MOLECULAR AND CELLULAR BIOLOGY, May 1986, p. 1529–1534
0270-7306/86/051529-06$02.00/0

Vol. 6, No. 5

Anomalous Placement of Introns in a Member of the Intermediate Filament Multigene Family: an Evolutionary Conundrum

SALLY A. LEWIS AND NICHOLAS J. COWAN*

Department of Biochemistry, New York University School of Medicine, New York, New York 10016

Received 13 December 1985/Accepted 7 February 1986

The origin of introns and their role (if any) in gene expression, in the evolution of the genome, and in the generation of new expressed sequences are issues that are understood poorly, if at all. Multigene families provide a favorable opportunity for examining the evolutionary history of introns because it is possible to identify changes in intron placement and content since the divergence of family members from a common ancestral sequence. Here we report the complete sequence of the gene encoding the 68-kilodalton (kDa) neurofilament protein; the gene is a member of the intermediate filament multigene family that diverged over 600 million years ago. Five other members of this family (desmin, vimentin, glial fibrillary acidic protein, and type I and type II keratins) are encoded by genes with six or more introns at homologous positions. To our surprise, the number and placement of introns in the 68-kDa neurofilament protein gene were completely anomalous, with only three introns, none of which corresponded in position to introns in any characterized intermediate filament gene. This finding was all the more unexpected because comparative amino acid sequence data suggest a closer relationship of the 68-kDa neurofilament protein to desmin, vimentin, and glial fibrillary acidic protein than between any of these three proteins and the keratins. It appears likely that an mRNA-mediated transposition event was involved in the evolution of the 68-kDa neurofilament protein gene and that subsequent events led to the acquisition of at least two of the three introns present in the contemporary sequence.

Multigene families encoding groups of structurally or functionally related polypeptides are widespread in the genomes of eucaryotes. This is true both for proteins that are expressed in most if not all cell types (e.g., histones, actins, and tubulins) and proteins that show restricted patterns of tissue-specific expression (e.g., globins and immunoglobulins). In all cases, structural homologies evident at either the protein or DNA level (or both) point to the divergence of members of a given multigene family from a single primordial sequence. The extent of divergence depends upon selective constraints. Where function has remained similar (if not identical) over evolutionary time, the extent of divergence has been small (e.g., actins and tubulins [3, 6]), with only relatively minor differences distinguishing one expressed member of the multigene family from another. Where functional demand has required diversity (e.g., immunoglobulin C regions), the extent of divergence has been greater (29).

The role of introns (if any) in the vast majority of expressed eucaryotic genes is uncertain, although their existence points to considerable reorganizational events that have occurred throughout evolutionary history. In particular, the placement of intervening sequences in individual members of multigene families must be explained either by their introduction since the time of initial divergence or by their maintenance or loss (or both) over time after duplication of an ancestral sequence that already contained one or more intervening sequences. The diversity of intron placement among various members of the actin multigene family is testimony to the complex nature of the events that seem to have transpired since divergence from a presumptive common ancestor (7, 9, 34). Although there is evidence that in some cases the DNA contained in exons encodes a clearly delineated functional domain (e.g., globin [4, 5]) and immunoglobulin [2]), this concept (15) does not seem to apply in other cases, where introns exist but do not appear to delineate functional regions (7, 9, 17, 21, 25, 34). In short, the nature of the events leading to either intron location or intron excision are only poorly understood.

The intermediate filament multigene family is a particularly favorable model for the study of this problem because of the extensive amino acid sequence diversity among its members and accumulating structural data regarding the sequence and organization of the genes (1, 10–14, 24–26, 33). Data on the genes encoding vimentin (25), desmin (26), and glial fibrillary acidic protein (GFAP) (1) show the presence of eight introns at identical positions, whereas a type I and a type II keratin gene have seven and six homologously placed introns, respectively, and one additional intron at a unique location (24, 33). Here we report the complete sequence of the gene encoding the 68-kilodalton (kDa) mouse neurofilament polypeptide. Surprisingly, only three introns are present, all of which occur at positions different from those of the introns in all other members of the intermediate filament multigene family. The evolutionary implications of this observation are discussed.

MATERIALS AND METHODS

Isolation of a 2.4-kb cDNA clone encoding the mouse 68-kDa neurofilament protein. The 1.2-kilobase (kb) cDNA probe, NF68, encoding the mouse 68-kDa neurofilament protein was used to screen a mouse brain library (constructed in bacteriophage λgt11 [22]) in a search for longer cloned cDNA copies. Approximately 40 positively hybridizing plaques were picked as a 5-mm plug and eluted into 1 ml of buffer (0.1 M NaCl, 10 mM Tris hydrochloride [pH 8.0], 10 mM $MgSO_4$), and a portion (30 μl) of the bacteriophage was amplified in a small (1 ml) culture without further purification. Bacteriophage DNA was prepared from these cultures as follows. Debris was removed from each culture by centrifugation at 15,000 ×

* Corresponding author.

```
   GAAAGCTGGGAGTGGGGTGGGGTGGGGTGGGGTGAAGGTGTCTTTGTGACATTGGATGTTAAAAGTTGATGTGTCAGTAGGGATCGCGAACCTTTTTATTTGGAACATTTAAGAGT

CCTAGTGATGTGGGCTTTAGGTATTAAGAGGCAGGATGGGACGGGGGAGTAGGAAAGGAAAAGAATAAGCAGGGAAGTTATGGGGGTCTCAGAGAAAGTAGGGGTCCTCATAGAAAAGC

GGCCAGGCTGGAGCCGCAGCTCGCTAGCCTCCTGTCGTTCGCCCCCGCCTTTGCTTTTGCGCAGAATCCTCCCCTTGGCTGCAGCAACGCGCTGCCCCCTACTGGCCTGGCCACGGCGA

GTCGATCACAGTCTGCGTCAGAGTCCCGGCGTATAAATAGAGGTGGCAGGACCCGCCGAGCCGCACACAGCCGATCCATCCTCCCCTTTCCCTCTTCTCCCCCGTTCCTCTCTCTAGGT

                                        1                                  10
                                       met ser ser phe ala ser asp pro ile phe ser thr ser tyr lys arg arg tyr val
CCCCCATCTCCCCCCGGGCCTGAGGGGCGCACCGGCCGCCACC ATG AGT TCG TTC GGC TCG GAT CCG ATC TTT TCG ACC TCC TAC AAG CGG CGC TAT GTG

 20                                 30                                  40
glu thr pro arg val his ile ser ser val arg ser gly tyr ser thr ala arg ser ala tyr ser ser tyr ser ala pro val ser ser
GAG ACG CCC CGG GTG CAC ATC TCC AGC GTG CGC AGC GGC TAC AGC ACG GCC CGC TCC GCG TAC TCC AGC TAC TCC GCG CCG GTC TCC TCC

 50                                 60                                  70
ser leu ser val arg arg ser tyr ser ser ser ser gly ser leu lys pro ser leu glu asn leu asp val ser gln val ala ala ile
TCG CTG TCC GTG CGC CGC AGC TAC TCG TCC AGC TCC GGC TCG TTG AAG CCC AGC CTG GAG AAT CTC GAT GTC AGC CAG GTA GCC GCC ATC

 80                                 90                                 100
ser asn asp leu lys ser ile arg ile gln glu lys ala gln leu gln asp leu asn asp arg phe ala ser phe ile glu arg val his
AGC AAC GAC CTC AAG TCT ATC CGC ACA CAA GAG AAG GCA CAG CTG CAG GAC CTC AAC CAT CGC TTC GCC AGC TTC ATC GAG CGC GTG CAC

110                                120                                 130
glu leu glu gln gln asn lys val leu glu ala gly leu leu val leu arg gln lys his ser gly pro ser arg phe arg ala leu tyr
GAG CTG GAG CAG CAG AAC AAG GTC CTG GAA GCC GAG CTG TTG GTG CTG CGC CAG AAA CAC TCT GAG CCT TCC CGC TTC CGC GCC CTG TAC

140                                150                                 160
glu gln glu ile arg asp leu arg leu ala ala glu asp ala thr asn glu lys gln ala leu glu gly glu arg glu gly leu glu glu
GAG CAG GAG ATC CGC GAT CTG CGG CTG GCA GCG GAA GAC GCC ACT AAC GAG AAG CAG GCG CTG CAG GGG GAG CGC GAG GGG CTG GAG GAG

170                                180                                 190
thr leu arg asn leu gln ala arg tyr glu glu glu val leu ser arg glu asp ala glu gly arg leu met glu ala arg lys gly ala
ACT CTG CGC AAC CTG CAG GCT CGC TAT GAG GAA GAA GTG CTG AGC CGC GAG GAC GCC GAG GGC CGG CTG ATG GAA GCG CGC AAA GGT GCC

200                                210                                 220
asp glu ala ala leu ala arg ala glu leu glu lys arg ile asp ser leu met asp glu ile ala phe leu lys lys val his glu glu
GAT GAG GCC GCG CTC GCC CGC GCC GAG CTG GAG AAG CGC ATC GAC AGC CTG ATG GAC GAG ATA GCT TTC CTG AAG AAG GTG CAC GAG GAA

230                                240                                 250
glu ile ala glu leu gln ala gln ile gln ile ala gln ile ser val glu met asp val ser ser lys pro asp leu ser ala ala leu
GAG ATC GCC GAG CTG CAG GCT CAG ATC CAG ATT GCT CAG ATC TCC GTG GAG ATG GAC GTG TCC TCC AAG CCC GAT CTC TCC GCC GCT CTC

260                                270                                 280
lys asp ile arg ala gln tyr glu lys leu ala ala lys asn met gln asn ala glu glu trp phe lys ser arg phe thr val leu thr
AAG GAC ATC CGC GCT CAG TAC GAG AAG CTG GCC GCC AAG AAC ATG CAG AAC GCC GAA GAG TGG TTC AAG AGC CGC TTC ACC GTG CTA ACC

290                                300                                 310
glu ser ala ala lys asn thr asp ala val arg ala ala lys asp glu val ser glu ser arg arg leu leu lys ala lys thr leu glu
GAG AGC GCC GCC AAG AAC ACC GAC GCT GTG CGC GCT GCC AAG GAC GAG GTG TCG GAA AGC CGC CGC CTG CTC AAG GCT AAG ACC CTG GAG

320                                330                                 340
ile glu ala cys arg gly met asn glu ala leu glu lys gln leu gln glu leu glu asp lys gln asn ala asp ile ser ala met gln
ATC GAA GCC TGC CGG GGT ATG AAC GAA GCT CTG GAG AAG CAG CTG CAG GAG CTA GAG GAC AAG CAG AAT GCA GAC ATT AGC GCC ATG CAG

GTATAGTACGGTTTAAGCTCAGTGGGGGTGGGGAGGTCCACGGTAGGGGCGGCGGCAGGGGGGGGGGGAGGAGTCAGGGTGTACCCAGAGAGTGAACGCTAGGGGGTTCTAGAGCACAG

AGCAGCTATCAGAGGGCTTGGCTTCCGAGAGGTTGTGGAAAGGATGGGGCTTTTGGAGGGGGCGGGGCGCGACTGCTGCCTGGGGATTGCTTGCTCTGGTTGAATCAGGTTGTTAGGGG

TGAGTTGGGAGGAAGGGGCGGAGGGAGTCGGAGTTGGGCGTTGCCAAGTAATGCCTTCTCTTTACTGACAACCCTTAACTCCCAAGTTACCCTCAGTACTTCCTGGATCTCCTGGTTAA

TAAATATGTCTGCTAGCTCTCACTAGTTTTGGGCCACCCAGCCTTGGGATAGTTTTTTTTTTTAATAGGGACACAAGAATACAGTAACAATATTTTAGCAATTAATCACTGCAGTCCAT

TAAAAAAAAAGTTCTTCATCTTTGGCGGGGGGAAGCATTTTAGACAAACAGCTACACAAATATATATAATTTAATTTTTATTTATATATTTATTTAGACAAACGAGAGAGATATATATA

TATATATATATATATATATATCCCCTTCCTAAATGCGACCCTGGCTAAAAGAATGTGACAAGGGTGGTTAATAATAATAATAATATCATCATCAATTTGGTAGTTTTAAAAGACAGC

TTACCATAATCGATTGAGAACAATTTTCAAAATTGTTCACATCAGTGTACTCCCCTCCTCTTCCTCCTCTCCTTCTCTTGCTTCACTAGACTGGGCATTTTTAGGCCAACCCCTCCCCC

TATGCCCCACCCCACCCCATTATTTTAAGGGAGATTTCTGGAAGGGTGTACAGTTTGGAGTTAACTAAGACCAGAACTTATCATGGATACTACAGTAAAACAAAAGCTGCAGTAGAGAT
```

g for 5 min, and DNAse was added to a final concentration of 1 μg/ml to digest liberated host DNA. After 30 min at room temperature, 0.2 ml of a solution containing 2.5% sodium dodecyl sulfate, 0.5 M Tris (pH 9.0), and 0.25 M EDTA was added, and the mixture was heated to 70°C for 20 min. Ammonium acetate was then added to a final concentration of 1.4 M. After 30 min on ice, precipitated material was removed by centrifugation at 15,000 × *g* for 5 min, and nucleic acid was precipitated by the addition of 0.6 volume of isopropanol. After recovery by centrifugation, the pellets were rinsed with cold 70% ethanol, air dried, dissolved in 40 μl of 10 mM Tris (pH 7.5)–1 mM EDTA, and stored at −20°C.

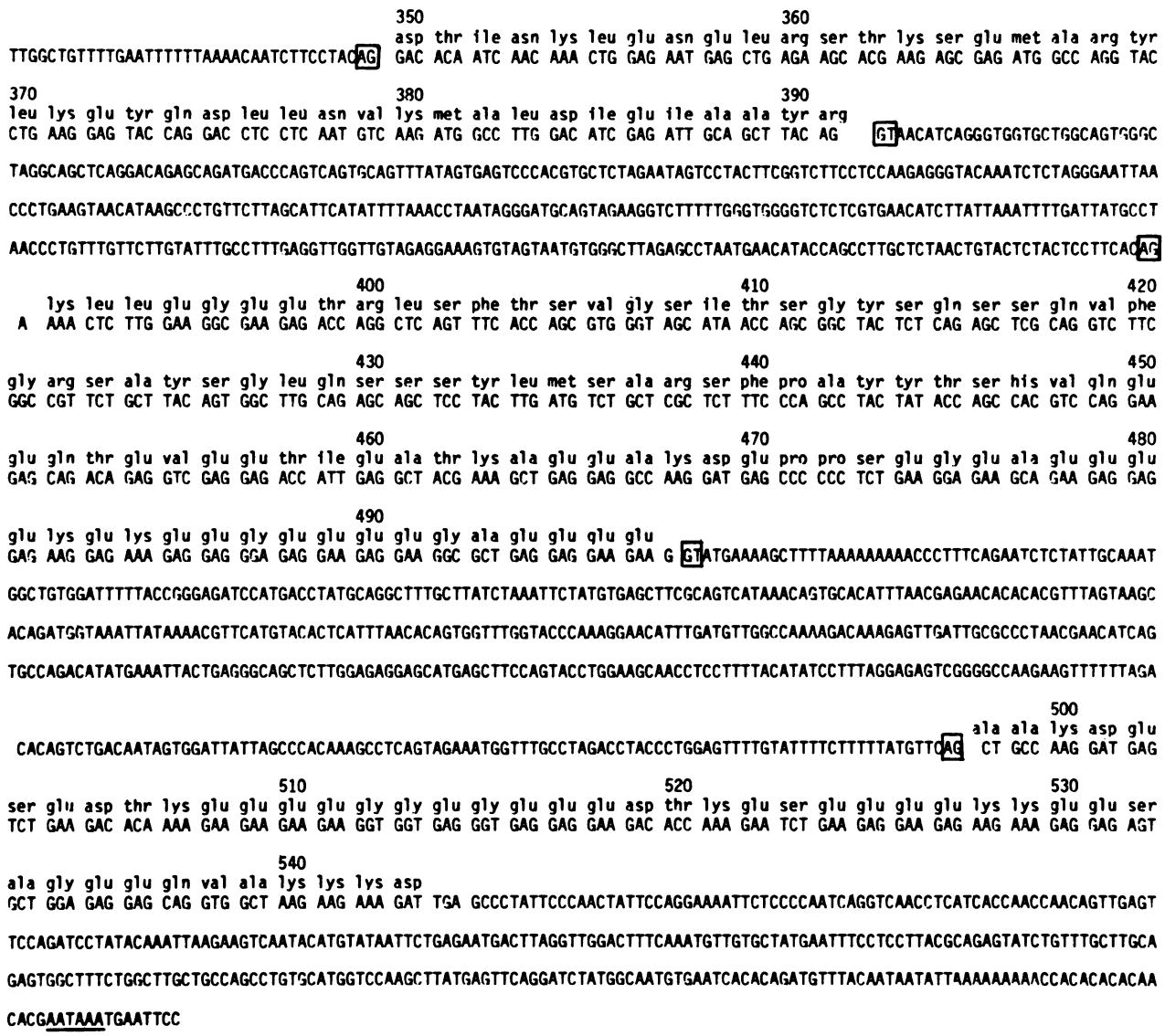

FIG. 1. Complete sequence of the gene encoding the mouse 68-kDa neurofilament protein. A possible ATA box and polyadenylation signals are underlined; splice signal dinucleotides are boxed. The 5′ end of the cDNA clone NF68L is marked by a vertical bar.

Aliquots (20 μl) of this material were digested with *Eco*RI (New England BioLabs, Inc.) under conditions recommended by the manufacturer and the digestion products were resolved on a 1% agarose gel. After transfer of the gel content to nitrocellulose (31), the blots were hybridized with approximately 10^7 cpm of the excised, gel-purified insert from NF68 (20) (^{32}P-labeled by nick translation [27]), washed to a final stringency of 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate at 68°C, and exposed to film. Impure bacteriophage preparations containing the largest recombinant inserts detected by the NF68 probe were then subjected to two successive rounds of plaque purification, and the inserts were subcloned into the bacterial plasmid pUC8.

Isolation and sequencing of the gene encoding the mouse 68-kDa neurofilament protein. A 2.2-kb cDNA clone, NF68L, encoding all but the 19 N-terminal amino acids of the 68-kDa neurofilament protein was selected (see above) and used to screen (23) a mouse genomic library cloned in bacteriophage λ (generously provided by P. D'Eustachio). The recombinant insert of a single positively hybridizing bacteriophage was subcloned into plasmid pUC8 and restriction mapped, and the location of gene sequences encoding the 68-kDa neurofilament protein were determined by hybridization with the NF68L probe which had been ^{32}P-labeled by nick translation (27). A 6.0-kb *Hin*dIII-*Eco*RI restriction fragment that contained all of the detectable hybridizing sequences was fully sequenced by the method of Sanger et al. (28). The data were analyzed with the computer programs of Staden (32).

RESULTS AND DISCUSSION

We recently reported the isolation of a cloned cDNA probe, NF68, which encodes the mouse 68-kDa neurofilament polypeptide (22). The sequence of NF68 revealed that it encodes the 303 carboxy-terminal amino acids of the protein, including the second α-helical domain, but lacks the N-terminal sequences (including those for the first α-helical domain) (22). To extend the sequence of the mouse 68-kDa neurofilament protein and to obtain a more extensive probe

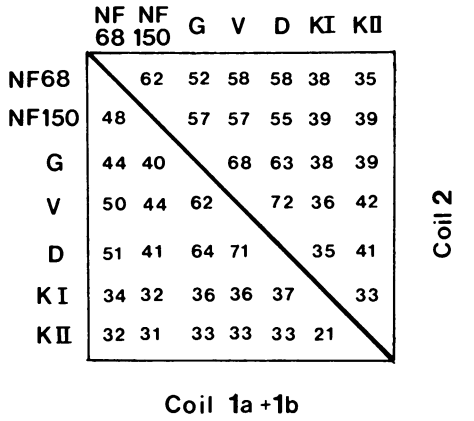

FIG. 2. Paired comparisons of amino acid homologies within the α-helical regions of various intermediate filament polypeptides. Amino acid sequences within the conserved α-helical coils of human keratin (types I and II), hamster vimentin, hamster desmin, mouse GFAP, and the mouse 68-kDa and porcine 150-kDa neurofilament proteins were aligned for maximum homology, and paired comparisons were made. The data were from references 24 (human type I keratin [KI]), 33 (human type II keratin [KII]), 25 (hamster vimentin [V], 26 (hamster desmin [D]), 1 (mouse GFAP [G]), and 12 (porcine 150-kDa neurofilament protein [NF150]) and from this study (mouse 68-kDa neurofilament protein [NF68]). The results are expressed as percent homology.

that could be used to map 5′ genomic sequences, the NF68 cDNA was used to screen a mouse brain cDNA library cloned in bacteriophage λgt11. Several positively hybridizing plaques were obtained, and their cDNA insert size was measured by Southern blot analysis after digestion with *Eco*RI (data not shown). Clones containing the largest cDNA inserts were selected for further study. Sequence analysis revealed that the longest of these clones, NF68L, encompassed almost the entire 2.6-kb mRNA encoding the 68-kDa neurofilament polypeptide, including the first α-helical region (Fig. 1).

Under conditions of high stringency, NF68 detects a single copy sequence in the mouse genome (22). To determine the structural organization of the gene encoding the 68-kDa neurofilament polypeptide and to thereby elucidate evolutionary relationships within the intermediate filament multigene family, we used the NF68L probe to screen a bacteriophage λ mouse genomic library. A single positively hybridizing plaque was picked and amplified for further study. The region within the recombinant fragment containing the NF68 68-kDa polypeptide gene was determined by restriction mapping and Southern blot analysis with the NF68L cDNA as the probe. A restriction fragment (*Hin*dIII-*Eco*RI; ~6 kb) that contained the entire region of hybridization was subcloned and fully sequenced; the data are presented in Fig. 1.

The mouse 68-kDa neurofilament gene encodes a polypeptide of 543 amino acids and has three introns at the positions shown. Paired comparisons of the homologies within the two α-helical regions common to all intermediate filament proteins are shown in Fig. 2. The α-helices are homologous and highly conserved because of their roles in the formation of coils and in the polymerization of these coils into filaments (8). These roles are similar in all intermediate filaments (though less so in the keratins, which cannot form homopolymers). The amino acid homologies may reflect evolutionary relationships among genes encoding the different subunits or may be the result of selective pressure working to create small functional differences.

The data in Fig. 2 show that desmin, vimentin, and GFAP are the most homologous, implying a relatively recent divergence from a presumptive common ancestral sequence. Consistent with this observation, the sequences of the genes encoding these three polypeptide shows that they contain all eight introns at identical locations (1, 25, 26). It seems likely, therefore, that the immediate progenitor of the vimentin, desmin, and GFAP genes also contained eight introns at the same positions. On the other hand, the relationship of the keratins to the other intermediate filaments is relatively distant, with only 32 to 42% homology within the α-helical regions (Fig. 2). Nonetheless, the keratin genes currently characterized (24, 33) contain six or seven introns at locations homologous to those of the introns in the desmin, vimentin, and GFAP genes. These data have therefore interpreted in terms of an ancient and primordial intermediate filament gene that predated the divergence of the keratin and nonkeratin sequences and contained the majority of its introns at locations identical to those of the introns in contemporary intermediate filament gene sequences (8, 33). This ancestral sequence could then have given rise to the intermediate filament multigene family by a series of gene duplication events. On the basis of this model, it was expected that the gene encoding the 68-kDa neurofilament polypeptide (which is more closely related to desmin, vimentin, and GFAP than it is to the keratins) would conform to the general pattern of intermediate filament gene structure.

The actual structure of the gene encoding the 68-kDa neurofilament protein (Fig. 1) is clearly at odds with this expectation. The gene contains only three introns, none of which occurs at a position corresponding to an intron position in any characterized intermediate filament gene (Fig. 3). Two of these introns (I and II) interrupt the most highly conserved regions of coil 2. Therefore, if we are to assume the existence of a primordial intermediate filament gene containing most if not all of its introns at positions identical to those found in the contemporary keratin, desmin, vimentin, and GFAP genes, we must explain the loss of these introns from the 68-kDa neurofilament protein gene and the introduction of three additional introns at novel locations within the gene.

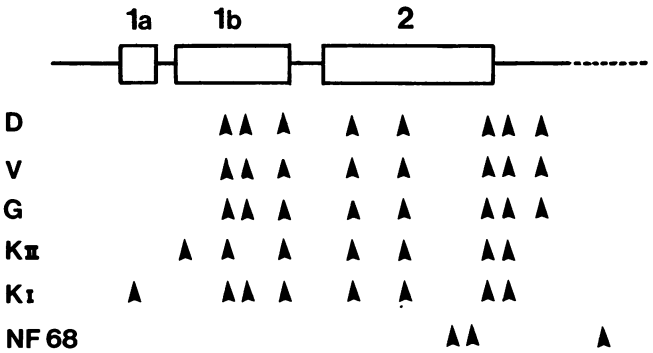

FIG. 3. Schematic diagram showing the location of introns in intermediate filament genes of known structure. The α-helical domains (coils 1a, 1b, and 2) are boxed; the extended carboxy-terminal sequence present in the 68-kDa neurofilament protein is shown (broken line). The data were from the sources listed in the legend to Fig. 2. The abbreviations are defined in the legend to Fig. 2.

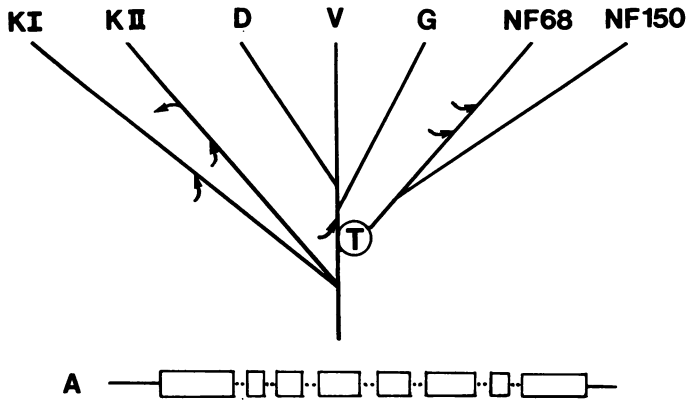

FIG. 4. Model for the evolution of the intermediate filament multigene family. Divergence was assumed to have occurred from an ancestral sequence containing seven introns. T, mRNA-mediated transposition event leading to the integration of an expressed intermediate filament gene sequence that later acquired two new introns in the second α-helical coil and a third intron (not shown) possibly as part of a process of recruitment of adjacent sequences that came to encode the acidic carboxy terminus; A, ancestral gene. The arrows denote acquisition or loss of an intron. We predict an intron arrangement in the gene encoding the 150-kDa neurofilament protein similar to that described here for the gene encoding the 68-kDa neurofilament protein (see text). The relative times of divergence are based on the data in Fig. 2 and are only intended as a rough guide. Other possible models are discussed in the text. The abbreviations are defined in the legend to Fig. 2.

One mechanism whereby introns may be lost involves the transposition of transcribed sequences to a new genetic locus via an mRNA intermediate (16, 19, 20). During such events, introns are lost as part of the splicing process that occurs in the production of cytoplasmic mRNA molecules. However, because mRNA transcripts in general lack promoter sequences, the integration of an intronless cDNA molecule into the host chromosome usually results in the generation of a nontranscribed processed pseudogene that, with the passage of time, acquires mutations and thereby loses homology with its parental transcribed sequence (19). For mRNA-mediated transposition to result in the generation of an expressed intronless sequence, either integration would have to occur adjacent to a functional promoter (with appropriate regulatory controls) or the initial mRNA transcript would have to include upstream promoter sequences that could then be carried through the transposition process. Evidence has been presented for precisely such a mechanism occurring in the generation of rat preproinsulin gene I, which lacks two of the introns present in the otherwise homologous preproinsulin gene II (30). The two genes are located on different chromosomes; gene I is flanked by a long direct repeat (a hallmark of integration events) and contains the remnants of a downstream poly(A) tract. Although it is conceivable that such a mechanism might have been involved in the generation of an intronless but nonetheless transcribed intermediate filament gene sequence that subsequently came to encode the 68-kDa neurofilament polypeptide (Fig. 4), the time at which this event might have occurred is sufficiently distant (at least 600 million years ago, based on its presence in vertebrates, annelids, and molluscs [18]) so that all traces of the molecular event [flanking direct repeats and a poly(A) tract] would have been obliterated by neutral drift. In addition, the original intronless sequence would have had to acquire three new introns, presumably by insertional transposition, subsequent to its own creation.

One alternative explanation for the unexpected placement of introns in the gene encoding the 68-kDa neurofilament protein would involve its divergence from a primordial sequence before the divergence of the keratin gene subfamily. In that event, the closer homology of the 68-kDa neurofilament polypeptide to desmin, vimentin, and GFAP than to the keratins would have to be explained by selective constraints leading to parallel evolution. In addition, after divergence of the 68-kDa neurofilament protein gene, the primordial sequence would have to have lost its three introns which would have been replaced (by whatever mechanism) with the eight introns present in contemporary nonneurofilament intermediate filament genes.

Various patterns of intron placement occur within different multigene families. In general, the genes of lower eucaryotic species seem to have fewer introns than do corresponding genes in vertebrate species (e.g., see references 3 and 7). In vertebrate genes, however, the placement of introns appears to be more rigidly conserved; for example, four intron positions are common to all vertebrate actin genes, with several other conserved intron positions present within each actin subfamily (7, 34). For the vertebrate tubulin multigene families, the conservation of intron position is even more evident (3).

The three genes encoding the neurofilament triplet proteins share two structural features in addition to the α-helices common to all intermediate filament proteins. These features are (i) neurospecific expression and (ii) an encoded long and highly acidic carboxy-terminal domain. Therefore, it might be predicted that these three genes themselves arose from the multiplication of a primordial neurofilament gene. Based on the extent of divergence of the known amino acid sequences in the conserved coils (Fig. 2; [11]) and the diversity of phyla containing two or three neurofilament peptides (18), this multiplication must have happened very soon after the creation of the primordial neurofilament sequence. If that primordial sequence arose via an mRNA-mediated transposition event, as we have argued, this would be reflected by nonhomologous intron placement in the genes encoding the 150- and 200-kDa neurofilament proteins.

In view of the anomalous structure of the gene encoding the 68-kDa neurofilament polypeptide, it is hard to conceive of a scheme for the divergence of the intermediate filament multigene family that did not involve both the loss of preexisting introns and the creation of new ones. It appears likely, therefore, that both kinds of events can contribute to the remarkable plasticity of the eucaryotic genome.

ACKNOWLEDGMENTS

We thank Argiris Efstratiadis for amusing discussions.

This work was supported by a grant from the National Institutes of Health and a grant from the Muscular Dystrophy Association.

LITERATURE CITED

1. **Balcarek, J. M., and N. J. Cowan.** 1985. Structure of the mouse glial fibrillary acidic protein gene: implications for the evolution of the intermediate filament multigene family. Nucleic Acids Res. **13**:5527–5543.
2. **Calame, K., J. Rogers, P. Early, M. Davis, D. Levant, R. Wall, and L. Hood.** 1980. Mouse C heavy chain immunoglobulin gene segment contains three intervening sequences separating domains. Nature (London) **284**:452–455.

3. **Cowan, N. J.** 1985. Tubulin genes and the diversity of microtubule function, p. 36–60. In N. McLean, (ed.), Oxford surveys on eukaryotic genes, vol. 1. Oxford University Press, Oxford.
4. **Craik, C. S., S. R. Buchman, and S. Beychok.** 1980. Characterization of globin domains: heme binding to the central exon product. Proc. Natl. Acad. Sci. USA **77**:1384–1388.
5. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. K. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
6. **Engel, J., P. Gunning, and L. Kedes.** 1982. Human actin proteins are encoded by a multigene family, p. 107–118. *In* M. L. Pearson and H. F. Epstein (ed.), Muscle development, molecular and cellular control. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **Fornwald, J. A., G. Kuncio, I. Peng, and C. P. Ordahl.** 1982. The complete nucleotide sequence of the chick α-actin gene and its evolutionary relationship to the actin gene family. Nucleic Acids Res. **10**:3861–3875.
8. **Fuchs, E., and I. Hanukoglu.** 1983. Unravelling the structure of the intermediate filaments. Cell **34**:332–334.
9. **Fyrberg, E. A., B. J. Bond, N. D. Hershey, K. S. Mixter, and W. Davidson.** 1981. The actin genes of Drosophila: protein coding regions are rigidly conserved but intron positions are not. Cell **24**:107–116.
10. **Geisler, N., S. Fisher, J. Vanderkerckhove, U. Plessmann, and K. Weber.** 1984. Hybrid character of a large neurofilament protein (NF-M): intermediate filament-type sequences followed by a long and acidic carboxyterminal extension. EMBO J. **3**:2701–2706.
11. **Geisler, N., S. Fisher, J. Vanderkerckhove, J. Van Damme, U. Plessmann, and K. Weber.** 1985. Protein-chemical characterization of NF-H, the largest mammalian neurofilament component; intermediate filament-type sequences followed by a unique carboxyterminal extension. EMBO J. **4**:57–63.
12. **Geisler, N., E. Kaufman, S. Fisher, U. Plessmann, and K. Weber.** 1983. Neurofilament architecture combines structural principles of intermediate filaments with carboxyterminal extensions increasing in size between triplet proteins. EMBO J. **2**:1295–1302.
13. **Geisler, N., and K. Weber.** 1982. The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins. EMBO J. **2**:1649–1656.
14. **Geisler, N., and K. Weber.** 1985. The complete amino acid sequence of the major mammalian neurofilament protein (NF-S). FEBS Lett. **182**:475–478.
15. **Gilbert, W.** 1978. Why genes in pieces? Nature (London) **271**:501.
16. **Karin, M., and R. I. Richards.** 1982. Human metallothionein genes: primary structure of the metallothionein II gene and a related processed gene. Nature (London) **299**:797–802.
17. **Karn, J., S. Brenner, and L. Barnett.** 1983. Protein structural domains in the *C. elegans* UNC-54 myosin heavy chain are not separated by introns. Proc. Natl. Acad. Sci. USA **80**:4253–4257.
18. **Lasek, R. J., L. Phillips, M. J. Katz, and L. Autilio-Gambetti.** 1985. Function and evolution of neurofilament proteins. Ann. N.Y. Acad. Sci. **455**:462–478.
19. **Lee, M. G. S., S. A. Lewis, C. D. Wilde, and N. J. Cowan.** 1983. Evolutionary history of a multigene family: an expressed human β-tubulin gene and three processed pseudogenes. Cell **33**:477–487.
20. **Lemischka, I., and P. A. Sharp.** 1982. The sequences of an expressed rat α-tubulin gene and a pseudogene with an inserted repetitive element. Nature (London) **300**:330–335.
21. **Lewin, R.** 1982. On the origin of introns. Science **217**:921–922.
22. **Lewis, S. A., and N. J. Cowan.** 1985. Genetics, evolution and expression of the 68,000 mol wt neurofilament protein: isolation of a cloned cDNA probe. J. Cell Biol. **100**:843–850.
23. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**:687–701.
24. **Marchuk, D., S. McCrohan, and E. Fuchs.** 1984. Remarkable conservation of structure among intermediate filament genes. Cell **39**:491–498.
25. **Quax, W., W. V. Egberts, W. Hendriks, Y. Quax-Jenken, and H. Bloemendal.** 1983. The structure of the vimentin gene. Cell **35**:215–223.
26. **Quax, W., L. Van der Broeck, W. V. Egberts, F. Ramaekers, and H. Bloemendal.** 1985. Characterization of the hamster desmin gene: expression and formation of desmin filaments in nonmuscle cells after gene transfer. Cell **43**:327–338.
27. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling DNA to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:231–251.
28. **Sanger, F., A. R. Coulsen, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
29. **Shimizu, A., N. Takahashi, Y. Yaoita, and T. Honjo.** 1982. Organization of the constant region gene family of the mouse immunoglobin heavy chain. Cell **28**:499–506.
30. **Soares, M. B., E. Schon, A. Henderson, S. K. Karathanasis, R. Cate, S. Zeitlin, J. Chirgwin, and A. Efstratiadis.** 1985. RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol. Cell Biol. **5**:2090–2103.
31. **Southern, E.** 1975. Detection of specific sequence homologies among fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
32. **Staden, R.** 1980. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. **8**:3673–3694.
33. **Tyner, A. L., M. J. Eichman, and E. Fuchs.** 1985. The sequence of a type II keratin gene expressed in human skin: conservation of structure among all intermediate filament genes. Proc. Natl. Acad. Sci. USA **82**:4683–4687.
34. **Ueyama, H., H. Hamada, N. Battula, and T. Kakunaga.** 1984. Structure of a human smooth muscle actin gene (aortic type) with a unique intron site. Mol. Cell. Biol. **4**:1073–1078.