Genetics, Evolution, and Expression of the 68,000-mol-wt Neurofilament Protein: Isolation of a Cloned cDNA Probe

SALLY A. LEWIS and NICHOLAS J. COWAN
Department of Biochemistry, New York University School of Medicine, New York 10016

ABSTRACT A 1.2-kilobase (kb) cDNA clone (NF68) encoding the mouse 68,000-mol-wt neurofilament protein is described. The clone was isolated from a mouse brain cDNA library by low-stringency cross-hybridization with a cDNA probe encoding mouse glial fibrillary acidic protein (Lewis et al., 1984, *Proc. Natl. Acad. Sci. USA.*, 81:2743–2746). The identity of NF68 was established by hybrid selection using mouse brain polyA$^+$ mRNA, and cell-free translation of the selected mRNA species. The cell-free translation product co-migrated with authentic 68,000-mol-wt neurofilament protein on an SDS/polyacrylamide gel, and was immunoprecipitable with a monospecific rabbit anti-bovine neurofilament antiserum. In addition, DNA sequence analysis of NF68 showed 90% homology at the amino acid level compared with the sequence of the porcine 68,000-mol-wt neurofilament protein. At high stringency, NF68 detects a single genomic sequence encoding the mouse 68,000-mol-wt neurofilament protein. Two mRNA species of 2.5 kb and 4.0 kb are transcribed from the single gene in mouse brain. The level of expression of these mRNAs remains almost constant in postnatal mouse brains of all ages and, indeed, in the adult. At reduced stringency, NF68 detects a number of mRNAs that are expressed in mouse brain, one of which encodes the 150,000-mol-wt neurofilament protein. The NF68 probe cross-hybridizes at high stringency with genomic sequences in species as diverse as human, chicken, and (weakly) frog, but not with DNA from *Drosophila* or sea urchin.

Intermediate filaments, together with microtubules and actin filaments, form the structural network of the cytoskeleton in higher eucaryotes. For reasons as yet unclear, different cell types require different types of intermediate filament. Five such types exist: the cytokeratins in epithelial cells, desmin in muscle cells, vimentin in cells of mesenchymal origin and in most cultured cells, glial filaments in astrocytes, and neurofilaments in neurons (for reviews, see references 16 and 20). The polypeptide subunits that polymerize to form these filaments have been extensively investigated by peptide mapping and amino acid sequencing, as well as by immunological and physicochemical methods. All five types of intermediate filament protein have a long, homologous central helical rod domain with a hydrophobic backbone that causes the aggregation of the subunits into the coiled-coil structure of the 10-nm filament (7, 12, 29).

Among the five intermediate filament classes, neurofilaments are of particular interest because of their central role as a structural matrix in neuronal axons, dendrites, and perikarya. Additionally, they are markers and, perhaps, determining factors in the development of the central nervous system and may play a role in degenerative neurological disorders such as Alzheimer's disease (27). Mammalian neurofilaments contain three related polypeptides with molecular weights of 68,000, 150,000, and 200,000 (14). The 68,000-mol-wt protein is the most abundant of the three and, unlike the higher molecular weight neurofilament polypeptides, can polymerize on its own in vitro to form 10-nm filaments (19). The three neurofilament polypeptides differ mainly in their lysine- and glutamic acid–rich carboxyterminal tails, which are thought to extend from the filament and interact with other cellular components (8).

The construction of cloned cDNA probes represents a crucial step for a detailed study of the function, genetics, expression, developmental regulation, and evolution of intermediate filament proteins. Such probes have been generated for epidermal keratins (12, 29), vimentin (23, 32), desmin (22), and the glial fibrillary acidic protein (GFAP)[1] (17). Here

[1] *Abbreviations used in this paper:* GFAP, glial fibrillary acidic protein; kb, kilobases.

we describe the isolation and characterization of a cDNA probe encoding the 68,000-mol-wt neurofilament protein. This probe has been used to investigate the number and evolutionary conservation of sequences encoding the 68,000-mol-wt neurofilament polypeptide, and the evolutionary relationships between the genes encoding the neurofilament triplet polypeptides. In addition, the probe has been used to measure the level of expression of the mRNA encoding the 68,000-mol-wt protein in developing and adult mouse brain.

MATERIALS AND METHODS

Preparation and Screening of a Mouse Brain cDNA Library: Mouse brain polyA$^+$ mRNA was prepared from 12–21-d-old mice as described (4). Double-stranded cDNA was synthesized with reverse transcriptase (Life Science Instruments, Elkhart, IN) according to the method of Helfman et al. (13), except that the cDNA was methylated with *Eco* RI methylase (New England Biolabs, Beverly, MA) under the conditions specified by the manufacturer before the addition and cutting back of *Eco* RI synthetic oligonucleotide linkers (Collaborative Research Inc., Lexington, MA). That fraction of double-stranded cDNA larger than 500 base pairs was ligated into DNA prepared from bacteriophage λgt11 (31) that had been cleaved with *Eco* RI and dephosphorylated with alkaline phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, IN). The resulting recombinant DNA was packaged into bacteriophage using a commercial packaging extract purchased from Promega Biotech, Inc (Madison, WI). The library was not amplified, so as to avoid any possibility of making it less representative as a consequence of differential growth rates of individual bacteriophage. Host cells (strain Y1090) (see reference 31) were infected and plated out, and the fusion protein was induced with isopropyl-β-D-thiogalactopyranoside as described (31). Approximately 5×10^5 recombinant plaques were screened using an anti-neurofilament antiserum (18). In a separate experiment, $\sim 5 \times 10^5$ plaques were replicated onto nitrocellulose (1), and the replicas hybridized with coding region fragment derived from the GFAP cDNA clone G1 (17) labeled with ^{32}P by nick-translation (25). After hybridization, the filters were washed to a final stringency of 2 × SSC (SSC is 150 mM NaCl, 15 mM Na citrate), 50°C. Clones giving positive signals in either of the screening experiments described above were purified to homogeneity and the cDNA inserts subcloned into either the replicative form of bacteriophage M13mp8 or into plasmid pUC8 for further amplification and study.

Analysis of cDNA Clones and DNA and RNA Blot Transfer Experiments: Hybrid selection/translation experiments were carried out with representative subcloned cDNAs by methods previously described (11). Rabbit reticulocyte cell-free translation products were characterized by analysis on stacking SDS polyacrylamide gels (15), both directly and after immunoprecipitation with the anti-neurofilament protein antiserum (18) as described (9). The NF68 cDNA clone, which contained an insert of ~1.2 kilobases (kb), was sequenced on both strands by the dideoxy chain terminator method (26). RNA prepared from the dissected brains of developmentally staged mice by the guanidine isothiocyanate procedure (2) was resolved on 1.0% agarose gels containing 2.2 M formaldehyde (3) and transferred to nitrocellulose. DNA prepared from human placenta, mouse liver, chicken embryos, frog erythrocytes, *Drosophila* embryos, and sea urchin sperm was digested with a fourfold excess of restriction endonuclease, resolved on 0.8% agarose gels, and blotted onto nitrocellulose (28). The blots were hybridized (20 h) with either cloned cDNA inserts or restriction fragments derived from G1 (17) or NF68, and ^{32}P-labeled by nick-translation (25). After hybridization, blots were washed to various final stringency conditions (see Results).

RESULTS

Isolation of a cDNA Clone, NF68, Encoding the Mouse 68,000-mol-wt Neurofilament Protein

Two independent approaches were taken in our search for a cloned probe encoding the 68,000-mol-wt neurofilament protein. The first depended on the construction of a mouse brain cDNA library in a vector (bacteriophage λgt11) designed to express the cDNA inserts as fusion proteins in conjunction with a portion of the *Escherichia coli* lac Z gene (31). Given that the cDNA library contains molecules of interest that are cloned in the correct orientation and reading frame, a fusion protein will be synthesized in the bacterial host. Upon induced lysis, bacteriophage plaques containing the fusion protein of interest can be detected using a specific antibody (in this case, rabbit anti-bovine 68,000-mol-wt neurofilament protein, [reference 18]). Though on the average only one in six cDNA clones will yield a recognizable translation product, the method offers the advantage that a very large number of bacteriophage plaques can be screened in a single experiment. Therefore, the availability of a specific polyclonal antiserum should allow the detection of cDNA clones present in the library at relatively low abundance.

A monospecific rabbit anti-bovine neurofilament antiserum (18) was used to screen about 5×10^5 recombinant λgt11 plaques containing cDNA inserts generated using unfractionated polyA$^+$ mRNA prepared from 12-d-old mouse brain. The average size of cDNA inserts contained in the vector was about 1.0 kb (data not shown). Bacteriophage from three plaques each giving positive signals with the anti-neurofilament antiserum were processed through three successive cycles of plaque purification, at each stage and in each case yielding positive signals with the specific antibody. The cDNA inserts were subcloned from amplified bacteriophage preparations into the bacteriophage vector M13, and the subclones were tested for their ability to select 68,000-mol-wt neurofilament protein-specific mRNA from the unfractionated mouse polyA$^+$ mRNA used as substrate for the generation of the λgt11 expression library. Surprisingly, none of the three subcloned cDNAs selected an mRNA species that translated to yield a neurofilament protein. One clone yielded no detectable translation product, while the remaining two selected mRNAs that translated to give unidentified low molecular weight proteins that failed to immunoprecipitate with the anti-neurofilament antiserum (data not shown). The apparent paradox presented by this data is discussed below (see Discussion).

The second approach to the isolation of a cDNA clone encoding the 68,000-mol-wt neurofilament protein depended on the established structural features of all intermediate filament proteins thus far studied, namely, the presence of conserved α-helical regions within the polypeptide (7, 12, 29). For example, comparison of the amino acid sequences of mouse GFAP (17) with porcine 68,000-mol-wt neurofilament protein (8) reveals significant homology within the second α-helical region. Based on the existence of this homology, we reasoned that the corresponding nucleic acid sequences contained in a cloned mouse GFAP cDNA might be expected to cross-hybridize at low stringency with the homologous sequences contained in a cloned 68,000-mol-wt neurofilament-specific cDNA.

To test the feasibility of this approach, we performed a mouse genomic Southern blot experiment using the GFAP cDNA probe. After hybridization, the blot was washed to low stringency (2 × SSC, 50°C) so as to detect sequences with partial homology to GFAP in addition to GFAP sequences themselves. As a control to identify fully homologous GFAP sequences, we rewashed the same blot to high stringency (0.2 × SSC, 68°C). At low stringency, a number of weaker bands are detected by the GFAP probe that do not appear after washing to high stringency (Fig. 1). Thus, the mouse genome contains a small, though uncertain, number of sequences that share partial homology with GFAP coding sequences.

On the basis of this result, we screened $\sim 5 \times 10^5$ recombinant λgt11 bacteriophage at low stringency (2 × SSC, 50°C) using the mouse GFAP cDNA probe. Approximately 120

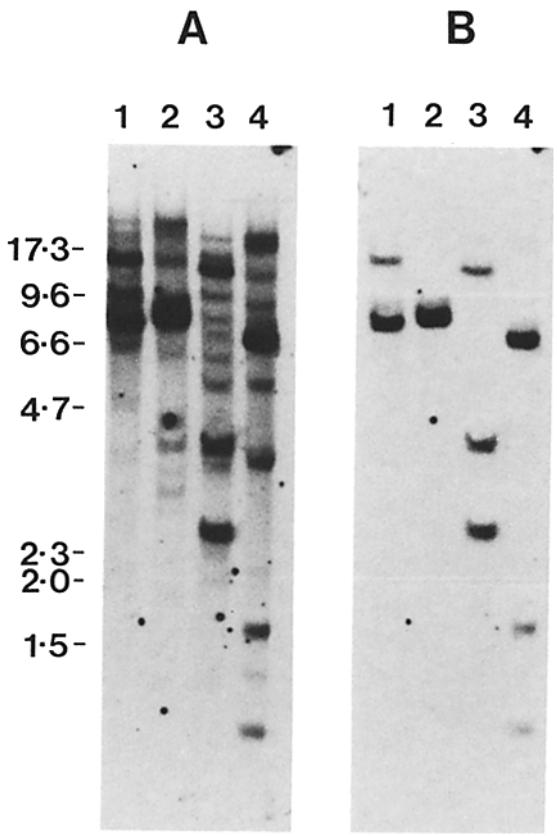

FIGURE 1 Detection of mouse genomic sequences with partial homology to a cloned GFAP probe. Mouse genomic DNA (10 μg) was digested with restriction endonuclease, then resolved on a 0.8% agarose gel, and the gel content was transferred to nitrocellulose (28). (Lane *1*) *Eco* RI; (lane *2*) *Bam* HI; (lane *3*) *Bgl* II; (lane *4*) *Hind* III. The blot was hybridized with a *Sal* 1/*Hind* III restriction fragment from the GFAP encoding cDNA clone G1 (17). ^{32}P-labeled by nick-translation (25). After hybridization, the blot was washed to a stringency of 2 × SSC, 50°C, and exposed to film (*A*). The same blot was then rewashed to a stringency of 0.2 × SSC, 68°C, and re-exposed (*B*).

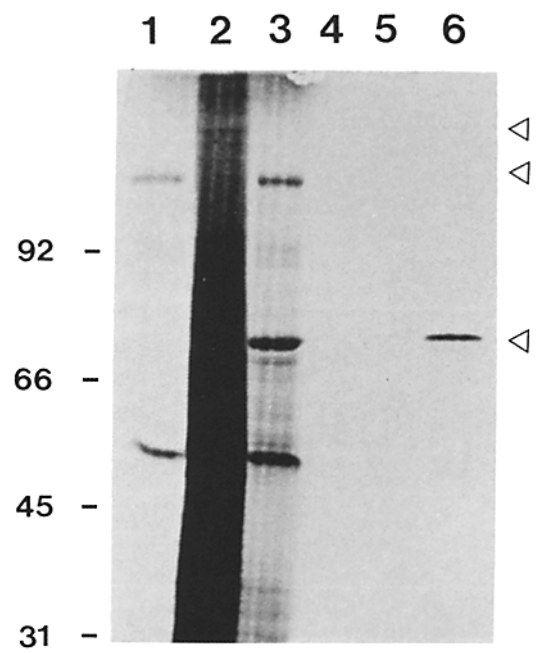

FIGURE 2 The NF68 cDNA clone contains sequences that encode the 68,000-mol-wt neurofilament protein. Approximately 30 μg of the NF68 cDNA clone was linearized by restriction digestion with *Sal* 1, denatured, and spotted onto a nitrocellulose filter (see Materials and Methods and Reference 11). This filter was prehybridized and hybridized at 41°C for 12 h with 30 μg of total mouse brain polyA$^+$ mRNA in a mixture containing 50% formamide, 0.4M NaCl, 20 mM PIPES (pH 7.0), 2 mM EDTA, and 10 μg/ml *Escherichia coli* tRNA. After hybridization, the filter was washed to a final stringency of 0.1 × SSC, 60°C. Hybridized material was eluted by boiling, recovered by ethanol precipitation, and translated in an enriched cell-free rabbit reticulocyte lysate containing [^{35}S]methionine (21). Translation products were analyzed either directly or after immunoprecipitation with an anti-neurofilament antiserum (9, 18) on a stacking 8% polyacrylamide gel (15). (Lane *1*) H_2O control (no added mRNA); (lane *2*) translation products from 2 μg total mouse brain polyA$^+$ mRNA; (lane *3*) translation products from material selected by the NF68 clone. (Lanes *4–6*) Immunoprecipitates obtained from the material shown in lanes *1–3* using the anti-neurofilament protein antiserum. Arrows show the position of (unlabeled) authentic rat neurofilament protein run on the same gel. The position of unlabeled molecular weight markers (× 10^{-3}) is shown on the left.

positively hybridizing plaques were obtained. Sixteen of these plaques were selected at random and purified, and the bacteriophage was amplified. To determine sequence relationships among these cloned cDNAs (if any), cloned DNA was excised by restriction digestion from five of the sixteen recombinants (again selected randomly), ^{32}P-labeled by nick-translation, and used as probes in a dot-blot experiment to classify the remaining eleven bacteriophage clones which contain partial homology to mouse GFAP cDNA. After washing to high stringency, the dot-blot experiment revealed that each of the sixteen recombinants could be classified as belonging to one of five groups (data not shown). The recombinant fragment from that member of each group containing the largest insert was subcloned into the plasmid vector pUC8. These five subcloned cDNAs were linearized by restriction digestion, bound to nitrocellulose filters, and used to select mRNA from unfractionated mouse brain polyA$^+$ mRNA. Selected mRNA was translated in a rabbit reticulocyte lysate containing [^{35}S]-methionine in parallel with total mouse brain polyA$^+$ mRNA and a negative control. Of the five classes represented by each subclone, one (clone NF68) selected an mRNA that translated to yield a product that co-migrated with authentic 68,000-mol-wt neurofilament protein (Fig. 2, lane *3*). This translation product was specifically immunoprecipitated by the antineurofilament antiserum (Fig. 2, lane *6*). On the basis of this evidence, plus the extensive homology (within the coding region) of the peptide encoded by NF68 with the known amino acid sequence of the porcine 68,000-mol-wt neurofilament protein (see below), we conclude that NF68 represents a cDNA clone encoding the mouse 68,000-mol-wt neurofilament protein.

Sequence of the NF68 cDNA Clone

To confirm that the NF68 clone indeed encodes the 68,000-mol-wt neurofilament protein, and to determine the extent of amino acid homology with known 68,000-mol-wt neurofilament protein sequences (8), both strands of the cDNA insert of NF68 were fully sequenced. The data (Fig. 3) reveal extensive (>90%) amino acid homology between mouse and porcine sequences. The NF68 clone encodes amino acids that span the entire second α-helical region and the glutamic acid–

```
    gln ile ser val glu met asp val ser ser lys pro asp leu ser ala ala leu lys
 CT CAG ATC TCC GTG GAG ATG GAC GTG TCC TCC AAG CCC GAC CTC TCC GCC GCT CTC AAG

asp ile arg ala gln tyr glu lys leu ala ala lys asn met gln asn ala glu glu trp
GAC ATC CGC GCT CAG TAC GAG AAG CTG GCC GCC AAG AAC ATG CAG AAC GCC GAA GAG TGG

phe lys ser arg phe thr val leu thr glu ser ala ala lys asn thr asp ala val arg
TTC AAG AGC CGC TTC ACC GTG CTA ACC GAG AGC GCC GCC AAG AAC ACC GAC GCT GTG CGC

ala ala lys asp glu val ser glu ser arg arg leu leu lys ala lys thr leu glu ile
GCT GCC AAG GAC GAG GTG TCG GAA AGC CGC CGC CTG CTC AAG GCT AAG ACC CTG GAG ATC

glu ala cys arg gly met asn glu ala leu glu lys gln leu gln glu leu glu asp lys
GAA GCC TGC CGG GGT ATG AAC GAA GCT CTG GAG AAG CAG CTG CAG GAG CTA GAG GAC AAG

gln asn ala asp ile ser ala met gln asp thr ile asn lys leu glu asn glu leu arg
CAG AAT GCA GAC ATT AGC GCC ATG CAG GAC ACA ATC AAC AAA CTG GAG AAT GAG CTG AGA

ser thr lys ser glu met ala arg tyr leu lys glu tyr gln asp leu leu asn val lys
AGC ACG AAG AGC GAG ATG GCC AGG TAC CTG AAG GAG TAC CAG GAC CTC CTC AAT GTC AAG

met ala leu asp ile glu ile ala ala tyr arg lys leu leu glu gly glu glu thr arg
ATG GCC TTG GAC ATC GAG ATT GCA GCT TAC AGA AAA CTC TTG GAA GGC GAA GAG ACC AGG

leu ser phe thr ser val gly ser ile thr ser gly tyr ser gln ser ser gln val phe
CTC AGT TTC ACC AGC GTG GGT AGC ATA ACC AGC GGC TAC TCT CAG AGC TCG CAG GTC TTC

gly arg ser ala tyr ser gly leu gln ser ser ser tyr leu met ser ala arg ser phe
GGC CGT TCT GCT TAC AGT GGC TTG CAG AGC AGC TCC TAC TTG ATG TCT GCT CGC TCT TTC

pro ala tyr tyr thr ser his val gln glu glu glu gln thr glu val glu glu thr ile glu
CCA GCC TAC TAT ACC AGC CAC GTC CAG GAA GAG CAG ACA GAG GTC GAG GAG ACC ATT GAG

ala thr lys ala glu glu ala lys asp glu pro pro ser glu gly glu ala glu glu glu
GCT ACG AAA GCT GAG GAG GCC AAG GAT GAG CCC CCC TCT GAA GGA GAA GCA GAA GAG GAG

glu lys glu lys glu glu gly glu glu glu glu gly ala glu glu glu glu ala ala lys
GAG AAG GAG AAA GAG GAG GGA GAG GAA GAG GAA GGC GCT GAG GAG GAA GAA GCT GCC AAG

asp glu ser glu asp thr lys glu glu glu glu gly gly glu gly gly glu glu asp thr
GAT GAG TCT GAA GAC ACA AAA GAA GAA GAA GAA GGT GGT GAG GGT GAG GAG GAA GAC ACC

lys glu ser glu glu glu glu lys lys glu glu ser ala gly glu glu gln val ala lys
AAA GAA TCT GAA GAG GAA GAG AAG AAA GAG GAG AGT GCT GGA GAG GAG CAG GTG GCT AAG

lys lys asp
AAG AAA GAT TGA GCCCTATTCCCAACTATTCCAGGAAAAGTTCTCCCCAATCAGGTCCAACCTCATCACCAACCA

ACCAGTTGAGTTCCAGATCCTATACAAATTAAGAAGTCAATACATGTATAATTCTGAGAATGACTTAGGTTGGACTTTC

AAATGTTGTGCTATGAATTTCCTCCTTACGCAGAGTATCTGTTTGCTTGCAGAGTGGCTTTCTGGCTTGCTGCCAGCCT

GTGCATGGTCCATGCTTATGAGTTCAGGATCTATGGCAATGTGAATCACACAGATGTTTAGAATAATAATAAAAAAAAA

ACCACACACACAACACGAATAAATGAATTCC
```

FIGURE 3 Sequence of the NF68 cDNA clone. The sequence of the insert contained in NF68 is shown together with the amino acids encoded within the coding region. Possible polyadenylation signals within the 3′-untranslated region of NF68 are underlined. The boundaries of the second α-helical region are delineated by parentheses. A *Sac* I restriction site used in the preparation of probes for blot analyses (see text) is indicated (dashes).

lysine-rich tailpiece previously identified in the porcine protein (8). A remarkable feature of the DNA sequence encoding this latter region is that it consists almost entirely (80%) of purine residues. Indeed, within this region there is an extensive sequence (encoding 14 contiguous amino acids) that is entirely devoid of pyrimidine residues.

A Single Genomic Sequence Encoding the Mouse 68,000-mol-wt Neurofilament Protein Is Transcribed to Yield Two mRNAs of 2.5 and 4.0 kb

We performed a genomic Southern blot to determine the number of sequences in the mouse genome that are homologous to the 68,000-mol-wt neurofilament mRNA. Four restriction enzymes—*Eco* RI, *Bam* HI, *Bgl* II, and *Hind* III—were used in this experiment; none of these enzymes cleave the cDNA insert in NF68. A single band was detected per restriction digest (Fig. 4*A*). Thus, in mouse, the 68,000-mol-wt neurofilament protein is encoded by a single gene.

The size of the mRNA transcribed from this gene was investigated in an RNA blot transfer experiment using polyA$^+$ mRNA prepared from adult mouse brain. Unexpectedly, two RNA species of 2.5 and 4.0 kb were detected after washing of the blot to high stringency (Fig. 4*B*). Because a single sequence encodes the mouse 68,000-mol-wt neurofilament protein (Fig. 4*A*), these two mRNAs must be transcribed from a single gene.

Expression of the 68,000-mol-wt Neurofilament Protein mRNA in Developing and Adult Mouse Brain

The temporal expression of RNA sequences encoding the mouse 68,000-mol-wt neurofilament protein was studied in staged developing mouse brains. Total RNA from the brains of stage-specific mice was analyzed in an RNA blot transfer experiment, and the blot probed with the NF68 clone ^{32}P-labeled by nick-translation. mRNA encoding the 68,000-mol-wt neurofilament protein is present in late embryo brain, and is expressed at a fairly constant level after about postnatal day 5, and, indeed, in adult mouse brain (Fig. 5).

Interspecies Conservation of the 68,000-mol-wt Neurofilament Protein Gene and the Detection of Homologous Expressed Sequences Including the 150,000-mol-wt Neurofilament Protein

To determine the extent to which sequences encoding the 68,000-mol-wt neurofilament protein have been evolutionarily conserved, we performed a Southern blot experiment with the NF68 probe using genomic DNA from mouse,

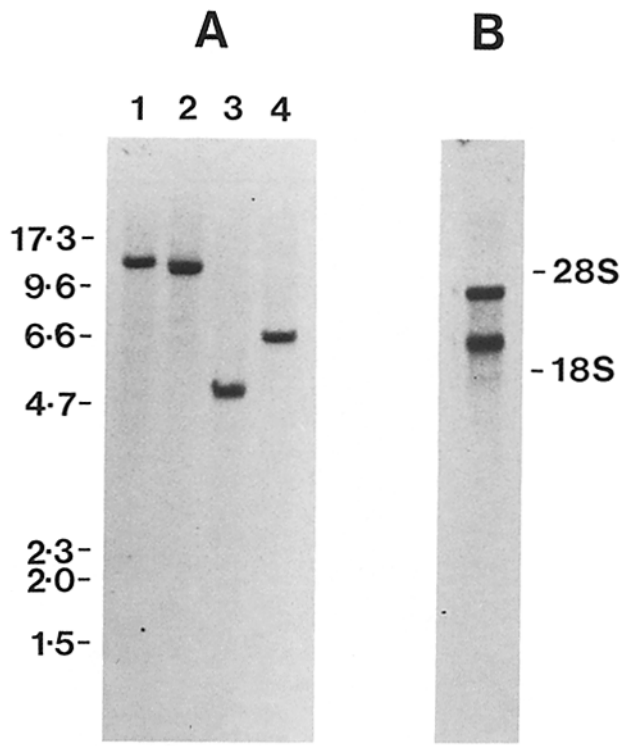

FIGURE 4 (*A*) A single sequence encodes the 68,000-mol-wt neurofilament polypeptide in the mouse genome. Aliquots (10 μg) of mouse genomic DNA were digested with a fourfold excess of restriction enzyme. Fragments were resolved on a 0.8% agarose gel and the gel content transferred to nitrocellulose (28). The blot was hybridized with 2×10^7 cpm of the cloned insert from NF68 labeled with ^{32}P by nick-translation (25). After hybridization, the blot was washed to a final stringency of 0.2 × SSC, 68°C. (Lane *1*) *Eco* RI; (lane *2*) *Bam* HI; (lane *3*) *Hind* III; (lane *4*) *Sac* I. Size markers (in kilobases) are shown at left. (*B*) Two mRNAs are expressed in mouse brain from a single 68,000-mol-wt neurofilament polypeptide gene. About 5 μg of adult mouse brain polyA$^+$ mRNA was resolved on a denaturing 1.0% agarose/formaldehyde gel (3) and the gel content was transferred to nitrocellulose. The blot was hybridized and washed as described in *A*. The position of unlabeled 28S and 18S ribosomal RNA markers run in an adjacent slot is shown.

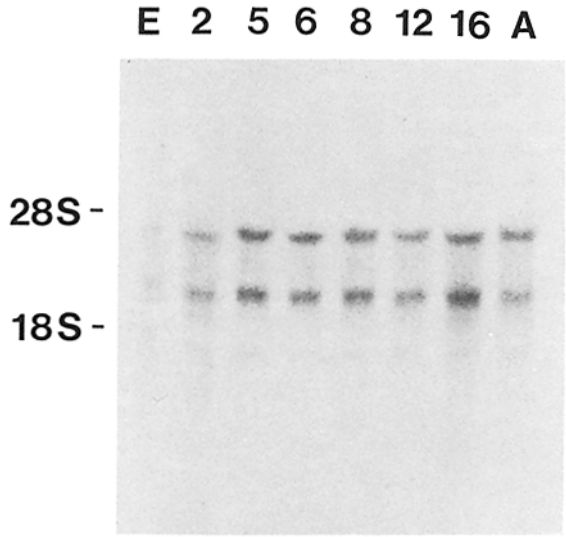

FIGURE 5 Expression of 68,000-mol-wt neurofilament protein-specific mRNA in developing and adult mouse brain. 20 μg of total mouse brain RNA from the brains of stage-specific mice was resolved on a 1% formaldehyde/agarose gel (3) and the gel content was transferred to nitrocellulose. The blot was hybridized with ~2 × 10^7 cpm of the cloned insert in NF68 labeled with ^{32}P by nick-translation (25). After hybridization, the blot was washed to a final stringency of 2 × SSC, 68°C. Numbers denote postnatal age of animals used for RNA preparation. *E*, 17-d embryo; *A*, adult. Positions of 28S and 18S ribosomal RNA markers are shown.

human, chicken, frog, *Drosophila*, and sea urchin. Because some degree of mismatching is to be expected between homologous genes from different species, the blot was initially washed to relatively low stringency (2 × SSC, 52°). Unexpectedly, the result was a smear of intense labeling covering the entire range of restriction fragment size in the homologous (mouse) DNA lanes, irrespective of the restriction enzyme used (data not shown). This continuum of hybridizing fragments, which is characteristic of repetitive sequences in the genome, was retained unless the blot was washed to high stringency (0.2 × SSC, 68°C) (see Fig. 4*A*). The NF68 cDNA clone includes an unusual and lengthy region that consists almost exclusively of purines encoding lysine and glutamic acid residues towards the COOH-terminal of the 68,000-mol-wt neurofilament protein (Fig. 3). With the possibility in mind that such purine-rich regions might cross-hybridize with repetitive sequences in the mammalian genome, we repeated the Southern blot experiment using a 300-base pair ^{32}P-labeled *Sac* I fragment derived from NF68 that encodes sequences 5′ to the lysine- and glutamic acid purine-rich tailpiece (see Fig. 3). In contrast to the result obtained with the intact NF68 probe, the data (Fig. 6) show a set of bands in the mouse genome indistinguishable from those obtained with the intact NF68 probe at high (0.2 × SSC, 68°C) stringency (Fig. 4*A*). The experiment also shows cross-hybridization of the NF68 cDNA probe with sequences in human, chicken, and frog genomic DNA but not the *Drosophila* or sea urchin DNA.

The isolation of a cloned probe encoding the 68,000-mol-wt neurofilament protein provides the basis for an investigation of evolutionary relationships at the genetic level between sequences encoding the neurofilament triplet proteins. Previous experiments (8) based on peptide mapping and amino acid analysis suggested that the glutamic acid- and lysine-rich tailpiece contained in the 68,000-mol-wt neurofilament protein is a feature shared by the 150,000- and 200,000-mol-wt neurofilament proteins. This suggests the possibility that cDNAs encoding these proteins could be identified with the NF68 probe. To examine this possibility, and to determine which mRNA species are most closely related to NF68, we performed a hybrid selection/translation experiment at reduced stringency using the NF68 probe. The success of this approach depends upon the absence in general of repetitive sequences in mRNA species selected by virtue of their polyA$^+$ tracts. The result is shown in Fig. 7. As the stringency conditions are lowered, an increased number of proteins translated from additionally selected mRNAs appear. In particular, unidentified proteins with molecular weights of 62,000 and 90,000 are apparent. When these translation products were immunoprecipitated with the neurofilament protein-specific antiserum, both the 68,000- and 150,000-mol-wt polypeptides were detected (Fig. 7, lane *8*). We conclude, therefore, that NF68 contains sequences that share homology with sequences encoding the 150,000-mol-wt neurofilament protein, as well as a number of other sequences encoding unidentified proteins that are expressed in mouse brain.

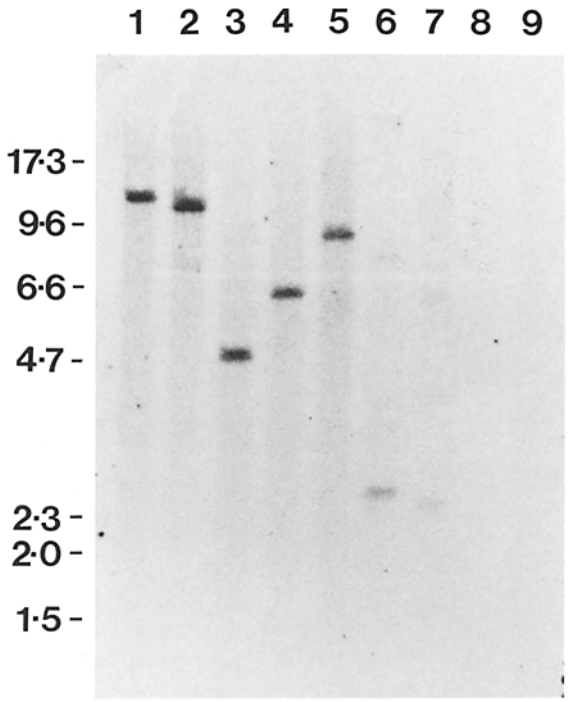

FIGURE 6 Evolutionary conservation of the 68,000-mol-wt neurofilament protein gene. Aliquots (10 μg) of DNA from mouse liver, human placenta, chicken embryo, frog reticulocytes, *Drosophila* embryo, and sea urchin sperm were digested with restriction enzyme, resolved on a 0.8% agarose gel, and blotted onto nitrocellulose. To avoid the detection of repetitive sequences (see text), we hybridized the blot with ~2 × 10^7 cpm of an *Eco* RI–*Sac* I fragment encoding the second α-helix of the 68,000-mol-wt neurofilament polypeptide, but not the lysine- and glutamic acid-rich carboxyterminal domain or the 3′-untranslated region (see text and Fig. 3). The blot was washed to a final stringency of 2 × SSC, 68°C. (Lanes *1–4*) Mouse DNA digested with *Eco* RI, *Bam* HI, *Hind* III, and *Sac* I, respectively. (Lanes *5–9*) *Eco* RI digests of DNA from human, chicken, frog (*Rana pipiens*), *Drosophila melanogaster*, and sea urchin. Size markers (in kilobases) are shown at left.

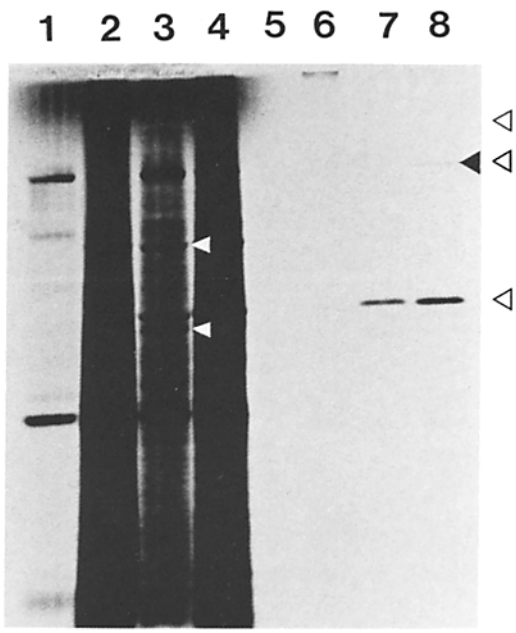

FIGURE 7 Partial homology between mRNA sequences encoding the 68,000- and 150,000-mol-wt neurofilament polypeptides. Two hybrid selection translation experiments were carried out as described in the legend to Fig. 2, except that the concentration of formamide in the prehybridization/hybridization mixture was reduced to 40% and the stringency of the final wash was reduced. (Lane *1*) H_2O control (no added mRNA); (lane *2*) translation products from 2 μg of total mouse brain polyA$^+$ mRNA; (lane *3*) translation products selected by an NF68-containing filter washed to a stringency of 0.1 × SSC, 52°C; (lane *4*) translation products from an identical filter washed to 0.1 × SSC, 45°C; (lanes *5–8*) analysis of translation products shown in lanes *1–4* after immunoprecipitation with the anti-neurofilament antiserum (18). Positions of unlabeled marker neurofilament polypeptides are marked (open arrows); the translated, immunoprecipitated 150,000-mol-wt neurofilament polypeptide is also marked (solid arrow). Unidentified translation products selected at reduced stringency (but not at high stringency) (see Fig. 2) are also marked (white arrows). The gel has been overexposed to show immunoprecipitated translation products.

DISCUSSION

This paper describes the isolation of a cDNA probe for the 68,000-mol-wt neurofilament gene from a mouse brain cDNA library cloned in bacteriophage λgt11. The isolation procedure depended upon the partial homology between the genes encoding the 68,000-mol-wt neurofilament protein and GFAP, another member of the loosely related intermediate filament multigene family. We first attempted to isolate a neurofilament cDNA from the λgt11 library by virtue of its expression as a fusion protein from the lac Z gene of the bacteriophage (31). Screening of 5 × 10^5 recombinant plaques using a polyclonal anti-neurofilament protein antiserum (18) resulted in the detection and plaque purification of three bacteriophage, each synthesizing a lac Z-cDNA fusion protein that was consistently recognized by the anti-neurofilament antiserum. Paradoxically, two of these three clones selected an mRNA species from total mouse brain polyA$^+$ mRNA that translated to give a protein immunologically unrelated to the neurofilament triplet polypeptides. Though the anti-neurofilament antiserum might recognize epitopes other than those presented by the neurofilament triplet proteins, this could not explain the nature of the immunoreactive fusion proteins, because the translation products from mRNA selected by these two clones were not immunoprecipitable with the same antiserum. A more convincing explanation is that a cloned cDNA unrelated to neurofilament sequences could be translated in the wrong reading frame from the lac Z gene promoter. When inserted into the λgt11 vector, such a sequence, or a region of a cDNA molecule corresponding to an untranslated portion of an mRNA molecule, could adventitiously give rise to an epitope recognized by the anti-neurofilament antiserum. In common with most polyclonal antisera, the anti-neurofilament antiserum recognizes many different epitopes in the neurofilament triplet polypeptides (18). Because epitopes can be as small as 6–7 amino acids, and because our bacteriophage cDNA expression library was large and contained cDNA fragments cloned indiscriminately without regard to orientation or reading frame, the generation of adventitious epitopes is a very real possibility. This kind of artifact appears to be an unexpected pitfall of procedures involving antibody screening of cDNA expression libraries.

The 1.2-kb cDNA clone, NF68, was isolated by low-stringency cross-hybridization with a GFAP-specific cDNA, and was shown to encode the mouse 68,000-mol-wt neurofilament polypeptide by two independent criteria. First, hybrid selection and translation of mouse brain mRNA and immunopre-

cipitation of the translation product identified the encoded protein as authentic 68,000-mol-wt neurofilament protein (Fig. 2). Second, DNA sequence analysis and comparison of the predicted amino acid sequence with the known amino acid sequence of a portion of the corresponding porcine protein revealed a homology of 90%. At high stringency, NF68 detects only a single band in a mouse genomic DNA blot transfer experiment (Fig. 4*A*). Thus, in mouse, the 68,000-mol-wt neurofilament polypeptide is encoded by a single gene. Two sizes of mRNA, of 2.5 and 4.0 kb, are detectable in mouse brain RNA blot transfer experiments with the NF68 probe. These mRNAs must therefore be transcribed from the same gene, probably via readthrough of a polyadenylation site, as has been shown for several other genes (see references 10 and 32). Since all three neurofilament polypeptides share similar α-helical domains, but differ in the length of their lysine- and glutamic acid–rich carboxyterminal ends, the possibility existed that the 150,000-mol-wt neurofilament protein could arise from this same gene by an alternative splicing mechanism, and be encoded by the larger (4.0-kb) mRNA. However, the hybrid selection experiment (Fig. 2) discussed above rules this out: when mRNA is selected by hybridization with NF68 at high stringency and translated in vitro, only the 68,000-mol-wt polypeptide is synthesized. On the other hand, at reduced stringency, the NF68 clone selects, in addition to mRNA encoding the 68,000-mol-wt protein, an mRNA that translates to yield the 150,000-mol-wt neurofilament polypeptide (among others) (Fig. 7). The partial homology implied by this experiment should be useful for the isolation of a cloned cDNA encoding the 150,000-mol-wt protein.

Neurofilaments begin to appear in the central and peripheral nervous system of the mouse after 9–10 d of embryonic development, coincident with the time of initial axonal extension (5). They are features of terminally differentiated nerve cells, and replace vimentin filaments in the cytoskeleton of dividing neuronal precursors, though both types of intermediate filament may be co-expressed for a short time (30). The mRNA encoding the 68,000-mol-wt neurofilament protein is present in mouse brain at a constant level after postnatal day 2, and at a lower level in late stage embryonic brain (Fig. 5). The expression of 68,000-mol-wt neurofilament–specific mRNA throughout postnatal development and its continued synthesis at a similar level in adult brain is consistent with a significant degree of turnover of neurofilament protein.

At reduced stringency, the NF68 clone hybridizes to a repetitive element in the mouse genome. For this reason, the intact probe could not be used in Southern blot experiments at low stringency. Therefore, to investigate the conservation of neurofilament sequences within the mouse genome and across species boundaries, we prepared two fragments from the cDNA clone, one which encompasses sequences encoding the second α-helical domain, and the other which contains the remainder, including the region encoding the lysine- and glutamic acid–rich carboxyterminus and the 3′ untranslated region. The latter fragment proved to be responsible for the detection of mouse repetitive sequences. It seems possible that the long purine tracts in the cDNA might cross-hybridize with mouse satellite DNA, which consists of short sequence repeats of mostly A and G residues (24) such as GAAAATGA, GAGAAATG, GAAAACTGA, etc. In any event, mouse brain RNA is not repetitive for any portion of the NF68 cDNA sequence under these stringency conditions, and therefore, the repetitive DNA responsible for extensive cross-hybridization is not transcribed in mouse brain.

The purine-rich sequence contained in the NF68 cDNA clone also detects repetitive sequences in the chicken and sea urchin genomes, but not in frog or *Drosophila* DNA (data not shown). It hybridizes to a single band in *Drosophila* DNA at moderate stringency, but whether this band represents a neurofilament gene is uncertain, especially in view of the unusual sequence properties of the probe. The nonrepetitive portion of NF68 hybridizes to a single band in genomic blots of human and chicken DNA, and weakly to frog DNA, at high (0.2 × SSC, 68°C) stringency (Fig. 6). No cross-hybridization is detectable with *Drosophila* or sea urchin DNA at any stringency. Thus, the 68,000-mol-wt neurofilament gene appears to be highly conserved among vertebrate species. This conclusion is reinforced by the amino acid sequence encoded by the mouse cDNA, which shows 90% identity with that portion of the porcine 68,000-mol-wt neurofilament protein that has been sequenced.

Between the 68,000-mol-wt neurofilament cDNA and other sequences encoding the intermediate filament multigene family, evolutionary relationships are more distant. For example, the mouse GFAP cDNA probe, G1, cross-hybridizes only at very low stringency with the 68,000-mol-wt neurofilament cDNA, though it was precisely this cross-hybridization that made its isolation possible. All in all, at 2 × SSC, 50°C, G1 detects 14 *Bgl* II restriction fragments in the mouse genome (Fig. 1), of which three are accounted for by the GFAP gene (Balcarek, J. M., and N. J. Cowan, unpublished results). At least one additional band is ascribable to the 68,000-mol-wt neurofilament polypeptide gene (Fig. 4), and others probably represent desmin and vimentin genes. The keratins are yet more distantly related, and sequences encoding these proteins are probably not detected at this stringency (6). We expected that cross-hybridization between intermediate filament genes would occur between sequences encoding the second α-helical domain, which is the most highly conserved region at the amino acid level. DNA sequence analysis confirms this expectation: a 150-base pair region encoding this helix shows 67% homology between G1 and NF68 (Fig. 3 and reference 17).

The three neurofilament polypeptides appear to be no more closely related to one another than each is to the other nonkeratin intermediate filaments. This is in spite of the fact that all three proteins have similar helical rod domains typical of the intermediate filament family, and, in addition, share lysine- and glutamic acid–rich tails. However, the 150,000-mol-wt neurofilament gene unequivocally cross-hybridizes with NF68 at reduced stringency (Fig. 7): an mRNA is selected that translates *in vitro* to yield the authentic 150,000-mol-wt protein. No band is visible at the 200,000-mol-wt marker in the same experiment. This result might be due to a limitation of the cell-free system: an mRNA encoding such a large polypeptide may be inefficiently translated in vitro and, to compound this difficulty, the 200,000-mol-wt neurofilament protein is known to be poor in methionine (the input label used for cell-free translation). Alternatively, it is possible that sequences encoding this protein are too distantly related to cross-hybridize at any stringency with NF68.

Thus far, studies on the expression of the neurofilament 68,000-mol-wt protein have been largely limited to experiments using fluorescent antisera. The construction of a cloned cDNA probe provides the opportunity for a fuller investiga-

tion, at the genetic level, of the factors that regulate the expression of this important neuronal cell-specific protein.

Our thanks are due to Drs. R. Liem and M. Shelanski, who provided the neurofilament antiserum and critically read the manuscript; and to Dr. M. G-S Lee, who helped with initial screening experiments. Genomic DNAs from chicken, *Rana pipiens, Drosophila melanogaster,* and sea urchin were generous gifts of Drs. D. W. Cleveland, R. Roeder, and S. Crowther. We thank Elizabeth Delgado for her heroic typing efforts.

Received for publication 21 September 1984, and in revised form 21 November 1984.

REFERENCES

1. Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plagues in situ. *Science (Wash. DC).* 196:180–181.
2. Berk, A., and P. Sharp. 1978. Spliced early mRNAs of SV40. *Proc. Natl. Acad. Sci. USA.* 75:1274–1278.
3. Boedtker, H. 1971. Conformation independent molecular weight determination of RNA by gel electrophoresis. *Biochim. Biophys. Acta.* 240:448–453.
4. Cleveland, D. W., M. A. Lopata, P. Sherline, and M. W. Kirschner. 1981. Unpolymerized tubulin modulates the level of tubulin mRNAs. *Cell.* 25:537–546.
5. Cochard, P., and D. Paulin. 1984. Initial expression of neurofilaments and vimentin in the central and peripheral nervous system of the mouse embryo in vivo. *J. Neurosci.* 4:2080–2094.
6. Fuchs, E. V., and D. Marchuk. 1983. Type I and type II keratins have evolved from lower eukaryotes to form the epidermal intermediate filaments in mammalian skin. *Proc. Natl. Acad. Sci. USA.* 80:5857–5861.
7. Geisler, N., and K. Weber. 1982. The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins. *EMBO (Eur. Mol. Biol. Organ.) J.* 1:1649–1656.
8. Geisler, N., E. Kaufmann, S. Fischer, V. Plessman, and K. Weber. 1983. Neurofilament architecture combines structural principles of intermediate filaments with caboxyterminal extensions increasing in size between triplet proteins. *EMBO (Eur. Mol. Biol. Organ.) J.* 2:1295–1302.
9. Goldman, B. M., and G. Blobel. 1978. Biogenesis of peroxisomes: intracellular site of synthesis of catalase and uricase. *Proc. Natl. Acad. Sci. USA.* 75:5066–5070.
10. Gwo-Shu Lee, M., 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.* 33:774–487.
11. Hall, J. L., L. Dudley, P. R. Dobner, S. A. Lewis, and N. J. Cowan. 1983. Identification of two human β-tubulin isotypes. *Mol. and Cell. Biol.* 3:854–862.
12. Hanukoglu, I., and E. V. Fuchs. 1983. The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structure domains among keratins. *Cell.* 33:915–924.
13. Helfman, D. M., J. R. Feramisco, J. C. Fiddes, G. P. Thomas, and S. H. Hughes. 1983. Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. *Proc. Natl. Acad. Sci. USA.* 80:31–35.
14. Hoffman, P. N., and R. Lasek. 1975. The slow component of axonal transport. Identification of major structural polypeptides of the axon and their generality among mammalian neurons. *J. Cell Biol.* 66:351–366.
15. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680–686.
16. Lazarides, E. 1982. Intermediate filaments. *Annu. Rev. Biochem.* 51:219–245.
17. Lewis, S. A., J. M. Balcarek, V. Krek, M. L. Shelanski, and N. J. Cowan. 1984. Sequence of a cDNA clone encoding mouse glial fibrillary acidic protein: structural conservation of intermediate filaments. *Proc. Natl. Acad. Sci. USA.* 81:2743–2746.
18. Liem, R. K. H., S. Yen, G. D. Salomon, and M. L. Shelanski. 1978. Intermediate filaments in nervous tissue. *J. Cell Biol.* 79:637–645.
19. Liem, R. K. H., and S. B. Hutchinson. 1982. Purification of the individual components of the neurofilament triplet: filament assembly for the 70,000 dalton subunit. *Biochemistry.* 21:3221–3226.
20. Osborn, M., N. Geisler, G. Shaw, G. Sharp, and K. Weber. 1982. Intermediate filaments. *Cold Spring Harbor Symp. Quant. Biol.* 46:413–423.
21. Pelham, H. R. B., and R. J. Jackson. 1976. An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* 67:247–256.
22. Quax, W., Y. Quax-Jeuken, W. Vree Eaberts, W., R. van den Heuvel, W. Hendriks, and H. Bloemendal. 1984. The genes for vimentin and desmin. *In* The Molecular Biology of the Cytoskeleton. D. W. Cleveland, D. Murphy and G. Borisy, editors. Cold Spring Harbor Pub. In press.
23. Quax-Jeuken, Y. E. F. M., W. J. Quax, and H. Boemendal. 1983. Primary and secondary structure of hamster vimentin predicted from the nucleotide sequence. *Proc. Natl. Acad. Sci. USA.* 80:3548–3552.
24. Reis, R. J. S., and P. A. Biro. 1978. Sequence and evolution of mouse satellite DNA. *J. Mol. Biol.* 1231:357–374.
25. Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling DNA to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237–251.
26. Sanger, F., A. R. Coulsen, B. G. Barrell, A. J. H. Smith, and B. Roe. 1980. Cloning single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161–178.
27. Selkoe, D. J., Y. Ihara, C. Abraham, G. Rasool, and A. McCluskey. 1982. Biochemical and immunocytochemical studies of Alzheimer paired helical filaments. *In* Banbury Report 12, Biological Aspects of Alzheimer's Disease. R. Katzman, editor. CSH Publications.
28. Southern, E. 1975. Detection of specific sequences among fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.
29. Steinert, P. M., R. H. Rice, D. R. Roop, B. L. Trus, and A. C. Steven. 1983. Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments. *Nature (Lond.).* 302:794–800.
30. Tapscott, S. J., G. S. Bennett, and H. Holtzer. 1981. Neuronal precursor cells in the chick neural tube express neurofilament proteins. *Nature (Lond.).* 282:853–855.
31. Young, R. A., and R. W. Davis. 1983. Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA.* 80:1194–1198.
32. Zehner, Z. E., and B. M. Paterson. 1983. Characterization of the chicken vimentin gene: single copy gene producing multiple mRNAs. *Proc. Natl. Acad. Sci. USA.* 80:911–915.