash conditions were as described [4]. The positions of molecular mass markers are shown at left

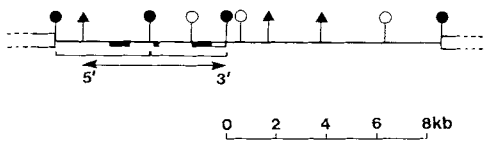

Fig. 2. *Restriction map of gNF-M.* The restriction map of the insert contained in the recombinant bacteriophage EMBL4 clone (gNF-M), isolated using the NF-M-encoding cDNA [29] as probe, is shown. (●) *Eco*RI; (▲) *Bam*HI; (○) *Hin*dIII. Arrow delinates the region sequenced. The two *Eco*RI fragments subcloned from the λ vector are indicated by brackets. The positions of NF-M coding sequences and 3′ non-coding sequences are indicated by filled and open boxes respectively. The direction of transcription 5′→3′) is shown

fragments reactive with both probes. The only mouse chromosome present in these hybrid cell lines and absent from the others was number 14 (Table 1), indicating that the genes encoding NF-L and NF-M are both located on this chromosome in the mouse. A similar experiment was performed to determine the chromosomal assignment of the gene encoding mouse glial filament protein [2], which resulted in the localization of this IF gene to mouse chromosome 11 (data not shown).

Though the NF-H polypeptide appears to be under separate developmental control during brain development [36], the expression of the NF-L and NF-M polypeptides is essentially coordinate. However, comparison of upstream sequences in the two genes reveals only short stretches of regional homology. Both share a 'TATA' motif that is common to a large number of eucaryotic promotors. Two additional short upstream homologies are also evident: one (CTGCCCCCT) that appears 50 bp and 90 bp, respectively, upstream from the putative TATA box in the NF-L and NF-M genes, and a second near-perfect obverse homology

Table 1. *Chromosomal mapping of NF-L and NF-M genes in somatic cell hybrids*
Chromosomal contents of somatic cell hybrids were determined by karyotypic and isoenzyme analysis. Mouse chromosomes observed at sequences greater than 0.15 copy/cell are noted as present. Presence or absence of mouse DNA sequences reactive with the NF-L and NF-M cDNA probes was determined by Southern blotting, as described [4]

Hybrid	Mouse chromosomes present	NF68	NF150
BEM1-4	1,2,3, 5,6, 8, 12,13,14,15,16,17,18,19,X	+	+
MACH4B31Az3	2, 7,8, 12, 15,16,17, 19,	−	−
MACH2A2B1	1,2,3,4, 6,7,8,9,10, 12, 14,15,16,17, X	+	+
MACH2A2C3	1,2,3, 7,8,9,10, 12,13, 15,16,17, 19,X	−	−
MACH2A2H3	1, 3, 6,7,8,9,10, 12,13,14,15,16,17,18,19,X	+	+
MAE28	12, X	−	−
R44-1	17,	−	−
ECm4e	14,15,	+	+

[GGGGTGGGGTGAAGG (NF-L); CCTTCCCCCCACCCC (NF-M)] between regions that lie 346 bp and 78 bp upstream from the respective TATA boxes (Fig. 3).

Amino acid comparisons of mouse NF-M (this paper), rat NF-M [29] and mouse NF-L [22] are shown in Fig. 4. The calculated molecular mass of mouse NF-M is 96 kDa. The marked difference between the authentic molecular mass and that estimated by SDS-PAGE is probably attributable to an extended helix conformation in the carboxy-terminal tailpiece, conferred by long stretches of glutamic acid residues [9, 26, 29]. Comparison of NF-M in rat and mouse reveals extensive conservation of these sequences; the coiled-coil regions (amino acids 99–244, 267–411) have changed by 1%, and the tailpiece (amino acids 412–849) by 5%, since the divergence of these two species. On the other hand, mouse NF-M and mouse NF-L are dramatically divergent: in the coiled-coil regions they are 54% homologous, similar to the homology found among all non-keratin intermediate filament coiled-coil domains [6–9, 22]. In the tail regions, however, there is no significant amino acid homology between the two sequences in spite of the great similarity in amino acid composition.

DISCUSSION

Here we report the complete structure of the gene encoding the mouse NF-M polypeptide. The data reveal the presence of only two introns, at locations homologous to the first two introns present in the gene encoding the mouse NF-L polypeptide [22]; there is no intron in the NF-M gene corresponding to the third intron present in the NF-L gene. Because the three mammalian NF proteins share similar (though not identical) patterns of tissue-specific expression and are distinct from other IF proteins by virtue of their highly acidic C-terminal tail regions, the overall structural similarity between the genes encoding the NF-L and NF-M polypeptides is not entirely surprising. Nonetheless, in accounting for the structure of contemporary IF genes we must explain both the radical difference between the structures of NF and non-NF genes, and the presence of only two introns in the gene encoding NF-M (Fig. 5). The fact that the most distantly related members of the IF multigene family (based on a comparison of homologies within the α-helical regions) share a very similar or identical gene structure has led us to argue (a) that the ancestral IF gene, from which contemporary sequences were derived, contained many (if not all) of its introns at their current locations; and (b) that the genes encoding NF polypeptides evolved from this ancestral sequence via an mRNA-mediated transposition event [22]. Such an event would result in the insertion of an intronless IF-encoding sequence at a new genetic locus that, under appropriate circumstances (i.e. the nearby location of promotor elements and appropriate regulatory signals) could result in the expression and evolution of a novel IF sequence. Though this sequence of events may appear unlikely, a precedent for the generation of an expressed sequence following an RNA-mediated transposition event does exist in the case of rat preproinsulin gene I [37]. This gene lacks two of the introns present in the otherwise homologous preproinsulin gene II, is flanked by a direct repeat (a hallmark for chromosomal integration events) and possesses a vestigial downstream poly(A) tract that could well derive from a polyadenylated mRNA intermediate. In the case of NF-encoding sequences such characteristic hallmarks of mRNA-mediated transposition events would be lost from the contemporary genes, as a result of neutral drift during the time (about 600 million years) that has elapsed since divergence of those phyla (vertebrates, annelids and molluscs) that contain NF polypeptides [15].

Given the existence of a single-copy intronless sequence encoding a primordial NF polypeptide, three kinds of events would be required to generate the three contemporary NF polypeptide genes: the acquisition of the encoded carboxy-terminal tailpiece, the insertion of introns and at least two rounds of gene duplication. The observation that the genes encoding NF-L and NF-M contain two introns at homologous locations suggests that these introns were acquired (presumably by insertional transposition) prior to the gene duplication event. The fact that the genes encoding NF-L and NF-M are in the same linkage group (Table 1) is consistent with their origin via gene duplication. In contrast, the mouse gene encoding glial filament protein is unlinked to the two NF genes (see above), as is a cluster of at least three keratin genes that has been assigned to mouse chromosome 15 (D. Meruelo, P. D'Eustachio and M. Nesbitt, unpublished data). In addition, cosmid clones have been isolated containing up to three type I or type II keratin genes, demonstrating their very close linkage [30]. It will be of interest to see if the NF genes are also clustered in this way. Thus far it appears that distantly related members of the IF multigene family are dispersed, while gene subfamilies may be linked.

Sequences encoding the highly charged tailpieces of the NF-L and NF-M proteins contain long tracks of purines, and overall are composed of more than 72% purine residues. We previously showed that these sequences are repetitive in the mouse genome, in that a probe derived from one of them

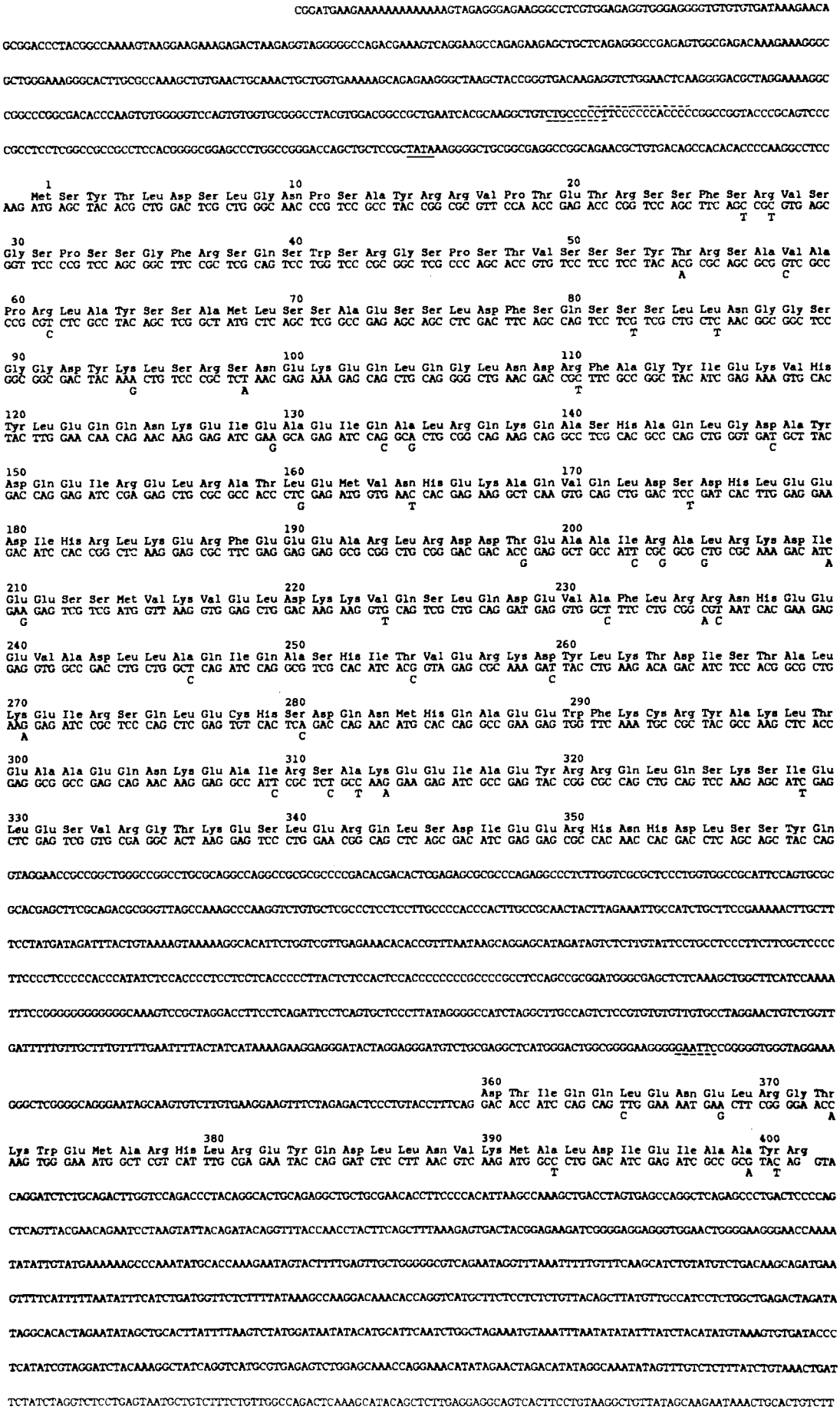

Fig. 3

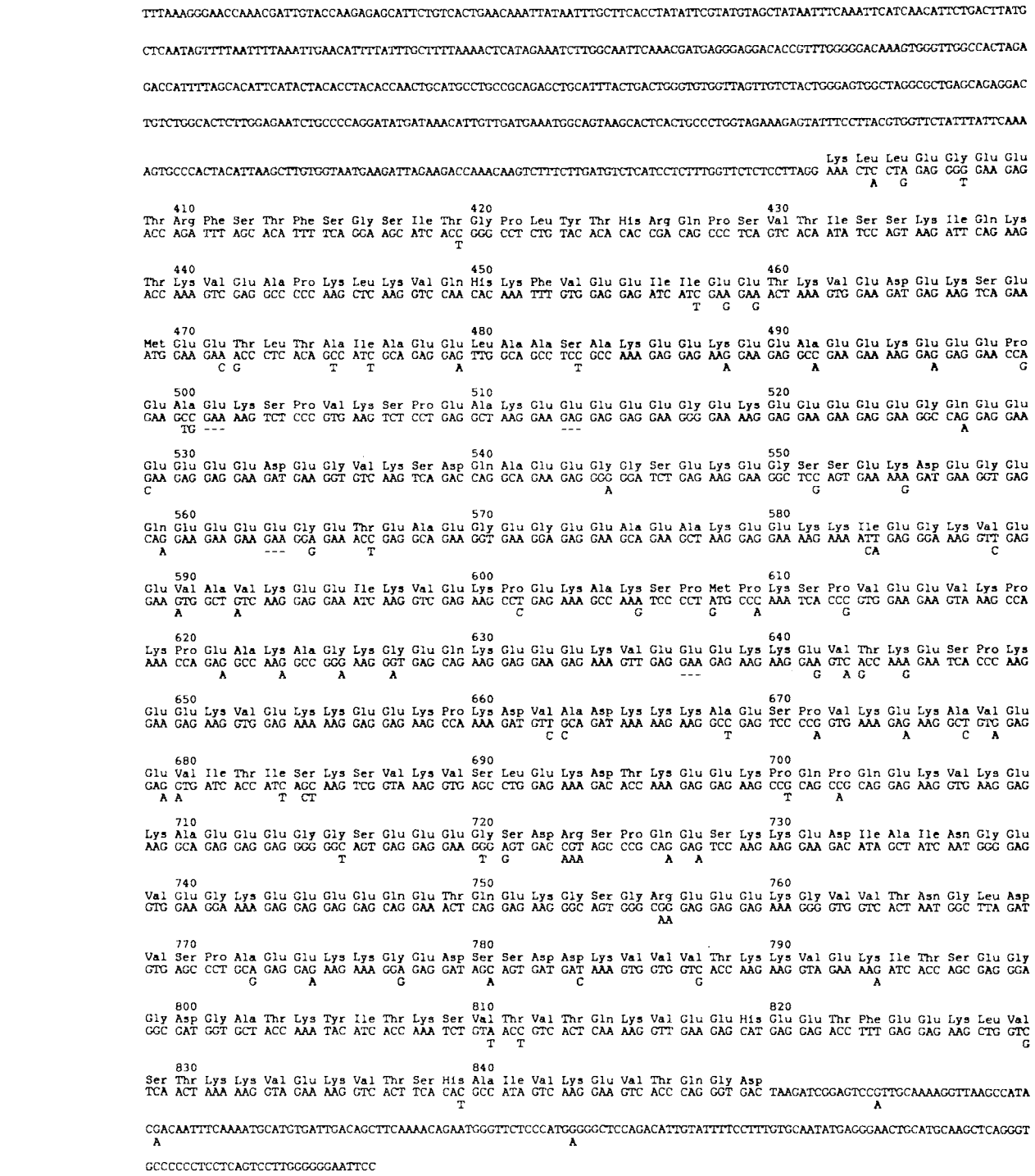

Fig. 3. *Complete sequence of the gene encoding mouse NF-M.* The sequence of the mouse NF-M gene is shown. An upstream 'TATA' motif is underlined. Differences between the mouse genomic sequence and the sequence of the rat NF-M cDNA, used to screen the library, are noted; blank spaces denote sequence identity. Sequences homologous to upstream regions in the gene encoding mouse NF-L [22] are indicated by dashed lines, as is an *Eco*RI restriction site used for subcloning (see Fig. 2)

cross-hybridizes at relatively high stringency (68 °C, 2 × standard saline/citrate) with a continuum of restriction fragments present in mouse genomic DNA [17], perhaps in part because this purine-rich probe bears significant homology to mouse satellite DNA [17]. Thus, if the hypothesized reverse transcript led to the primordial neurofilament gene integrated at a chromosomal locus adjacent to one of these abundant purine-rich repetitive sequences, a portion of these sequences might then have been recruited to form the carboxy terminus of the new protein.

In any case it is clear from the unique intron/exon structure shared by the genes encoding NF-M and NF-L (Fig. 5) that these genes must derive from a common ancestral gene and that they form a distinct branch of the IF-gene family

```
NF-M mouse    MSYTLDSLGNPSAYRRVPTETRSSFSRVSGSPSSGFRSQSWSRGSPSTVS
NF-M rat
NF-L mouse     SFAS PIFST YK  YVETP VHI S RSGY TARSAYSSYSA* V**
               S
NF-M mouse    SSYTRSAVAPRLAYSSAMLSSAESSLDFSQSSSLLNGGSGGDYKLSRSNE
NF-M rat        K   L
NF-L mouse      LS*****V RS   SS***GSLKPSLENLDVSQVAAISN L SI IQ

NF-M mouse    KEQLQGLNDRFAGYIEKVHYLEQQNKEIEAEIQALRQKQASHAQLGDAYD
NF-M rat                                      H
NF-L mouse     A   D      SF  R  E      VL  GLLV    HSGPSRFRAL E

NF-M mouse    QEIRELRATLEMVNHEKAQVQLDSDHLEEDIHRLKERFEEEARLRDDTEA
NF-M rat
NF-L mouse        D  LAA DATN  QALEGEREG   TLRN QA Y   VLS E A G

NF-M mouse    AIRALRKDIEESSMVKVELDKKVQSLQDEVAFLRRNHEEEVADLLAQIQA
NF-M rat        V                            S
NF-L mouse    RLMEA  GAD AALARA  E RID  M  I   KKV    I E Q    I

NF-M mouse    SHITVERKDYLKTDISTALKEIRSQLECHSDQNMHQAEEWFKCRYAKLTE
NF-M rat
NF-L mouse    AQ S  MDVSS P L A   D  A Y KLAAK  QN      S FTV

NF-M mouse    AAEQNKEAIRSAKEEIAEYRRQLQSKSIELESVRGTKESLERQLSDIEER
NF-M rat
NF-L mouse    S AK TD V A  D VS S  L KA TL I AC  MN A  K  QEL DK

NF-M mouse    HNHDLSSYQDTIQQLENELRGTKWEMARHLREYQDLLNVKMALDIEIAAY
NF-M rat
NF-L mouse    Q A I AN    NK      S  S    Y K

NF-M mouse    RKLLEGEETRFSTFSGSITGPLYTHRQPSVTISSKIQKTKVEAPKLKVQH
NF-M rat
NF-L mouse                 L FT ]  SGYSQSS VFGRSAYSGLQSSSYLMSARSFP

NF-M mouse    KFVEEIIEETKVEDEKSEMEETLTAIAEELAASAKEEKEEAEEKEEEPEA
NF-M rat                          DA  V                       V
NF-L mouse    AYYTSHVQ EQT V ETIEATKAEEAKD PPSEGEA E  K KE G E E

NF-M mouse    EKSPVKSPEAKEEEEEGEKEEEEEGQEEEEEEDEGVKSDQAEEGGSEKEG
NF-M rat      *          *
NF-L mouse    GAEEEEAAKDES DTKE E GG GEE DTK SE EE KEESAGEEQVAKK

NF-M mouse    SSEKDEGEQEEEEGETEAEGEGEEAEAKEEKKIEGKVEEVAVKEEIKVEK
NF-M rat                *                        T      M I
NF-L mouse    KDJ

NF-M mouse    PEKAKSPMPKSPVEEVKPKPEAKAGKGEQKEEEKVEEEKKEVTKESPKEE
NF-M rat            V                    D        *      A

NF-M mouse    KVEKKEEKPKDVADKKKAESPVKEKAVEEVITISKSVKVSLEKDTKEEKP
NF-M rat                    P                  M   T

NF-M mouse    QPQEKVKEKAEEEGGSEEEGSDRSPQESKKEDIAINGEVEGKEEEEQETQ
NF-M rat       Q                   VG K

NF-M mouse    EKGSGREEEKGVVTNGLDVSPAEEKKGEDSSDDKVVVTKKVEKITSEGGD
NF-M rat           Q                        R

NF-M mouse    GATKYITKSVTVTQKVEEHEETFEEKLVSTKKVEKVTSHAIVKEVTQGD
NF-M rat
```

Fig. 4. *Comparison of NF-M and NF-L amino acid sequences.* The amino acid sequences of mouse NF-M (this paper), rat NF-M [29] and mouse NF-L [22] are compared. Blank spaces denote homology; asterisks have been introduced to denote 'deletions' so as to maximize overall homology. The location of the carboxy terminus of NF-L is indicated by a square bracket

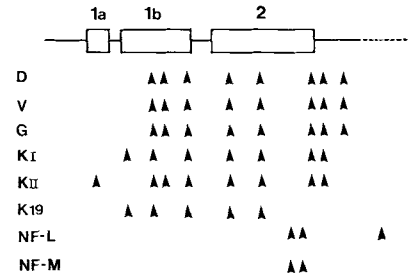

Fig. 5. *Comparison of intron locations in intermediate filament genes of known structure.* The α-helical domains are boxed (coils 1a, 1b, and 2); the extended carboxy-terminal tail region characteristic of the NF polypeptides is shown (– – –). D = hamster desmin [31], V = hamster vimentin [32], G = mouse glial fibrillary acidic protein [2], KI = human type I cytokeratin [24], KII = human type II cytokeratin [39], K19 = a bovine type I cytokeratin [1], mouse NF-L [22] and mouse NF-M (this paper). The position of the final intron varies somewhat among the keratin gene family members (see for example [1, 18]) but this can be explained by the fact that this intron interrupts the sequences encoding the tailpiece, a region of great variability. Similarly, the absence of the final two introns in K19 is not surprising, as this keratin is truncated and terminates at the end of coil 2 [1]

evolutionary tree. Following the triplication of the putative primordial NF-encoding gene, the regions encoding the carboxy-terminal tailpieces of the NF-M and NF-L proteins (Fig. 4) must have diverged to such an extent that there is no remaining homology between them over and above that which can be accounted for solely on the basis of the purine-rich nature of both sequences (Figs 3, 4). Furthermore, the gene encoding NF-L has a third intron in this region [22] not present in the gene encoding NF-M, which must either have been lost from the NF-M gene or inserted into the NF-L gene subsequent to the triplication of the primordial gene. Because of the tendency of repeat-containing sequences to drift, it is surprising to note the stability of the sequences encoding the carboxy-terminal regions of NF-M: the amino acid sequences of mouse and rat (Fig. 4) are 95% identical over this region. Thus, the tailpiece is almost as conserved as the coiled-coils, and must therefore be under a similar (though somewhat lesser) degree of selective pressure.

REFERENCES

1. Bader, B. L., Magin, T. M., Hatzfeld, M. & Franke, W. W. (1986) *EMBO J. 5*, 1865–1875.
2. Balcarek, J. M. & Cowan, N. J. (1985) *Nucleic Acids Res. 13*, 5527–5543.
3. Cleveland, D. W. & Sullivan, K. F. (1986) *Annu. Rev. Biochem. 54*, 331–365.
4. D'Eustachio, P., Kristensen, T., Wetsel, R. A., Riblet, R., Taylor, B. A. & Tack, B. F. (1987) *J. Immunol.*, in the press.
5. Engel, J., Gunning, P. & Kedes, L. (1982) in *Muscle development, molecular and cellular control* (Pearson, M. L., & Epstein, H., eds) pp. 107–118, Cold Spring Harbor Laboratory, NY.
6. Fuchs, E. & Hanukoglu, I. (1983) *Cell 34*, 332–334.
7. Geisler, N. & Weber, K. (1982) *EMBO J. 1*, 1649–1656.
8. Geisler, N., Kaufman, E., Fisher, S., Plessmann, U. & Weber, K. (1983) *EMBO J. 2*, 1295–1302.
9. Geisler, N., Fisher, S., Vandekerckhove, J., Plessmann, U. & Weber, K. (1984) *EMBO J. 3*, 2701–2706.
10. Hanukoglu, I. & Fuchs, E. (1982) *Cell 31*, 243–252.
11. Hirokawa, N., Glicksman, M. A. & Willard, M. (1984) *J. Cell Biol. 98*, 1523–1536.
12. Hoffman, P. N. & Lasek, R. J. (1975) *J. Cell Biol. 66*, 351–366.
13. Hoffman, P. N., Griffin, J. W. & Price, D. L. (1984) *J. Cell Biol. 99*, 705–714.
14. Lasek, R. J., Oblinger, M. H. & Drake, P. F. (1983) *Cold Spring Harbor Symp. Quant. Biol. 48*, 731–744.
15. Lasek, R. J., Phillips, L., Katz, M. J. & Autilio-Gambetti, L. (1985) *Ann. N.Y. Acad. Sci. 455*, 462–478.
16. Lazarides, E. (1980) *Nature (Lond.) 283*, 249–256.
17. Lazarides, E. (1980) *Annu. Rev. Biochem. 51*, 219–250.
18. Lehnert, M. E., Jorcano, J. L., Zentgraf, H., Blessing, M., Franz, J. K. & Franke, W. W. (1984) *EMBO J. 3*, 3279–3287.
19. Lewis, S. A., Balcarek, J. M., Krek, V., Shelanski, M. & Cowan, N. J. (1984) *Proc. Natl Acad. Sci. USA 81*, 2743–2746.
20. Lewis, S. A. & Cowan, N. J. (1985) *J. Cell Biol. 100*, 843–850.
21. Lewis, S. A., Gilmartin, M. E., Hall, J. L. & Cowan, N. J. (1985) *J. Mol. Biol. 182*, 11–20.
22. Lewis, S. A. & Cowan, N. J. (1986) *Mol. Cell Biol. 6*, 1529–1534.
23. Liem, R. K. H., Yen, S.-H., Salomon, G. D. & Shelanski, M. L. (1978) *J. Cell Biol. 79*, 637–645.
24. Marchuk, D., McCrohan, S. & Fuchs, E. (1984) *Cell 39*, 491–498.

25. Maniatis, T., Hardison, E., Lacy, E., Laver, J., O'Connell, D., Quon, D., Sun, G. K. & Efstratiadis, A. (1978) *Cell 15*, 686–701.
26. Maxfield, F. R., Alter, J. E., Taylor, G. T. & Scheraga, H. A. (1975) *Macromolecules 8*, 479–491.
27. McKeon, F. D., Kirshner, M. W. & Caput, P. (1986) *Nature (Lond.) 319*, 463–468.
28. Morris, J. R. & Lasek, R. J. (1982) *J. Cell Biol. 92*, 192–198.
29. Napolitano, E. W., Chin, S. M., Colman, D. R. & Liem, R. K. H. (1986) *J. Neurosci.*, in the press.
30. Powell, B. C., Cam, G. R., Fietz, M. J. & Rogers, G. A. (1986) *Proc. Natl Acad. Sci. USA 83*, 5048–5052.
31. Quax, W., Van der Broeck, L., Egberts, W. V., Ramaekers, F. & Bloemendal, H. (1985) *Cell 43*, 327–338.
32. Quax, W., Egberts, W. V., Hendriks, W., Quax-Jueken, Y. & Bloemendal, H. (1983) *Cell 35*, 215–223.
33. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol. 113*, 231–251.
34. Sanger, F., Coulsen, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. (1980) *J. Mol. Biol. 143*, 161–178.
35. Shaw, G., Osborn, M. & Weber, K. (1981) *J. Cell Biol. 26*, 68–82.
36. Shaw, G. & Weber, K. (1981) *Nature (Lond.) 298*, 277–279.
37. Soares, M. B., Scherr, E., Henderson, A., Karathanasis, S. K., Cate, R., Zeitlin, S., Chirgwin, J. & Efstratiadis, A. (1985) *Mol. Cell Biol. 5*, 2090–2103.
38. Southern, E. (1975) *J. Mol. Biol. 98*, 503–517.
39. Tyner, A. L., Eichman, M. J. & Fuchs, E. (1985) *Proc. Natl Acad. Sci. 82*, 4683–4687.
40. Weber, K., Shaw, G., Osborn, M., Debus, E. & Geisler, N. (1983) *Cold Spring Harbor Symp. Quant. Biol. 48*, 717–729.