

Intermediate filament cDNAs from BHK-21 cells: Demonstration of distinct genes for desmin and vimentin in all vertebrate classes

(cytoskeleton/protein evolution/ α -helical domain/single-copy gene/tissue-specific expression)

WIM QUAX, RICHARD VAN DEN HEUVEL, WILMA VREE EGBERTS, YVONNE QUAX-JEUKEN,
AND HANS BLOEMENDAL

Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands

Communicated by L. N. M. Duysens, June 25, 1984

ABSTRACT Recombinant cDNA plasmids for the intermediate filament proteins desmin and vimentin were constructed from baby hamster kidney (BHK-21) mRNA. Analysis of four desmin clones gave a sequence of 1574 nucleotides, which is 75% of the total mRNA length. The derived amino acid sequence for hamster desmin shows 92% overall homology with chicken desmin; the homology with hamster vimentin is highest in the α -helical middle part (74%). The 3'-noncoding region of desmin mRNA is found to be 677 nucleotides long. With the aid of 5'- and 3'-specific probes, it has been established that there is a single gene for desmin in the hamster genome. This gene expresses a single mRNA species of 2.2 kilobases. Hybridization experiments of a number of DNAs with desmin and vimentin probes show that there are distinct restriction enzyme fragments carrying vimentin and desmin sequences in the genome of representatives of all vertebrate classes.

The intermediate filaments (IF) along with microfilaments and microtubules are the major constituents of the cytoskeleton of higher eukaryotic cells. Although all IF show a rather similar morphology, their subunits have been divided into five chemically distinct classes, each characteristic of a particular cell type: keratin filaments in epithelial cells, neurofilaments in neurons, glial filaments in astrocytes, vimentin filaments in cells of mesenchymal origin, and desmin filaments in muscle cells (for review, see ref. 1). In the past few years, sequence data have been obtained for some representatives of these classes, and it appeared that long α -helical regions in the middle part of the polypeptide chain are common structures shared by all IF (2-4). On the other hand, the nonhelical COOH and NH₂ parts are identified as the variable structures of IF, because they show large variations in both sequence and size (5-8).

The functional significance for the existence of different IF proteins in different cell types is still unclear. For studies of the molecular basis of this tissue-specific expression, knowledge of the responsible genes is needed. So far, the gene organization of only two IF classes has been studied with the aid of cDNA probes: the epidermal keratin genes, which form a multigene family (9, 10), and the single-copy vimentin gene, whose structure has been completely elucidated (11).

To obtain further insight into the organization, evolution, and expression of the IF genes, it is desirable to have cDNA probes for all five classes. We present here the cloning and characterization of desmin cDNA and the detection of the corresponding sequences in the genome of all vertebrate classes.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Construction of Recombinant Plasmids. Poly(A)⁺ RNA was isolated from baby hamster kidney cells (BHK-21) as described (12). Double-stranded cDNAs were synthesized on this RNA by a one-step procedure adopted from Wickens *et al.* (13) as described (12). The only change that we introduced for the construction of the BHK cDNA was the consecutive use of both reverse transcriptase and *Escherichia coli* DNA polymerase I (nuclease-free) for the second strand synthesis. cDNAs were size-selected on 3.5% polyacrylamide gels, and the fraction above 700 base pairs was electroeluted, C-tailed, and annealed into *Pst*-cleaved G-tailed pBR327. *E. coli* HB101 cells were transformed by the recombinant plasmids and plated on HATF filters as described (14). Replica filters were hybridized with the *Sau*3A fragment of pVim-1 that covers amino acids 240-453 (4) under the conditions described below, except that 40% formamide was used and the last wash with 2 × NaCl/Cit/0.5% NaDodSO₄ was at 50°C (1 × NaCl/Cit = 0.15 M NaCl/0.015 M Na citrate). Isolation of plasmid DNA from positive clones was done using standard methods (15).

Sequence Analysis and Blot Analysis. For sequence analysis, suitable fragments were ligated into M13 mp8 and mp9 vectors, grown in JM103 (16), and subjected to dideoxy reactions as described (17). RNA blots were prepared and hybridized essentially as described by Thomas (18). Total genomic DNA was isolated from diverse tissues of all species described using standard methods (19). Ten-microgram amounts were electrophoresed on 0.7% agarose gels, blotted, and hybridized under the following conditions: 50% formamide/5 × NaCl/Cit/20 mM sodium phosphate, pH 6.8/5 mM EDTA/0.06% bovine serum albumin/0.06% Ficoll/0.06% polyvinylpyrrolidone/100 μ g of herring sperm single-stranded DNA per ml/0.1% NaDodSO₄ at 42°C. Washings were done with hybridization mix and, finally, with NaCl/Cit and NaDodSO₄. The stringency of the last washing is described in the figure legends. M13 probes were prepared by carrying out a standard T-reaction of dideoxy sequencing with ATP of >3000 Ci/mmol (1 Ci = 37 GBq) (17, 20).

RESULTS

Isolation of Desmin cDNA Clones. We constructed a cDNA library on poly(A)⁺ RNA isolated from BHK-21 cells. These cells, which express both vimentin and desmin (21, 22), are derived from a Syrian gold hamster. Lens mRNA from the same species has been used for the construction of vimentin clones (12). About 5000 colonies were plated and screened under conditions of low stringency with a ³²P-labeled probe derived from pVim-1 (4). This probe covers amino acids 240-453 of vimentin. Within that region, there is a sequence

Abbreviations: IF, intermediate filaments; kb, kilobase(s); GFA, glial filament acidic protein; NaCl/Cit, 0.15 M NaCl/0.015 M Na citrate.

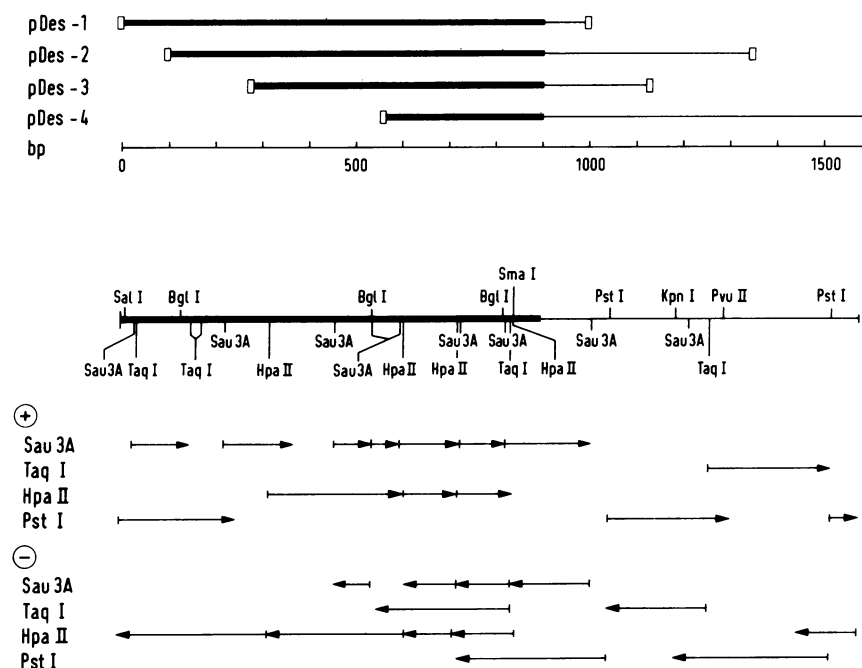


FIG. 1. Restriction map and sequencing strategy for pDes-1, pDes-2, pDes-3, and pDes-4. Upper lines indicate the length and relative positions of the inserts; thick bars represent coding region, and thin bars are the 3'-noncoding parts. Open bars at the ends represent the G-C tails that were introduced during the cloning procedure. Arrows indicate the direction and extent of the dideoxy sequence as it was determined after subcloning of the corresponding fragments in mp8 or mp9.

of 35 amino acids that is nearly identical to the corresponding stretch in desmin (4, 23). Thirteen more or less positive clones were identified, 11 of which showed a restriction map that agreed with the map of pVim-1 (4). The other two plasmids had a different restriction pattern. Partial sequence analysis showed that within these clones there is a region that encodes a sequence that is identical to the partial porcine desmin sequence (23). Rescreening of the library with the insert of these two clones yielded two additional desmin clones. The restriction map and relative position of the four desmin cDNAs are shown in Fig. 1.

The Sequence of Desmin cDNA. Suitable restriction fragments were isolated from all four recombinant plasmids (Fig. 1), and their nucleotide sequence was determined by the M13 dideoxy method (16, 17). For pDes-1 and pDes-4, both strands were completely determined; the sequences of pDes-2 and pDes-3 are completely located within pDes-1 and pDes-4. No differences between the corresponding parts of the four desmin clones could be found. The combined sequence shown in Fig. 2 covers 1599 nucleotides. There is one open reading frame of 894 bases, which encodes 298 amino acids. The TAA stop codon is followed by a noncoding region of 677 nucleotides. At the 3' end, a stretch of 25 adenines is found, which is preceded by the consensus A-A-T-A-A-A polyadenylation signal (24). This led us to conclude that in pDes-4 the complete 3' nontranslated region is present. The rather unusual length of this noncoding region is consistent with the estimated size of 2200 bases for desmin mRNA (see below). This implies that the cDNAs represent $\approx 75\%$ of the total length of the mRNA. From a comparison with the known sequence of chicken desmin (25), we estimate that the information for 160 amino acids is lacking at the 5' end of pDes-1.

The Structure of Hamster Desmin Compared with Chicken Desmin, Vimentin, and Glial Filament Acidic Protein (GFA). It has been shown that intermediate filaments share homology predominantly in the helical middle part of the protein. Therefore, these regions are called the constant domains and the NH₂ and COOH ends are considered as the variable domains. For the regions that span amino acids 1-67 and

114-244 of the sequence in Fig. 2, α -helical conformation is predicted. This was done by the application of the heptade convention, which has been used for secondary structure predictions of other IF proteins (2, 4). The open circles in Fig. 3 indicate the "a" and "d" positions of the heptade; 82% of these residues are hydrophobic. To test the variability in the COOH domain, we aligned the sequence of hamster desmin to chicken desmin (25) and to the two most closely related other IF subunits, vimentin (4) and GFA (8). In Fig. 3, the hamster desmin sequence is shown at the bottom line; from the aligned sequences in the upper lines, only the deviating residues are drawn. The arrowhead marks the border between the helical region and the nonhelical carboxyl terminus. The homology of hamster desmin with the other IF subunits is highest in the helical part; in this region, the homology with vimentin is 74%; with the partial GFA sequence, it is 73%. The 52 COOH-terminal residues show much less homology: 44% with vimentin and 42% with GFA. Between hamster desmin and chicken desmin, there is only 8% sequence divergence. These differences are, however, not predominantly found in the COOH-terminal piece, but they are evenly spread over helical and nonhelical regions, indicating that all domains are equally well conserved comparing one type of IF subunit in different species.

Size and Specific Expression of Desmin mRNA. The size of desmin mRNA was determined by RNA blot analysis. Poly(A)⁺ RNA from BHK-21 cells and from hamster lens cells was glyoxylated, electrophoresed, and transferred to a nitrocellulose filter (18). Two identical blots were hybridized to pDes-1 and pVim-1, respectively (Fig. 4A). Obviously, desmin mRNA is only expressed in BHK-21 cells, in contrast to vimentin mRNA, which is detected both in BHK-21 and in lens cells. This is in concert with the observation that lens cells contain only IF of the vimentin type (26) and that IF of BHK-21 cells comprise vimentin and desmin (21, 22). This latter finding has led to the suggestion that BHK-21 cells are probably derived from an embryonal vascular smooth muscle cell (27).

The detection of a single desmin mRNA class shows that no alternative polyadenylation signals are used for tran-

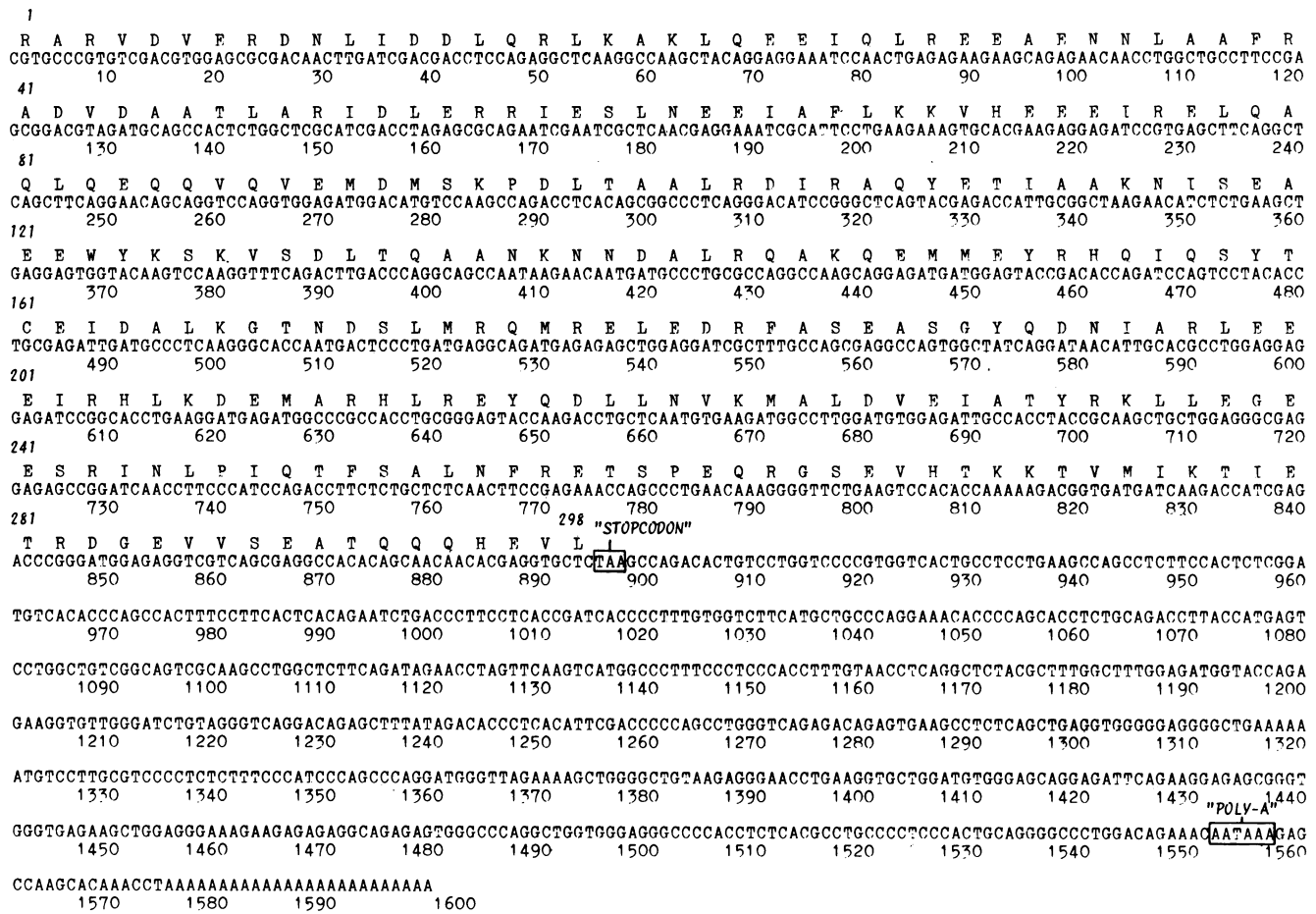


Fig. 2. Combined nucleotide sequence of pDes-1, pDes-2, pDes-3, and pDes-4. The predicted amino acids (represented by standard one-letter abbreviations) are shown above the middle of their triplets. The polyadenylation signal and the stop codon are indicated.

scription, as in the case of the chicken vimentin gene (28, 29). With a length of ≈ 2.2 kb, desmin mRNA is 0.3 kb longer than vimentin mRNA. This difference is consistent with the longer 3'-nontranslated region of the desmin cDNA (677 bases; Fig. 2) as compared to the vimentin cDNA (320 bases; see ref. 11).

Evidence for a Single-Copy Desmin Gene. To test genomic hamster DNA for the presence of desmin sequences, we used as hybridization probe M13 recombinant phages, which

were constructed for sequencing the cDNA. Two identical Southern blots of hamster DNA digested with two restriction enzymes were hybridized to a 5'-specific and a 3'-specific probe, respectively. The 5' probe (covering nucleotides 315-609 of the cDNA, as presented in Fig. 2) hybridizes exclusively to a 7.4-kb *EcoRI* fragment and to a 3.4-kb *BamHI/EcoRI* fragment (Fig. 4B). The 3' probe (covering nucleotides 1061-1257) hybridizes exclusively to another *EcoRI* fragment, 8.2 kb long, and to a 5.4-kb *BamHI/EcoRI*

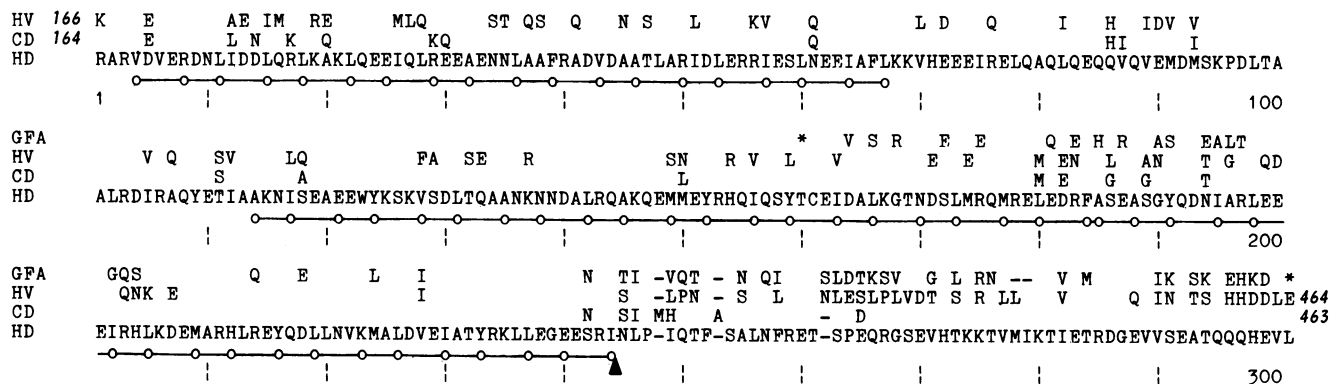


FIG. 3. Comparison of hamster desmin with chicken desmin (25), hamster vimentin (4), and the partial GFA sequence (8). The residues of the aligned sequences are only printed when they differ from hamster desmin. Hamster desmin residues are numbered on the bottom line; the other sequences are numbered at the beginning and end. Borders of the partial GFA sequence are marked by asterisks. The information that corresponds to the first 164 residues of chicken desmin is not present in our clones. The line under the sequences represents the α -helical regions; open circles in this line represent the "a" and "d" positions of the heptade convention. Arrow marks the border between the helical core part and the nonhelical COOH region.

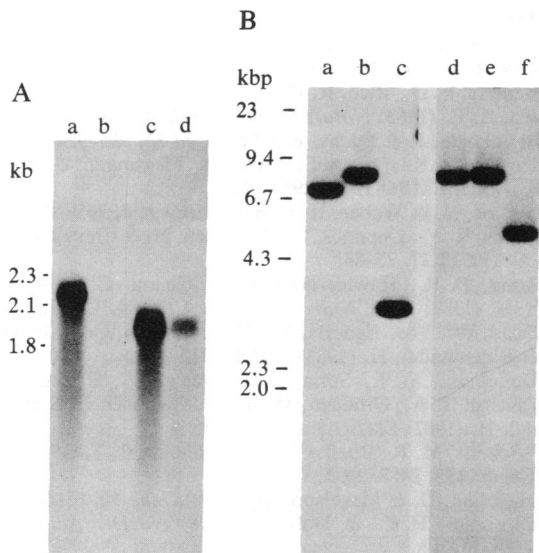


FIG. 4. (A) Size and expression of desmin and vimentin mRNA. RNA blots of 10 μ g of BHK-21 poly(A)⁺ RNA (lanes a and c) and 3 μ g of lens poly(A)⁺ RNA (lanes b and d) were hybridized with nick-translated pDes-1 (lanes a and b) and pVim-1 (lanes c and d). Desmin mRNA is only detected in BHK cells; vimentin mRNA is in both cell types. Ribosomal 18S and *Pvu* II/*Eco*RI-digested pBR322 were run on parallel lanes as markers. (B) Desmin gene number in hamster DNA. Ten-microgram amounts of total hamster DNA were digested with *Eco*RI (lanes a and d), *Bam*HI (lanes b and e), and *Eco*RI/*Bam*HI (lanes c and f), run on 0.7% agarose gels, and blotted on nitrocellulose filters. Lanes a, b, and c were hybridized with a 5'-specific desmin probe; lanes d, e, and f were hybridized with a 3'-specific probe. Final washing was done with 0.1 \times NaCl/Cit/0.1% NaDodSO₄ at 63°C for 30 min.

fragment. An 8.5-kb *Bam*HI band is detected by both probes. This implies that there is a single desmin gene that resides (at least for the major part) on the 8.5-kb *Bam*HI fragment; within this gene, there is an *Eco*RI cut that separates the 5' and 3' parts of the gene. Since the cDNA does not carry an *Eco*RI site, this cut must lie in an intronic region.

Detection of Desmin and Vimentin Sequences in the DNA of Vertebrates. The use of cloned cDNA probes allows us to test different vertebrate genomes for the presence of sequences homologous to desmin and vimentin. *Eco*RI-digested DNAs from representatives of all vertebrate classes were hybridized with the inserts of pDes-1 (Fig. 5A) and pVim-1 (Fig. 5B). Representatives of all vertebrate classes, including fish, amphibia, reptiles, birds, and mammals, all showed clear bands under the medium stringent conditions used. Although as a result of cross-hybridization in some DNAs the desmin and the vimentin probe seem to recognize a common band under these conditions, it is clear that in all species pDes-1 and pVim-1 show the strongest hybridization with different fragments. This means that there are different genes that code for desmin and vimentin in all vertebrates. A precise estimation of the number of genes in all these species cannot be made from this experiment. However, the presence of one or two prominent bands in all lanes suggests that the single-copy nature of the vimentin and desmin gene, as it is found in the hamster genome, holds true for all species. In the same experiment, we also tested DNA from *Drosophila* and yeast from hybridization with pDes-1 and pVim-1, but no signal could be detected under these conditions. Comparison of Fig. 5A (lane a) (hamster DNA hybridized with pDes-1) with Fig. 4B (lane a) shows that under conditions of lower stringent washings more bands are detected in addition to the already mentioned 7.4-kb *Eco*RI fragment. The disappearance of the weaker bands upon more stringent washing was also observed with the other mammalian DNAs with both the

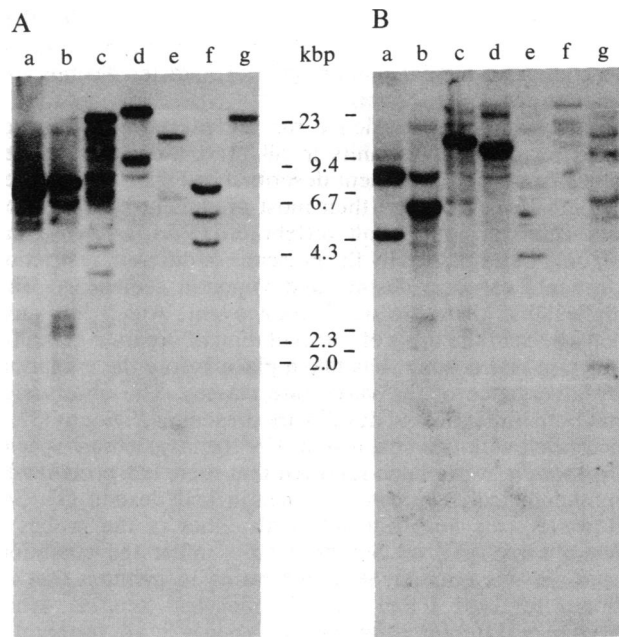


FIG. 5. Detection of desmin and vimentin sequences in vertebrate DNAs. Ten-microgram amounts of DNA were digested with *Eco*RI, run on 0.7% agarose gels, blotted onto nitrocellulose, and hybridized with the nick-translated insert of pDes-1 (A) and pVim-1 (B). The DNAs were extracted from the following species: lane a, Syrian gold hamster; lane b, mouse; lane c, man; lane d, chicken; lane e, Varan (lizard); lane f, *Xenopus laevis*; and lane g, *Tilapia mossambica*, a cichlid teleost fish. Hybridization was under conditions described and final washing was at 55°C with 1 \times NaCl/Cit/0.2% NaDodSO₄ for 30 min.

desmin and the vimentin probe (data not shown). Candidates for these additional bands are other IF genes and/or desmin pseudogenes. The observation that most weak bands in the desmin lane do not show a counterpart in the vimentin lane is an argument against the first possibility, because one would expect the same comigrating weak bands with both probes as they share the same degree of homology with other IF genes (7, 8). Therefore, the possibility of pseudogenes cannot be excluded.

DISCUSSION

The amino acid sequence derived from the nucleotide sequences of the four desmin cDNA clones described allows the prediction of a secondary structure for hamster desmin similar to other IF proteins (2-5). Comparison of the structure of hamster desmin with both chicken desmin and other closely related IF proteins leads to some interesting conclusions: (i) the interspecies differences (at least between mammals and birds) for one and the same IF protein are much less than the differences with other IF proteins, and (ii) the COOH-terminal domain that has been recognized to be variable upon comparison of different IF protein sequences (2, 3, 5, 7, 8) is not variable when we consider corresponding IF proteins in different species.

On the basis of paleontological and biochemical data, it is assumed that the evolutionary divergence of the ancestors of chicken and hamster took place some 270 million yr ago (30, 31). Supposing a constant evolution rate for IF sequences in the helical domains, we can estimate from the differences between hamster desmin and chicken desmin (8%) on the one hand and hamster desmin and vimentin (26%) on the other hand that the divergence of the precursor desmin and vimentin genes took place before the origin of the earliest vertebrates (500-600 million yr ago; see ref. 30). Another

indication for this statement stems from the observation that there is no homology in the nucleotides at silent positions of the codons of similar amino acids of hamster desmin and vimentin (data not shown).

More convincing evidence for the presence of distinct genes for both IF subunits in all present-day vertebrates comes from the experiment described in Fig. 5. The probes for both IF genes show their most prominent hybridization with different bands in all vertebrate DNAs, a finding that can only be explained by the presence of different restriction fragments carrying desmin and vimentin sequences. This implies that the gene duplication event, which has most likely been at the basis of the evolution of separate genes for vimentin and desmin, has taken place before the evolutionary divergence of the vertebrate classes. The observation that both vimentin and desmin are present in *Xenopus* (32) is in concert with this conclusion. For the invertebrate species *Drosophila*, it has been reported that there is a protein with immunological relations to vimentin and desmin (33, 34). However, we were not able with either of the probes to detect any signal on Southern blots under the conditions used. In this context, it is interesting to mention that the clones for type I and type II epidermal keratins, which recognize different restriction fragments in all vertebrates, also do not hybridize with invertebrate DNA under conditions comparable to those we used (35). Screening of a gene library of invertebrate species under low stringent conditions and characterization of homologous sequences could shed more light on the early evolution of IF genes.

The availability of IF cDNA clones is a prerequisite for the study at the gene level of the tissue-specific expression of IF proteins. Desmin cDNA clones in combination with vimentin cDNAs open the possibility of investigating RNA expression of IF during embryogenesis and, in particular, during myogenesis. An important conclusion from the results presented here is that for desmin there is only a single gene, as has been evidenced recently also for vimentin (11). Both IF genes, which are strongly related in sequence, show a totally different expression pattern: the desmin gene is active in muscle tissue, while the vimentin gene is predominantly expressed in cells of mesenchymal origin. A comparison of structure and *in vitro* expression of the cloned genes for these IF can help to discover which regulatory elements are responsible for this tissue specificity.

We thank Dr. W. de Jong for critically reading the manuscript and Mrs. S. Bruisten and Mr. L. van den Broek for help with the isolation of some DNAs. We also acknowledge the assistance of Messrs. Ger Austen and Jack Leunissen with computer analysis. The BHK-21 cell line was kindly supplied by Dr. A. J. van der Eb. The present investigations have been carried out in part under the auspices of the Netherlands Foundation for Chemical Research (SON) and with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

1. Lazarides, E. (1982) *Annu. Rev. Biochem.* **51**, 219-250.
2. Geisler, N., Kaufmann, E. & Weber, K. (1982) *Cell* **30**, 277-286.

3. Hanukoglu, I. & Fuchs, E. (1982) *Cell* **31**, 243-252.
4. Quax-Jeuken, Y., Quax, W. & Bloemendal, H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3548-3552.
5. Steinert, P. D., Rice, R. H., Roop, D. B., Trus, B. L. & Steven, A. C. (1983) *Nature (London)* **302**, 794-800.
6. Hanukoglu, I. & Fuchs, E. (1983) *Cell* **33**, 915-924.
7. Geisler, N., Kaufmann, E., Fisher, S., Plessmann, U. & Weber, K. (1983) *EMBO J.* **2**, 1295-1302.
8. Geisler, N. & Weber, K. (1983) *EMBO J.* **2**, 2059-2063.
9. Fuchs, E. V., Coppock, S. M., Green, H. & Cleveland, D. W. (1981) *Cell* **27**, 75-84.
10. Roop, D. R., Hawley-Nelson, P., Cheng, C. K. & Yuspa, S. H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 716-720.
11. Quax, W., Vree Egberts, W., Hendriks, W., Quax-Jeuken, Y. & Bloemendal, H. (1983) *Cell* **35**, 215-223.
12. Dodemont, H. J., Soriano, P., Quax, W., Ramaekers, F., Lenstra, J. A., Groenen, M. A. M., Bernardi, G. & Bloemendal, H. (1982) *EMBO J.* **1**, 167-171.
13. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253**, 2483-2495.
14. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63-67.
15. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513-1523.
16. Messing, J., Crea, R. & Seeburg, P. (1981) *Nucleic Acids Res.* **9**, 309-321.
17. Sanger, F., Coulson, A. R., Barrel, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161-178.
18. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201-5205.
19. van der Putten, H., Terwindt, E., Berns, A. & Jaenisch, R. (1979) *Cell* **18**, 109-116.
20. Hu, N.-t. & Messing, J. (1982) *Gene* **17**, 271-277.
21. Tuszynski, G. P., Frank, E. D., Damsky, C. H., Buck, C. A. & Warren, L. (1979) *J. Biol. Chem.* **254**, 6138-6143.
22. Gard, D. L., Bell, P. B. & Lazarides, E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3894-3898.
23. Geisler, N. & Weber, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4120-4123.
24. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211-214.
25. Geisler, N. & Weber, K. (1982) *EMBO J.* **1**, 1649-1656.
26. Ramaekers, F. C. S., Osborn, M., Schmid, E., Weber, K., Bloemendal, H. & Franke, W. W. (1980) *Exp. Cell Res.* **127**, 309-327.
27. Quinlan, R. A. & Franke, W. W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3452-3456.
28. Zehner, Z. E. & Paterson, B. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 911-915.
29. Capetanaki, Y. G., Ngai, J., Flytzanis, C. N. & Lazarides, E. (1983) *Cell* **35**, 411-420.
30. Romer, A. S. (1966) *Vertebrate Paleontology* (Univ. of Chicago Press, Chicago), 3rd Ed.
31. Wilson, A. C., Carlson, S. S. & White, T. J. (1977) *Annu. Rev. Biochem.* **46**, 573-639.
32. Franz, J., Gall, L., Williams, M. A., Picheral, B. & Franke, W. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6254-6258.
33. Falkner, F.-G., Saumweber, H. & Biessmann, H. (1981) *J. Cell Biol.* **91**, 175-183.
34. Walter, M. F. & Biessmann, H. (1984) *Mol. Cell. Biochem.* **60**, 99-108.
35. Fuchs, E. & Marchuk, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5857-5861.