Biochimica et Biophysica Acta 825 (1985) 398–404
Elsevier

BBA 91491

Cloning of a cDNA encoding the smallest neurofilament protein from the rat

Jean-Pierre Julien [a], Kuzhalmannam Ramachandran [b] and Frank Grosveld [a]

[a] *Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA (U.K.) and* [b] *Biogen Research Corporation Inc., 14, Cambridge Centre, Cambridge, MA 02142 (U.S.A.)*

(Received April 18th, 1985)

Key words: Neurofilament protein; Gene structure; Cloning; Nucleic acid sequence

We have cloned a cDNA coding for the smallest rat neurofilament protein. The cDNA is 861 nucleotides long coding for 287 amino acids from the internal α-helical region and the carboxy-terminal tail domain of the neurofilament protein. Comparison of the porcine, mouse and rat neurofilament protein sequences shows that the protein is highly conserved (> 93% identity). Blot analysis indicates that the cDNA is derived from a single neurofilament gene that codes for two different poly(A)$^+$ mRNA species.

Introduction

Each class of intermediate filaments is composed of biochemically distinct proteins which are differentially expressed in different tissues: cytokeratin filaments are found in epithelial cells, vimentin filaments in cells of mesenchymal origin, desmin in muscle cells, glial filaments in glial cells and astrocytes and, finally, neurofilaments in neurons [1,2].

Amino acid sequence comparison shows homologous α-helical regions comprising a so-called rod domain in the middle part of all intermediate filament proteins, while amino- and carboxy-terminal domains are hypervariable in sequence and size [3–13]. In addition, a comparison between keratin genes [14,15] which form a multigene family and the single-copy vimentin gene [10] has shown that the intron positions are highly conserved [14,15], in spite of substantial sequence divergence between these intermediate filament genes.

The structure of other classes of intermediate filament genes is not yet known, although genomic blot analysis with cloned cDNA probes indicates the existence of a single gene for desmin [11] and, at most, two genes for the glial fibrillary acidic protein [13].

Mammalian neurofilaments are composed of three proteins with apparent molecular weight of 68 000 (NF-L), 145 000 (NF-M) and 200 000 (NF-H) on SDS-gel electrophoresis [16–19]. Like other types of intermediate filament proteins, neurofilament proteins share a homologous rod domain, but contain long tail extensions at their carboxy terminals, responsible for their different molecular weights [5,6,7,20]. The α-helical rod domain confers to the proteins the ability to assemble into a filamentous structure, while the tail domain forms a projection at the periphery of the filament. The latter may modulate interactions between neurofilaments and other neuronal components [20,21].

We report here the isolation of a cDNA clone encoding the rat NF-L protein by using synthetic oligonucleotides as probes. The amino acid sequences deduced from the rat cDNA sequence are compared with the sequences of the corresponding porcine subunit [5,22] and other closely related intermediate filament proteins. Lastly, we provide evidence that at least two mRNA species are derived from the single rat NF-L gene.

Results

Isolation of a cDNA clone

Fig. 1 shows two oligonucleotide probes which were synthesized from known amino acid sequences of the procine NF-L protein [5]. Probes I and II are mixtures of nucleotide sequences complementary to the mRNA of peptides located in helix II of the rod domain and the tail region of NF-L protein, respectively. Poly(A)$^+$ RNA was prepared from rat brain and converted into cDNA following the procedure of Gubler and Hoffman [23], with the exception that probe II was used as a primer for reverse transcriptase in the first strand synthesis. The cDNA molecules were tailed with oligo(dC), annealed to dG-tailed pBR322 and the recombinant plasmids were used to transform *Escherichia coli* HB101. Only one of the 40 000 colonies hybridized with both probes and contained a plasmid DNA (p567c) with a 900 bp insert. A restriction map of the cDNA insert is shown in Fig. 2 Neither rescreening of the cDNA library described above nor screening of another library of 25 000 colonies made of cDNA primed with oligo(dT) yielded any cDNA clones extending further in either the 5′ or 3′ direction.

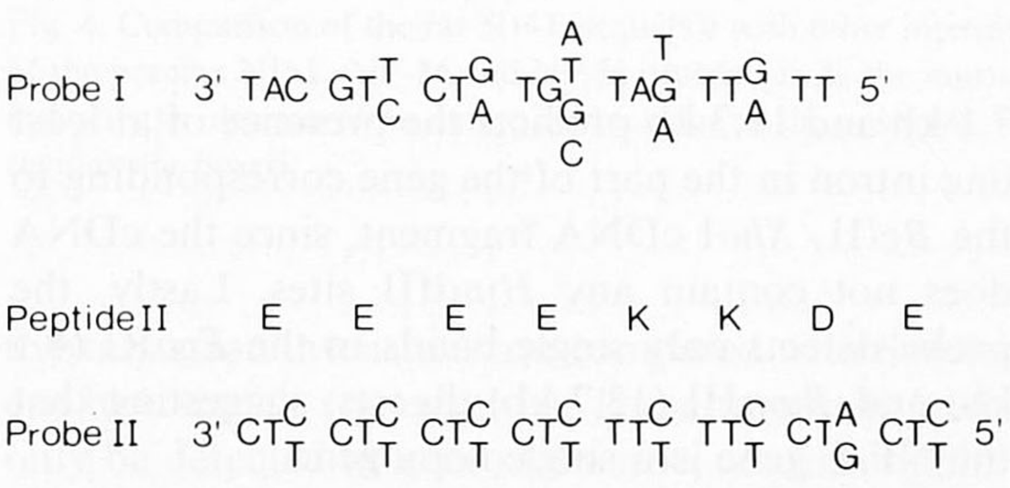

Fig. 1. Synthetic oligonucleotides used as probes in screening the cDNA library. The oligonucleotide mixtures were synthesized according to the amino acid sequences of peptide I in helix Ib and peptide II in the carboxy-terminal domain of the porcine NF-L protein [5]. All possible anti-coding sequences were synthesized as one pool for each of the two peptides. The positions of probe I and probe II in Fig. 3 correspond to nucleotides 451–470 and 838–861, respectively.

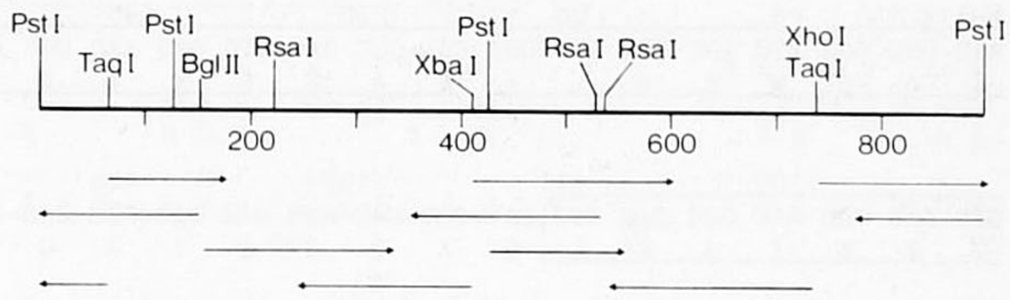

Fig. 2. Restriction cleavage map and sequencing strategy for NF-L cDNA (p567c). The cDNA was sequenced as described by Maxam and Gilbert [24]. The direction and length of each sequence determination is shown by horizontal arrows, at appropriate restriction sites. Only restriction sites used for sequence analysis are shown. The numbers are in basepairs.

Sequence analysis

The 900 nucleotide insert of p567c was sequenced by the method of Maxam and Gilbert [24], according to the strategy shown in Fig. 2. There is only one open reading frame of 861 nucleotides, encoding 287 amino acids, flanked by oligo(dG) and dC generated by the tailing procedure (Fig. 3). Comparison with the known amino acid sequence data shows that the cDNA spans helix Ib and helix II as well as part of the tail domain of NF-L. The boxed regions in Fig. 3 indicate the boundaries of the α-helical domains predicted by the heptade repeats of hydrophobic residues which are typical of intermediate filament proteins [4–13]. Intermediate filament proteins have been shown to share homology predominantly within the coil-rod domain, whereas the amino- and carboxy-terminals vary in size and composition. Fig. 4 shows a protein sequence comparison between the rat NF-L protein and the aligned sequences in the upper lines of the mouse NF-L [22], the porcine NF-L, NF-M and NF-H proteins [5–7], and the two other most closely related intermediate proteins, desmin [25] and vimentin [11]. Throughout the rod domain (residues 1–205) NF-L shares a 59%, 59% and 57% sequence identity with NF-M, desmin and vimentin, respectively. A comparison of NF-L sequences with the NF-H sequences in part of the same area (residues 1–102) reveals 40% identity. As expected, the rat NF-L shows a high homology with the murine NF-L (3% amino acid divergence and 3.5% DNA sequence divergence). Both of these show no homology in the tail domain with the other IF proteins. A comparison between the rat and porcine NF-L sequences which span residues

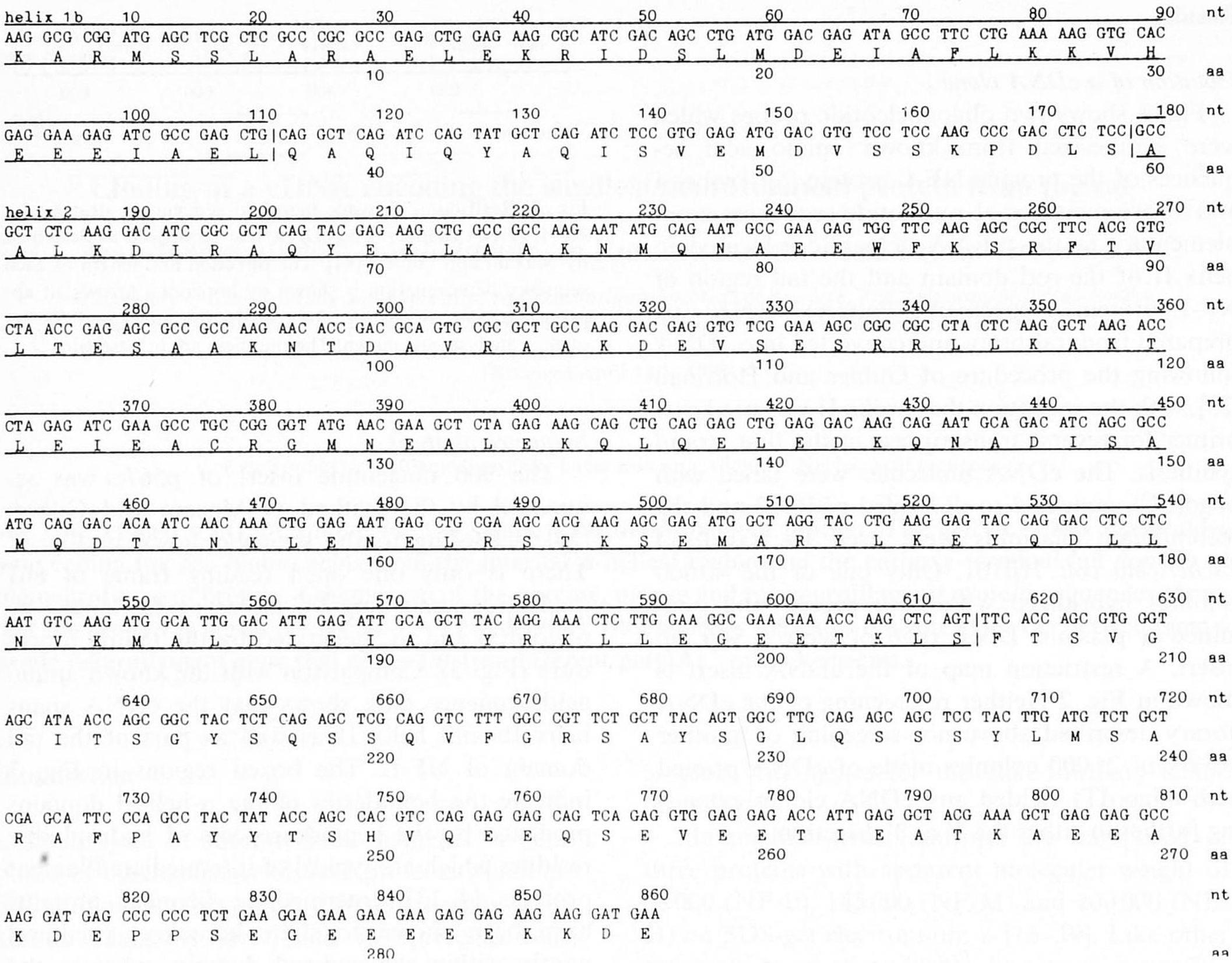

Fig. 3. The nucleotide sequence of NF-L cDNA (p567c). The nucleotide sequence was determined according to Maxam and Gilbert [24]. The predicted amino acid sequence of the open reading frame is shown below the DNA sequence. The helix Ib and helix II regions predicted by the heptade repeats of hydrophobic residues typical of intermediate filament proteins [5] are boxed.

83–287 shows only 7% interspecies divergence, indicating that NF-L is highly conserved in several species.

Multiple mRNA transcripts derived from a single copy gene

Southern blots of rat brain DNA, digested with various restriction enzymes, were hybridized to a ^{32}P-labelled *Bgl*II/*Xho*I fragment (Fig. 2) which contains 585 bp of the central part of the neurofilament cDNA (Fig. 5A). We identified two *Pst*I fragments of 7.9 and 0.3 kb, which would agree with the presence of *Pst*I sites in the coding region (Fig. 2). The presence of two *Hin*dIII fragments of 7.1 kb and 11.3 kb predicts the presence of at least one intron in the part of the gene corresponding to the *Bgl*II/*Xho*I cDNA fragment, since the cDNA does not contain any *Hin*dIII sites. Lastly, the probe detects only single bands in the *Eco*RI (4.5 kb) and *Bam*HI (13.7 kb) digests, suggesting that the NF-L gene is a single copy gene.

Northern blot hybridization to the *Xho*I/*Bgl*II probe (Fig. 5B) of RNA from several different tissues shows that the cDNA detects two different sized, but equally abundant, poly(A)$^+$ mRNA species of 3.4 and 2.4 kb, as well as a minor poly(A)$^-$ RNA species. The abundance of each mRNA is at least 100 copies per cell when the

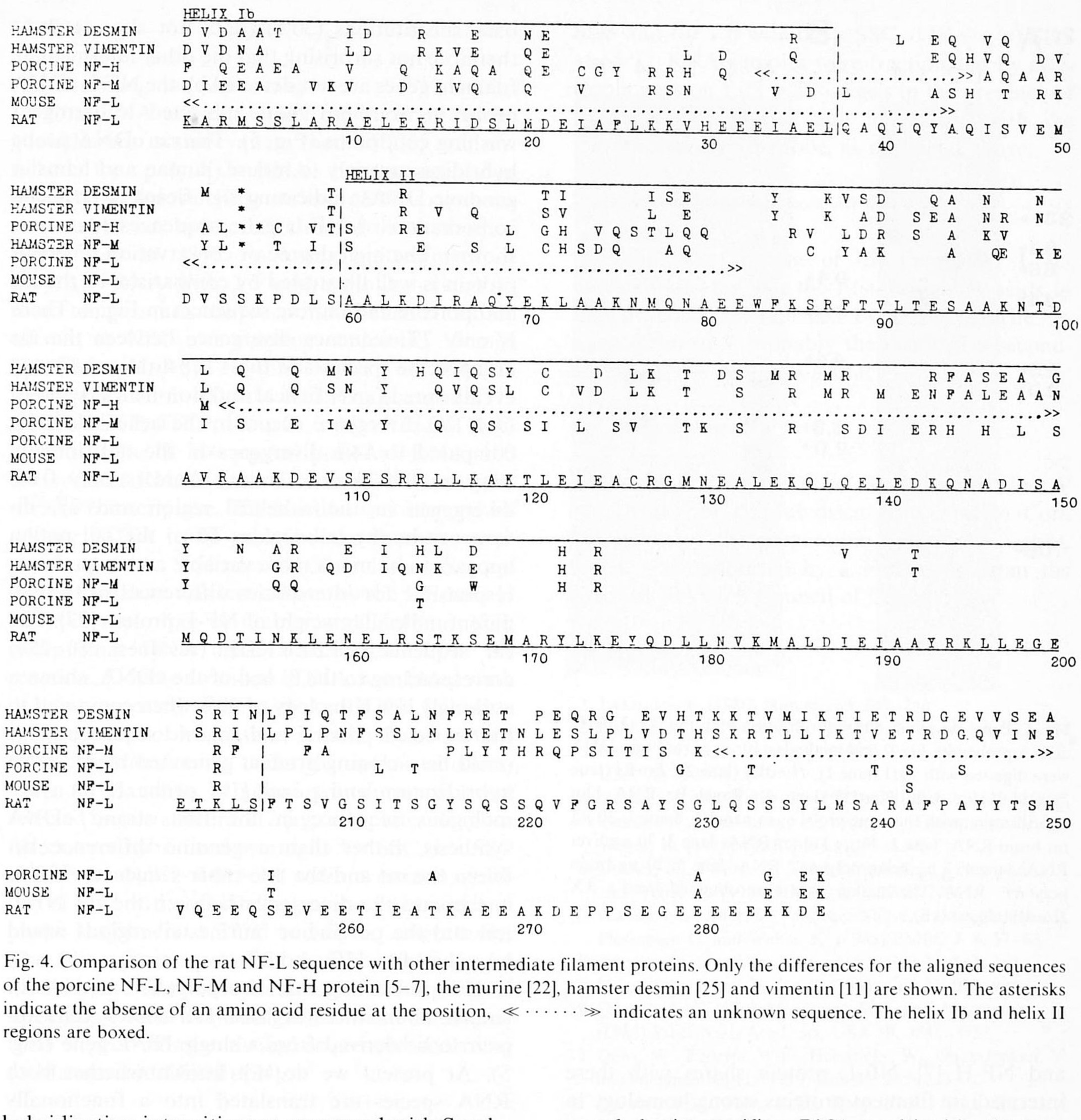

Fig. 4. Comparison of the rat NF-L sequence with other intermediate filament proteins. Only the differences for the aligned sequences of the porcine NF-L, NF-M and NF-H protein [5–7], the murine [22], hamster desmin [25] and vimentin [11] are shown. The asterisks indicate the absence of an amino acid residue at the position, ≪ · · · · · ≫ indicates an unknown sequence. The helix Ib and helix II regions are boxed.

hybridization intensities are compared with Southern blot signals (not shown). These transcripts can only be detected in neuronal tissue, confirming the expected tissue-specific expression of neurofilament genes.

Discussion

In this paper we report the cloning of a cDNA for the rat NF-L protein. The cDNA was found to encode brain-specific mRNAs and its identity was confirmed by comparing the predicted amino acid sequence (Fig. 4) with the known sequence of the corresponding porcine and murine proteins [5,22]. The rat NF-L cDNA clone provides additional amino acid sequence data covering the rod domain of the NF-L protein. As shown in Fig. 4, this domain can be aligned unambiguously with the corresponding domain of desmin [25], vimentin [11] and the other two neurofilament proteins [6]

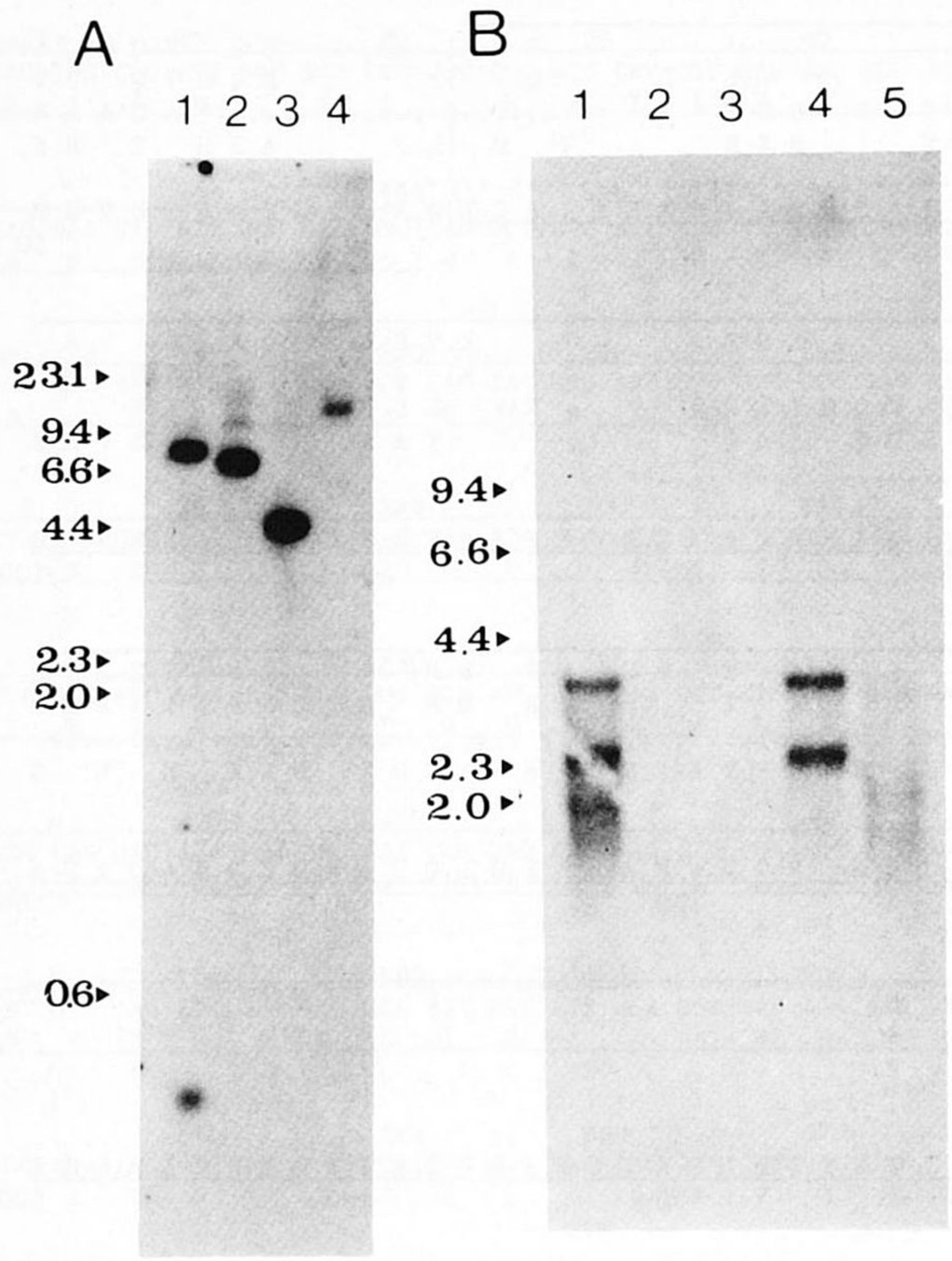

Fig. 5. Panel A: DNA blot hybridization with the *Xho*I-*Bgl*II cDNA probe (see Fig. 2 and methods). 10 μg of rat brain DNA were digested with *Pst*I (lane 1), *Hin*dIII (lane 2), *Eco*RI (lane 3) and 2 μg with *Bam*HI (lane 4). Panel B: RNA blot hybridization with the same probe as in panel A. Lane 1, 30 μg rat brain RNA; lane 2, 30 μg kidney RNA; lane 3, 30 μg liver RNA; lane 4, 3 μg brain poly(A)$^+$ RNA; lane 5, 20 μg brain poly(A)$^-$ RNA. The marker sizes were obtained from a λX *Hin*dIII digest (kb).

and NF-H [7]. NF-L protein shares with these intermediate filament proteins strong homology in helix Ib (residues 1–37) and helix II (residues 60–205), as well as in the interruption stretch (residues 38–59). There is a region (residues 168–201) which corresponds to the most highly conserved consensus sequence of intermediate filament proteins and appears to be recognized by a general intermediate filament antibody [5,26]. However, a comparison of this region at the nucleic acid level of NF-L with desmin [25] and vimentin [11] reveals substantial divergence, mostly by silent base substitutions (30%) (data not shown). It is therefore not surprising that the other intermediate filament genes are not detected by the NF-L cDNA probe in Southern blots subjected to stringent washing conditions (Fig. 5). The rat cDNA probe hybridizes strongly to mouse, human and hamster genomic DNAs, indicating significant interspecies conservation of NF-L gene sequences (data not shown). The high degree of conservation of NF-L protein is well illustrated by comparison of the rat and porcine and murine sequences in Fig. 4. There is only 7% sequence divergence between the rat and porcine species and these substitutions are not evenly spread over helical and non-helical regions. Only 2% divergence occurs in the helical domain, compared to 14% divergence in the tail domain. Between the rat and mouse there is only 0.6% divergence in the α-helical region and 7% divergence in the tail region. Thus, the tail region appears to be much more variable and is, perhaps, responsible for interspecies differences in the apparent molecular weight of NF-L protein [27]. The rat sequence, EEEEKKDE (residues 280–287) corresponding to the 3′ end of the cDNA, shows a strikingly low homology of 50% when compared to the expected porcine and murine sequence. This could be a cloning artefact generated by incorrect hybridization and priming of probe II to a homologous sequence in the first strand cDNA synthesis, rather than a genuine difference between the rat and the two other sequences. If this is the case the divergence between the rat NF-L tail and the porcine or murine tail regions would be reduced to 11% and 2%, respectively.

It is interesting that two poly(A)$^+$ mRNAs of similar abundance (at least 100 copies/cell) appear to be derived from a single NF-L gene (Fig. 5). At present we do not known whether both RNA species are translated into a functionally active protein. The transcripts could be derived from alternative polyadenylation sites at the 3′ end of the gene analogous to the chicken vimentin gene [28,39] or other unrelated genes [30–32]. Multiple transcripts may also be generated from different initiation sites of transcription [33,34], or alternative splicing of the same mRNA precursor [31–33,35–37]. Analysis of the genomic copy of the NF-L gene should enable us to explain the identity of these multiple transcripts.

Methods

cDNA cloning and sequencing

Total RNA was extracted from rat brain by the guanidinium isothiocyanate method [38] and enriched for poly(A)$^+$ RNA by passage over oligo(dT)-cellulose. First strand cDNA synthesis was carried out by modification of the method of Wickens et al. [39]. Following pretreatment of poly(A)$^+$ RNA with 2.5 mM methyl mercury at room temperature, the reaction was carried out for 60 min at 42°C in a solution containing 50 mM Tris-Hcl (pH 8.3), 10 mM $MgCl_2$, 30 mM 2-mercaptoethanol, 100 mM KCl, 1 mM eachdATP, dCTP, dGTP, dTTP, 100 μCi/ml [γ-^{32}P]dATP, 10 μg/ml probe II, 100 μg/ml poly(A)$^+$ RNA, 100 units/ml of RNAase inhibitor and 1000 units/ml reverse transcriptase. The reaction was stopped by adding EDTA to 20 mM and the products were purified on a Sephadex G-50 column. Second strand synthesis was carried out by the RNAase H method, according to Gubler and Hoffman [23], except that DNA ligase was omitted from the reaction. Double-stranded cDNA was tailed with dC and annealed to dG-tailed pBR322, following the procedure of Maniatis et al. [40]. *E. coli* HB101 cells were transformed under the optimized conditions of Hanahan [41] and transformants were selected on tetracycline (12 μg/ml) plates. Replica filters were hybridized overnight 5 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.1% (w/v) SDS with ^{32}P-labelled probe I (14mer) and probe II (24mer) at 37°C and 55°C, respectively. The filters were washed in 6 × SSC, 0.1% (w/v) SDS at 45°C for probe I and at 55°C for probe II screenings. Restriction fragments were labelled at their 5′ protruding ends and sequenced according to Maxam and Gilbert [24].

DNA and RNA blot analysis

DNA extracted from rat brain was digested with various restriction endonucleases, and fractionated on 0.7% agarose gels. The DNA was transferred to nitrocellulose [42] and the blots were hybridized at 65°C overnight with ^{32}P-labelled cDNA [43] in a solution containing 3 × SSC, 10 × Denhart's solution [44], 10% (w/v) Dextran, 50 μg/ml denaturated salmon sperm DNA. The blots were washed for 1 h with 3 × SSC, 0.1% (w/v) SDS and for 1 h with 0.3 × SSC, 0.1% (w/v) SDS at 65°C. RNA samples were fractionated by electrophoresis on 1.0% agarose gels in the presence of formaldehyde, blotted and hybridized with the ^{32}P-labelled cDNA probe, as described above.

Note added in proof (Received June 27th, 1985)

Preliminary analysis of the mouse NF-L sequence indicates that the first six amino acids in Fig. 3 should be replaced by GADEA. The sequence shown is probably the result of a second-strand cDNA synthesis artifact.

Acknowledgements

We are grateful to Dr. M. Kurkinen and Dr. W Mushynski for helpful discussions, and to Cora O'Carroll for the preparation of the manuscript. J.-P.J. was supported by a Fellowship from the Medical Research Council of Canada.

References

1 Lazarides, E. (1980) Nature 283, 249–256
2 Osborn, M. and Weber, K. (1983) Lab. Invest. 48, 372–394
3 Geisler, N. and Weber, K. (1981) Proc. Natl. Acad. Sci. USA 78, 4120–4123
4 Geisler, N. and Weber, K. (1982) EMBO J. 1, 1649–1656
5 Geisler, N., Kaufmann, E., Fisher, S. Plessmann, U. and Weber, K. (1983) EMBO J. 2, 1295–1302
6 Geisler, N., Fisher, S., Vandekerckhove, J., Plessmann, U. and Weber, K. (1984) EMBO J. 3, 2701–2706
7 Geisler, N., Fisher, S., Vandekerckhove, J., Van Damme, J., Plessmann U. and Weber, K. (1985) EMBO J. 4, 57–63
8 Hanukoglu, I. and Fuchs, E. (1982) Cell 31, 243–252
9 Hanukoglu, I. and Fuchs, E. (1983) Cell 33, 915–924
10 Quax-Jenken, Y.E.F.M., Quax, W.J. and Bloemendal, H. (1983) Proc. Natl. Acad. Sci. USA 80, 3548–3552
11 Quax, W., Egberts, W.V., Hendricks, W., Quax-Jenken, Y. and Bloemendal, H. (1983) Cell 35, 215–223
12 Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.C. (1983) Nature 302, 794–800
13 Lewis, S.A., Balcarek, J.M., Krek, V., Shelanski, M. and Cowan, N.J. (1984) Proc. Natl. Acad. Sci. USA 81, 2473-2746
14 Lehnert, M.E., Jorcano, J.L., Zentgraf, H., Blessing, M., Franz, J.K. and Franke, W.W. (1984) EMBO J. 3, 3279–3287
15 Marchuk, D., McCrohon, S. and Fuchs, E. (1984) Cell 39, 491–498
16 Hoffman, P.N. and Lasek, R.J. (1975) J. Cell Biol. 66, 351–366
17 Liem, R.K.H., Yen, S.-H., Salomon, G.D. and Shelanski, M.L. (1978) J. Cell. Biol. 79, 637–645

18 Sharp, G.A., Shaw, G. and Weber, K. (1982) Exp. Cell Res. 137, 403–413
19 Julien, J.-P. and Mushynski, W.E. (1982) J. Biol. Chem. 257, 10467–10470
20 Julien, J.-P. and Mushynski, W.E. (1983) J. Biol. Chem. 258, 4019–4025
21 Hirokawa, N., Glicksman, M.A. and Willard, M.B. (1984) J. Cell Biol. 98, 1523–1536
22 Lewis, S. and Cohen, N. (1985) J. Cell. Biol. 100, 843–850
23 Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269
24 Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
25 Quax, W., Van Den Heuvel, R., Vree Egberts, W., Quax-Jenken, Y. and Bloemendal, H. (1984) Proc. Natl. Acad. Sci. USA 81, 5970–5974
26 Pruss, R.M., Mirsky, R., Raff, M.C., Thorpe, R., Dowding, A.J. and Anderton, B.H. (1981) Cell 27, 419–428
27 Shaw, G., Debus, E. and Weber, K. (1984) Eur. J. Cell Biol. 34, 130–136
28 Dodemont, H.J., Soriano, P., Quax, W.J., Ramaekers, F., Lenstra, J.A., Groenen, M.A., Bernardi, G. and Bloemendal, H. (1982) EMBO J. 1, 167–171
29 Zehner, Z.E. and Paterson, B.M. (1983) Proc. Natl. Acad. Sci. USA 80, 911–915
30 Setzer, D.R., McGrogan, M., Nunber, J.H. and Schimke, R.T. (1980) Cell 22, 361–370
31 Leonard, W.J., Depper, J.M., Crabtree, G.R., Rudikoff, S., Pumphrey, J., Robb, R.J., Kronke, M., Svetlik, P.B., Peffer, N.J., Waldmann, T.A. and Greene, W.C. (1984) Nature 311, 626–631
32 Nikaido, T., Shimizu, A., Ishida, N., Sabe, H., Teshigawara, K., Maeda, M., Uchiyama, T., Yodoi, J. and Honjo, T. (1984) Nature 311, 631–635
33 Young, R.A., Hagenbuchle, O. and Schibler, U. (1981) Cell 23, 451–458
34 Nabeshima, Y., Fukii-Kuriyama, Y., Muramatsu, M. and Ogata, K. (1984) Nature 308, 333–336
35 Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Cell 20, 313–319
36 Crabtree, G.R. and Kaut, J.A. (1982) Cell, 31, 159–166
37 Rosenfeld, M.G. (1983) Nature 304, 129–135
38 Lehrach, H., Diamond, D., Woeney, J.M. and Boadtker, H. (1977) Biochemistry 16, 4743–4747
39 Wickens, M.P., Buell, G.N. and Schimke, R.T. (1978) J. Boil. Chem. 253, 2483–2495
40 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, New York
41 Hanahan, D. (1983) J. Mol. Biol. 166, 557–580
42 Southern, E. (1975) J. Mol. Biol. 98, 503–517
43 Jeffreys, A. and Flavell, R.A. (1977) Cell 12, 429–439
44 Denhardt, D. (1966) Biochem, Biophys. Res. Commun. 23, 641–646