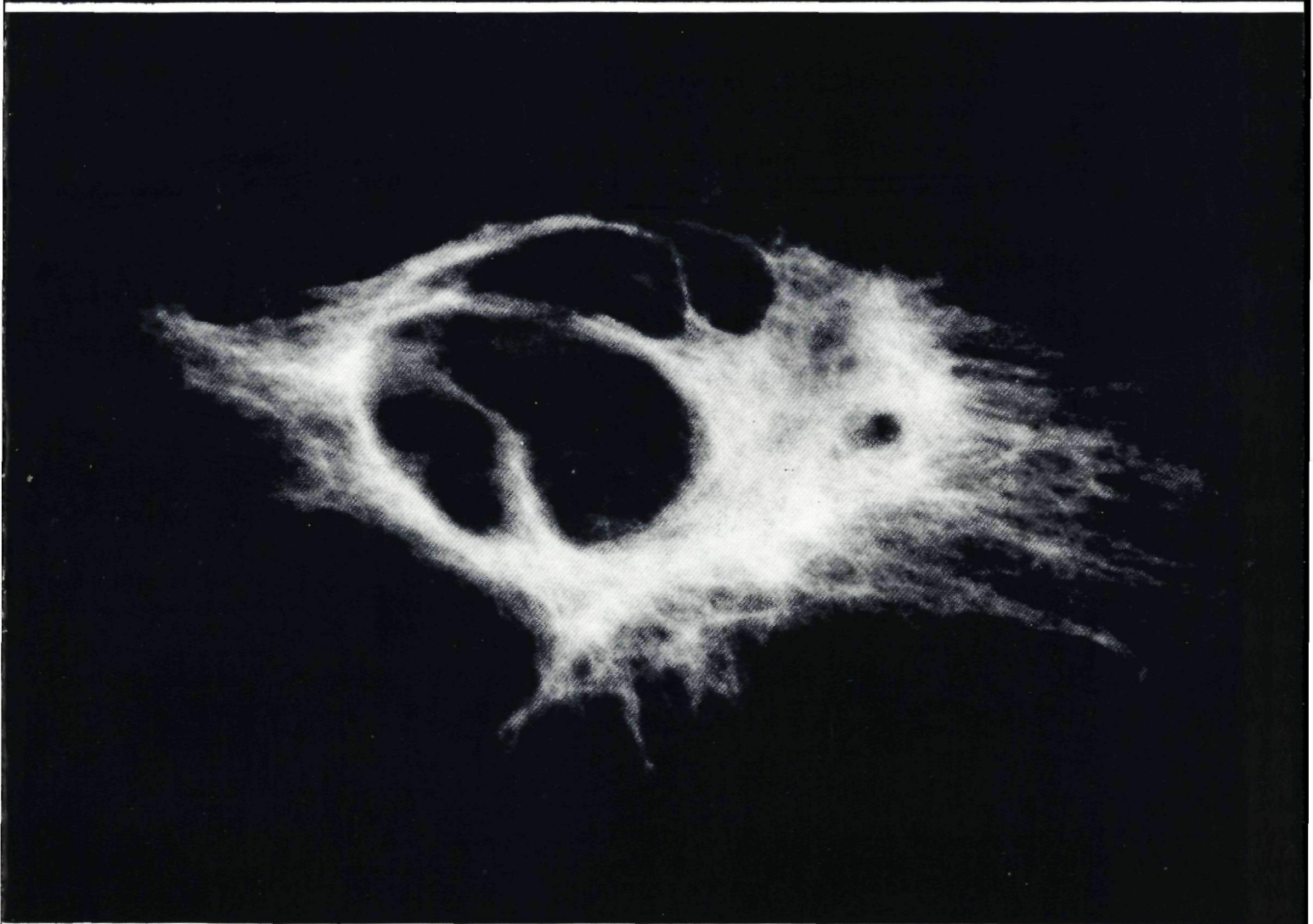
VIMENTIN AND DESMIN: FROM GENE TO FILAMENT



Wim Quax

VIMENTIN AND DESMIN:
FROM GENE TO FILAMENT

VIMENTIN AND DESMIN:
FROM GENE TO FILAMENT

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR
IN DE WISKUNDE EN NATUURWETENSCHAPPEN
AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN
OP GEZAG VAN DE RECTOR MAGNIFICUS
PROF. DR. J.H.G.I. GIESBERS
VOLGENS BESLUIT VAN HET COLLEGE VAN DEKANEN
IN HET OPENBAAR TE VERDEDIGEN
OP WOENSDAG 19 JUNI 1985
DES NAMIDDAGS TE 4.00 UUR

DOOR

WILHELMUS JOHANNES QUAX

GEBOREN TE GELEEN

NIJMEGEN, 1985

PROMOTOR: PROF. DR. H. BLOEMENDAL

aan mijn ouders

DANKWOORD

Het voltooien van een proefschrift is een goede aanleiding eens terug te blikken op de totstandkoming ervan. Op deze plaats wil ik dan ook al die velen bedanken, die een bijdrage hebben geleverd aan het verzamelen en verwerken van de resultaten zoals die in de navolgende hoofdstukken beschreven staan. Allereerst is dat de gehele afdeling Biochemie WNF, die al die jaren het vruchtbare klimaat voor het verrichten van moleculair biologisch onderzoek vormde. Mijn bijzondere dank wil ik graag uitspreken aan Wilma Vree Egberts, die met haar enthousiaste en opgewekte aanpak een onschatbare bijdrage heeft geleverd, niet alleen aan het opzetten en uitvoeren van de beschreven experimenten, maar ook aan de werksfeer binnen onze groep. Dank gaat ook uit naar Richard van den Heuvel, Wiljan Hendriks, Leon van den Broek, Martin Pelgrom en Zjak van Eupen, wier bijdragen in het kader van hun doctoraal-stages gedeeltelijk in het proefschrift verwerkt zijn. De discussies en samenwerking met Huub Dodemont, Hans Lenstra, Frans Ramaekers en Marielle van Workum tijdens de beginfase van mijn promotie-onderzoek hebben mij belangrijk geholpen om in het cytoskelet-onderzoek thuis te geraken. Ook wil ik op deze plaats de prettige samenwerking met Dr. Meera Khan noemen, welke de chromosoom localisatie experimenten mogelijk gemaakt heeft. De afdeling Fotografie van de faculteit WNF dank ik voor het fotografisch vastleggen van de experimentele gegevens. Erkentelijkheid ben ik verder verschuldigd aan de medewerkers van de afdeling IVV en aan Jack Leunissen binnen de afdeling Biochemie voor het installeren en vooral het draaiend houden van de computerprogramma's, die een flinke tijdsbesparing bij het analyseren van de DNA-volgordes en bij het verwerken van de tekst voor dit proefschrift hebben opgeleverd. Tot slot wil ik jou, Yvonne, danken voor je samenwerking en steun, zowel op het lab als thuis.

Het in dit proefschrift beschreven onderzoek is uitgevoerd onder auspiciën van de Stichting Scheikundig Onderzoek in Nederland (SON) met financiële steun van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO).

TABLE OF CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

Molecular Biology of the Intermediate Filaments: an Overview.

The Discovery of Intermediate Filaments

The detailed view of the electron microscope has enabled the identification of a number of filamentous structures in the eucaryotic cell. Differences in general morphology and diameter in addition to the susceptibility to certain drugs have been used to classify the observed cytoskeletal filaments. Actin filaments (6 nm in diameter) and microtubules (25 nm) in nonmuscle cells and myosin filaments (15 nm) in muscle cells were readily characterized. However, the filaments with a diameter of 9-10 nm (intermediate between the diameter of actin-containing microfilaments and those of myosin containing thick filaments and microtubules) were, in the beginning, not regarded as a different class of fibers.

A first step towards the appreciation of these filaments as a different type, was the observation, that intermediate filaments (or 10 nm filaments) do not stain with heavy meromyosin (in contrast to actin-filaments) and that they are persistent to a treatment with colchicine (Ishikawa et al.,1968). This drug, which dissociates microtubules into its subunits caused the intermediate filaments to collapse into a perinuclear organization, but did not desintegrate the individual filaments. From these observations it was anticipated that intermediate filaments (IF) form a distinct cytoskeletal structure, which interacts with microtubules (Ishikawa et al.,1968). Despite the clear distinction to other cytoskeletal fibers progress in the biochemical characterization of IF was slow, mainly because of the insolubility of the filaments and the resulting difficulties in the isolation of their subunits (Shelanski et al.,1971; Eng et al.,1971; Skerrow et al.,1973; Hoffman & Lasek,1975). Moreover, it was puzzling that number and sizes of subunits isolated from diverse sources as nerve cells, smooth muscle cells, epithelia, and astrocytes were very different, although the 10 nm filaments from different tissues showed a similar morphology in electron microscopy studies (Hoffman & Lasek,1975; Steinert & Idler,1975; Small and Sobieszek,1977). The full understanding of the diversity and wide distribution of 10 nm filaments was only achieved when suitable antibodies directed to the different subunits were used in immunofluorescence studies (Bignami et al.,1972; Lazarides & Huberd,1976; Liem et al.,1978; Hynes & Destree,1978; Franke et al.,1978a,b; Sun & Green,1978). In the immunofluorescence microscope bright filamentous structures were observed after treatment of permeabilized cells with antibodies directed against IF proteins. The localization of these fluorescent fibers and the collapse into a perinuclear cap upon treatment with colchicine were the main arguments to link the brightly fluorescent network with the electron microscopic observations of 10 nm filaments (Osborn et al.,1977; Hynes & Destree,1978). With the aid of different antibodies it was now rapidly understood that intermediate filaments showed tissue-specificity: for instance antibodies raised against IF from epithelial cells were able to decorate filaments in all types of epithelial cells, but not in muscle cells or nerve cells (Franke et al,1978a; Lazarides & Hubbard,1976; Bennet et al.,1978).

During the last 5 years enormous progress has been made in the inventarization and characterization of different IF subunits thanks to the application of new techniques as high resolution two-dimensional gel-electrophoresis, production of monoclonal antibodies, immunoblotting and recently DNA cloning.

The diversity of IF subunits

Isolation procedures for IF-subunits merely depended on a common chemical property of 10 nm filaments, namely their insolubility in buffers containing Triton X-100 (1%) and high or low salt concentrations (for references see Franke et al.,1982). Solubilization of IF-proteins is only achieved under denaturing conditions, e.g. 6 M urea. Surprisingly it was found that after removal of urea, purified proteins reassembled to form authentic 10 nm filaments in vitro (Steinert et al.,1976). This shows that the polymerization of IF-subunits to filaments is a reversible process that takes place spontaneously without the need of energy or accessory-proteins. By carrying out subsequent disassembly/reassociation steps some people used this phenomenon to develop a procedure to purify IF proteins (for references see Steinert et al.,1982). On the other hand SDS-polyacrylamide gel electrophoresis of the Triton X-100 insoluble fraction was in most cases sufficient to obtain a pure antigen for the production of antibodies. Antisera obtained in this way have been used to classify the different IF proteins. In the past few

CELL TYPE	SUBUNITS	#	MOLECULAR WEIGHT	SEQUENCE	REFERENCE
MUSCLE	DESMIN	1	52 KD	CHICKEN (100 %)	GEISLER & WEBER, 1982
				CHICKEN (COOH 20%)	CAPETANAKI ET AL., 1984
			54 KD	HAMSTER (COOH 70 %)	QUAX ET AL, 1984
				HAMSTER (100 %)	QUAX ET AL., 1985
MESENCHYMAL	VIMENTIN	1	53.5 KD	HAMSTER (COOH 95 %)	QUAX-JEUKEN ET AL., 1983
				HAMSTER (100 %)	QUAX ET AL., 1983
				PIG (COOH 30 %)	GEISLER & WEBER, 1981
GLIAL	GFAP	1	50 KD	PIG (NH2 + COOH 30 %)	GEISLER & WEBER, 1983
				MOUSE (COOH 90 %)	LEWIS ET AL., 1984
NEURONAL	NEURO-FILAMENTS	3	68 KD+	PIG (50 %)	GEISLER ET AL., 1983
				MOUSE (COOH 60 %)	LEWIS & COWAN., 1985
			160 KD+	PIG (NH2 30%)	GEISLER ET AL., 1984
			200 KD+	PIG (30 %)	GEISLER ET AL., 1983, 1985
EPHITHELIAL	KERATINS	MANY*			
	MAN: BASIC	8	52.5-68 KD	HUMAN 56 KD (COOH 60%)	HANUKOGLU & FUCHS, 1983
	ACIDIC	11	40-64 KD	HUMAN 50 KD (COOH 85 %)	HANUKOGLU & FUCHS, 1982
				HUMAN 50 KD (100 %)	MARCHUK ET AL., 1984
	COW: BAS:	± 15		COW 1A (COOH 30%) COW 1B (COOH 30%)	JORCANO ET AL.,1984C JORCANO ET AL.,1984C
	AC:	± 11	43-57 KD	COW VIB (COOH 40%)	JORCANO ET AL.,1984B
				COW VII (COOH 20%)	JORCANO ET AL.,1984B
	MOUSE: BAS:	± 14	54-70 KD	MOUSE 59 KD (100 %)	STEINERT ET AL., 1983
	AC:	± 12	45-58 KD	MOUSE 69 KD (100 %)	STEINERT ET AL., 1984
	XENOPUS: BAS:	>3	63-65 KD		
	AC:	>3	49-53 KD	XEN 51 KD (COOH 25 %)	HOFFMANN & FRANZ, 1984

Table 1

An overview of the various intermediate filament subunits as they have been identified by immunological and biochemical techniques. The first column (Cell Type) refers to the tissue in which the corresponding subunit (column 2) is expressed. Column 3 (#) represents the number of primary gene products and column 4 the molecular weigth identified for that subunit type. Column 5 (Sequence) summarizes the elucidation of primary structures of various IF proteins as they have been performed by direct amino acid sequencing or by prediction from the corresponding DNA sequence.

+) There is an enormous discrepancy between the molecular weight estimations performed on SDS-PAGE gels and those determined by other methods (Kaufmann et al.,1984).

*) For the number and iso-electric point of the human cytokeratins see the human keratin catalogue (Moll et al.,1982); for other species see Franke et al.(1982).

years this classification has been confirmed and refined by data on the primary structure of the different proteins. Part of these data were obtained by direct amino acid sequencing of the purified proteins (Geisler & Weber,1981,1982,1983; Geisler et al.,1983,1984). Most of the data, however, were obtained by analysing the DNA information encoding the IF subunits (Hanukoglu and Fuchs,1982,1983; Steinert et al., 1983,1984; Quax-Jeuken et al.,1983; Quax et al.,1983,1984a,b; Hofmann & Franz,1984; Lewis et al.,1984; Jorcano et al.,1984b). Based on sequence data and immunological observations we have summarized the classification of IF in table 1.

As it will become clear from table 1 the expression pattern of the different IF shows close parallels to established patterns of embryological differentiation: In muscle cells we find desmin, in neurones neurofilaments, in cells of glial origin (astrocytes) glial filaments, in cells of mesenchymal origin vimentin and in epithelial tissue keratins. This tissue-specificity is one of the main causes for the growing interest in these proteins, because this renders them to very specific differentiation markers useful in developmental biology and tumor diagnosis (Osborn & Weber,1982; Ramaekers et al.,1982). In contrast to the non-epithelial IF-subunits, the keratins form a large group of related proteins encoded by a family of related genes. Subdivision of keratins has been performed on the basis of isoelectric point and mutual homology. This has resulted in the definition of a subclass of acidic (Type I) and basic (Type II) proteins. Within one group the keratins show about 60-70% homology (=identical amino acids at corresponding positions). Keratins of different groups are only about 30% homologous. The homology in primary structure between the keratins on the one hand and the non-epithelial IF subunits on the other hand is low, but significant (+/- 30% homology). The mutual homology of non-epithelial IF proteins is higher (> 50%) with vimentin, desmin and GFAP being the most related (> 65%). Although the divergence in primary structure of intermediate filament proteins suggests rather dramatic differences between the subunits, the secondary structure (as it can be predicted from the primary structure) and the physical properties of the proteins show striking similarities. It appears that all IF molecules have a central polypeptide domain of about 310 residues that is mainly in α-helical conformation (as was first observed from circular dichroism measurements). Within this α-helical region there is a regular heptade distribution of hydrophobic amino acids which is typical for α-helices that can interact to form coiled coils. An identical heptade distribution has thus far been found in all IF molecules suggesting the same lengths and interruptions for the coiled coils. These findings have resulted in the prediction of a general model for IF proteins. In figure 1 the model is outlined on the basis of the vimentin sequence (Quax-Jeuken et al.,1983). The central α-helical domain can be considered as the constant domain of IF, since all subunits show a nearly identical secondary structure in this region. Consequently this is also the region of highest homology in primary structure between the various IF sequences. The non-helical NH_2 and COOH part account for the diversity of IF. Homologous sequences in these domains are only found between very closely related members, - e.g. vimentin is about 45% identical to desmin and GFAP in the COOH domain, but homology with neurofilaments or keratins is completely absent in this region. The length of the NH_2 terminal domain varies among different IF molecules from short in GFAP (45 residues) to long in the 57 kD mouse keratin (125 residues). The COOH tail piece is extremely variable in length: from 45 in human keratin 50 kd to probably more than 700 amino acids in the highest molecular weight neurofilament subunit (Geisler et al.,1984; Kaufmann et al.,1984). The extreme long extensions of the two high-molecular weight neurofilament subunits (NF-M and NF-H) are probably not involved in filament formation. Amino acid composition data indicate that these tailpieces are extremely rich in charged residues, particularly glutamic acid and lysine (Geisler et al.,1983,1984). These charged residues are located surface-exposed at the outside of neurofilaments. Whether these charged polypeptide chains are involved in nerve conduction is unknown.

As already mentioned isolated IF subunits can polymerize in vitro. However, only the non-epithelial members are able to form homopolymer filaments constituted of a single subunit (Small & Sobieszek,1977; Geisler & Weber,1980). The keratins always form a heteropolymer, built up from at least one type I and one type II subunit (Franke et al.,1983; Quinlan et al.,1984). Purified monomers of one type do not form filaments in vitro (Steinert et al.,1982). In the case of the non-epithelial IF homopolymer filaments represent the common situation. However in cell types, which con-

tain more than one subunit heteropolymers can also form (Quinlan et al.,1982,1983). This concerns mostly cultured cells, since cultered cells as a rule start synthesizing vimentin in addition to the original IF subunit (for references see Franke et al.,1982). Polymerization of keratins and vimentin into the same filaments has never been observed.

The knowledge and abundance of data on the structure of IF subunits is in sharp contrast to the lack of understanding of the polymerization process by which individual subunits assemble into 10 nm filaments. It is now generally accepted that the first polymerization step is the formation of a dimer through a coiled coil interaction and the subsequent alignment of two dimers to form a tetramer. The number and the arrangment of these tetramers within 10 nm filaments are however still subject of speculation (Ip et al.,1984).

The genes encoding intermediate filaments

The introduction of recombinant DNA technology has also put it's mark on intermediate filament research. cDNA clones have not only allowed to derive the primary structure of many subunits from the nucleotide sequence, but they have also allowed to study the chromosomal information and gene expression of IF. Keratin cDNAs, which were the first to be constructed (Fuchs et al.,1981) enabled the distinction of two different keratin subclasses on the basis of hybridization experiments. On genomic DNA each cDNA hybridized with multiple restriction fragments indicating that there are many closely related genes encoding keratins: a multigene family. Vimentin cDNA, which was isolated next (Dodemont et al.,1982; Quax et al.,1982) hybridized only with a single chromosomal locus. Hence, vimentin represented the first cytoskeletal protein that was proven to be encoded by a single copy gene. Recently it was shown that also GFAP (Lewis et al.,1984) and desmin (Quax et al.,1984a) are encoded by single copy genes. More detailed studies of keratin cDNA (Jorcano et al.1984a; Lehnert et al.,1984) showed that keratin genes behave like single copy genes if one applies highly stringent hybridization conditions. This would mean that each intermediate filament protein, that can be visualized in a specific position on a two-dimensional gel, is encoded by a single gene detectable on genomic blots. These findings distinguish the intermediate filament proteins from actins and tubulins, the other cytoskeletal proteins. The latter proteins are encoded by multiple very closely related genes (e.g. >20 actin genes in man, Soriano et al.,1982; Quax et al.,1982) that encode identical or nearly identical proteins with nearly equal migration behavior in 2-dimensional gel electrophoresis.

Despite the diversity of IF proteins, the homology at the DNA- and protein level favors the idea that IF emerged from a common ancestral gene. In search for the evolutionary origin of intermediate filament genes, cDNA probes have been used to detect homologous sequences in the genomic DNA of vertebrate and invertebrate species. In all vertebrates tested vimentin cDNA (Quax et al.,1982; Quax et al.,1984a), keratin cDNA (Fuchs & Marchuk,1983) and desmin cDNA (Quax et al.,1984a) were able to detect hybridizing fragments under normal stringent conditions. (50% formamide, 0.6 M NaCl, 42 ^{o}C). However, detection of positive hybridizing fragments in invertebrates was not possible under the same conditions. This implies that the coding information for IF is evolutionary lesser conserved than the genes for actin and tubulin, a finding that could be anticipated because of the diversity of IF proteins.

As far as the structure of IF genes is concerned data are only recently available (Quax et al.,1983,this thesis chapter 3; this thesis chapter 5; Lehnert et al.,1984; Johnson et al.,1984). The hamster vimentin gene was the first IF gene whose structure was unraveled (this thesis chapter 3). Nine exons, which are spread over +/- 9 kb of DNA , are carrying the coding capacity for vimentin mRNA (1848 bases). Relating the intron positions to the general model of IF (figure 1) it becomes clear that some introns match with a protein domain border (intron 3 and 6), but most do not. This correlation of gene and protein structure, which is found in many genes, is thought to be a remainder of processes of exon shuffling and gene duplication during evolution (Gilbert,1978; Blake,1978). From the structure of the vimentin gene one could anticipate that those intron positions that do not map at domain borders of the IF model are unique for the vimentin gene and therefore not evolutionary conserved. Surprisingly, however, precisely the same borders are also found in the hamster desmin gene (this thesis chapter 5) showing that the intron positions of these two genes have been conserved eversince the divergence of these two genes. The moment of divergence of these two IF genes is thought to have taken place amply before the evolution of vertebrate species (Quax et al.,1984a).

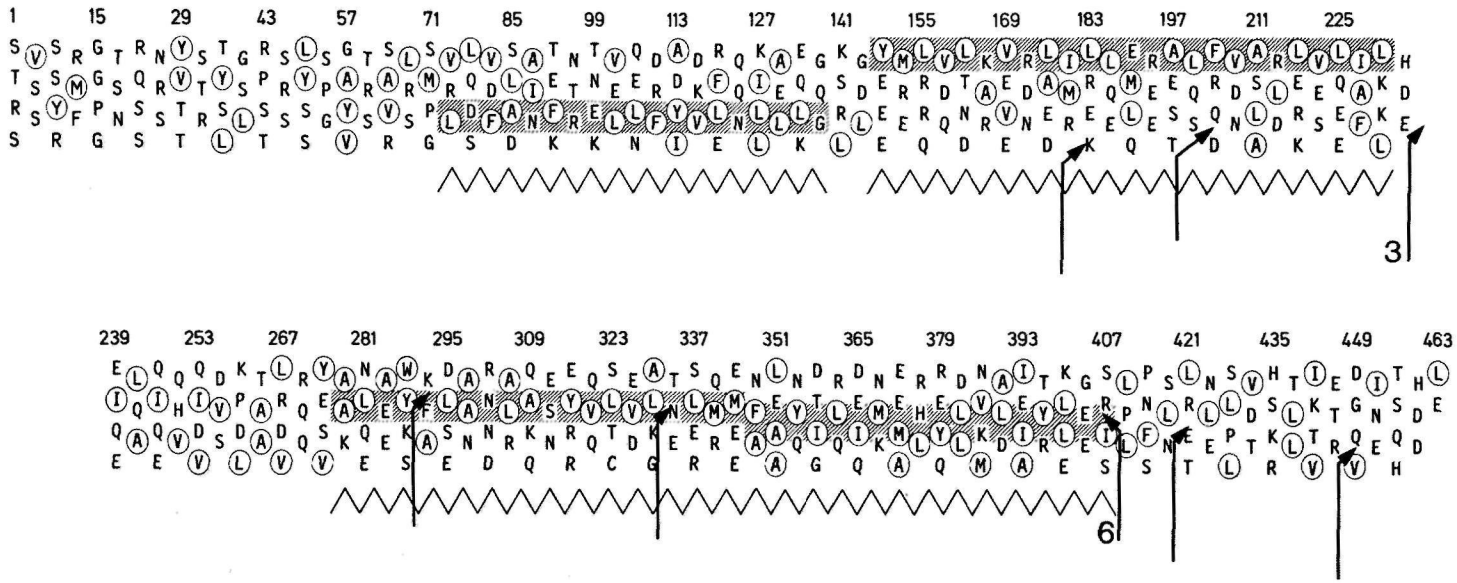


Figure 1

The General IF-Model and Intron Positions.

The complete primary structure of vimentin as it has been deduced from the nucleotide sequence of the gene (Quax et al.,1983) has been arranged in a periodic distribution with a regularity of seven. Nonpolar amino acids [alanine (A), valine (V), methionine (M), isoleucine (I), leucine (L), tyrosine (Y), phenylalanine (F), and tryptophan (W)] are encircled. Shaded regions show a heptade distribution of hydrophobic residues and are therefore thought to be in α-helical conformation (Quax-Jeuken et al.,1983). It can be seen that the arrows, representing the intron positions in the vimentin gene, do not match with protein domain borders apart from intron 3 and 6. These latter two intervening sequences are the only ones that comply with the terms of the Gilbert hypothesis (Gilber,1978).

Recently some data on the exon/intron pattern of keratin genes have been reported. R-looping analysis of four bovine keratin genes suggests that the number of introns in the keratin genes is the same (8) or one less (7) as in the vimentin and desmin genes (Lehnert et al.,1984). Although it seems that not all intron positions are equal to the vimentin gene, some of them are. One of the exon/intron borders was sequenced (λKB1a, Lehnert et al.,1984) and this correlates exactly with vimentin and desmin intron 6: the intron that interrupts the gene precisely behind the information for the coiled-coil α-helical domain. Another preliminary report (Johnson et al.,1984) states that the gene for mouse keratin 67 kD is also interrupted precisely behind the information encoding the coiled coil. The most extensive information on keratin gene structure has recently become available for the human gene encoding the 50 kD keratin (Marchuk et al.,1984). A very remarkable conservation of intron positions among IF genes was revealed by this study. Five out of seven introns of the human keratin gene are located at positions that fully correspond to intron positions of the vimentin gene. This evidences once more that IF genes emerged from a common ancestral gene.

The expression of IF genes seems to be regulated at the transcriptional level. mRNA quantities have been found to correlate correctly with the presence of proteins. The mRNA expression of the vimentin gene was puzzling in chicken cells, because two mRNA size bands could be detected in northern blots, although only one gene was found (Dodemont et al.,1982; Zehner & Paterson,1983; Capetanaki et al.,1983). A comparison of the structure of the 3′ part of the chicken vimentin gene and the vimentin cDNA showed that multiple mRNAs originated from the alternative use of more polyadenylation sites at the 3′ end of the gene (Zehner & Paterson,1983).

One of the most revolutionary impacts of the availability of cloned IF genes is the possibility to transfer and express genes into heterologous cells. This enables one to study the behavior of newly synthesized IF proteins of a certain type in cells in which the corresponding gene is normally inactive (this thesis, chapter 5). Furthermore genetic engineering of the cloned genes would allow one to alter the primary structure of IF proteins and to investigate the effect of such alterations on the ability of IF subunits to assemble into filaments. This may lead to a better understanding of the function of the different IF protein domains in filament formation.

The cloned IF genes in combination with suitable expression systems form also the obvious way to study the molecular mechanism that is at the basis of the tissue-specific transcription of intermediate filaments. Promoter regions, "activators" or "enhancers" which might be responsible for the onset or shut-off of gene expression can be characterized by testing parts of the cloned gene inserted into suitable eucaryotic vectors in different cell lines. The genes as they are described in this thesis are a good starting material for those future experiments.

REFERENCES

- Bennet, G.S., Fellini, S.A., Croop, J.M., Otto, J.J., Bryan, J., & Holzer, H. (1978) Proc. Natl. Acad. Sci. USA 75, 4364.

- Bignami, A., Eng, L.F., Dahl, D., & Uyeda, C.T. (1972) Brain Res. 43, 429.

- Blake, C.C.F. (1978) Nature 273, 267.

- Capetanaki, Y.G., Ngai, J., Flytzanis, L.N. & Lazarides, E. (1983) Cell 35, 411.

- Capetanaki, Y.G., Ngai, J., Lazarides, E. (1984) Proc. Natl. Acad. Sci. USA 81, 6909-6913.

- Dodemont, H.J., Soriano, P., Quax, W.J., Ramaekers, F., Lenstra, J.A., Groenen, M.A., Bernardi, G., & Bloemendal, H. (1982) EMBO J. 1, 167.

- Eng, L.F., Vanderhaegen, J.J., Bignami, A., & Gerstl, B. (1971) Brain Res. 28, 351.

- Franke, W.W., Weber, K., Osborn, M., Schmid, E., & Freudenstein, C. (1978a) Exp. Cell Res. 116, 429.

- Franke, W.W., Schmid, E., Osborn, M., & Weber, K. (1978b) Proc. Natl. Acad. Sci. USA 75, 5034.

- Franke, W.W., Schmid, E., Schiller, D.L., Winter, S., Jarash, E., Moll, R., Denk, H., Jackson, B.W., & Illmensee, K. (1982) Cold Spring Harbor Symp. Quant. Biol. 46, 413.

- Franke, W.W., Schiller, D.L., Hatzfeld, M., Winter, S. (1983) Proc. Natl. Acad. Sci. USA 80, 7113-7117.

- Fuchs, E.V., Coppock, S.M., Green, H., & Cleveland, D.W. (1981) Cell 27, 75.

- Fuchs, E., & Marchuk, D. (1983) Proc. Natl. Acad. Sci. USA 80, 5857.

- Geisler, N., & Weber, K. (1980) Eur. J. Biochem. 111, 425.

- Geisler, N., & Weber, K. (1981) Proc. Natl. Acad. Sci. USA 78, 4120.

- Geisler, N., & Weber, K. (1982) EMBO J. 1, 1649.

- Geisler, N., & Weber, K. (1983) EMBO J. 2, 2059.

- Geisler, N., Kaufmann, E., Fisher, S., Plessmann, U., & Weber, K. (1983) EMBO J. 2, 1295.

- Geisler, N., Fisher, S., Vandekerckhove, J., Plessmann, U., & Weber, K. (1984) EMBO J. 3, 2701.

- Geisler , N. et al. (1985) EMBO J., in press.

- Gilbert, W. (1978) Nature 271, 501.

- Hanukoglu, I & Fuchs, E. (1982) Cell 31, 243.

- Hanukoglu, I & Fuchs, E. (1983) Cell 33, 915.

- Hoffman, P.N., & Lasek, R.T. (1975) J. Cell Biol. 66, 351.

- Hoffman, W., Franz, J.K. (1984) EMBO J. 3, 1301.

- Hynes, R.O., & Destree, A.T. (1978) Cell 13, 151.

- Ip, W., Pang, S., Harzer, M.K., & Robson, R.M. (1985) Proc. New York Acad. Sci., in press.

- Ishikawa, A., Bischoff, R, & Holtzer, H. (1968) J. Cell Biol. 38, 538.

- Johnson, L., Idler, W., Zhou, X-M., & Steinert, P. (1984) J. Cell Biol. 99, 317a.

- Jorcano, J.L., Magin, T.M., & Franke, W.W. (1984a) J. Mol. Biol. 176, 21.

- Jorcano, J.L., Rieger, M., Franz, J.K., Schiller, D.L., Moll, R., & Franke, W.W. (1984b) J. Mol. Biol. 179, 257.

- Jorcano, J.L., Franz, J.K., & Franke, W.W. (1984c) Differentiation 28, 155.

- Kaufmann, E., Geisler, N., & Weber, K. (1984) FEBS Lett. 170, 81.

- Lazarides, E., & Hubbard, B.D. (1976) Proc. Natl. Acad. Sci. USA 73, 4344.

- Lehnert, M.E., Jorcano, J.L., Zentgraf, H., Blessing, M., Franz, J.K., & Franke, W.W., (1984) EMBO J. 3, 3279.

- Lewis, S.A., Balcarek, J.M., Krek, V., Shelanski, M., & Cowan, N.J. (1984) Proc. Natl. Acad. Sci. USA 81, 2743.

- Lewis, S.A., & Cowan, N.J. (1985) J. Cell. Biol., in press

- Liem, R.K.H., Yen, S.H., Salomon, G.D., & Shelanski, M.L. (1978) J. Cell. Biol. 78, 637.

- Marchuk, D., McCrohon, S., & Fuchs, E. (1984) Cell 39, 491-498.

- Moll, R., Franke, W.W., Schiller, D.L., Geiger, B., & Krepler, R. (1982) Cell 31, 11.

- Osborn, M., Franke, W.W., & Weber, K. (1977) Proc. Natl. Acad. Sci. USA 74, 2490.

- Osborn, M., & Weber, K. (1982) Cell 31, 303.

- Quax-Jeuken, Y., Quax, W., & Bloemendal, H. (1983) Proc. Natl. Acad. Sci. USA 80, 3548.

- Quax, W.J., Dodemont, H,J., Lenstra, J.A., Ramaekers, F., Soriano, P., van Workum, M.E.S., Bernardi, G. & Bloemendal, H. (1982) Adv. Exp. Medicine and Biology 158, 349.

- Quax, W., Vree Egberts, W., Hendriks, W., Quax-Jeuken, Y., & Bloemendal, H. (1983) Cell 35, 215.

- Quax, W., van den Heuvel, R., Vree Egberts, W., Quax-Jeuken, Y., & Bloemendal, H. (1984a) Proc. Natl. Acad. Sci. USA 81, 5970.

- Quax, W., Quax-Jeuken, Y., van den Heuvel, R., Vree Egberts, W., & Bloemendal, H. (1984b) Cold Spring Harbor Laboratory. Molecular Biology of the Cytoskeleton, in press.

- Quinlan, R.A., & Franke, W.W. (1982) Proc. Natl. Acad. Sci. USA 79, 3452.

- Quinlan, R.A., & Franke, W.W. (1983) Eur. J. Biochem. 132, 477.

- Quinlan, R.A., Cohlberg, J.A., Schiller, D.L., Hatzfeld, M., & Franke, W.W. (1984) J. Mol. Biol. 178, 365.

- Ramaekers, F.C.S., Puts, J.J.G., Kant, A., Moesker, O., Jap, P.H.K., & Vooys, G.P. (1982) Cold Spring Harbor Symp. Quant. Biol. 46, 331.

- Shelanski, M.L., Albert, S, DeVries, G.H., & Norton, W.T. (1971) Science 174, 1242.

- Skerrow, D., Matoltsy, A.G., & Matoltsy, M.N. (1973) J. Biol. Chem. 248, 4820.

- Small, J.V., & Sobieszek, A. (1977) J. Cell Sci. 23, 243.

- Soriano, P.,Szabo, P, & Bernardi, G. (1982) EMBO J. 1, 579.

- Steinert, P.M., & Idler, W.W. (1975) Biochem. J. 151, 603.

- Steinert, P.M., Idler, W.W., & Zimmerman, S.B. (1976) J. Mol. Biol. 108, 547.

- Steinert, P., Idler, W., Aynardi-Whitman, M., Zackroff, R., & Goldman, R.D. (1982) Cold Spring Harbor Symp. Quant. Biol. 46, 465.

- Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L., & Steven, A.C. (1983) Nature 302, 794.

- Steinert, P.M., Parry, P.A.D., Racoosin, E.L., Idler, W.W., Steven, A.C., Trus, B.L., & Roop, D.R. (1984) Proc. Natl. Acad. Sci. USA 81, 5709.

- Sun, T.-T. & Green, H. (1978) Cell 14, 469.

- Zehner, Z.E., Paterson, B.M. (1983) Proc. Natl. Acad. Sci. USA 80, 911.

CHAPTER 2

THE GENES CODING FOR THE CYTOSKELETAL PROTEINS ACTIN AND VIMENTIN IN WARM-BLOODED VERTEBRATES

EMBO J. 1, 167-171 (1982)

The EMBO Journal, Vol. 1 No. 2, pp.167–171, 1982

The genes coding for the cytoskeletal proteins actin and vimentin in warm-blooded vertebrates

Huub J. Dodemont[1], Philippe Soriano[2], Wim J. Quax[1], Frans Ramaekers[1], Johannes A. Lenstra[1], Martien A. M. Groenen[1], Giorgio Bernardi[2], and Hans Bloemendal[1]*

[1]Department of Biochemistry, University of Nijmegen, Geert Grooteplein N21, 6525 EZ Nijmegen, The Netherlands, and [2]Laboratoire de Genetique Moleculaire, Institut de Recherche en Biologie Moleculaire, 2 Place Jussieu, 75005 Paris, France

Communicated by H Bloemendal
Received on 21 December 1981

Recombinant plasmids were made containing cDNAs synthesized on hamster mRNAs coding for cytoskeletal (β- or γ-) actins and for vimentin. Hybridization of the actin probe on restriction digests of one avian and five mammalian DNAs yielded multiple bands; the vimentin probe revealed only one band (accompanied by 2–3 faint bands in some DNAs). The results obtained with the vimentin probe indicate that the corresponding coding sequences: (a) are highly conserved in warm-blooded vertebrates like the actin sequences; (b) have strongly diverged from those coding for other intermediate filament proteins, since hybridization of the vimentin probe does not lead to a diagnostic multiband pattern; and (c) most likely contribute a single gene, in contrast to the sequences coding for other cytoskeletal proteins. Hybridization of the probes on mRNAs from the different sources used showed that the non-coding sequences of both vimentin and actin genes are conserved in length.

Key words: cytoskeleton/recombinant DNA/multigene families

Introduction

Intermediate-sized filaments (IF), microfilaments and microtubules are essential structural constituents of the cytoskeleton of eukaryotic cells (Lazarides, 1980). IF can be visualized by electron microscopy as 7–11 nm thick fibrils (Lazarides, 1981). Gel electrophoresis and immunofluorescence studies can distinguish between different types of IF, and five major classes have thus been defined: (1) vimentin filaments, found in a variety of cells of mesenchymal origin or in cultured cells of any origin; (2) keratin filaments, found in cells of epithelial origin; (3) desmin filaments, found essentially in muscle cells; (4) neurofilaments, found in neurones; and (5) glial filaments, which are specific to astroglia.

Protein sequences and immunological cross-reactivities have shown that actins and tubulins, the major constituents of microfilaments and microtubules, respectively, and, to a lesser extent, IF proteins, are conserved throughout evolution (Lazarides, 1980; Lazarides, 1981; Vandekerckhove and Weber, 1979; Franke *et al.*, 1979; Geisler and Weber, 1981). The organization of the genes coding for actins and tubulins has been mainly investigated using plasmids carrying specific cDNAs (for review, see Firtel, 1981); such studies have shown that these proteins are encoded by multigene families, that their biosynthesis is developmentally regulated (McKeown and Firtel, 1981), and that the corresponding genes are conserved in sequence in different species (Cleveland *et al.*, 1980). In mammals, two-dimensional gel analysis and direct protein sequencing have revealed three different forms of actin; α-actin, found in skeletal and cardiac muscle cells, and encoded by at least four genes, and β- and γ-actin, found in microfilaments of non-muscle cells (Vandekerckhove and Weber, 1979). It has been assumed that there are two β-actin genes per haploid human genome (Vandekerckhove *et al.*, 1980), and at least four human β-actin genes have been found by Kedes (personal communication). Similarly, both α- and β-tubulin are encoded by multiple genes (Cleveland *et al.*, 1981). No results on these points have yet been reported in the case of IF proteins. Protein sequences have shown, however, considerable homology between desmin, vimentin and, to a certain extent, wool keratin and tropomyosin; this suggests that IF proteins are encoded by a multigene family (Geisler and Weber, 1981). In this connection, it is of interest that intermediate filaments of BHK cells and bovine epidermal keratinocytes appear to fit with a general model composed of a similar three-chain unit which contains regions of coiled α-helix interspersed with regions of non-α-helix, the former having the same size, the latter different sizes (Steinert *et al.*, 1980).

We have constructed plasmids containing cDNAs synthesized on hamster mRNAs coding for vimentin and β- or γ-actin. Hybridization of our actin probe on restriction digests of hamster, mouse, rat, human, calf, and chicken DNAs yielded multiple bands in accord with previous results (Cleveland *et al.*, 1980). However, the vimentin probe yielded only a single major band with most DNA sources and additional faint bands with some species. These results have several interesting implications: first, the sequences coding for vimentin are highly conserved in warm-blooded vertebrates, resembling those sequences coding for actins, which are also highly conserved in evolution; second, the sequences coding for other IF proteins have diverged considerably from those coding for vimentin, as shown by the absence of the typical multiband hybridization patterns, like that obtained with the actin probe; third, the sequences coding for vimentin very likely correspond to a single gene. It has also been shown that the non-coding sequences of both vimentin and actin mRNAs are conserved in length.

Results

Construction and characterization of actin and vimentin cDNA plasmids

As a convenient source of mRNAs coding for cytoskeletal proteins, we chose an SV40-transformed epithelial cell line from hamster lens that grows in suspension; amongst the major proteins synthesized by these cells are vimentin and the cytoskeletal actins; lens crystallins are not synthesized in detectable amounts (Bloemendal *et al.*, 1980). Poly A^+ cytoplasmic RNA was denatured with 10 mM methylmercury hydroxide and fractionated on an isokinetic sucrose density gradient (Figure 1A). Aliquots from gradient fractions were directly assayed by *in vitro* protein synthesis, and the translation products were analyzed by SDS-polyacrylamide gel elec-

*To whom reprint requests should be sent.

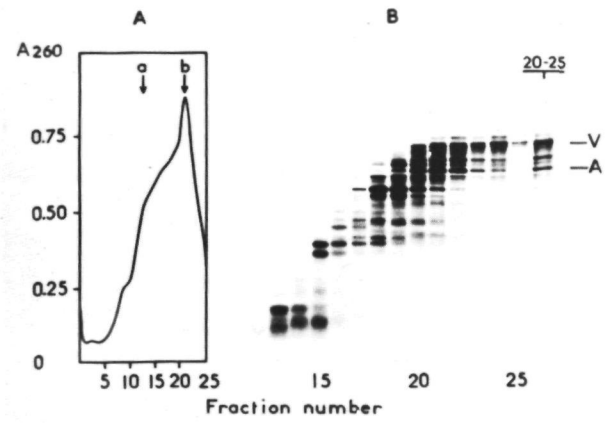


Fig. 1. Fractionation of poly A^+ RNA from a hamster lens epithelial cell line. Poly A^+ RNA (160 μg) was denatured with 10 mM methylmercury hydroxide, centrifuged on a sucrose density gradient, and collected in 0.4 ml fractions. **(A)** Absorption profile at 260 nm. The positions of rabbit globin mRNA (a), and of rat 18S rRNA (b) run on parallel gradients are indicated; 1 μl aliquots from fractions 8–25 were directly assayed for *in vitro* translation and products were run on a 13% SDS-polyacrylamide gel. **(B)** Fluorogram of the gel. Actin (A) and vimentin (V) are indicated. Pooled fractions 20–25, enriched in actin and vimentin mRNAs, were used as templates for cDNA synthesis.

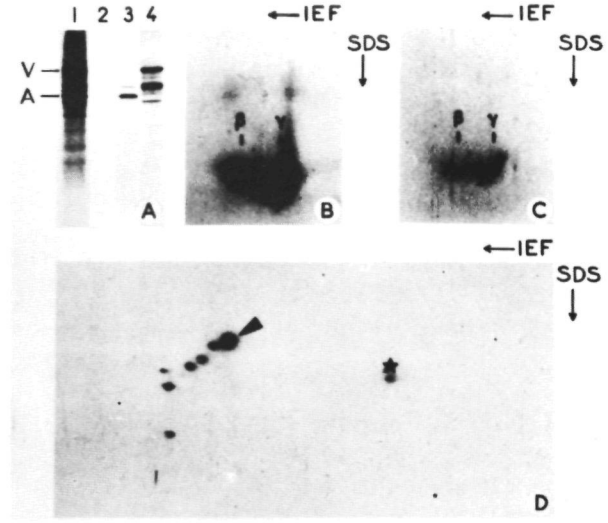


Fig. 2. Identification of *in vitro* translation products of mRNAs hybridizing to recombinant plasmids. Plasmid DNAs from individual clones were subjected to positive hybridization-translation assays as detailed in Materials and methods. **(A)** *In vitro* translation products: from the initial size-fractionated mRNAs **(lane 1)**; from mRNA hybridized to pBR322 **(lane 2)**; from mRNA hybridized to pAct-1 **(lane 3)**; from mRNA hybridized to pVim-1 **(lane 4)** (this includes characteristic proteolytic cleavage products of vimentin). Actin (A) and vimentin (V) are indicated. The common band present in the four lanes is endogenous to the reticulocyte lysate; differences in its intensity are due to differences in loads. **(B)** and **(C)** correspond to Coomassie Blue staining of unlabeled bovine brain actin **(B)** and fluorography of translation products synthesized by mRNAs hybridizing to pAct-1 **(C)**, run on the same gel. **(D)** Autoradiograph of translation products derived from mRNA hybridizing to pVim-1. Vimentin was identified by comigration on the same gel with unlabeled vimentin extracted from hamster lens "epithelial" cells, and is indicated here by an arrowhead; a ladder of specific breakdown products can be seen. The spot marked with an asterisk is endogenous to the reticulocyte lysate. IEF and SDS indicate the direction of electrophoretic migration under isoelectrofocusing and in the presence of SDS, respectively.

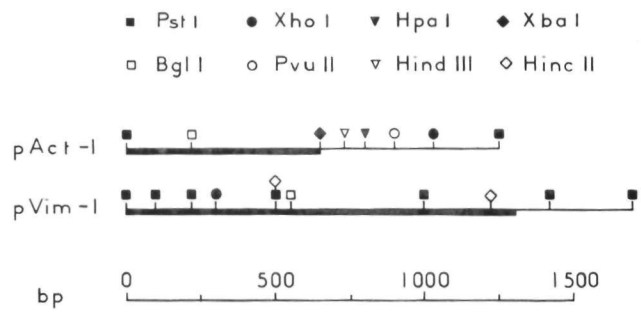


Fig. 3. Restriction maps of pAct-1 and pVim-1 inserts. Regions coding for the C-terminal sections of actin and vimentin were identified on the basis of the correspondence between restriction sites and amino acid sequences (see text); they are indicated by thick lines. The 3′ non-coding regions are indicated by thin lines. The *Eco*RI site of pBR322 is located 750 bp to the left of the left end of the map.

trophoresis (Figure 1B). mRNAs directing the synthesis of polypeptides having the same molecular weight as actin and vimentin were present in fractions 20–25; these were pooled and used as templates for cDNA synthesis. Recombinant cDNA plasmids were constructed as described in Materials and methods.

cDNA plasmids were screened for those containing actin and vimentin specific sequences using two different methods. Since *in vitro* translation of mRNA preparations from the hamster cells had shown that actin and vimentin were major products (Bloemendal *et al.*, 1980), it was conceivable that the corresponding mRNAs were highly represented in the total mRNA population. Therefore, a preliminary colony hybridization with a ^{32}P-labeled single-stranded cDNA probe made from the original mRNAs (fractions 20–25) was performed to select for cDNA clones corresponding to abundant mRNA transcripts. Forty clones giving strong colony hybridization signals were selected and mRNAs hybridizing to each one of these plasmids were translated *in vitro*. Among these, the mRNAs species hybridizing to one plasmid, pAct-1, directed the synthesis of a protein comigrating with actin (Figure 2A, lane 3), whereas mRNAs hybridizing to other plasmids directed the synthesis of relatively abundant proteins, but not of vimentin. In particular, eight plasmids specific for a 45-kd protein, one of the major translation products, were detected.

Vimentin-specific plasmids were isolated by hybridizing mRNAs to 288 plasmids distributed in 24 groups of 12 clones each; mRNAs hybridizing to two groups directed the synthesis of a product comigrating with vimentin. The two plasmids responsible for this hybridization, pVim-1 (Figure 2A, lane 4) and pVim-2 (not shown), were then isolated.

The identity of the polypeptides synthesized by mRNAs hybridizing to pAct-1 and pVim-1 was established by two-dimensional gel electrophoresis (Figure 2B, C, D). The translation products of the mRNA selected by pAct-1 comigrated with both unlabeled β- and γ-actin run on the same gel, suggesting that this plasmid is able to hybridize to both corresponding mRNAs. The major translation product obtained from pVim-1 comigrates with unlabeled vimentin run on the same gel. In addition, there is a ladder of specific breakdown products similar to that found in Triton-KCl extracts of chicken myoblasts (Gard and Lazarides, 1980).

From the length of the actin polypeptide chain, 375 amino acids (Vandekerckhove and Weber, 1979), and the molecular weight (mol. wt.) of vimentin 57 000 (Franke *et al.*, 1979), the lengths of the coding regions in the corresponding mRNAs can be calculated as ~1100 and 1450 bases, respectively;

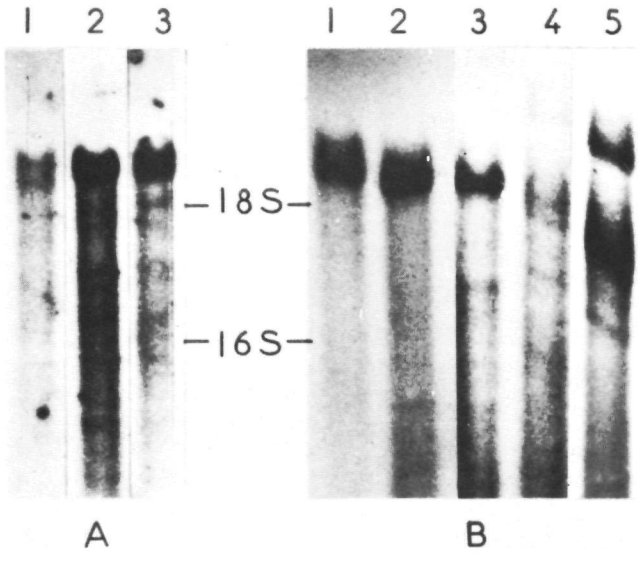


Fig. 4. RNA blot hybridization with the actin and vimentin probes. Polyribosomal poly A$^+$ RNA (20 μg) from calf lenses (1), the hamster cell line (2), rat (3), duck (4), and chicken (5) lenses, were fractionated by electrophoresis on a 1.5% agarose gel in the presence of 10 mM methylmercury hydroxide, transferred to nitrocellulose filters, and hybridized with the probes. **(A)** Hybridization with the pAct-1 probe. **(B)** Hybridization with the pVim-1 probe. Markers run on the gel were rat 28S rRNA (5000 bases), *E. coli* 23S rRNA (2940 bases), rat 18S rRNA (1940 bases), *E. coli* 16S rRNA (1570 bases) (McMaster and Carmichael, 1977).

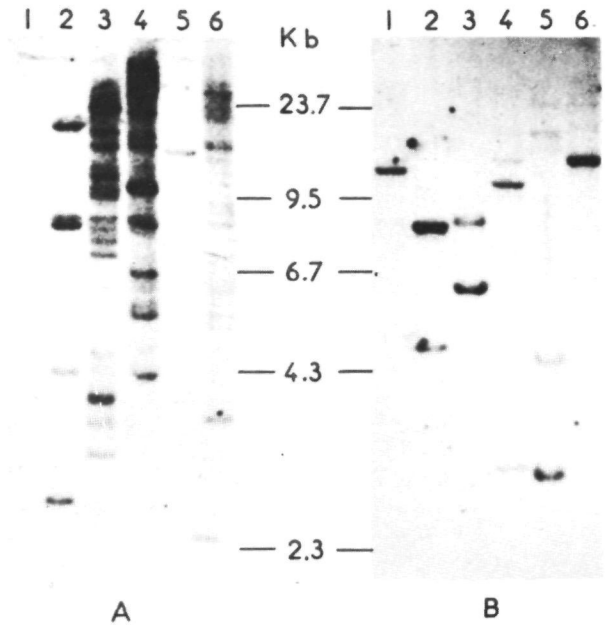


Fig. 5. DNA blot hybridization with the actin and vimentin probes. DNA from chicken erythrocytes (1), hamster liver (2), mouse liver (3), rat liver (4), calf thymus (5), and human liver (6), were degraded by *Eco*RI, fractionated on 0.6% agarose gels, transferred to nitrocellulose filters, and hybridized with **(A)** pAct-1, and **(B)** pVim-1. Loads were 8 μg of DNA in all cases, except for chicken where load was only 4 μg. *Hind*III digests of λ DNA were used as mol. wt. markers.

since both mRNAs are ~2100 bases long in hamster cells (see below), non-coding regions, including the poly A tract, must be 1000 bases and 600 bases long, for actin and vimentin mRNAs, respectively. The cDNA inserts in pAct-1 (1250 bp) and pVim-1 (1700 bp) contain, therefore, a substantial portion of the coding sequence of actin and most of that of vimentin. In fact, a comparison of the restriction maps of the cDNA inserts of pAct-1 and pVim-1 with predictions based on the protein sequences established for bovine β- and γ-actin (Vandekerckhove and Weber, 1979) and for the 141 amino acids from the C-terminal section of porcine vimentin (Geisler and Weber, 1981) allow us tentatively to locate the coding and non-coding regions present in the inserts (Figure 3). The comparison was made using a computer program (Roizès and Pelaquier, 1980) and was possible because of the conserved amino acid sequence of these proteins in mammals. This analysis also suggests that the cDNA inserts correspond to the region coding for the C-terminal section of the proteins and to the 3′ non-coding regions. Finally, it should be mentioned that the restriction map of pVim-2 (whose insert length was 1200 bp) overlapped with that of pVim-1 (data not shown).

Size determination of actin and vimentin mRNAs

Nick-translated, ^{32}P-labeled pAct-1 and pVim-1 were hybridized to poly A$^+$ RNA from the hamster lens epithelial cell line and from rat, calf, chicken, and duck lenses. Both probes cross-hybridized to the RNAs from these five different sources. Hybridization of pAct-1 to mRNAs from three mammals revealed a band, corresponding to 2100 bases; in the case of calf, a distinct additional band of 2000 bases could be detected (Figure 4A). Since our probe hybridizes to both β- and γ-actin mRNAs (Figure 2C), this result suggests that both mRNAs have virtually the same size in hamster and rat, but not in calf. With the vimentin probe, an mRNA species 2070 bases long could be revealed in all cases; in addition, the two avian mRNAs showed a strongly hybridizing band corresponding to 2400 bases and several weaker bands of lower mol. wts. (Figure 4B).

Detection of actin and vimentin sequences in genomic DNAs

*Eco*RI restriction digests of human, calf, rat, hamster, mouse, and chicken DNAs were hybridized with nick-translated ^{32}P-labeled actin and vimentin plasmids. With the actin probe, all DNAs displayed multiple bands of widely different intensities (Figure 5A). Calf and chicken DNAs exhibit simpler multiband patterns of lower intensity than the other DNAs; hamster DNA displayed a rather simple pattern with strong bands. The vimentin probe, by contrast, hybridized to only one strong band in chicken, rat, calf, and human DNAs or to two strong bands in hamster and mouse DNAs; one or two additional faint bands appeared in rat, calf, and human DNAs, but not in chicken, hamster, or mouse DNAs (Figure 5B). One or two strong hybridization bands were also obtained with restriction digests of mouse, human, rat, calf, and hamster DNAs produced with *Bam*HI, *Hind*III, *Xho*I, *Sal*I, and *Cla*I (data not shown). Interestingly, only one strong hybridization band was found in the *Bam*HI digest of hamster DNA, and in the *Xho*I digest of mouse DNA. Therefore, all six DNAs showed a single hybridization band with at least one restriction digest.

Discussion

This work provides information mainly bearing on the mRNAs for cytoskeletal actins and for vimentin, and on the corresponding genes in some warm-blooded vertebrates.

Hybridization of the actin probe pAct-1 with mRNAs from hamster, rat, and calf revealed an actin mRNA species 2100 bases long, in agreement with results previously reported for both mammals and birds (Cleveland *et al.*, 1980; Hunter and Garrels, 1977; Katcoff *et al.*, 1980; Minty *et al.*, 1981). Only

in the case of calf, two distinct mRNA species, differing in length by 100 bases, were found; these presumably correspond to β- and γ-actin mRNAs.

Hybridization of the vimentin probe pVim-1 to mRNAs from hamster, rat, calf, chicken, and duck revealed a vimentin mRNA species 2070 bases long. This was accompanied, in the case of the two avian RNAs, by several additional mRNA species. Some of these species were of lower mol. wt. and could represent specific degradation products, but one had a higher mol. wt. (2400 bases). This species might correspond to a nuclear precursor, to a polyribosomal precursor, to the product of another vimentin gene, or to a product arising by different processing of a single primary transcript. A nuclear precursor is, however, unlikely to contaminate the polyribosomal RNA preparation; a polyribosomal precursor is purely hypothetical and thus not known for any eukaryotic gene; a second vimentin gene is unlikely to exist (see below). The most plausible explanation, therefore, remains that the 2400-base mRNA species is the result of different processing of the same primary transcript. There is a precedent in the α-amylase mRNAs, although in this case different processing pathways are found in different organs (Young *et al.*, 1981).

The lengths of the mRNAs coding for cytoskeletal actins and vimentin are essentially the same in all mammals tested. Since the coding regions are conserved in length (Vandekerckhove and Weber, 1979; Franke *et al.*, 1979), this implies that the lengths of the non-coding regions, 1000 and 600 bases, respectively, are also conserved. In addition, the 5′ and 3′ non-coding regions should be individually conserved in length, unless differences in the two regions compensate each other. This length conservation, while not precluding a sequence divergence, which was in fact observed (Shani *et al.*, 1981), is an unexpected finding which may suggest a "structural" role for the non-coding regions.

The hybridization results obtained with the pAct-1 and pVim-1 probes on DNAs from warm-blooded vertebrates indicate a conservation of the sequences of the corresponding actin and vimentin genes. Such a conservation has been found for the genes of other structural proteins (Cleveland *et al.*, 1980) and is not necessarily limited to warm-blooded vertebrates. In fact, both our probes hybridize to other eukaryotic DNAs, but less strongly (data not shown). It should be pointed out, however, that the hybridization bands obtained with pAct-1 are much stronger with rodent DNAs than with DNAs of the other species, whereas those obtained with pVim-1 are much closer in intensity with all the different DNAs. This suggests greater sequence conservation of the vimentin gene than the actin genes.

As far as the gene copy number is concerned, the situation appears to be different for actin and for vimentin. With the actin probe, all DNAs display multiple bands, indicating that the actin genes form a multigene family (which includes, in fact, the muscle α-actin genes as well), in agreement with previous reports (Firtel, 1981; Cleveland *et al.*, 1980; Engel *et al.*, 1981). Rather surprisingly, the hamster probe revealed many more hybridization bands in mouse and rat than in hamster DNA, suggesting that the copy number of actin is different in these different rodents. A high number of hybridization bands was also found in human DNA, whereas this number was smaller in chicken and calf. In this case, however, the sequence divergence might be responsible for the difference seen.

Hybridization of the vimentin probe shows a pattern quite different from that obtained with the actin probe since, in all six genomes examined, only single, strong hybridization bands, 3 – 12 kb in size, were found, in at least one restriction digest. These findings suggest that vimentin is encoded by either a single gene or by a tight cluster of genes. In calf DNA the band size, 3 kb, precludes the existence of more than one vimentin gene, since the corresponding mRNA is already 2.1 kb in size (1.7 kb of which are represented in the probe). This being the case, it is highly unlikely that the vimentin gene is present in multiple tightly clustered copies in the other genomes examined. A detailed analysis of genomic clones is, however, needed to provide a final confirmation of such a general conclusion.

As far as the faint hybridization bands are concerned, they may be due to small gene fragments split off the main gene segment by restriction enzyme cut(s) in one or more introns, and/or to other genes of the IF multigene family. The first explanation appears to be more likely since the second one is difficult to reconcile with the absence of such bands in restriction digests from chicken, mouse, and hamster DNAs. The latter case is particularly striking since a hamster probe was used and this should have more homology with the other IF genes of hamster than with those of other species.

An implication of this interpretation is that genes from the IF multigene family have diverged enough so as not to give any evident hybridization with pVim-1, at least under the rather stringent conditions used in this work. It should be noted that the large sequence divergence of the genes of the IF family is accompanied by a high sequence conservation in the corresponding proteins, as vimentin shares a 64% homology over the last 141 C-terminal amino acids with another IF protein, desmin; in particular, a stretch of 37 amino acids in this region is common to the two proteins with only one amino acid change (Geisler and Weber, 1981). Such a situation is not a new one, since it has already been found in the case of other structural proteins.

Our conclusion that, in all likelihood, there is just one vimentin gene in the species examined is at variance with indications to the contrary based on preliminary characterization of the chicken vimentin gene (Zehner and Paterson, 1981). However, after the work described here was completed, a report appeared (Zehner *et al.*, 1981) confirming our finding of only one vimentin gene and two vimentin mRNA species in chicken.

Materials and methods

Isolation and translation of mRNA

Polyribosomes from an SV40-transformed hamster lens epithelial cell line (Bloemendal *et al.*, 1980) and from whole lenses from rat, chicken, and duck, were prepared according to Palmiter (1974), polyribosomes from calf lenses were prepared according to Bloemendal *et al* (1966). Poly A^+ RNA was purified by oligo(dT)-cellulose chromatography, and fractionated in sucrose gradients after disaggregation in 10 mM methylmercury hydroxide (Dodemont *et al*, 1981); this denaturing agent, used to date only in gel electrophoresis, was found to produce much better separations than formamide or dimethyl sulfoxide. Fractions containing actin and vimentin mRNAs were identified by assaying 1 μl aliquots directly in a nuclease-treated rabbit reticulocyte lysate (Pelham and Jackson, 1976). Translation products were analyzed by one- and two-dimensional electrophoresis (Laemmli, 1970; O'Farrell, 1975), and revealed by fluorography (Bonner and Laskey, 1974).

Construction of recombinant plasmids

Double-stranded cDNAs were synthesized using a one-step procedure derived from Wickens *et al* (1978). Reaction mixtures for first strand synthesis (20 μl) consisted of 50 mM Tris, pH 8.3, 8 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 1 mM of each dNTP (including trace [^{32}P]dCTP; Radiochemical Center Amersham, UK), 100 μg/ml oligo$(dT)_{12-18}$ (Collaborative Research, Waltham, MA), 50 μg/ml mRNA and 1000 units/ml reverse

transcriptase (a generous gift of J Beard) After incubation at 42°C for 15 min, reaction mixtures were placed in a boiling water bath for 3 min, quickly chilled in ice, and supplemented with an equal volume of the reaction mixture for second strand synthesis, this contained 200 mM Hepes pH 6.9, 12 mM $MgCl_2$, 10 mM DTT, 1 mM each dNTP, and 200 units/ml of the large fragment of *Escherichia coli* DNA polymerase I (Boehringer, Mannheim, FRG). Reaction mixtures were incubated at 15°C for 90 min, heat-inactivated, and filtered on Sephadex G50 cDNAs were precipitated with ethanol in the presence of *E coli* tRNA carrier, treated with S1 nuclease (BRL, Gaithersburg, MD), and fractionated on 1.5% agarose gels Fragments 1200 – 2600 bp long were eluted electrophoretically, extracted with phenol, dialyzed against water, and lyophilized. cDNAs and *Pst*I-cleaved pBR322 were tailed with dCTP and dGTP, respectively, using terminal transferase (P L Biochemicals, Milwaukee, WI), hybrid molecules were constructed by annealing and used to transform competent *E. coli* HB101 as described (Therwath *et al* , 1980). Colony hybridization was performed according to Grunstein and Hogness (1975) Plasmid DNAs were extracted according to Birnboim and Doly (1979), or purified by CsCl density gradients in the presence of ethidium bromide All experiments were performed under the containment conditions recommended by the French Committee on Recombinant DNA

Positive hybridization-translation assays

2 μg of vector and recombinant plasmid DNAs (in 10 mM Tris, pH 7.5, 1 mM EDTA) were heated at 100°C for 5 min, chilled in ice, brought to a final concentration of 6 x SSC (standard saline citrate; 0.15 M NaCl, 0 015 M trisodium citrate), and spotted on nitrocellulose filters (BA85, Schleicher and Schull, Dassel, FRG) previously treated with 6 x SSC Filters were rinsed in 6 x SSC, dried, and heated at 80°C for 2 h *in vacuo* Prehybridization was performed at 50°C for 2 h in 50% formamide, 10 mM Pipes, pH 6.4, 0 4 M NaCl, 4 mM EDTA, and 100 μg/ml poly (rA) Hybridization was carried out in the same buffer with poly (rA) substituted by 50 – 100 μg/ml mRNA, with the temperature gradually lowered from 55°C to 45°C over a period of 3 h Filters were washed with the hybridization solution for 30 min at 45°C, followed by five washings, each for 1 min, in 1 x SSC, 0 5% SDS at room temperature, five washings in 0 1 x SSC, 0.1% SDS, at 55°C, and two washings in 10 mM Tris, pH 7.5, 1 mM EDTA, at 55°C Hybridized mRNAs were extracted by boiling the filters in 100 μl of water for 60 s, and precipitated with ethanol in the presence of 1 μg of *E coli* tRNA carrier. The precipitates were collected, washed with 75% ethanol, dried, dissolved in 2 μl of water, and assayed for *in vitro* translation (see above)

Electrophoresis of RNA and DNA and blot-hybridization

Electrophoresis of RNA was performed on 1 5% agarose gels containing 10 mM methylmercury hydroxide (Bailey and Davidson, 1976) DNA preparation, digestion with restriction endonucleases, and electrophoresis in horizontal agarose gels were as described (Van der Putten *et al* , 1979). Transfer of RNA and DNA to nitrocellulose filters was performed according to Thomas (1980), and Southern (1975), respectively. Prehybridization and hybridization conditions with nick-translated plasmid probe ($1 – 5\ 10^8$ c p.m./μg) was performed as described by Wahl *et al* (1979), except that dextran sulfate was omitted from the hybridization solution After hybridization, filters were washed twice with the hybridization solution at 42°C for 1 h, once with 2 x SSC, 0 1% SDS at 55 – 60°C for 15 min and twice with 0.1 x SSC, 0 1% SDS at 55 – 60°C. Autoradiography was for 1 – 5 days at – 70°C with Kodak XR1 film, using intensifying screens

Acknowledgements

We thank Francois Bastie-Sigeac and Gerard Roizes for the computer analysis of actin and vimentin amino acid sequences, Ap Groeneveld for maintaining the hamster lens epithelial cell line, Marielle Van Workum for technical assistance, Martine Brient for typing the manuscript, and Philippe Breton for the photographic work. This work was supported in part by the Netherlands Foundation for Chemical Research (SON) and the Netherlands Foundation for the Advancement of Pure Research (ZWO) H J D was the recipient of a FEBS Fellowship J A L was a fellow of the Dutch Society against Cancer (KWF)

References

Bailey,J M , and Davidson,N (1976) *Anal Biochem* , **70**, 75-85

Birnboim,H C., and Doly,J (1979) *Nucleic Acids Res* , **7**, 1513-1523

Bloemendal,H., Lenstra,J A., Dodemont,H., Ramaekers,F.C.S., Groeneveld,A A , Dunia,I , and Benedetti,E.L. (1980) *Exp Eye Res.*, **31**, 513-525.

Bloemendal,H., Schoenmakers,J., Zweers,A , Matze,R , and Benedetti,E L (1966) *Biochim Biophys Acta*, **123**, 217-220

Bonner,W M , and Laskey,R.A (1974) *Eur J Biochem* , **46**, 83-88

Cleveland,D.W., Hughes,S.H., Stubblefield,E , Kirschner,M.W , and Varmus,H E. (1981) *J Biol Chem.*, **256**, 3130-3134.

Cleveland,D.W., Lopata,M A., MacDonald,R.J., Cowan,N.J., Rutter, W.J., and Kirschner,M.W (1980) *Cell*, **20**, 95-105.

Dodemont,H.J., Andreoli,P.M , Moormann,R.J.M., Ramaekers,F.C.S , Schoenmakers,J G.G., and Bloemendal,H. (1981) *Proc Natl Acad Sci USA*, **78**, 5320-5324

Engel,J.N., Gunning,P.W., and Kedes,L. (1981) *Proc. Natl. Acad Sci. USA*, **78**, 4674-4678.

Firtel,R.A. (1981) *Cell*, **24**, 6-7.

Franke,W W , Schmid,E , Winter,S , Osborn,M., and Weber,K. (1979) *Exp Cell Res.*, **123**, 25-46.

Gard,D.L., and Lazarides,E (1980) *Cell*, **19**, 263-275.

Geisler,N., and Weber,K (1981) *Proc Natl Acad Sci. USA*, **78**, 4120-4123

Grunstein,M., and Hogness,D S. (1975) *Proc Natl. Acad. Sci USA*, **72**, 3961-3965

Hunter,T , and Garrels,J I. (1977) *Cell*, **12**, 767-781.

Katcoff,D., Nudel,U , Zevin-Sonkin,D., Carmon,Y., Shani,M., Lehrach,H., Frischauf,A M , and Yaffe,D (1980) *Proc. Natl. Acad Sci. USA*, **77**, 960-964.

Laemmli,U.K. (1970) *Nature*, **227**, 680-685.

Lazarides,E. (1980) *Nature*, **283**, 249-256

Lazarides,E. (1981) *Cell*, **23**, 649-650.

McKeown,M , and Firtel,R A (1981) *Cell*, **24**, 799-807.

McMaster,G K , and Carmichael,G.C. (1977) *Proc. Natl. Acad. Sci. USA.*, **74**, 4835-4838

Minty,A J., Caravatti,M., Robert,B , Cohen,A., Daubas,P., Weydert,A., Gros,F , and Buckingham,M E (1981) *J Biol Chem.*, **256**, 1008-1014.

O'Farrell,P.H (1975) *J Biol Chem* , **250**, 4007-4021.

Palmiter,R D (1974) *Biochemistry (Wash)*, **13**, 3606-3615

Pelham,H R B , and Jackson,R J (1976) *Eur. J Biochem* , **67**, 247-256.

Roizes,G., and Pelaquier,E. (1980) *Nucleic Acids Res* , **8**, 351-360.

Shani,M , Nudel,U., Zevin-Sonkin,D., Zakut,R., Givol,D., Katcoff,D., Carmon,Y , Reiter,J., Frischauf,A.M., and Yaffe,D. (1981) *Nucleic Acids Res.*, **9**, 579-589.

Southern,E M (1975) *J. Mol Biol.*, **98**, 503-517

Steinert,P M., Idler,W.W., and Goldman,R.D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4534-4538.

Therwath,A., Soriano,P., and Scherrer,K. (1980) *Biochem. Int.*, **1**, 32-40

Thomas,P.S (1980) *Proc Natl. Acad Sci USA*, **77**, 5201-5205.

Vandekerckhove,J , and Weber,K. (1979) *Differentiation*, **14**, 123-133.

Vandekerckhove,J., Leavitt,J , Kakunaga,T., and Weber,K. (1980) *Cell*, **22**, 893-901.

Van der Putten,H., Terwindt,E., Berns,A., and Jaenisch,R. (1979) *Cell*, **18**, 109-116.

Wahl,G M., Stern,M., and Stark,G.R. (1979) *Proc. Natl. Acad Sci. USA*, **76**, 3683-3687.

Wickens,M P., Buell,G N., and Schimke,R.T. (1978) *J. Biol Chem.*, **253**, 2483-2495

Young,R.A., Hagenbuchle,O., and Schibler,U. (1981) *Cell*, **23**, 451-458.

Zehner,Z E., O'Neill,M.C , and Paterson,B M. (1981) in *Molecular and Cellular Control of Muscle Development*, Cold Spring Harbor Laboratory Press, NY, Abstracts p 52.

Zehner,Z.E., and Paterson,B.M (1981) *J. Supramol. Struct., Cell Biochem.*, **Suppl. 5**, p. 402.

CHAPTER 3

THE STRUCTURE OF THE VIMENTIN GENE

Cell 35, 215-223 (1983)

Cell Vol 35 215-223 November 1983

The Structure of the Vimentin Gene

Wim Quax, Wilma Vree Egberts, Wiljan Hendriks, Yvonne Quax-Jeuken, and Hans Bloemendal
Department of Biochemistry
University of Nijmegen
Geert Grooteplein N21
6525 EZ Nijmegen, The Netherlands

Summary

The structure of the chromosomal gene encoding the intermediate filament protein vimentin is described. This gene, which is present as a single copy in the hamster genome, comprises about 10 kb of DNA and contains more than 80% of intron sequences. S1 mapping and sequence analysis reveal nine exons with a total length of 1848 nucleotides. For the complete primary structure of hamster vimentin, 464 amino acids are predicted, giving a molecular weight of 53,500 daltons. The intron positions are at codons 186, 206/207, 238/239, 292/293, 334/335, 408, 423, and 451/452. The overall homology with chicken desmin is 60% and is even higher in the central (α-helical) regions of both molecules. Cross-hybridization at the DNA level, however, is low. Comparison of the amino acid sequence of vimentin with prekeratin sequences shows that there is lesser homology of primary structure, but both the position and size of α-helical regions are strongly conserved. At the 5′ end of the gene there is a consensus promoter sequence. The first AUG start codon is found 132 nucleotides downstream of the estimated cap site. The 3′ nontranslated sequence shows homologies with the chicken vimentin gene. An interesting feature of the vimentin gene is a stretch of 44 nucleotides of alternating dC and dA within intron 2 that may form left-handed Z-DNA.

Introduction

Intermediate filaments (IF) represent major elements of the cytoskeleton of most eucaryotic cells From the morphological viewpoint these filaments in a variety of cell types such as epithelium, muscle, or nerve cells are virtually identical Nevertheless, immunofluoresence has revealed that the composite proteins are quite different (reviewed by Lazarides, 1980) Both biochemical and immunological data have allowed the distinction of five different classes of intermediate filament subunits whose expression parallels embryological differentiation (Franke et al , 1981, for review see Osborn and Weber, 1982) From sequence studies on prekeratin, desmin, and vimentin, it can be concluded that long α-helical regions are important structural characteristics shared by at least three classes of intermediate filaments (Hanukoglu and Fuchs, 1982, Geisler and Weber, 1982, Steinert et al , 1983, Quax-Jeuken et al , 1983) The role of the helices in the formation of the 10 nm filaments, however, is still poorly understood The reason why tissues from different embryological origins have different IF subunits is even less clear The control of expression seems to be affected when cells are brought under in vitro conditions, since most cell cultures express vimentin in addition to their original subunit type (Franke et al , 1979)

As an approach to resolve the regulation mechanism of the tissue specific IF expression, we report the isolation and characterization of the vimentin gene This is the first description of the complete structure of an IF gene

Results

Isolation of the Vimentin Gene

Southern blot hybridization studies of total genomic hamster DNA have suggested that there is only one vimentin gene per haploid genome (Quax et al , 1982) In total genomic DNA only one Bam HI restriction fragment, 15 kb in length, was detected by the vimentin cDNA clone, pVim-1 This led us to construct a charon 28 gene library carrying hamster Bam HI restriction fragments ranging from 13 to 18 kb From this library 15 phages with vimentin sequences could be isolated All these phages showed an identical restriction pattern and therefore originated from a single chromosomal locus One of these, λ-havim, was studied in detail The insert, 15 kb in length, was isolated and subjected to restriction enzyme and blotting analysis Figure 1 shows that all of the vimentin hybridization bands that can be found in total hamster DNA are also present in λ-havim, providing additional evidence for a single chromosomal locus for the vimentin gene To determine the position and transcriptional orientation of the gene we isolated the outermost Hinf I restriction fragments from pVim-2 and pVim-1 (Quax-Jeuken et al , 1983) to serve as 5′ and 3′ specific probes, respectively. The two probes hybridized at positions that were more than 9 kb apart (see Figure 3a) and this made us conclude that the vimentin mRNA, which is about 2000 nucleotides long (Dodemont et al , 1982), is encoded by a gene that contains more than 80% of intron sequences

Detection of Vimentin Exons

S1 mapping experiment combined with sequence analysis have shown that the vimentin gene consists of nine exons. The S1 nuclease digestion was carried out on hybrids formed between λ-havim DNA and hamster lens mRNA The samples were hydrolyzed with alkaline and analyzed by blotting hybridization Seven exon bands with approximate sizes of 700, 365, 225, 165, 125, 100, and 85 nucleotides could be detected (Figure 2). The two additional smaller exons were revealed only by sequence analysis (see Figure 4) The blotting procedure enabled us to distinguish the 5′ and 3′ terminal exons by hybridizing with the 5′ and 3′ specific probes In this way we found that the 700 bp exon is the proximate 5′ and the 365 exon is the proximate 3′ exon

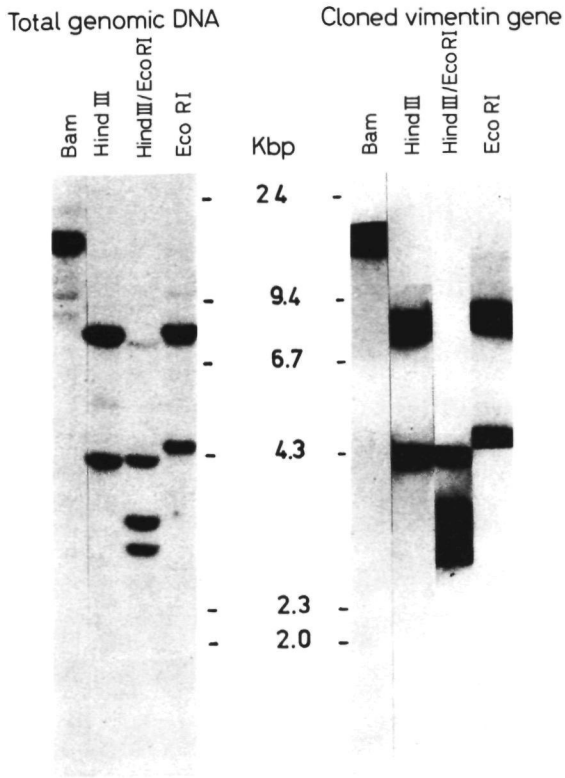


Figure 1. Comparison of λ-Havim and Total Hamster DNA

Total hamster DNA (left panel) and the λ-havim insert (right panel) were digested and double digested by the indicated enzymes. Southern blots were hybridized with ^{32}P-labeled pVim-1. Hind III digested λ-DNA was the marker.

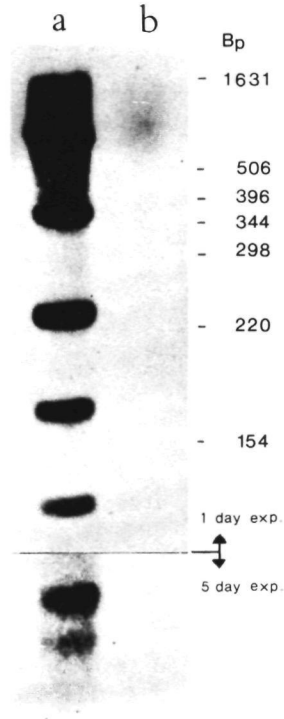


Figure 2. S1 Mapping of the Exons for the Vimentin Gene

Lane a: 300 ng of λ-havim insert DNA was hybridized to hamster lens culture mRNA (5 μg), treated with S1 nuclease, separated on a polyacrylamide ureum gel, blotted to Genescreen and hybridized to labeled λ-havim DNA. The lower part of the autoradiogram was exposed 5 days. Lane b: The same experiment without addition of mRNA.

Nucleotide Sequence and Complete Structure of the Vimentin Gene

To elucidate the detailed exon–intron structure and to unravel the complete coding information of the vimentin gene, we determined the nucleotide sequence of all exons and parts of their flanking introns. With this aim in mind we constructed recombinants between M13 vectors and λ-havim restriction fragments. Those phages that contained exon sequences were selected by hybridization with the cDNA plasmid, pVim-1, and subsequently subjected to dideoxy sequencing reactions (Sanger et al., 1980; Messing et al., 1981). To ensure a maximum of accuracy, sequencing was performed at least twice and for more than 90% on both strands. Figure 3b shows the position of the sequenced regions relative to the physical map and Figure 4 depicts the sequences. The nine exons making up the mature mRNA sequence and the eight introns are indicated. The amino acids encoded by the exons are shown above their triplet. The precise lengths of the exons in transcriptional 5′–3′ direction are: exon I: 690; II: 61; III: 96; IV: 162, V: 126; VI: 221; VII: 44; VIII: 86, and IX: 362 base pairs. This is in agreement with the results of the S1 mapping (Figure 2). The sequences of introns 3 and 6 are also shown, with sizes of 91 and 416 nucleotides, respectively. The size of the other introns can be derived from Figure 3c, where the precise positions of the exons relative to the physical map are presented. Intron 1 is 0.8 kb; intron 2: 2.6 kb; intron 4: 0.7 kb; intron 5: 0.5 kb; intron 7: 0.8 kb; and intron 8: 0.7 kb. The first exon is preceded by a TTATAAA sequence that is homologous to the consensus promoter for RNA transcription in eucaryotic genes (M. Goldberg, Ph. D. thesis, Stanford University, Palo Alto, California, 1979). At position -95 relative to the presumed cap site there is a sequence CCATT that might represent an equivalent of the CCAAT box found in many eucaryotic genes (Efstratiadis et al., 1980). The first AUG codon is found 132 nucleotides downstream of the presumed cap-site, and a second one is found at position +171. Both codons are in the same reading frame and comparing them with the consensus translation start codon, $\overset{A}{G}$XXAUGG (Kozak, 1981), we cannot distinguish which of the two is used. Comparison of the sequence with the N-terminal amino acids of the desmin polypeptide (Geisler and Weber, 1982) shows that the homology extends to the first AUG codon, suggesting that this one is actually used, as is found for most eucaryotic genes (Kozak, 1978). The last exon ends with the sequence AATAAA, which is considered a signal for poly (A) addition (Proudfoot and Brownley, 1976). We found only one such signal in the

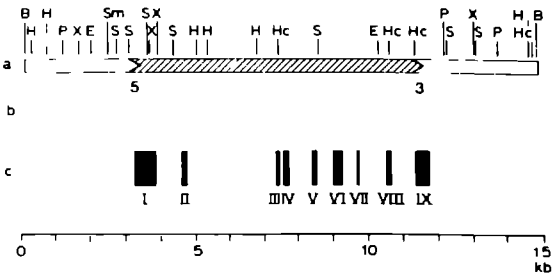


Figure 3 Physical Map of the Vimentin Gene

(a) The sites for some often used restriction enzymes in the λ havim insert are shown B = Bam HI E = Eco RI H = Hind III Hc = Hinc II P = PVU II S = Sst I Sm = Sma I X = Xho I 5′ and 3′ correspond to the positions where the 5′ and 3′ specific probes derived from the cDNA hybridized, respectively (b) shaded bars = regions whose sequence is presented in Figure 4 (c) The black bars represent the position and size of the vimentin exons They are numbered with Roman numerals

hamster vimentin gene, in contrast to what has been reported for the chicken vimentin gene (Zehner and Paterson, 1983) In Figure 5 all exon–intron donor and acceptor sites are compared to a proposed consensus sequence (Mount, 1982) It can be seen that the bordering intron dinucleotides AG and GT are always found However, the other nucleotides do not strictly obey the consensus rule We have also indicated the positions of the splice points relative to the amino acid number of the protein (Figure 5) There are no introns within the 5′ and 3′ noncoding region of the mRNA

Complete Primary Structure of Vimentin and Its Homology to Other Intermediate Filaments

The coding capacity of the exons is in agreement with the reported sequence of vimentin cDNA (Quax-Jeuken et al , 1983) There are only a few differences, three of which represent differently predicted amino acids (marked in Figure 4) These differences can be explained either by mutation events, because the culture from which the cDNA was derived has undergone many cell devisions, or by polymorphism, because the gene was isolated from a randomly bred hamster

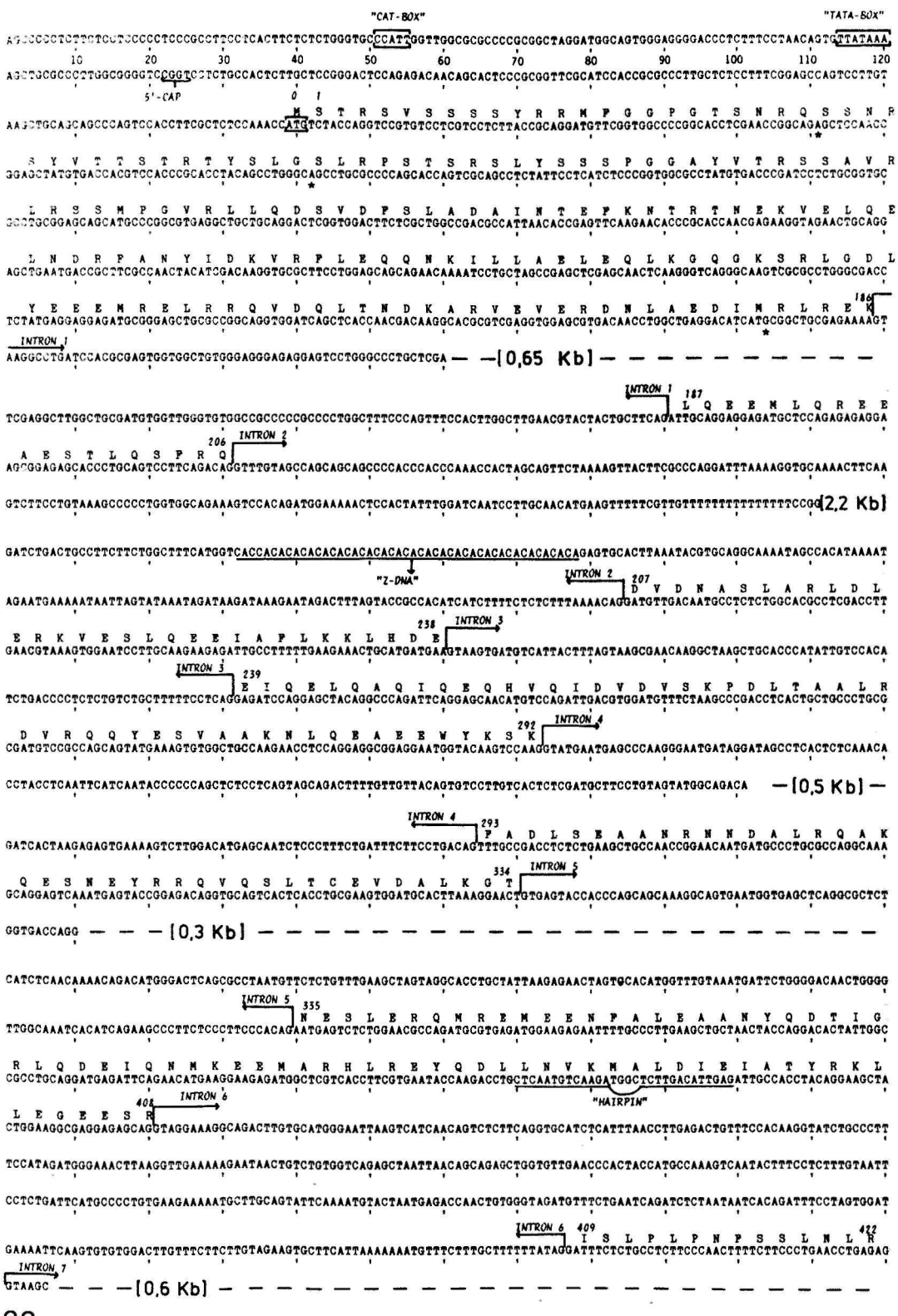
Because the 5′ end of the gene has been sequenced, the prediction of the primary structure for the N-terminal region of the vimentin polypeptide, which was not possible from our cDNA clones (Quax-Jeuken et al , 1983), can now be completed Two potential start codons, 132 and 171 nucleotides downstream of the presumed cap site, are found If we assume that the first ATG codon is used (see above) and that the methionine is removed, then the first amino acid present in the vimentin polypeptide would be serine Our observation that vimentin is N-terminally blocked (unpublished results) presumably by an acetyl group (Steinert et al , 1980) and the data that serine is the most frequently found N-α-acetylated amino acid (H Driessen, Ph D thesis, University of Nijmegen, Nijmegen, The Netherlands, 1983) are in concert with this assumption The sequence derived from the cDNA starts at the 15th codon downstream of the first ATG codon

Because the complete vimentin sequence is now available, we can make a comparison with other IF proteins This is shown in Figure 6 where the complete chicken desmin (Geisler and Weber, 1982), the complete mouse epidermal prekeratin (Steinert et al , 1983), and the almost complete human prekeratin primary structure (Hanukoglu and Fuchs, 1982) are aligned to the vimentin sequence Identical residues in vimentin, desmin, and prekeratin are indicated by a hatched bar, common residues in only vimentin and desmin by an open bar We also marked the position of α-helix forming sequences in the vimentin polypeptide (Quax-Jeuken et al , 1983) The homology between vimentin and prekeratin is obvious within the regions of the proposed α-helices, but low or practically zero in the NH2 and COOH-terminal regions The homology with desmin is much more pronounced and not only restricted to the helical regions, although these regions show the highest homology Only from amino acid 20 to 80 is the homology very low This may mean that most of the immunological differences between desmin and vimentin are situated in this region

Potential Left-handed Z-DNA in an Intron

In the second intron, at position -170 to 126 before the start of the third exon, there is a striking sequence of 22 alternating CA copolymers This sequence was detected by both the dideoxy and the Maxam–Gilbert procedure Therefore a sequence artefact can be excluded It has been shown that synthetic poly(CA–GT) can adopt a left handed (Z) conformation (Zimmer et al , 1982, Vorlickova et al , 1982) and the presence of these sequences in vivo raises the question of whether left handed DNA fulfils an essential role in eucaryotic DNA First this sequence was found in an immunoglobulin κ V gene region (Nishioka and Leder, 1980) and intergenic between the δ- and β globin genes (Miesveld et al , 1981), but recently Hamada and Kakunaga (1982) reported its presence in an intron of a human actin gene We found a stretch of 44 nucleotides in the second intron of the vimentin gene The orientation relative to the transcriptional direction of the gene, however, is in our case (CA) opposite to the sequence found in the actin gene (GT) We used the (CA)22 containing M13 recombinant as a probe to search for other CA polymers Within λ-havim no other fragment hybridized, but total genomic hamster DNA gave rise to a hybridization smear, indicating that the CA polymer is repetitive as in other species (Meisveld et al , 1981, Hamada and Kakunaga, 1982) A positive hybridization signal could only be found when E coli DNA instead of salmon DNA was used as a carrier This observation confirms that E coli DNA, in contrast to salmon DNA, lacks poly (dC-dA, dG-dT) fragments that interfere with the labeled probe (Miesveld et al , 1981)

We also searched for the presence of other repetitive sequences within λ havim Therefore, we hybridized restriction fragments with labeled total genomic hamster DNA, a technique that enables detection of highly repetitive DNA (Heilig et al , 1980) All fragments in λ havim that are


"CAT-BOX"
"TATA-BOX"
5'-CAP
INTRON 1
INTRON 2
INTRON 3
INTRON 4
INTRON 5
INTRON 6
INTRON 7
"Z-DNA"
"HAIRPIN"
[0,65 Kb]
[2,2 Kb]
[0,5 Kb]
[0,3 Kb]
[0,6 Kb]

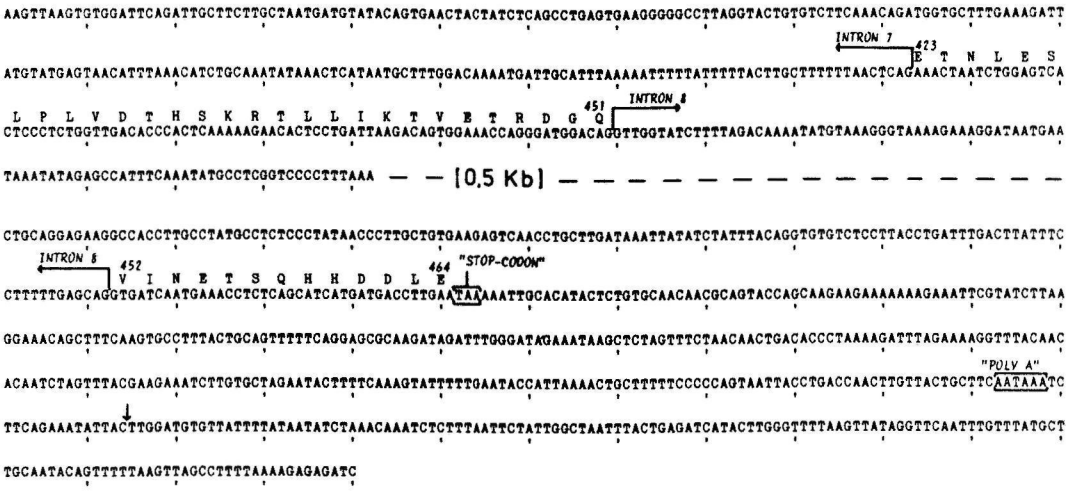


Figure 4. Sequence of the Hamster Vimentin Gene

The sequence of all coding regions and adjacent nucleotides are shown with 120 bases per line. The introns are marked by arrows and the amino acid sequence encoded by the exons is given in single letter codes above the nucleotide sequence. The exons were identified by comparison with the cDNA sequence (Quax-Jeuken et al., 1983). Lengths of (intron) fragments that are not sequenced are in brackets. "TATA box" and "CAT box" are transcriptional control signals (M. Goldberg, Ph. D. thesis, Standord University, Palo Alto, California, 1979; Efstratiadis, 1980). 5′ CAP: estimated location of the capped nucleotide. Poly-A: polyadenylation signal, the arrow 15 bases ahead marks the probable addition site. At the splice positions the numbers for the corresponding amino acids of the vimentin polypeptide are shown. "Z-DNA" marks a stretch of potential left-handed Z-DNA. "Hairpin": position where a perfect hairpin with 13 nucleotides stem and 4 nucleotides loop can be formed in the resulting RNA. Asterisks indicate changes that represent differently predicted amino acids with respect to the cDNA sequence (see Results and Quax-Jeuken et al., 1983).

Consensus	C/A AGgt a/g agt	ttttttttttt t ccccccccccc n c agG
	186	187
	Ly	sLeu
Intron 1	AAAAgtaagg	gtactactgcttcagATTG
	206	207
	Gln	Asp
Intron 2	ACAGgtttgt	ctctctttaaaacagGATG
	238	239
	Glu	Glu
Intron 3	TGAAgtaagt	ctgctttttcctcagGAGA
	292	293
	Lys	Phe
Intron 4	CAAGgtatga	atttcttcctgacagTTTG
	334	335
	Thr	Asn
Intron 5	AACTgtgagt	tctcccttcccacagAATG
	408	409
	Ar	gIle
Intron 6	GCAGgtagga	tttgctttttttatagGATT
	422	423
	ArgG	Lu
Intron 7	AGAGgtaagc	gctttttttaactcagAAAC
	451	452
	Gln	Val
Intron 8	ACAGgttggt	ttccttttgagcagGTGA

Figure 5. Exon–Intron Junctions of the Vimentin Gene

Nucleotides around the splice sites are compared to a consensus donor (left) and acceptor (right) sequence (Mount, 1982). Exon sequences are in capital letters, the number on top of the amino acids is the relative number in the vimentin polypeptide. The "ag" at the beginning and the "gt" at the end of an intron are always present, the other nucleotides frequently differ from the consensus.

downstream of the AATAAA poly (A) addition signal also appeared to contain highly repetitive sequences (data not shown).

Discussion

Vimentin Gene Organization

Our results provide evidence that vimentin is encoded by a single copy gene in the hamster genome as has been suggested previously (Quax et al., 1982). This gene is 10 kb in length and shows a complex organization of nine exons encoding the 1848 nucleotides that compose the unique sequences of the vimentin mRNA. Although there is no evidence for a vimentin gene other than the one present in λ-havim, a careful examination of Figure 1 shows that there are several faint bands visible with total genomic DNA that do not originate from the vimentin gene. We assume that these bands result from cross-hybridization of our cDNA probe with other IF gene sequences. Desmin must especially share a reasonable nucleotide sequence homology in view of the great homology in primary structure (Geisler and Weber, 1982; Quax-Jeuken et al., 1983; Geisler and Weber, 1981).

Whereas actins and tubulins are encoded by multigene families (Cleveland et al., 1980; reviewed by Firtel, 1981), the situation with vimentin is clearly different. If we regard the IF genes as a group of evolutionary related genes, we must conclude that individual members of this family have diverged much more than actins or tubulins. In this connection one should consider that in mesenchymal tissue only one member, the vimentin gene, is expressed, whereas in epidermis, an epithelial tissue, a number of

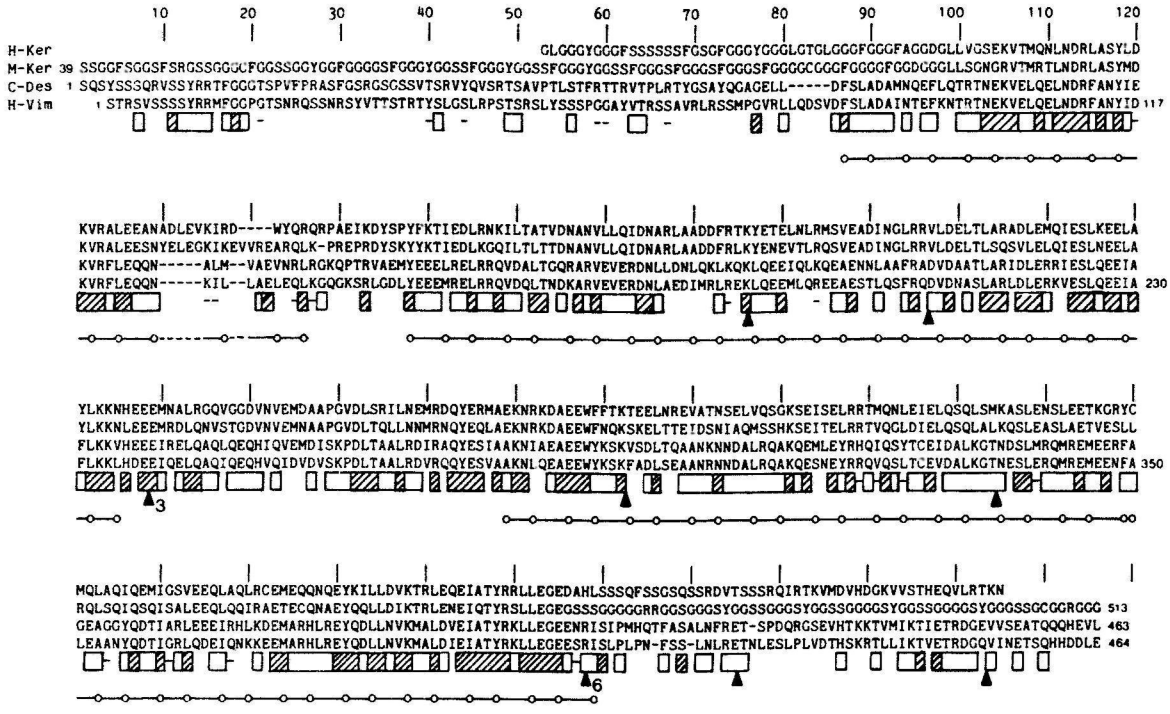


Figure 6. Comparison of Vimentin with Other IF Proteins

H-Ker: a human prekeratin sequence derived from a cDNA clone (Hanakoglu and Fuchs, 1982). M-Ker: a mouse prekeratin sequence derived from a cDNA clone (Steinert et al., 1983). C-Des: chicken desmin sequenced directly (Geisler and Weber, 1982). H-Vim: the hamster vimentin primary structure derived from Figure 4. The sequences are, where possible, numbered from the N-terminal amino acid. Alignment was achieved by the program DIAGON (Staden, 1982) and as little as possible deletion or insertion events are assumed. The partial amino acid sequences for wool α-keratins (Crewther et al., 1978; Gough et al., 1978) showing homology to the helical parts of intermediate filaments (Hanukoglu and Fuchs, 1982; Geisler and Weber, 1982) have not been included in this figure. Open bar: identical residues in vimentin and desmin. Hatched bar: identical residues in vimentin, desmin and one or both prekeratins. Dashes: residues only identical between vimentin and prekeratin. O——O: α-helical conformation of vimentin (Quax-Jeuken et al., 1983). The open circles mark position (a) and (d) of the heptade. ▲: position of introns in the corresponding part of the vimentin gene.

members, the prekeratin genes (Fuchs et al., 1981), may be responsible for IF formation. How tissue specificity is determined by signals at the DNA level (a control at the translation level is very unlikely, Kreis et al., 1983) will be the subject of future studies in which we will test in vitro expression of vimentin sequences.

The hamster vimentin gene has never been shown to express more than one mRNA. In contrast, two mRNA species are found for chick vimentin (Dodemont et al., 1982; Zehner and Paterson, 1983) that originate from multiple poly (A) addition signals at the 3′ end of the gene (Zehner and Paterson, 1983). We found only one AATAAA signal in the hamster gene and the S1 experiment (Figure 2) suggests that only this is used as a transcription stop signal. In spite of this difference at the polyadenylation site, the remaining nucleotides of hamster display at several regions a high degree of homology with the chick nontranslated 3′ nucleotides (Figure 7), suggesting that these nucleotides have to be conserved for the functioning of the vimentin gene. However, the open reading frames that have been assumed to play a role in chick vimentin expression (Zehner and Paterson, 1983) are not present in the hamster nontranslated region.

A computer analysis revealed a striking sequence of 27 nucleotides capable of forming a strong hairpin with a stem of 13 nucleotides and a loop of 4 nucleotides within exon VI (Figure 4). This hairpin, with a calculated energy of -19 kcal, ends up in that part of the mRNA that encodes amino acids 385 to 394, which are near the COOH terminal region of the protein. In other genes hairpins have been found mostly in the 5′ nontranslated region where they may play a role in ribosome attachment (Shine and Dalgarno, 1975). The observation of a hairpin in the protein coding region is to our knowledge unprecedented. A possible role in mRNA translation and/or processing should be tested.

IF Protein Domains and Evolution

For years a model consisting of two nearly equally long α-helices separated by a nonhelical spacer has been used to represent IF protein structure (Steinert et al., 1980). Since sequence information has become available (Geisler and Weber, 1981; Hanukoglu and Fuchs, 1982; Steinert et al., 1983; Quax-Jeuken et al., 1983) it is now possible to construct a better and more detailed model. For vimentin this model can be completed, and when we apply the heptade convention to identify α-helical regions the follow-

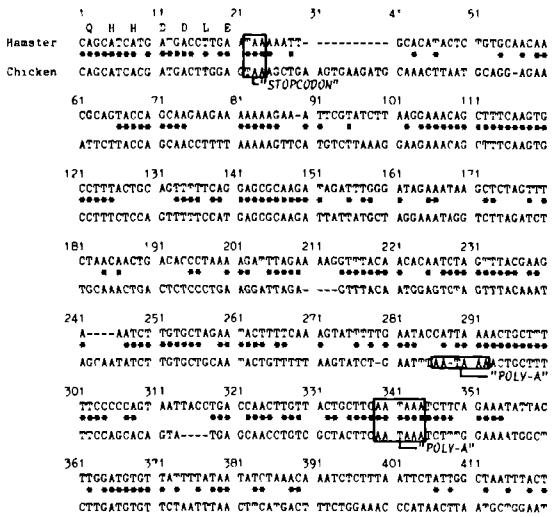


Figure 7 3' Nontranslated Region of Hamster and Chicken Vimentin Gene

The 3' sequences of chicken (Zehner and Paterson, 1983) and hamster vimentin are aligned Common residues are marked by an asterisk From the stop codon to the hamster poly (A) addition signal the homology is 83%, however, many deletions or insertions have taken place In contrast to the single poly (A) addition signal of the hamster gene, the chicken has got four, of which only two are included in this figure

ing picture emerges (see Figure 6) The first 84 amino acids are presumably not in α-helix conformation and contain many β-turns Residues 85 to 136 and 148 to 236 clearly show the distribution of hydrophobic amino acids at the first (a) and fourth (d) position of a heptade, typical for a coiled-coil α-helix. Positions 236 to 278, which include one proline residue, form an interruption in front of another long helical region ranging from positions 278 to 409 This latter coiled-coil shows a discontinuity in the hydrophobic backbone at position 350. The carboxy terminus is not in helical conformation

These sizes and interruptions of coiled-coil regions are found at nearly identical positions in the sequences of desmin, mouse prekeratin, and human prekeratin and therefore must play an essential role in 10 nm filament formation Sequence homology between vimentin and desmin on the one hand and prekeratins on the other is predominantly found within these coiled-coil regions, albeit the homology is concentrated in distinct regions. It can also be seen from Figure 6 that not only the hydrophobic residues at positions (a) and (d) of the heptade necessary for maintenance of α-helix conformation, but also acidic and basic residues are often conserved. This might imply that the periodic distribution of charged residues (McLaghlan and Stewart, 1982) is a common feature among intermediate filaments that has been conserved throughout evolution From amino acid 236 to 278 there are interruptions of the α-helices, but without a substantial decrease of homology, suggesting that the amino acids surrounding these interruptions are also essential for the 10 nm filament formation. If we compare only vimentin and desmin we see that apart from the C- and N-terminus, a strong homology is found in coiled-coil regions as well as in the interrupting stretches

It has been proposed that exons play a role in the evolution of different protein domains and that therefore the gene exon-intron pattern might show a relation to protein structure (Gilbert, 1978, Blake, 1978). In Figure 6 the positions of the introns are shown If we look for a relation with the indicated regions of coiled-coils, which are considered the important domains, we can see that intron 6 is positioned at the border of a coiled-coil and the COOH-terminus At precisely this position (residue 408) the glycine serine-rich region of prekeratins starts, suggesting that sequences homologous to vimentin exon 7, 8, and 9 are not present in prekeratins Moreover intron 3 can be considered to be located between two α-helices However, other introns are located within the helical regions at positions where homology is very high and can hardly be considered to separate different domains Two introns are found in the COOH-terminus coding region where domain structures of various IF proteins differ Furthermore, the first exon (186 amino acids) encodes a nonhelical headpiece as well as a very conserved α-helix Thus a correlation between exon-intron structure and protein domains is not evident It will be necessary to resolve the structure of other IF genes before a possible role for exons in IF evolution can be fully understood Although the Gilbert hypothesis has proved to be correct for some genes (Artymiuk et al , 1981, Blake, 1981, Inana et al , 1983) it has become clear that genes lacking recent duplications or evolutionary highly conserved genes such as actins (Nudel et al , 1982, Fryberg et al , 1981) do not show a clear relation between gene structure and protein structure Probably IF genes must be reckoned among the latter group.

Z-DNA Conformation In Vivo?

The existence of a 44 nucleotides long stretch of alternating dC and dA residues within the second intron of the vimentin gene leads one to ascribe a possible role for this sequence in the genome It was shown that in vitro synthetical d(CA/GT) polymers can adopt a left-handed Z-DNA conformation under certain high salt conditions (Zimmer et al , 1981, Vorlicka et al , 1981) The observation that DNA can also adopt the left-handed conformation under more physiological conditions (Singleton et al., 1982; Nordheim and Rich, 1983) and the finding that antibodies against Z-DNA react with Drosophila polytene chromosomes (Nordheim et al , 1981) are indications for the existence of Z-conformation in native DNA.

It has been shown that d(GC/CG)n and d(GT/CA)n sequences are widely dispersed throughout the genome in a number of eucaryotic species (Hamada et al., 1982) Moreover there are sequence data on the occurrence of d(CA/GT)n sequences within intergenic and intronic regions of several genes (Miesveld, 1981, Slightom et al , 1980; Hamada and Kakanuga, 1982) As a possible function of these sequences it has been suggested that they

may play a role in gene rearrangement and recombination (Slighthorn et al., 1980). On the other hand it cannot be excluded that Z-DNA is involved in regulation of gene expression. Recently it has been speculated that eucaryotic Z-DNA may play a role in the organization of chromatin domains (Nordheim and Rich, 1983). The formation of Z-DNA and its influence on the superhelicity of DNA could then be reflected in the change of S1 nuclease and DNAase I sensitivity observed in the chromatin regions of actively transcribed genes (Weintraub, 1983) The proteins that bind specifically to Z-DNA (Nordheim et al , 1982) may affect the formation and/or stabilization of the secondary DNA structures that interact with the chromatin Such a Z-DNA control element would show similarities with sequences that have been described as "enhancer" elements in viruses (Yaniv, 1982) and recently in immunoglobulin genes (Boss, 1983). Both would be active irrespective of their orientation to the gene and both would be capable of exerting their influence at a rather large distance from the transcription start. A mechanism for enhancers has also been proposed where the secondary structure of the sequence would influence the chromatin structure and the DNA superhelicity (Yaniv, 1982). Further experimentation is needed to verify these parallels between Z-DNA and enhancer sequences.

Experimental Procedures

Construction and Screening of the Hamster Gene Library

Since there is only one Bam HI restriction fragment carrying vimentin sequences in the hamster genome (Quax et al , 1982) we constructed a partial gene library from a Syrian gold hamster Bam HI restriction fragments ranging from 13 to 18 kb were isolated by agarose gel electrophoresis and electroelution Ligation, in vitro packaging, and plating were performed with Charon 28 as vector (Rimm et al , 1980) Screening with pVim-1 was done under hybridization conditions as described below Positive plaques were purified and DNA isolated Restriction enzyme analysis and blotting hybridization to compare the phage DNA with total hamster DNA were done as described (Quax et al , 1982) Hybridization was carried out at 42°C during 16 h in 50% formamide, 5X SSC, 1X Denhart's solution, 20 mM sodium phosphate (pH 6 8), 5 mM EDTA, and 100 μg/ml single-stranded salmon sperm DNA Washings were at 42°C during 3 hr with the hybridization solution, at 55°C for 15 min in 2X SSC, 0 5% SDS, and at 55°C for 30 min in 0 1X SSC, 0 1% SDS The structure of the cDNA plasmids pVim-1 and pVim-2 that are used as probes have been described previously (Quax Jeuken et al , 1983)

S1 Nuclease Mapping

The phage insert (300 ng) was hybridized to hamster lens culture RNA (5 μg) at 53°C during 3 hr in 20 μl of the following solution 80% formamide, 400 mM NaCl, 40 mM Pipes (pH 6 4) and 1 mM EDTA Then the hybrids were directly diluted into 200 μl of S1 solution containing 2 5 U S1/μl (modification of Berk and Sharp, 1977). After 45 min at 45°C the samples were treated with alkali and the remaining DNA was separated on a 7% polyacrylamide/6M ureum gel and blotted electrophoretically onto genescreen (NEN) with 25 mM sodiumphosphate (pH 6 5) as transfer buffer Detection with different probes was under hybridization conditions as mentioned before

DNA Sequence Analysis

Initially fragments were end labeled and sequenced by the method of Maxam and Gilbert (1980) Later we applied the M13-dideoxy strategy Therefore the 15 kb insert of λ-havim was digested with the enzymes Sau 3A, Taq I, and Hpa II, respectively Resulting fragments were ligated into the Bam HI or Acc I site of Mp8, transfected on JM103, and plated (Messing et al 1981) Plaque filter replicas were prepared and hybridized with cDNA probes Most coding sequences could be sequenced in both directions in this way For some stretches known restriction fragments were directly subcloned into Mp8 and Mp9 Template isolation and dideoxy reactions were performed as described (Sanger et al , 1980) Analysis was on 40 and 85 cm 6% sequencing gels

The gel readings were recorded, edited, and compared by the programs of Staden (1979) Search for homology and optimal alignment was done by the program DIAGON (Staden, 1982) The energy for hairpin structures was calculated by the rules of Tinoco et al (1973)

Acknowledgments

We thank Dr Anton Berns for discussion and advice, Mrs Els Hulsebos for help with construction of the gene library, and Dr Peter van Wezenbeek for advice on M13 dideoxy sequencing The help of Messrs Ger Austen, Ruud Luiten, and Jack Leunissen with computer analysis is also highly acknowledged The present investigations have been carried out in part under the auspices of the Netherlands Foundation for Chemical Research (SON) and with the financial aid of the Netherlands Organization for the Advancement of Pure Research (ZWO)

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

Received June 28, 1983, revised August 2, 1983

References

Artymiuk, P J , Blake C C F , and Sippel, A E (1981) Genes pieced together—exons delineate homologous structures of diverged lysozymes Nature *290*, 287–288

Berk, A J , and Sharp, P A (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids Cell *12*, 721–732

Blake, C C F (1978) Do genes-in-piece imply proteins-in piece? Nature *273*, 267

Blake, C C F (1981) Exons and the structure, function and evolution of haemoglobin Nature *291*, 616

Boss M A (1983) Enhancer elements in immunoglobulin genes Nature *303*, 281–282

Cleveland, D W , Lopata, M A , MacDonald, R J , Cowan, N J , Rutter, W J , and Kirschner, M W (1980) Number and evolutionary conservation of α- and β-tubulin and cytoplasmic β and γ-actin genes using specific cloned cDNA probes Cell *20*, 95–105

Crewther, W G , Inglis, A S , and McKern, N M (1978) Amino acid sequences of α-helical segments from S-carboxymethylkeratene-A Complete sequence of a type II segment Biochem J *173*, 365–371

Dodemont, H J , Soriano, P , Quax, W J , Ramaekers, F , Lenstra, J A , Groenen, M A , Bernardi, G , and Bloemendal, H (1982) The genes coding for the cytoskeletal proteins actin and vimentin in warm-blooded vertebrates EMBO J *1*, 167–171

Efstratiadis, A , Posakony, J W , Maniatis, T , Lawn, R M , O'Connell, C , Spritz, R A , DeRiel, J K , Forget, B G , Weissman, S M , Slightom, J L , Blechl A E , Smithies, O , Baralle, F E , Shoulders, C C , and Proudfoot, N J (1980) The structure and evolution of the human β-globin gene family Cell *21*, 653–668

Firtel, R A (1981) Multigene families encoding actin and tubulin Cell *24*, 6–7

Franke, W W , Schmid, E , Osborn, M , and Weber, K (1979) Widespread occurrence of intermediate-sized filaments of the vimentin-type in cultured cells from diverse vertebrates Exp Cell Res *123*, 25–46

Franke W W , Schmid, E , Schiller D L., Winter, S , Jarasch, E -D , Moll, R , Denk, H , Jackson, B , and Illmensee, K (1981) Differentiation related patterns of expression of proteins of intermediate sized filaments in tissues

and cultured cells Cold Spring Harbor Symp Quant Biol *46*, 431–453

Fuchs, E V , Coppock, S M , Green, H , and Cleveland, D W (1981) Two distinct classes of keratin genes and their evolutionary significance Cell *27*, 75–84

Fyrberg, E A , Bond, B J , Hershey N D , Mixter, K S , and Davidson, N (1981) The actin genes of drosophila protein coding regions are highly conserved but intron positions are not Cell *24*, 107–116

Geisler, N , and Weber, K (1981) Comparison of the proteins of two immunologically distinct intermediate-sized filaments by amino acid sequence analysis desmin and vimentin Proc Nat Acad Sci USA *78*, 4120–4123

Geisler, N , and Weber, K (1982) The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins EMBO J *1*, 1649–1656

Gilbert, W (1978) Why genes in pieces? Nature *271*, 501

Gough, K H , Inglis, A S , and Crewther, W G (1978) Amino acid sequence of α-helical segments from S carboxymethylkerateine-A Complete sequence of a type I segment Biochem J *173*, 373–385

Hamada, H and Kakunaga, T (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome Nature *298*, 396–398

Hamada, H , Petrino, M , and Kakunaga, T (1982) A novel repeated element with Z DNA-forming potential is widely found in evolutionary diverse eukaryotic genomes Proc Nat Acad Sci USA *79*, 6465–6569

Hanukoglu, I , and Fuchs, E (1982) The cDNA sequence of a human epidermal keratin divergence of sequence but conservation of structure among intermediate filament proteins Cell *31*, 243–252.

Heilig, R , Perrin, F , Gannon, F , Mandel, J L , and Chambon, P (1980) The ovalbumin gene family structure of the X gene and evolution of duplicated split genes Cell *20*, 625–637

Inana, G , Piatigorsky, J , Norman, B , Slingsby, C , and Blundell, T (1983) Gene and protein structure of a β-crystallin polypeptide in murine lens relationship of exons and structural motifs Nature *302*, 310–315

Kozak, M (1978) How do eucaryotic ribosomes select initiation in messenger RNA? Cell *15*, 1109–1123

Kozak, M (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes Nucl Acids Res *9*, 5233–5252

Kreis, T E , Geiger, B , Schmid, E , Jorcano, J L , and Franke, W W (1983) De novo synthesis and specific assembly of keratin filaments in nonepithelial cells after microinjection of mRNA for epidermal keratin Cell *32*, 1125–1137

Lazarides, E (1980) Intermediate filaments as mechanical integraters of cellular space Nature *283*, 249–256

Maxam, A M , and Gilbert, W (1980) Sequencing end labeled DNA with base-specific chemical cleavages Meth Enzymol *65*, 499–560

McLachlan, A D , and Stewart, M (1982) Periodic charge distribution in the intermediate filament proteins desmin and vimentin J Mol Biol *162*, 693–698

Messing, J , Crea, R , and Seeburg, P (1981) A system for shotgun DNA sequencing Nucl Acid Res *9*, 309–321

Miesveld, R , Krystal, M , and Arnheim, N (1981) A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ and β globin genes Nucl Acid Res *9*, 5931–5947

Mount, S M (1982) A catalogue of splice junction sequences Nucl Acid Res *10*, 459–472

Nishioka, Y , and Leder, P (1980) The organization and complete sequence of identical embryonic and plasmacytoma κ V region genes J Biol Chem *255*, 3691–3694

Nordheim, A and Rich, A (1983) The sequence (dC-dA)n (dG-dT)n forms left handed Z DNA in negatively supercoiled plasmids Proc Nat Acad Sci USA *80*, 1821–1825

Nordheim, A , Pardue, M L , Lafer, E M , Moller, A , Stollar, B D , and Rich, A (1981) Antibodies to left handed Z DNA bind to interband regions of drosophila polytene chromosomes Nature *294*, 417–422

Nordheim, A , Tesser, P , Azorin, F , Kwon, Y H , Moller, A , and Rich, A (1982) Isolation of Drosophila proteins that bind selectively to left handed Z DNA Proc Nat Acad Sci USA *79*, 7729–7733

Nudel, U , Katcoff, D , Zakut, R , Shani, M , Carmon, Y , Finer, M , Czosnek, H , Ginsburg, I , and Yaffe, D (1982) Isolation and characterization of rat skeletal muscle and cytoplasmic actin genes Proc Nat Acad Sci *79*, 2763–2767

Osborn, M , and Weber, K (1982) Intermediate filaments cell type specific markers in differentiation and pathology Cell *31*, 303–306

Proudfoot, N J , and Brownlee, G G (1976) 3′ non-coding region sequences in eukaryotic messenger RNA Nature *263*, 211–214

Quax, W J , Dodemont, H J , Lenstra, J A , Ramaekers, F , Soriano, P , van Workum, M E S , Bernardi, G , and Bloemendal, H (1982) Genes coding for vimentin and actin in mammals and birds Adv Exp Medicine and Biology *158*, 349–357

Quax Jeuken, Y , Quax, W , and Bloemendal, H (1983) Primary and secondary structure of hamster vimentin predicted from the nucleotide sequence Proc Nat Acad Sci USA *80*, 3548–3552

Rimm, D L , Horness, D , Kucera, J , and Blattner, F R (1980) Construction of coliphage lambda Charon vectors with Bam HI cloning sites Gene *12*, 301–309

Sanger, F , Coulson, A R , Barrel, B G , Smith, A J H , and Roe, B A (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing J Mol Biol *143*, 161–178

Shine, J , and Dalgarno, L (1975) Determinant of cistron specificity in bacterial ribosomes Nature *254*, 34–38

Singleton, C K , Klysik J , Stirdivant, S M , and Wells, R D (1982) Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions Nature *299*, 312–316

Slighlom, J L , Blecht, A E , and Smithies, O (1980) Human fetal $^{G}\gamma$ and $^{A}\gamma$ globin genes complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes Cell *21*, 627–638

Staden, R (1979) A strategy of DNA sequencing employing computer programs Nucl Acids Res *6*, 2601–2610

Staden, R (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences Nucl Acids Res *10*, 2951–2961

Steinert, P M , Idler, W W , and Goldman, R D (1980) Intermediate filaments of baby hamster kidney (BHK 21) cells and bovine epidermal keratinocytes have similar ultrastructures and subunit domain structures Proc Nat Acad Sci USA *77*, 4534–4538

Steinert, P D , Rice, R H , Roop, D B , Trus, B L , and Steven, A C (1983) Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments Nature *302*, 794–800

Tinoco, I , Jr , Borer, P N , Dengler, B , Levine, M D , Uhlenbeck, O C , Crothers, D M , and Gralla, J (1973) Improved estimation of secondary structure in ribonucleic acids Nature New Biol *246*, 40–41

Vorlickova, M , Kypr, J , Stokrova, S , and Sponar, J (1982) A Z like form of poly(dA-dC) poly(dG-dT) in solution? Nucl Acid Res *10*, 1071–1080

Weintraub, H (1983) A dominant role for DNA secondary structure in forming hypersensitive structures in chromatin Cell *32*, 1191–1203

Yaniv, M (1982) Enhancing elements for activation of eukaryotic promoters Nature *297*, 17–18

Zehner, Z E , and Paterson, B M (1983) Characterization of the chicken vimentin gene single copy gene producing multiple mRNAs Proc Nat Acad Sci USA *80*, 911–915

Zimmer, C , Tymen, S , Marck, C , and Guschlbauer, W (1982) Conformational transition of poly(dA-dC) poly(dG-dT) induced by high salt or in ethanolic solution Nucl Acid Res *10*, 1081–1091

Note Added in Proof

In a recent paper, Nature *303*, 674–679, 1983, Nordheim and Rich showed that certain sequences of the SV40 enhancer region can form Z-DNA

CHAPTER 4

INTERMEDIATE FILAMENT cDNAs FROM BHK-21 CELLS: DEMONSTRATION OF DISTINCT GENES FOR DESMIN AND VIMENTIN IN ALL VERTEBRATE CLASSES

Proc. Natl. Acad. Sci. USA 81, 5970-5974 (1984)

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5970–5974, October 1984
Biochemistry

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 NH_2 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.

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% $NaDodSO_4$ 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 μg of herring sperm single-stranded DNA per ml/0.1% $NaDodSO_4$ at 42°C. Washings were done with hybridization mix and, finally, with NaCl/Cit and $NaDodSO_4$. 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 ^{32}P-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.

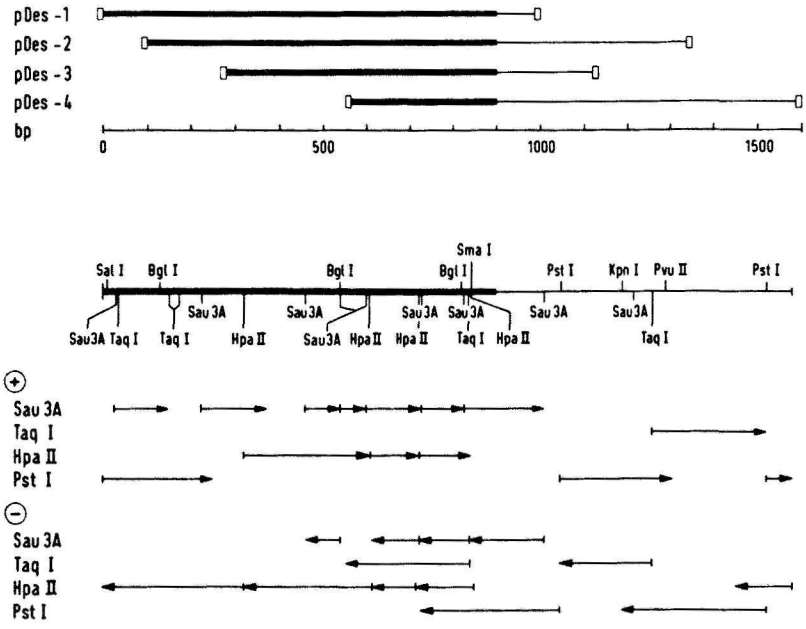


FIG. 1. Restriction map and sequencing strategy for pDes-1, pDes-2, pDes-3, and pDes-4. Upper lines indicate the length and relative positions of the inserts; thick bars represent coding region, and thin bars are the 3'-noncoding parts. Open bars at the ends represent the G-C tails that were introduced 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 polyadenylylation 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 ≈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 NH_2 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. 4*A*). 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 polyadenylylation signals are used for tran-

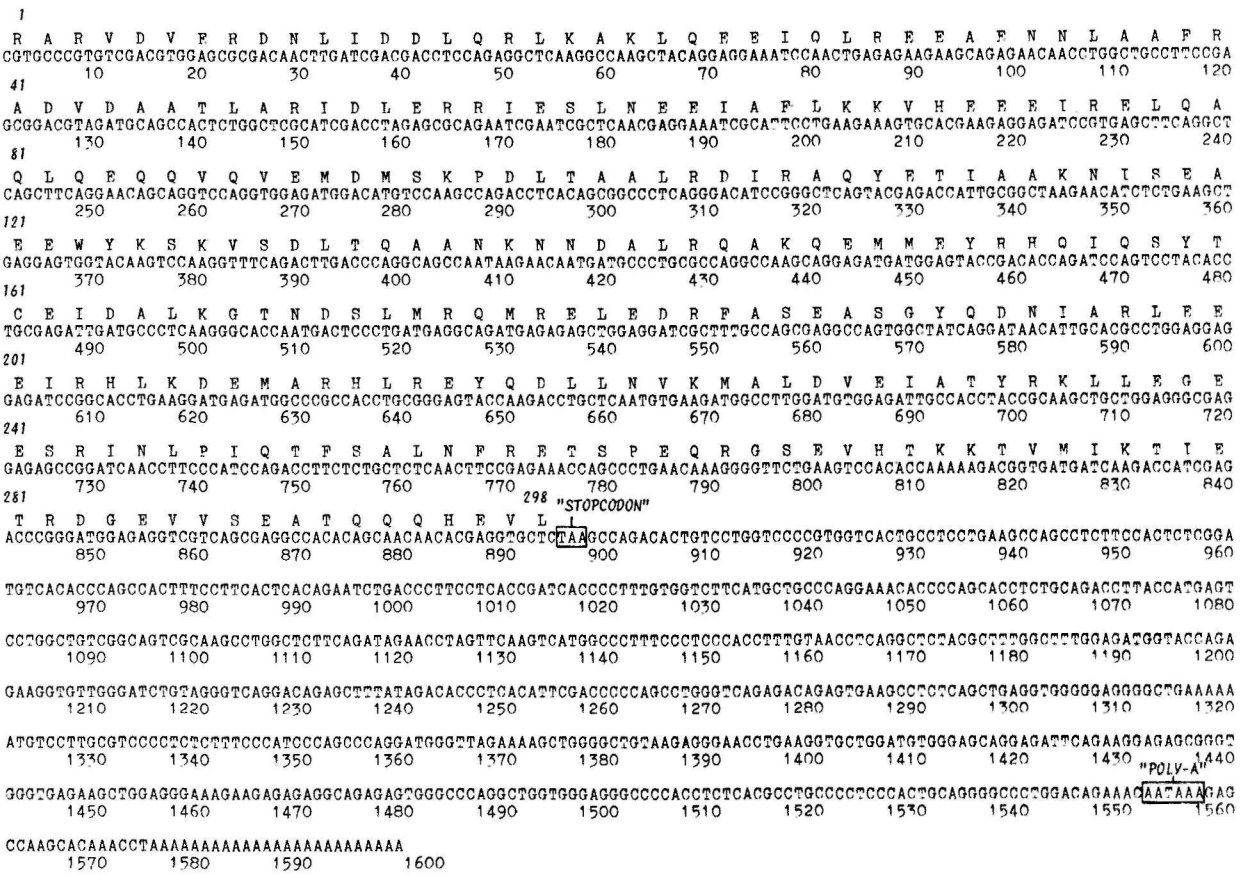


FIG. 2. Combined nucleotide sequence of pDes-1, pDes-2, pDes-3, and pDes-4. The predicted amino acids (represented by standard one-letter abbreviations) are shown above the middle of their triplets. The polyadenylylation signal and the stop codon are indicated.

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 *Eco*RI fragment and to a 3.4-kb *Bam*HI/*Eco*RI fragment (Fig. 4*B*). The 3′ probe (covering nucleotides 1061–1257) hyridizes exclusively to another *Eco*RI fragment, 8.2 kb long, and to a 5.4-kb *Bam*HI/*Eco*RI

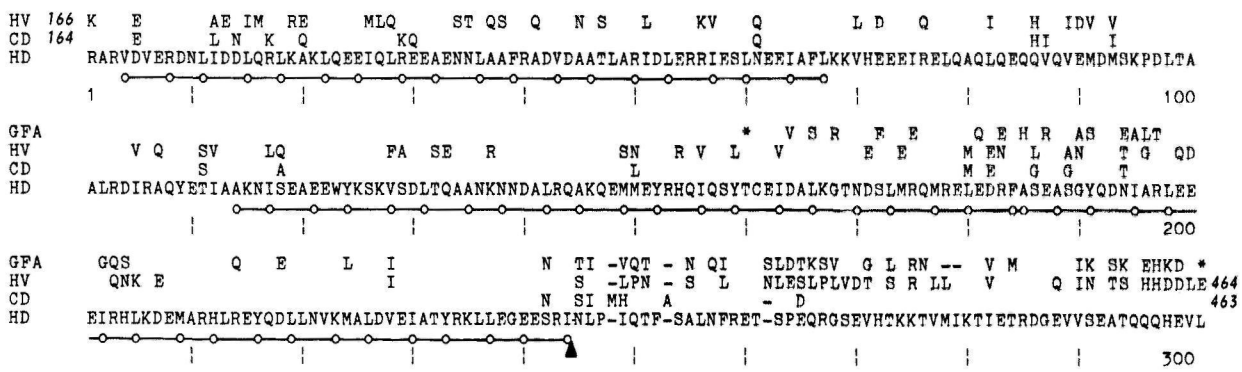


FIG. 3. Comparison of hamster desmin with chicken desmin (25), hamster vimentin (4), and the partial GFA sequence (8). The residues of the aligned sequences are only printed when they differ from hamster desmin. Hamster desmin residues are numbered on the bottom line; the other sequences are numbered 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.

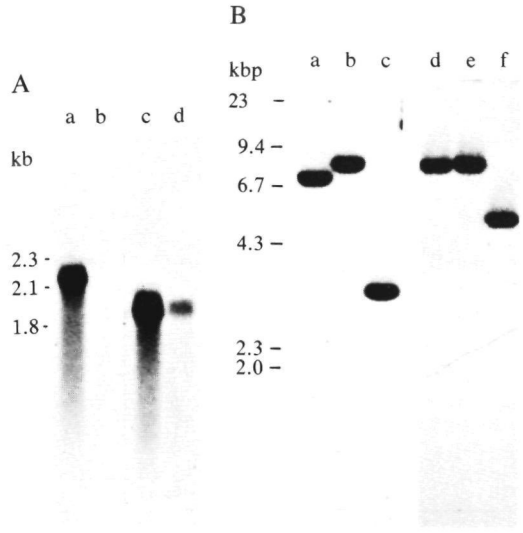


FIG. 4. (*A*) Size and expression of desmin and vimentin mRNA. RNA blots of 10 μg of BHK-21 poly(A)$^+$ RNA (lanes a and c) and 3 μg of lens poly(A)$^+$ RNA (lanes b and d) were hybridized with nick-translated pDes-1 (lanes a and b) and pVim-1 (lanes c and d). Desmin mRNA is only detected in BHK cells; 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× NaCl/Cit/0.1% $NaDodSO_4$ at 63°C for 30 min.

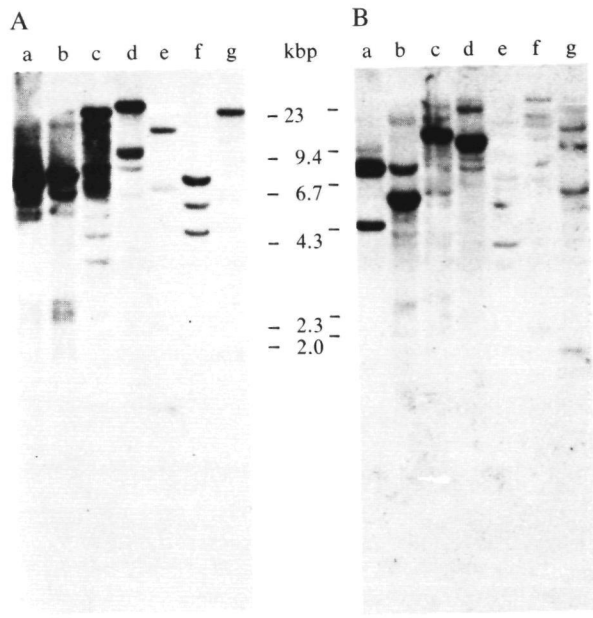


FIG. 5. Detection of desmin and vimentin sequences in vertebrate DNAs. Ten-microgram amounts of DNA were digested with *Eco*RI, run on 0.7% agarose gels, blotted onto nitrocellulose, and hybridized with the nick-translated insert of pDes-1 (*A*) and pVim-1 (*B*). The DNAs were extracted from the following 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× NaCl/Cit/0.2% $NaDodSO_4$ 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. 5*A*) and pVim-1 (Fig. 5*B*). 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. 5*A* (lane a) (hamster DNA hybridized with pDes-1) with Fig. 4*B* (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 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

CHAPTER 5

CHARACTERIZATION OF THE DESMIN GENE: EXPRESSION AND FORMATION OF DESMIN FILAMENTS IN NONMUSCLE CELLS AFTER GENE TRANSFER

Wim Quax, Leon van den Broek, Wilma Vree Egberts,
Frans Ramaekers* and Hans Bloemendal

*Department of Biochemistry and Pathology
University of Nijmegen

submitted for publication

SUMMARY

The structural organization of the gene encoding the intermediate filament protein desmin has been determined. The gene, 6.5 kb in length, consists of nine exons with a total length of 2169 nucleotides. The intervening sequences map at positions that fully correspond to those of the vimentin gene confirming the evolutionary relationship of these two intermediate filament genes. The derived complete primary structure for hamster desmin (468 amino acids; MW. 53250 daltons) reveals striking species variations in the NH_2 terminal domain of desmin. A plasmid containing the complete transcription unit of the desmin gene was transfected into hamster lens cells and into human epithelial (HeLa) cells. In both non-muscle cell lines the desmin gene was biologically active and the synthesized desmin assembled into authentic intermediate filaments as monitored by immunofluorescence using monoclonal and polyclonal antibodies to desmin. Double immunofluorescence staining with these antibodies and polyclonal or monoclonal antibodies to vimentin showed that the newly formed desmin filaments colocalize with the preexisting vimentin filaments. In a small subset of positive cells typical clumblike aggregates of staining, possibly originating from an excess synthesis of desmin, were observed.

INTRODUCTION

The intermediate filaments (IF) have been characterized as a unique type of cytoskeletal structures, which in different cell types are composed of different, hardly soluble, proteins. Immunological, biochemical and recently sequential data have allowed the distinction of five different classes of intermediate filament subunits, whose expression parallels embryological differentiation (for review see Franke et al.,1981; Osborn & Weber,1982). Epithelial cells contain cytokeratins, which can be subdivided into acidic (type I) and basic (type II) proteins; in neurones neurofilaments are found; glial filaments are present in astrocytes; vimentin is found in mesenchymally derived cells and in most cultured cell lines; finally desmin filaments are only found in muscle cells. Amino acid sequences derived for representatives of all classes (Hanukoglu & Fuchs,1982,1983; Geisler & Weber, 1982,1983; Geisler et al.,1983b,1984; Quax-Jeuken et al.,1983; Steinert et al.,1983) have allowed the definition of common structural principles for all IF. A mainly α-helical domain of about 320 residues capable of forming coiled coil interactions between two different molecules is flanked by non-helical NH_2 and COOH terminal domains, which are variable in length and sequence (for a review see Weber & Geisler,1984). The role of these variable domains in filament formation is hardly understood. In the α-helical core domain the non-epithelial IF proteins show a mutual amino acid homology of more than 50% (Geisler & Weber,1983; Quax et al.,1983,1984a; Lewis et al.,1984) while in comparison to the cytokeratins the homology is about 30%. Another aspect of distinction between epithelial and non-epithelial IF is the capability of the latter to form homopolymer filaments, whereas the formation of cytokeratin filaments requires at least one member of the acidic (type I) and one member of the basic (type II) subfamily (Quinlan et al.,1984).

In contrast to the vast information on protein structure, little is known about the structural organization of the genes encoding IF. Both the vimentin gene and the desmin gene have been shown to be present as a single copy in the hamster genome (Quax et al.,1983,1984a) and probably also in the chicken genome (Capetanaki et al.,1984). Structural analysis has revealed that the complete vimentin gene consists of nine exons. Two of the intervening sequences of the vimentin gene map at positions that delineate domains of the general IF model (Quax et al.,1983,1984b). To investigate the relation of the vimentin gene structure with other IF genes we have chosen to unravel the exon/intron pattern of the desmin gene. Furthermore as a first approach to resolve the transcription signals important for the tissue-specific desmin expression and with the aim to study the expression and assembly of desmin filaments in non-muscle cells or in cells of other species we have decided to use the cloned hamster desmin gene for transfer into heterologous cells. In addition the development of a system that enables the synthesis of desmin from the cloned gene is important in view of future experiments in which the effects of alterations in the coding information of IF genes on filament formation can be studied.

Isolation of the hamster desmin gene

DNA from an inbred strain of Syrian golden hamster was used to construct a gene library of partial Sau 3A fragments in the Charon 28 λ-vector (Rimm et al.,1980). Upon hybridization with the desmin cDNA clones pDes-1 and pDes-4 (Quax et al.,1984a) two positive phages could be isolated out of $5x10^5$ recombinants. Restriction enzyme and blotting analysis showed that the inserts of these two phages are overlapping (figure 1). Specific 5′ and 3′ probes (Quax et al.,1984a) were used to determine the transcriptional direction of the desmin gene. It was found that λ-HaDes1 contains the 5′ half of the desmin gene plus 11 kb of 5′upstream sequences. The second phage, λ-HaDes2, contains the complete desmin gene, about 6.5 kb in length, flanked by 7 kb 5′upstream and 0.5 kb 3′ downstream sequences. Since the desmin mRNA is about 2200 bases long (Quax et al.,1984a), one can conclude that about 70% of the 6.5 kb long hamster desmin gene consists of intervening sequences. To allow a more detailed study of the gene we made subclones of the large EcoR1 fragment (7.4 kb) and the flanking 3.8 kb EcoR1-Hpa1 fragment (Hpa1 cuts in the rightmost Charon 28 arm) of λ-HaDes2 in pBR322. These two subclones were used to investigate the transcriptional organization and the detailed intron/exon pattern of the desmin gene.

Nucleotide sequence of the desmin gene

To define the precise exon-intron organization of the gene and to unravel the structural basis of desmin expression we determined the nucleotide sequence of the major part of the gene. For this aim the above mentioned subclones of the 7.4 EcoR1-EcoR1 and the 3.8 EcoR1-Hpa1 fragments (figure 1) were submitted to a more detailed restriction enzyme analysis. A 3500 bp stretch of the 7.4 fragment and a 250 bp plus a 1300 base pair region of the 3.8 kb fragment (figure 1) were sequenced completely on both strands using the M13-dideoxy chain termination method (Sanger et al.,1980). Figure 2 depicts the sequences of the three regions. Comparison of the gene sequences with the previously reported desmin cDNA sequence (Quax et al.,1984a) allowed us to assign the different exons. The information of the mRNA was found to be interrupted by eigth intervening sequences in the genome. The precise localization of the exon-intron borders was determined by a detailed comparison of cDNA and

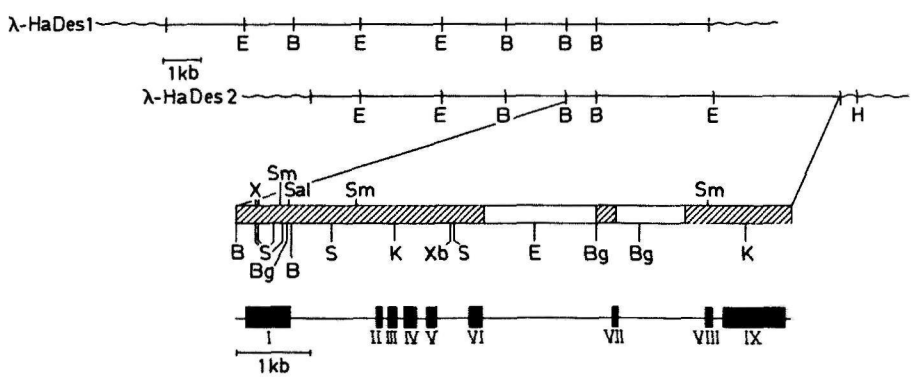


Figure 1

Restriction Map of the Hamster Desmin Gene

The two overlapping Charon phages carrying desmin gene sequences (λ-HaDes1 and λ-HaDes2) are shown in the upper two lines. The wavy lines represent Charon 28 sequences. The large (7.4 kb) central EcoR1-EcoR1 fragment of λ-HaDes2 and the 3.8 kb fragment ranging from the rigthmost EcoR1 site of λ-Hades2 to the Hpa1 site of Charon 28 were subcloned in pBR322. The thick bar is the region, which was analysed in detail with the striated parts representing the sequenced areas. The lowermost line is a schematic drawing of the desmin gene: the closed boxes representing the exons and the thin lines in between them the introns. Abbreviations of restriction enzyme names are: B= BamH1; Bg= BglII; E= EcoR1; H= Hpa1; K= Kpn1; S= Sac1; Sal= Sal1; Sm= Sma1; X= Xho1; Xb= Xba1.

genomic sequences. However, as the cDNA sequence lacked the 5′ portion of the mRNA, only the 3′ proximal nucleotides of the first exon were available. To circumvent this problem we compared the information preceding position 685 of figure 2A with the vimentin gene and with the chicken desmin amino acid sequence. This comparison made clear that exon I codes for the 180 NH_2 terminal residues of desmin. Furthermore S_1 nuclease mapping of hybrids formed between a ^{32}P-labeled fragment containing the BamH1-BamH1 fragment of 0.9 kb (figure 1) and desmin mRNA showed that the first exon is about 650 nucleotides in length (data not shown). The precise start of exon I (capping site) was determined by S_1 mapping of a ^{32}P-labeled M13 subclone, which covers the first 314 bases of the sequence as depicted in figure 2A, and desmin mRNA. The size of the protected fragment was 225 nucleotides (data not shown); we therefore marked the capping site at position 91 of figure 2A.

A schematic representation of the sizes of exons and introns and of their positions relative to the physical map is given in figure 1. The precise lengths of the exons in transcriptional 5′-3′ direction are: I: 656 bp; II: 61 bp; III: 96 bp; IV: 162 bp; V: 126 bp; VI: 221 bp; VII: 44 bp; VIII: 83 bp and IX: 720 bp. In figure 2 the amino acids encoded by the exons are shown above their triplets. Exon nine contains the information for the 677 nucleotides of the 3′ non-coding region of desmin mRNA (Quax et al.,1984a,b). As can be derived from figure 2 all introns with exception of number 6 and 7 have been sequenced completely. The sizes determined for the introns are: intron 1: 1179 bp; intron 2: 97 bp, intron 3: 119 bp; intron 4: 161 bp; intron 5: 443 bp; intron 6: 1.7×10^3 bp; intron 7: 1.25×10^3 bp and intron 8: 159 bp. No introns are located in the non-coding regions of the gene.

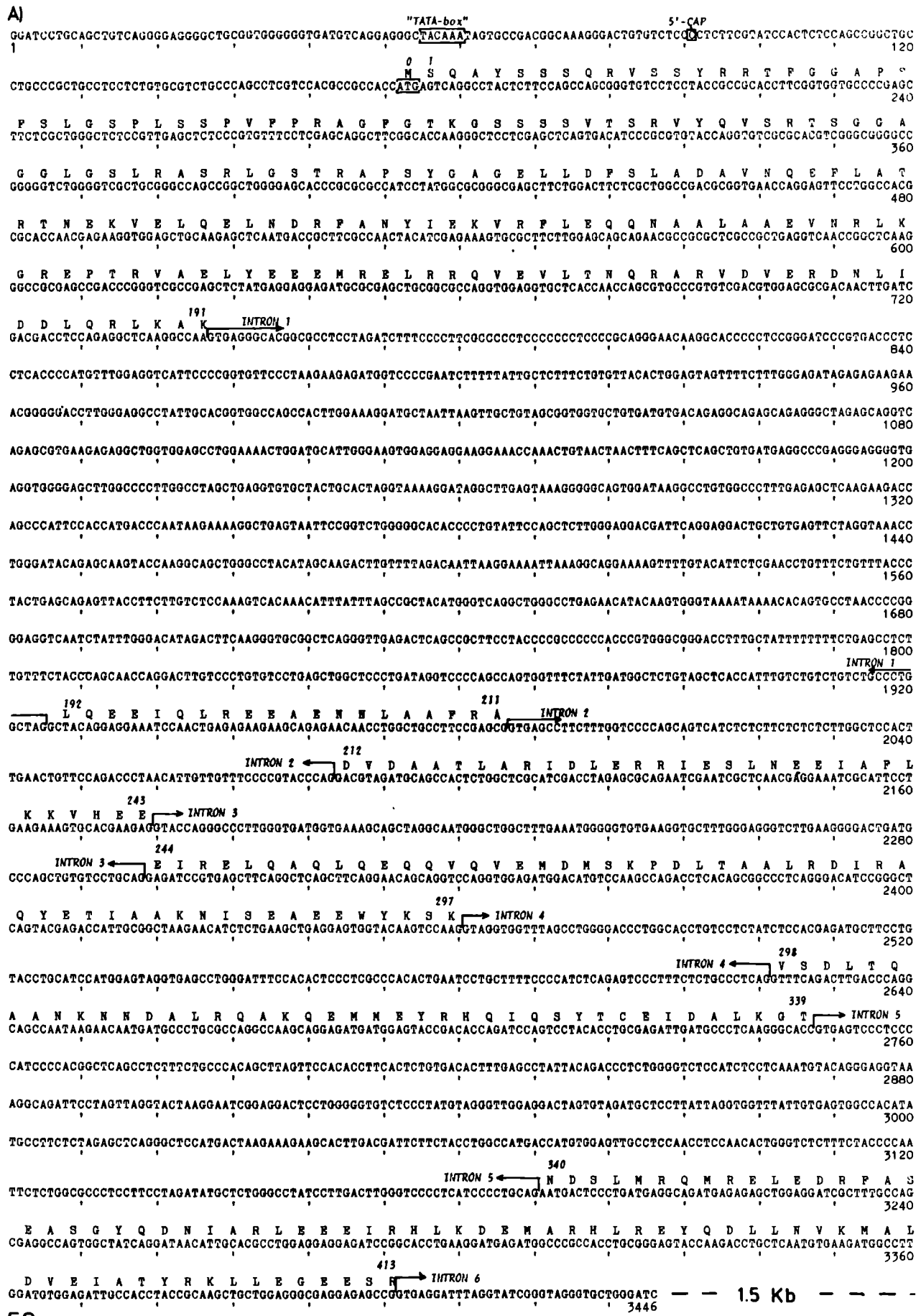
At position -36 to -31 upstream the cap site there is a sequence "TACAAA" which differs only in the third position from the consensus promotor sequence for eucaryotic genes (Breatnach & Chambon,1981). At position + 82 relative to the cap site the first AUG codon is found, which is the translation startcodon actually used by the desmin mRNA as can be judged from a comparison with the chicken desmin amino acid sequence (Geisler & Weber,1982). At the 3′ end of the gene we have found a consensus polyadenylation signal. The arrow at 17 nucleotide positions next to this sequence marks the position where in the desmin cDNA sequence a remnant polyA track is found (Quax et al.,1984a)

Full correspondence of position, but large variation in size of desmin and vimentin introns

The number of introns in the desmin gene is the same as has been reported for the vimentin gene (Quax et al.,1983). A detailed comparison of the exon-intron borders of these two genes shows that all intervening sequences map at precisely corresponding positions of the two proteins. In figure 3 the flanking sequences of the splice points are compared to each other and to a consensus 5′ and 3′ splice sequence (Mount,1982). The amino acids are shown above or below their corresponding exon sequence, respectively; numbering is relative to the first residue of the predicted complete primary structure (figure 2; Quax et al.,1983). As a consequence of the corresponding splice sites all exons of vimentin and desmin have the same size, apart from the first and the last exon, which harbor the non-coding regions, and exon 8, which encodes a single amino acid deletion of desmin as compared to vimentin (Quax et al.,1984a). A comparison of the splice points with the consensus splice border sequences shows that the intron bordering dinucleotides "GT" and "AG" are always present. The other nucleotides frequently differ from the consensus 5′ splice site, which is thought to interact with the 5′ end of U_1-snRNPs by basepairing (Kramer et al.,1984). In figure 3 the sizes of the vimentin and desmin introns are also compared. It is obvious that despite the correspondence of position, the sizes of the introns are completely different. Moreover a DIAGON based computer-search (Staden,1982) for homologies of intron sequences between the two genes failed to detect any significant sequence relationship.

The NH_2 terminal domain of desmin is higly variable in sequence

From a comparison of a number of different intermediate filament protein sequences it has emerged that the most pronounced common structural similarity of IF is formed by a mainly α-helical domain of about 320 residues, which enables coiled coil interactions between two polypeptide chains (for review see Weber & Geisler,1984). On both ends this α-helical domain is flanked by non-helical regions, which are called the variable domains of the intermediate filament proteins, be-


A)
"TATA-box"
5'-CAP
INTRON 1
INTRON 2
INTRON 3
INTRON 4
INTRON 5
INTRON 6
1.5 Kb

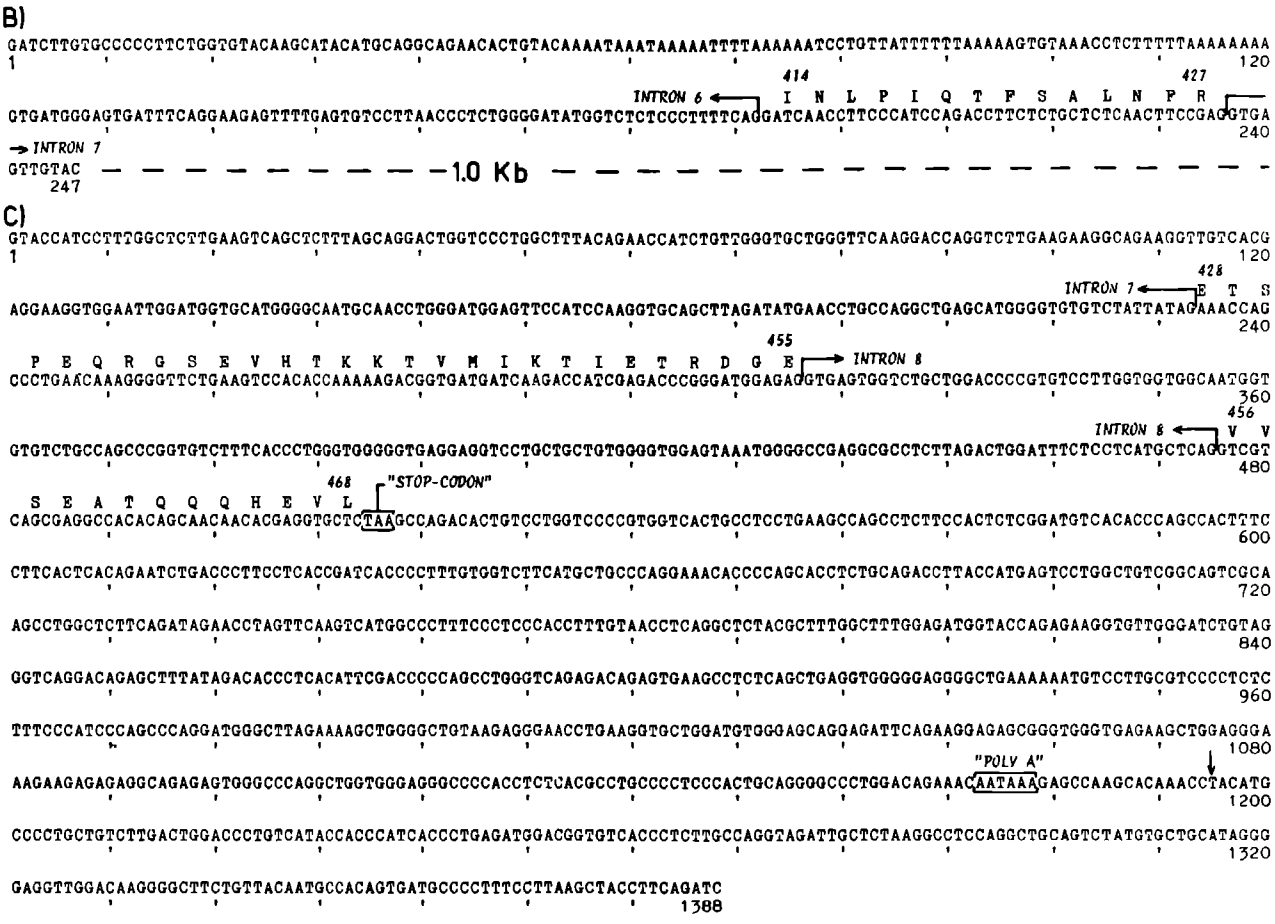


Figure 2

Nucleotide Sequence of the Desmin Gene

The complete sequence of the hamster desmin gene with exception of intron 6 and 7 is shown with 120 nucleotides per line. The intron borders are marked by arrows and the amino acids encoded by the exons are given in single letter codes above their codon. Lengths of intron fragments that were not sequenced are depicted between dotted lines. The putative promotor sequence TACAAA and the polyadenylation sequence are transcription signals. The 5′-cap was determined by S_1 analysis of hybrids formed between BHK-21 mRNA and an M13 clone carrying the first 314 nucleotides of sequence A). The arrow 17 bases next to the "poly A" signal marks the last unique nucleotide prior to the poly A track in the desmin cDNA sequence (Quax et al.,1984a,b). At the splice points the numbers of the corresponding amino acids in the desmin polypeptide are given. The C-residue at position 985 of C) (the 3′ non-coding region) represents the only difference between the coding part of the gene and the desmin cDNA sequence.

```
                C    a                                  ttttttttttt t
consensus        AGgt agt                                     n agG
                A    g                                  ccccccccccc c

                191                intron 1                        192
            Lys Ala Ly                                          s Leu Gln Glu
Des   CAAGGCCAAgtgagggcacgg          1.2Kb    gtctgccctggctagGCTACAGGAG
Vim   GCGAGAAAAgtaaggcctgat          0.8Kb    gtactactgcttcagATTGCAGGAG
            Arg Glu Ly                                          s Leu Gln Glu
                186                                             187

                211                intron 2                        212
            Phe Arg Ala                                         Asp Val Asp
Des   TTCCGAGCGgtgagccttctt          0.1Kb    tttccccgtacccagGACGTAGAT
Vim   TTCAGACAGgtttgtagccag          2.6Kb    ctctctttaaaacagGATGTTGAC
            Phe Arg Gln                                         Asp Val Asp
                206                                             207

                243                intron 3                        244
            His Glu Glu                                         Glu Ile Arg
Des   CACGAAGAGgtaccagggccc          0.1Kb    agctgtgtcctgcagGAGATCCGT
Vim   CATGATGAAgtaagtgatgtc          0.1Kb    ctgctttttcctcagGAGATCCAG
            His Asp Glu                                         Glu Ile Gln
                238                                             239

                297                intron 4                        298
            Lys Ser Lys                                         Val Ser Asp
Des   AAGTCCAAGgtaggtggttta          0.2Kb    tttctctgccctcagGTTTCAGAC
Vim   AAGTCCAAGgtatgaatgagc          0.7Kb    atttcttcctgacagTTTGCCGAC
            Lys Ser Lys                                         Phe Ala Asp
                292                                             293

                339                intron 5                        340
            Lys Gly Thr                                         Asn Asp Ser
Des   AAGGGCACCgtgagtccctcc          0.3Kb    cctcatcccctgcagAATGACTCC
Vim   AAAGGAACTgtgagtaccacc          0.5Kb    tctcccttcccacagAATGAGTCT
            Lys Gly Thr                                         Asn Glu Ser
                334                                             335

                413                intron 6                        414
            Glu Ser Ar                                          g Ile Asn Leu
Des   GGAGAGCCGgtgaggatttag          1.7Kb    tctctccctttttcagGATCAACCTT
Vim   GGAGAGCAGgtaggaaaggca          0.4Kb    tttgctttttttatagGATTTCTCTG
            Glu Ser Ar                                          g Ile Ser Leu
                408                                             409

                427                intron 7                        428
            sn Phe Arg G                                        Lu Thr Ser
Des   ACTTCCGAGgtgagttgtac          1.2Kb    gtgtgtctattatagAAACCAGCC
Vim   ACCTGAGAGgtaagc              0.8Kb    gctttttttaactcagAAACTAATC
            sn Leu Arg G                                        Lu Thr Asn
                422                                             423

                455                intron 8                        456
            Asp Gly Glu                                         Val Val Ser
Des   GATGGAGAGgtgagtggtctg          0.2Kb    tctcctcatgctcagGTCGTCAGC
Vim   GATGGACAGgttggtatcttt          0.7Kb    ttcctttttgagcagGTGATCAAT
            Asp Gly Gln                                         Val Ile Asn
                451                                             452
```

Figure 3

Exon-Intron Junctions of the Desmin Gene compared with the Vimentin Gene.

The nucleotides surrounding the splice points of the desmin gene (Des) and the vimentin gene (Vim) are compared to a consensus 5′ splice point (left) and a consensus 3′ splice sequence (right), which has been deduced by comparing a great number of intron borders (Mount,1982). Exon sequences are in capital letters, the number on top or below the amino acids represent the relative number in the desmin polypeptide or the vimentin polypeptide, respectively. The sizes of the introns are shown in the middle column. Apart from the dinucleotides "gt" and "ag" the homologies between intron border sequences of the two genes is low or absent.

cause they show little or no sequence homology between the different IF-subunits. Comparison of one and the same IF subunit in different species, however, showed that the COOH non-helical region can be equally well conserved during evolution as the helical region. For instance, the 55 COOH proximal residues of hamster and chicken desmin show only 10 % difference in sequence, which is about the same percentage as for the coiled coil domains (Quax et al.,1984a). The same phenomenon was observed by comparison of the carboxyterminal part of the bovine cytokeratin VII (Jorcano et al.,1984) with its human counterpart, the 50 kd cytokeratin (Hanukoglu and Fuchs,1983): number 14 of the human keratin catalog (Moll et al.,1982). With the completion of the hamster desmin sequence it is now possible to compare two complete sequences from one and the same IF subunit in two different species, namely those from hamster and chicken. In this comparison we paid special attention to the NH_2 regions of the two desmin sequences. In figure 4 we aligned the sequence encoded by exon I to the corresponding sequence of chicken desmin (Geisler & Weber,1982). In the α-helical region the two species show a high degree of homology (95%), but in the NH_2 terminal domain large differences exist. The most striking differences are represented by a stretch of 24 residues, which do not exhibit any homology at all and by a deletion of 9 amino acids in the bird protein as compared to the hamster desmin. These deviating regions of the two proteins might very well account for the observed differences in molecular weigth and antigenic behavior of chicken (463 residues) and mammalian (468 residues) desmin (Lazarides & Balzer,1978; Geisler & Weber,1980). In fact the comparison of the NH_2 termini of desmin and vimentin (Quax-Jeuken et al.,1983) also showed more differences than the alignment of the COOH region of these two different IF subunits. Interestingly, from figure 4 it can be deduced that despite the difference in primary structure, the distribution of charged residues in the NH_2 region, mainly arginines, is very similar in bird and hamster desmin, a finding that also holds true when desmin is compared with vimentin (Quax-Jeuken et al.,1983). In this context it should be mentioned that the head domain sequences of vimentin already showed substantial differences even between two mammalian species, namely pig and hamster (Geisler et al.,1983a).

Structural homologies in the 5′ upstream region of the hamster desmin and vimentin genes

To determine whether we could detect DNA sequences other than the "TATA" box that might play a role in the transcription of intermediate filament genes, we compared the upstream regions of the desmin and the vimentin gene. For this purpose we extended the sequence of the vimentin gene (Quax et al.,1983) and the sequence of figure 3 that far that we were able to compare the 200 proximal 5′ upstream nucleotides. In figure 5 they are aligned with the cap site of the desmin gene coinciding with the proposed cap site of the vimentin gene. A computer search revealed a number of short sequence identities (6-8 bases). However, one longer stretch of homology, 16 base pairs, was also found. In the desmin gene the 16 bp (boxed in figure 5) are located between the putative promoter "TACAAA" and the cap site; in the vimentin gene it is found at position -190 to -170 of the cap site. Another similarity in structure is stressed by a dotted line in figure 5. The region -70 to -50 from the cap is in both genes very rich in G residues. In addition positions -140 to -110 of both genes are mainly occupied by C residues. A consequence of this organization is the ability of both genes to form a strong secondary structure by binding of the -120 region to the -60 region due to a number of short inverted repeats present between these regions. Whether these observed structural similarities fulfill any function has to await future experiments. Especially the construction of suitable deletion mutants of this region and the subsequent influence on the expression of the cloned gene can give answers. The expression experiments described below may form an ideal model system to monitor the effect of different mutations of the 5′ upstream region.

Expression of the cloned desmin gene after transfer into hamster lens cells

The availability of a cloned gene not only allows one to determine its structure, but also opens the possibility to study the expression of the gene and the behavior of synthesized proteins after reintroduction of the DNA into living cells. With respect to the desmin gene we wondered whether the cloned DNA would be biologically active after transfection into cultured cells and if so, whether it would be possible to generate a desmin IF-cytoskeleton in cells that normally do not contain desmin filaments. We therefore constructed a

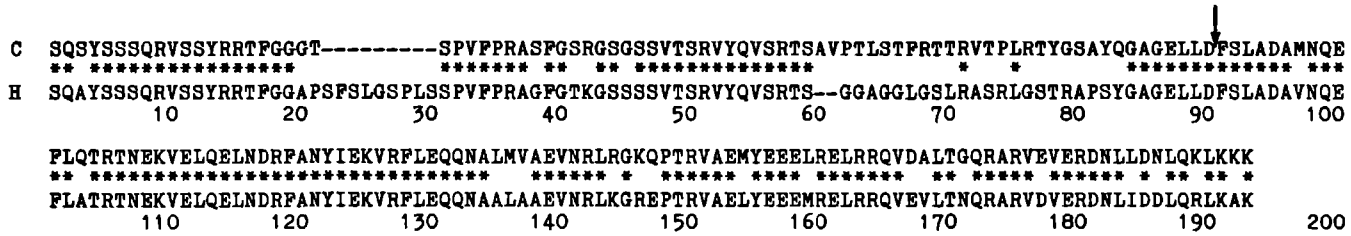


Figure 4

Comparison of the Desmin Head Domain in Hamster and Chicken

The amino acids encoded by the first exon of the hamster desmin gene (H) are aligned to the corresponding chicken desmin sequence (C). Optimal alignment was achieved with the aid of DIAGON (Staden,1982). Asterisks mark identical residues. The arrow indicates the border of the non-helical head domain and the α-helical middle domain of the general model of IF (Quax-Jeuken et al.,1983). Note that despite the large differences in the first 90 amino acids, the excess of arginine (R), serine (S) and threonine (T) residues has been maintained in both head domains.

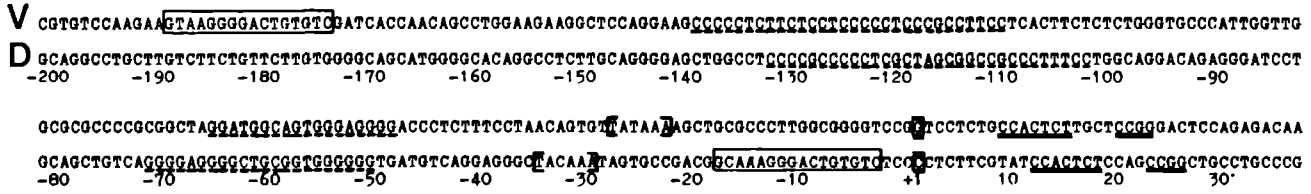


Figure 5

Similarities in the 5′ upstream Regions of the Desmin (D) and Vimentin (V) Genes

Of both genes the sequences of 200 nucleotides preceding the cap-site were determined, aligned and numbered relative to the capped nucleotide (position +1). The putative promotor sequences are placed in between brackets. Homologies in the 5′ non-coding region of the two IF mRNAs are stressed by double lines. An homologues 16-mer is emboxed. The dotted lines indicate the -120 region, which is rich in C-residues and the -60 region, which is rich in G-residues.

plasmid, which contains the 7.4 EcoR1-Eco1 and the 3.8 EcoR1-Hpa1 fragment of λ-HaDes2 together with the EcoR1-PvuII fragment carrying the origin and the ampicillin resistance gene of pBR322 (figure 6). The complete desmin transcription unit plus 3.5 kb of 5′ upstream sequences are present in this construct. A calciumphosphate precipitate containing this plasmid was transfected onto a monolayer of hamster lens cells, a culture previously shown to contain vimentin as their sole IF subunit (Bloemendal et al.,1980). Two days after transfection the cells were fixed on the cover slips, which were used as substratum, and incubated with antibodies directed against desmin. After staining with an appropriate fluorescein conjugated second antibody the monolayers were monitored for desmin positive cells. About 0.5-2 % of the cells showed the filament staining pattern typical for IF (figure 7). This percentage of positive cells is within the normal range for the calciumphosphate technique (Sompayrac & Danna,1981; Graham & van der Eb,1973). In control experiments without DNA or with plasmids without desmin sequences desmin positive cells were never observed. From the fotographs of figure 7b+c it is obvious that the newly synthesized desmin is perfectly able to form filaments. Using double immunofluorescence staining with polyclonal or monoclonal vimentin antibodies in combination with a monoclonal or polyclonal desmin antibody,respectively (figure 7d,e) we observed that the newly formed fibrillar desmin staining pattern colocalizes with the vimentin filaments, which are always present in the lens cells. This demonstrates that the newly synthesized desmin is fully able to form a heterogenous cytoskeleton with vimentin IF probably also existing of heteropolymer filaments as described for BHK-21 cells (Quinlan & Franke,1982). Next to the authentically looking desmin filaments in most cells, in a small subset of cells, however, we observed clumb-like aggregates which were intensively stained with the desmin antibody (figure 7f,g). The origin of these aggregates is unclear, but one could anticipate that the synthesis of desmin in those cells occurs at such high levels that the intermediate filament proteins precipitate. An indication for this hypothesis comes from the observation that the vimentin fluorescence of these cells showed exactly the same aggregated distribution (figure 7h,i). Furthermore, in some of the cells the aggregate was not concentrated in one clumb, but rather distributed in a number of dot-like structures (figure 7g).

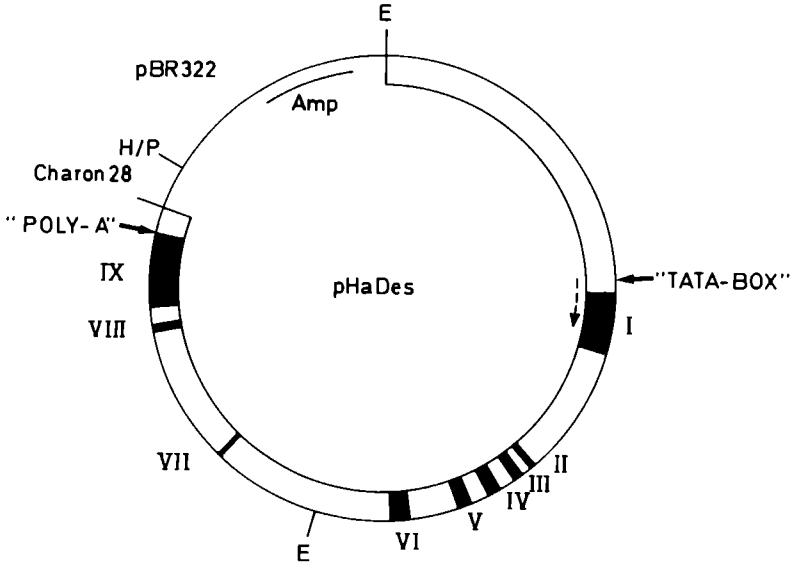


Figure 6

Structure of pHaDes: the Construction used for Gene Transfer

The 7.4 EcoR1-EcoR1 plus the 3.8 EcoR1-Hpa1 fragment of λ-HaDes2 were subcloned into pBR322, which was cut with the enzymes EcoR1 and PvuII. The hamster genomic fragment, represented by the thick bar, carries the complete desmin gene. "TATA-BOX" and "POLY-A" are the transcription start and stop signals of the desmin gene. The closed parts of the bar represent the exons (numbered I to IX). The pBR322 part of the construct carries the origin of replication and the ampicillin gene allowing the use of Escherichia Coli for propagation of the construct. "Charon 28" indicates the fragment of the rightmost arm of the gene cloning vector Charon 28, which was introduced by subcloning the 3.8 EcoR1-Hpa1 fragment of λ-HaDes2. The total size of pHaDes is 13.5 Kb.

Whether the dots in those cells originate from an excess synthesis or from an uncontrolled or ineffective filament assembly is unclear at this moment. The fact that desmin positive cells are in many instances found in clusters (figure 7d) in combination with the observation that occasionally cells seemed to go through some stages of mitosis, gives a strong indication that these hamster cells are still able to divide and multiply, independent of the synthesis and assembly of desmin polypeptides.

Expression of the hamster desmin gene in human epithelial (HeLa) cells

To test whether there is any species specificity in the transcription of the desmin gene or in the subsequent formation of desmin filaments, we decided to transfect the plasmid pHaDes into HeLa cells. These epithelial cells express cytokeratins number 7, 8, 17 and 18 (Moll et al.,1982) in addition to vimentin, but no desmin. Transfection of pHaDes resulted in about 0.5-2% of the cells to react positively with desmin antibodies after two days (figure 8a,b), indicating that there is no species barrier for the transcription signals of the hamster gene. The desmin staining reaction shows a similar fibrillar pattern as in the hamster lens cells (figure 8b). Double immunofluorescence of the same cell with a monoclonal desmin and a polyclonal vimentin antibody (figure 8c,d) shows that at least the major part of the newly synthesized desmin subunits become completely integrated in the preexisting vimentin cytoskeleton of HeLa cells. When aggregation or precipitation of desmin IF was seen, also the vimentin IF could be shown to occur (partly) in an aggregated form (figure 8e,f), strengthening the hypothesis that the aggregates in these cells result from precipitation of an overwhelming amount of synthesized desmin protein forming a coprecipitate with the already present vimentin molecules.

Since no cell type has been described to express both desmin and cytokeratins so far, we wondered how the two types of filaments would interact in the manipulated HeLa cells. Therefore we applied the double-label immunofluorescence technique with the polyclonal desmin antiserum and the monoclonal antibody to cytokeratin 18. As can be seen in figures 8g-j the desmin fluorescence does not show the same overall distribution pattern as the cytokeratin fluorescence, although in some regions both types of IF seem to occupy the same positions. Due to the small size of the cells, however, it is difficult to determine whether it concerns true colocalization in these regions. In additional experiments, where we used the monoclonal desmin antibody in combination with a polyclonal keratin antiserum, the two types of filaments did not seem to be precisely colocalized either. Finally, also desmin producing HeLa cells seem to be able to go through mitosis, as can be concluded on the basis of the same arguments as for lens cells.

DISCUSSION

Organizational similarities of the desmin gene and the vimentin gene

The most striking conclusion that can be drawn from the elucidaton of the desmin gene structure is the finding that the intron positions fully correspond to those of the vimentin gene. This similarity in organization between the two genes forms strong evidence for previous suppositions that both genes arose from divergent evolution of a duplicated ancestor gene. It can now be anticipated that the ancestor gene at the time of duplication, which is thought to have taken place before the divergence of vertebrate species (> 500 million years ago; Quax et al.,1984a), displayed the same exon-intron pattern as the desmin and vimentin gene occuring in the hamster species. Consequently it can be concluded that the exon-intron organization of desmin and vimentin genes in most, if not all, vertebrate species will be the same or very closely related to the two hamster genes. Moreover, since the glial filament acidic protein (GFAP) shows a homology to desmin and vimentin of 60 %, which is only sligthly less than the homology between desmin and vimentin (65%), we expect the exon-intron pattern of the GFAP gene to be comparable to that of the ancestor gene of desmin and vimentin. In fact some recent data on keratin genes point to a conservation of at least some intron positions among all IF genes. For instance, the bovine gene for keratin Ia contains 8 introns, one of which was precisely mapped by sequencing. This intron (number 7) was located exactly at the position that corresponds to intron 6 of the desmin gene (Lehnert et al.,1984). In another recent report it was mentioned that intron 9 of the gene for human keratin 67 kD maps at a position comparable to intron 6 of the desmin and vimentin gene (Johnson et al.,1984). Moreover in a report which appeared very shortly before submission of this paper the human keratin gene encoding the 50 kD keratin was shown to be in-

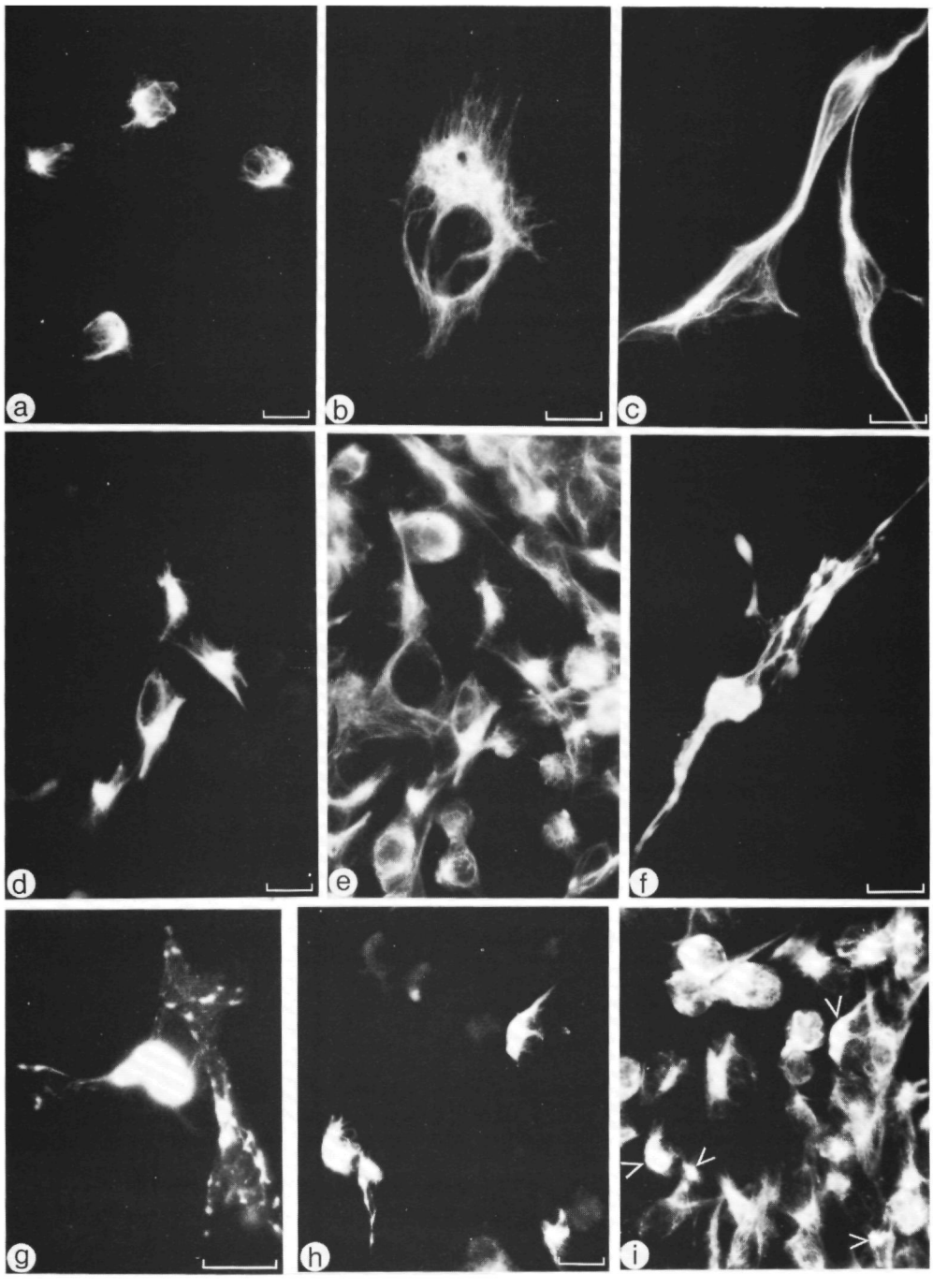


Figure 7

Expression and assembly into intermediate filaments of desmin in cultured hamster lens cells, 45 hours after transfection with the cloned desmin gene.

a-c) Immunofluorescence microscopy showing synthesis of desmin and assembly into authentic intermediate filaments, as monitored with the polyclonal desmin antiserum.

d,e) Double-label immunofluorescence microscopy with the monoclonal antiserum (RD301;d) and the polyclonal antivimentin (e) showing complete colocalization of both IF protein types.

f,g) Occasionally, desmin positive immunofluorescence is observed in large, clumb-like aggregates or in smaller dot-like (g) structures (polyclonal antidesmin).

h,i) Double-label immunofluorescence microscopy showing perfect matching of desmin (as stained with RD301) and vimentin (stained with polyVim) immunofluorescence in intermediate filament aggregates (indicated by arrowheads in i).

Bars: 20 µm.

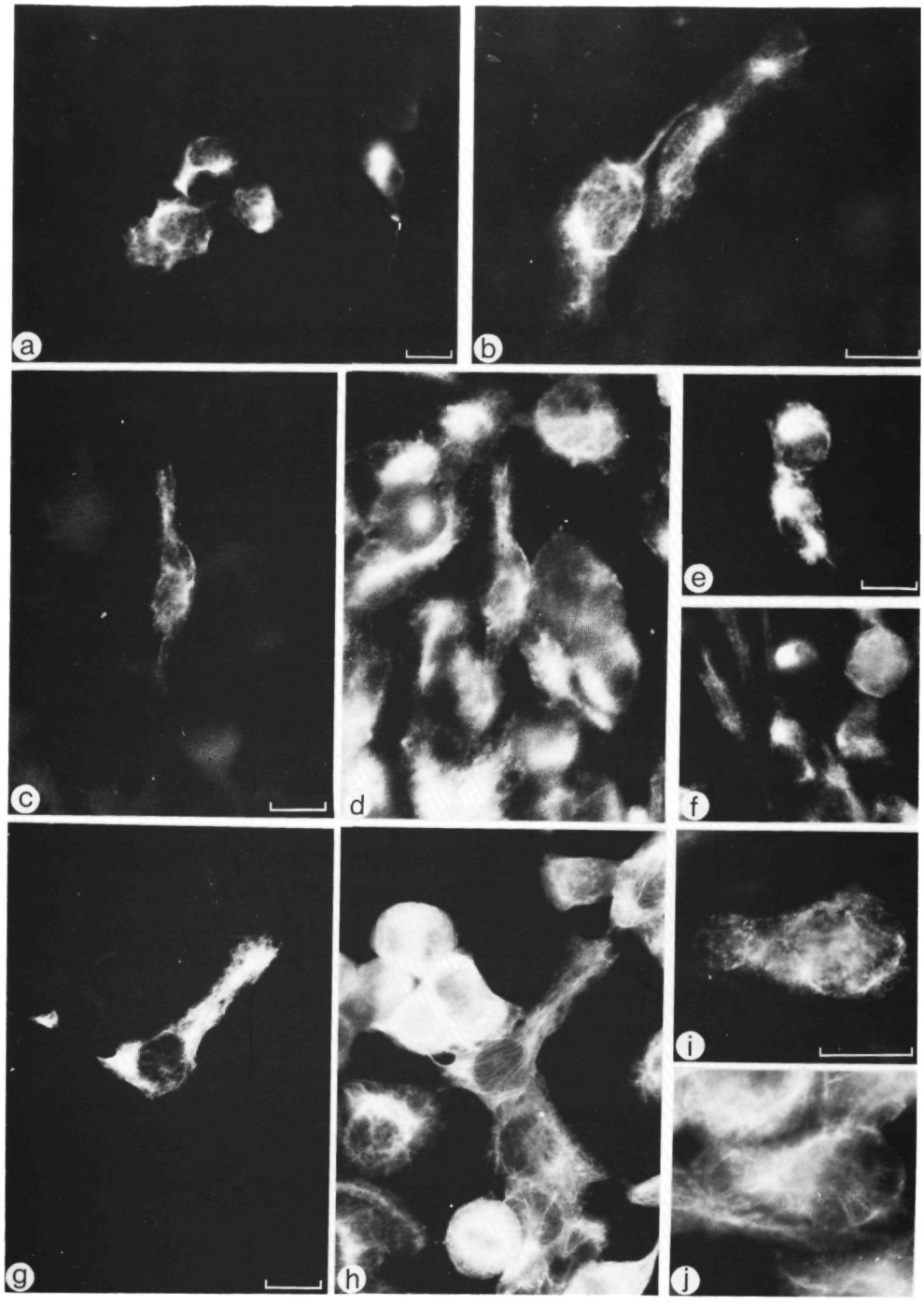


Figure 8

Expression and assembly of desmin into filaments in HeLa cells, 45 hours after transfection with pHaDes

a,b) Single-label immunofluorescence microscopy of desmin producing HeLa cells, as detected by the polyclonal desmin antiserum (polyDes). Note fibrillar organization in b).

c-f) Double-label immunofluorescence microscopy showing desmin intermediate filaments (c,e) and colocalization with vimentin (d,f) both in network structures (c,d) as well as in aggregates (e,f). Antibodies used were monoDes, RD301 (c,e) and polyVim (d,f).

g-j) Micrographs demonstrating that no real spatial relationship exists between the newly made desmin (g,i; stained with polyDes) and the HeLa cytokeratin filaments (h,j; stained with the monoclonal antibody to cytokeratin 18, RGE 53)

Bars: 20 µm.

terrupted by 7 introns with five of them occuring at positions that are fully corresponding to introns of the vimentin gene (Marchuk et al.,1984). A conservation of intron positions over a long period during evolution is also found in the globin gene family. The active α-globin and β-globin genes, which are thought to have diverged about 500 million years ago, all posses two intervening sequences at corresponding positions (Efstratiadis et al.,1980). On the other hand the intron positions of the closely related members of the actin gene family show variations among different species and among different isoforms within one species (Ueyama et al.,1984; Gallwitz & Sures,1980; Fyrberg et al.,1981; Nudel et al.,1982). However, despite the close sequence relationship of actins there is good evidence that the different actin genes diverged early in evolution and that they have evolved independently for a long time (for review see Firtel,1981). Therefore the diversity of intron positions in the actin gene family as compared to vimentin and desmin might be just a reflection of the longer evolutionary history of actin genes as compared to that of the two IF genes. In contrast to their positions, the lengths of the introns and the nucleotides comprised within them do not show apart from the splice points any relation between the two IF genes. This feature has been observed for other gene families (for a review see Breatnach & Chambon,1981).

All intervening sequences of the desmin gene are located in the protein coding region. If we look for a correlation between gene structure and protein domain structure (Gilbert,1978; Blake,1978) we find only introns 3 and 6 located at putative protein domain borders, as has been discussed for vimentin previously (Quax et al.,1983). Surprisingly precisely the same position as intron 6 is also occupied by an intron in a bovine (Lehnert et al.,1984), a mouse (Johnson et al.,1984) and a human cytokeratin gene (Marchuk et al.,1984).

Apart from the protein coding nucleotides, the only sequence similarities of the desmin gene compared with the vimentin gene are located in the 5′ upstream region with the most pronounced homology formed by a 16 bp fragment of which 14 nucleotides are identical. Such a conserved sequence element might be a good candidate for an enhancer type structure, although a comparison of the 16 bp with a consensus enhancer "core" sequence (Weiher et al.,1983) did not show homology. The similarities in the -60 region, which are very G rich, and the -120 region, which are very C rich, are noteworthy. Especially the capability of these two regions to interact with each other to form a strong secondary structure on one strand of the DNA could point to a mechanism through which this region might be involved in the start of transcription. It should, however, be kept in mind that not identities, but rather differences in regulation sequences must be responsible for the quite different expression patterns exhibited by the two genes.

The NH_2 domain of one and the same IF type shows many interspecies variations

The comparison of the complete primary structure of desmin from two representatives of different vertebrate classes (this report,figure 4; Quax et al.,1984a; Geisler and Weber, 1982) makes clear that the NH_2 non-helical domain shows much more interspecies variation than the rest of the protein, including the COOH non-helical domain. Since the same tendency can also be observed when comparing hamster vimentin with porcine vimentin (Geisler et al.,1983a) or mouse GFAP (Lewis et al.,1984) with porcine GFAP (Geisler and Weber,1983) one may wonder if there are any sequence requirements for this NH_2 domain of IF at all. However, a closer inspection of the head pieces of non-epithelial IFs shows that all of these proteins contain an excess of arginine residues and hydroxyl amino acids in combination with a virtual lack of acidic residues. Precisely these elements are not changed between chicken and hamster desmin. Interestingly filament formation of non-epithelial IFs seems to be dependent on the non-helical flanking domains (Geisler & Weber,1982) and consequently on the observed abundance of basic residues in the NH_2 sequence of different subunits. For the cytokeratins the sequence requirements in the head domain seem to be different (Steinert et al.,1983,1984). This might be reflected in the fact that cytokeratins can only form heteropolymer filaments in contrast to the non-epithelial IF.

Transfer of cloned IF genes as a novel system to manipulate the cytoskeleton

Several techniques have been developed to study the distribution and dynamics of a variety of cytoskeletal proteins in the living cell. Microinjection of antibodies against cytoskeletal proteins and microinjection of fluorescently labeled proteins have been used to manipulate the cytoskeleton (Lin & Feramisco,1981). Purified intermediate fi-

laments, however, are practically insoluble in physiological buffers and can therefore not be introduced into the cell by microinjection. Hence, the microinjection of total mRNA was chosen as a different approach to generate IF cytoskeletons in a heterologues cell type. It was shown that authentic cytokeratin filaments can arise from epidermal mRNA injected into lens cells (Kreis et al.,1983), which normally do not express cytokeratins, and into non-epidermal epithelial cells (Franke et al.,1984). A conclusion from these experiments was that the translation of epidermal keratin mRNA functions well in non-epidermal cells and that the assembly of intermediate filments is not dependent on the cell type once the proteins are formed.

We think that the experiments described in this report form an important extension to the methods to experimentally alter the IF-cytoskeleton. It can be concluded from the ability of the formed protein to react with both monoclonal and polyclonal desmin antibodies, in combination with the observed correct formation of desmin filaments that the product formed by the transferred gene is authenthic desmin. Furthermore, this conclusion is strongly supported by our observations in the double-label experiments with vimentin and cytokeratin antibodies. In accord with findings by other authors (see for example Quinlan & Franke,1982) we have found that the desmin formed in both cell types used in this study colocalizes almost completely with the vimentin filament network. In the HeLa cells, however, it can be stated that in general no or to a lesser degree colocalization of cytokeratin and desmin in filaments is observed. Focal alignments or intermingling of these two types of IF systems may occasionally be observed, but such an apparent spatial relationship may be explained by crowding of filaments of different types within a thin rim of cytoplasm as it is often observed in the cells used in this study.

One advantage of the use of a cloned gene is the ability to study the expression and filament formation of one single subunit at a time with the profitable circumstance of not introducing unknown and possibly interacting proteins, which is the case when using an impure mRNA mixture. Therefore we conclude that in our experiments the assembly of the de novo synthesized desmin molecules into filaments is not dependent on muscle-specific accessory proteins. Secondly the expression of the cloned desmin gene makes clear that there is no blockade on the transcriptional level as well as on the translational level in heterologues cells as well as in a heterologues species. This apparant notobeyance of a gene to the normally observed tissue-specificity has also been found for other genes in a transient expression assay (Gunning et al.,1984). Probably the regulation of gene expression is not functioning absolutely when a great number of episomal DNA molecules carrying the gene are present in the nucleus.

A third and in the future the most exciting advantage of using a cloned gene is the possibility to engineer the DNA in vitro prior to transfection. This will enable the construction of genes that express altered intermediate filament proteins and to study the effects of these alterations on filament formation. Ultimately the sequence requirements for the structural domains of IF, which now have been deduced merely from comparison of different sequences, can be experimentally tested.

MATERIALS AND METHODS

Construction and screening of the hamster gene library

Total DNA was isolated from the liver of an inbred Syrian golden hamster strain, which was obtained from the Animal Center of the University of Nijmegen. A partial Sau3A digestion was carried out for 5, 10 and 20 minutes and the resulting fragments were run in an agarose gel. A size selection from 15 to 20 kb was carried out on all three digests to ensure a good representation of all parts of the genome. The combined DNA was ligated, in vitro packaged and plated with Charon 28 as vector (Rimm et al.,1980). Replica filters were hybridized with ^{32}P-labeled pDes-1 and pDes-4 (Quax et al.,1984) under conditions of 50% formamide, 5xSSC, 1xDenhardt's solution, 20 mM sodium phosphate (pH 6.8), 5 mM EDTA, 100 µg/ml single stranded herring DNA, 0.1% SDS at 42 °C. The resultant positives were plaque purified and further analysed by restriction enzyme and blotting analysis of the recombinant phage DNA.

Sequence analysis and comparison

Inserts of the already mentioned subclones 7.4 EcoR1- EcoR1 and the 3.8 EcoR1- Hpa1 were digested with the enzymes Sau3A, Taq1, HpaII, HaeIII, AluI or RsaI. Resulting fragments were ligated into a suitable restriction site of Mp10 or Mp11 and transfected on JM101 or JM103 (Messing et al.,1981). Plaque filters were prepared and hybri-

dized with the cDNA probes. Most coding sequences could be detected in this way. Remaining gaps in the sequence were, if necessary, closed by guided subcloning of suitable restriction fragments into an M13 vector. Isolation of single stranded DNA and dideoxy chain termination reactions were performed as described (Sanger et al.,1980). Analysis was done on 40 cm 6% sequencing gels. The gel readings were recorded, edited and compared by the use of programs of Staden (1979). Search for optimal alignment was done by the program DIAGON (Staden,1982).

DNA transfections

Hamster lens cells (Bloemendal et al.,1980) and HeLa cells were grown for two days on coverslips in 6 well-dishes in DMEM supplemented with 10% foetal calf serum. Coverslips with a cell density of about 30 to 50% were exposed to 1 ml of a $Ca_3(PO_4)_2$ precipitate including the appropriate DNA at a density of 10 µg/ml. After 20 minutes of incubation cells were supplemented with 5 ml DMEM containing 10% foetal calf serum. Following a 6 hour incubation cells were glycerol shocked (15 % glycerol in HBS buffer) during 3 minutes and then allowed to grow for 24 to 48 hours. After this period cells had mostly grown to subconfluency. Prior to fixation in methanol, the dishes were washed 3 times with PBS.

Antibodies and immunofluorescence

The following antibody preparations were used in this study:

1). A polyclonal antiserum to chicken gizzard desmin (polyDes) raised in rabbits (Ramaekers et al.,1983a,b) absorbed on keratins. This antibody only stains striated- and smooth muscle cells and a filamentous network in cultured BHK-cells. Bovine lens cells and HeLa cells have been shown to be negative. Immunoblotting assays have demonstrated that this antiserum only reacts with desmin and not with any of the other IF-proteins.

2). A mouse monoclonal antibody to chicken gizzard desmin, RD301, which showed a staining pattern in BHK-cells virtually identicall to that of the polyclonal desmin antiserum. Bovine lens cells, fibroblasts and HeLa cells were negative. In immunoblotting assays this antibody was shown to react exclusively with the desmin protein band in cytoskeletal preparations from BHK-cells, smooth muscle and a human leiomyosarcoma.

3) A polyclonal antiserum to bovine lens vimentin (Ramaekers et al.,1981,1983c). This antibody does not stain striated muscle cells, while in immunoblotting it does not crossreact with desmin.

4) A mouse monoclonal antibody to bovine lens vimentin, monoVim (Euro-Diagnostics, Apeldoorn, The Netherlands), which in immunoblotting assays only reacts with vimentin and not with desmin.

5) An affinity purified rabbit antiserum to human skin keratins that reacts specifically with most epithelial tissues and epithelial cultured cells (Ramaekers et al.,1981).

6) A mouse monoclonal antibody to cytokeratin 18, RGE53 (Ramaekers et al.,1983d,1985), which in immunoblotting assays only recognizes cytokeratin 18 and not vimentin or desmin.

As second antibodies we have used: 1) Goat-anti-rabbit IgG conjugated with fluoresceine, FITC (Nordic, Tilburg, The Netherlands). 2) Sheep anti-mouse Ig conjugated with Texas-Red (New England Nuclear, Boston, Mass.). 3) Goat anti-mouse IgG conjugated with FITC (Nordic).

Single and double labelling indirect immnunofluorescence assays (including appropriate controls) were performed on fixed cells essentially as described before (Ramaekers et al.,1983a,e), however with the exception that for double labelling experiments primary antibodies were applied simultaniously to the cell preparations for about 1 hour.

ACKNOWLEDGEMENTS

We thank Drs. R. van den Heuvel and Y. Quax-Jeuken for stimulating discussions and commment on the manuscript. Anita Huysmans and Gert Schaart are acknowledged for excellent technical assistance in the preparation and testing of the antibodies. The present investigations have been carried out in part under the auspicies of the Netherlands Foundation for Chemical Research (SON) with financial aid of the Netherlands Organization for the Advancement of Pure Research (ZWO), and the Netherlands Cancer Foundation, The Queen Wilhelmina Fund (KWF).

REFERENCES

- Blake, C.C.F. (1978). Do genes-in-piece imply proteins-in-piece? Nature 273, 267.

- Bloemendal, H., Lenstra, J.A., Dodemont, H., Ramaekers, F.C.S., Groeneveld, A.A., Dunia, I., & Benedetti, E.L. (1980) SV-40 transformed hamster lens epithelial cells: a novel system for

the isolation of cytoskeletal messenger RNAs and their translation products. Exp. Eye Res. 31, 513-525.

- Breathnach, R., & Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

- Capetanaki, Y.G., Ngai, J., Lazarides, E. (1984) Characterization and regulation in expression of a gene coding for the intermediate filament protein desmin. Proc. Natl. Acad. Sci. USA 81, 6909-6913.

- Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connel, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., & Proudfoot, N.J. (1980). The structure and evolution of the human β-globin gene family. Cell 21, 653-668.

- Firtel, R.A. (1981). Multigene families encoding actin and tubulin. Cell 24, 6-7.

- Franke, W.W., Schmid, E., Schiller, D.L., Winter, S., Jarasch, E.D., Moll, R., Denk, H., Jackson, B.W., & Illmensee, K. (1981). Differentiation-related patterns of expression of proteins of intermediate-size filaments in tissues and cultured cells. Cold Spring Harbor Symp Quant Biol 46, 431-454.

- Franke, W.W., Schmid, E., Mittnacht, S., Grund, C., & Jorcano, J. (1984) Integration of different keratins into the same filament system after microinjection of mRNA for epidermal keratins into kidney epithelial cells. Cell 36, 813-825.

- Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S., & Davidson, N. (1981) The actin genes of Drosophila: protein coding regions are highly conserved but intron positions are not. Cell 24, 107-116.

- Gallwitz, D., & Sures, I. (1980). Structure of a split yeast gene: Complete nucleotide sequence of the actin gene in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 77, 2546-2550.

- Geisler, N., & Weber, K. (1980) Purification of smooth-muscle desmin and a protein-chemical comparison of desmins from chicken gizzard and hog stomach. Eur. J. Biochem. 111, 425-433.

- Geisler, N., & Weber, K. (1982). The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins. EMBO J. 1, 1649-1656.

- Geisler, N., & Weber, K. (1983). Amino acid sequence data on glial fibrillary acidic protein (GFA); implications for subdivision of intermediate filaments into epithelial and non-epithelial members. EMBO J. 2, 2059-2063.

- Geisler, N., Plessmann, U., & Weber, K. (1983a). Amino acid characterization of mammalian vimentin, the mesenchymal intermediate filament protein. FEBS Lett. 163, 22-24.

- Geisler, N., Kaufmann, E., Fisher, S., Plessmann, U., & Weber, K. (1983b). Neurofilament architecture combines structural principles of intermediate filaments with carboxy-terminal extensions increasing in size between triplet proteins. EMBO J. 2, 1295-1302.

- Geisler, N., Fisher, S., Vandekerckhove, J., Plessmann, U., & Weber, K. (1984). Hybrid character of a large neurofilament protein (NF-M): intermediate filament type sequence followed by long and carboxy-terminal extension. EMBO J. 3, 2701-2706.

- Gilbert, W. (1978). Why genes in pieces? Nature 271, 501.

- Graham, F.L., & van der Eb, A.J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52, 456-467.

- Gunning, P., Ponte, P, Kedes, L., Hickey, R.J., Skoultchi, A.I. (1984) Expression of human cardiac actin in mouse L cells: a sarcomeric actin associates with a nonmuscle cytoskeleton. Cell 36, 709-715.

- Hanukoglu, I., & Fuchs, E. (1982). The cDNA sequence of a human epidermal keratin: divergence of sequence but conservation of structure among intermediate filament proteins. Cell 31, 243-

252.

- Hanukoglu, I., & Fuchs, E. (1983). The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins. Cell 33, 915-924.

- Johnson, L., Idler, W., Zhou, X-M., & Steinert, P. (1984). Structure of the gene for the human epidermal keratin of 67 kD. J. Cell. Biol. 99, 317a.

- Jorcano, J.L., Rieger, M., Franz, J.K., Schiller, D.L., Moll, R. & Franke, W.W. (1984). Identification of two types of keratin polypeptides within the acidic cytokeratin subfamily I. J. Mol. Biol. 179, 257-281.

- Krämer, A., Keller, W., Appel, B., Luhrmann, R. (1984). The 5′ terminus of the RNA moiety of U_1 small nuclear ribonucleoprotein particles is requiered for the splicing of messenger RNA precursors. Cell 38, 299-307.

- Kreis, T.E., Geiger, B., Schmid, E., Jorcano, J.L., & Franke, W.W. (1983) De novo synthesis and specific assembly of keratin filaments in nonepithelial cells after microinjection of mRNA for epidermal keratin. Cell 32, 1125-1137.

- Lazarides, E., & Balzer, D.R.Jr. (1978) Specificity of desmin to avian and mammalian muscle cells. Cell 14, 429-438.

- Lehnert, M.E., Jorcano, J.L., Zentgraf, H., Blessing, M., Franz, J.K., & Franke, W.W. (1984) Characterization of bovine keratin genes: Similarities of exon patterns in genes coding for different keratins EMBO J. 3, 3279-3287.

- Lewis, S.A., Balcarek, J.M., Krek, V., Shelanski, M., & Cowan, N.J. (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.

- Lin, J.J.-C., & Feramisco, J.R. (1981) Disruption of the in vivo distribution of the intermediate filaments in fibroblasts through the microinjection of a specific monoclonal antibody. Cell 21, 185-193.

- Marchuk, D., McCrohon, S., & Fuchs, E. (1984) Remarkable conservation of structure among intermediate filament genes. Cell 39, 491-498.

- Messing, J., Crea, R., & Seeburg, P. (1981). A system for shotgun DNA sequencing. Nucl. Acids Res. 9, 309-321.

- Moll, R., Franke, W.W., Schiller, D.L., Geiger, B., & Krepler, R. (1982) The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell 31, 11-24.

- Mount, S.M. (1982). A catalog of splice junction sequences. Nucl. Acids Res. 10, 459-472.

- Nudel, U., Katcoff, D., Zakut, R., Shani, M., Carmon, Y., Finer, Y., Czosnek, H., Ginsburg, I., & Yaffe, D. (1982). Isolation and characterization of rat skeletal muscle and cytoplasmic actin genes. Proc. Natl. Acad. Sci. USA 79, 2763-2767.

- Osborn, M. & Weber, K. (1982). Intermediate filaments: Cell-type-specific markers in differentiation and pathology. Cell 31, 303-306.

- Quax, W., Vree Egberts, W., Hendriks, W., Quax-Jeuken, Y., & Bloemendal, H. (1983). The structure of the vimentin gene. Cell 35, 215-223.

- Quax, W., van den Heuvel, R., Vree Egberts, W., Quax-Jeuken, Y., & Bloemendal, H. (1984a). Intermediate filament cDNAs from BHK-21 cells: Demonstration of distinct genes for desmin and vimentin in all vertebrate classes. Proc. Natl. Acad. Sci. USA. 81, 5970-5974.

- Quax, W., Quax-Jeuken, Y., van den Heuvel, R., Vree Egberts, W., & Bloemendal, H. (1984b). Organization and sequence of the genes for desmin and vimentin. Cold Spring Harbor Laboratory. Molecular Biology of the Cytoskeleton, in press.

- Quax-Jeuken, Y., Quax, W., & Bloemendal, H. (1983). Primary and secondary structure of hamster vimentin predicted from the nucleotide sequence. Proc. Natl. Acad. Sci. USA 80, 3548-3552.

Quinlan, R.A., & Franke, W.W. (1982). Heteropolymer filaments of vimentin and desmin in vascular smooth muscle tissue and cultured baby hamster kidney cells demonstrated by chemical crosslinking. Proc. Natl. Acad. Sci. USA 79, 3452-3456.

- Quinlan, R.A., Cohlberg, J.A., Schiller, D.L., Hatzfeld, M., & Franke, W.W. (1984). Heterotypic tetramer (A_2D_2) complexes of non-epidermal keratins isolated from cytoskeletons of rat hepatocytes and hepatoma cells. J. Mol. Biol. 178, 365-388.

- Ramaekers, F.C.S., Puts, J.J.G., Moesker, O., Jap, P.H.K., & Vooijs, G.P. (1981) Use of antibodies to intermediate filaments in the characterization of human tumors. Cold Spring Harbor Symp. Quant. Biol. 46, 331-339.

- Ramaekers, F.C.S., Puts, J.J.G., Moesker, O., Kant, A., Huysmans, A., Haag, D., Jap, P.H.K., Herman, C., & Vooijs, G.P. (1983a) Antibodies to intermediate filament proteins in the immunohistochemical identification of human tumours; an overview. Histochemical Journal 15, 691-713.

- Ramaekers, F.C.S., Verheijen, R.H.M., Moesker, O., Kant, A., Vooijs, G.P., & Herman, C. (1983b) Mesodermal mixed tumor: diagnosis by analysis of intermediate filament proteins. Am. J. Surg. Path. 7, 381-385.

- Ramaekers, F.C.S., Puts, J.J.G., Moesker, O., Kant, A., Vooijs, G.P., & Jap, P.H.K. (1983c) Intermediate filaments in malignant melanomas. Identification and use as marker in surgical pathology. J. Clin. Inv. 71, 635-643.

- Ramaekers, F.C.S., Huysmans, A., Moesker, O., Kant, A., Jap, P.H.K., Herman, C. & Vooijs, G.P (1983d) Monoclonal antibodies to keratin filaments, specific for glanular epithelia and their tumors. Use in surgical pathology. Lab. Invest. 49, 353-361.

- Ramaekers, F.C.S., Haag, D., Kant, A., Moesker, O., Jap, P.H.K., & Vooijs (1983e) Coexpression of keratin and vimentin-type intermediate filaments in human metastatic carcinoma cells. Proc. Natl. Acad. Sci. USA 80, 2618-2622.

- Ramaekers, F., Huysmans, A., Moesker, O., Schaart, G., Herman, C., & Vooijs, P. (1985) Cytokeratin expression during neoplastic progression of human transitional cell carcinomas as detected by a monoclonal and a polyclonal antibody. Lab. Invest. 52, 31-38.

- Rimm, D.L., Horness, D., Kucera, J., & Blattner, F.R. (1980). Construction of coliphage lambda Charon vectors with BamH1 cloning sites. Gene 12, 301-309.

- Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H., & Roe, B.A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143, 161-178.

- Sompayrac, L.M., & Danna, K.J. (1981). Efficient infection of monkey cells with DNA of simian virus 40. Proc. Natl. Acad. Sci. USA 78, 7575-7578.

- Staden, R. (1979) A strategy of DNA sequencing employing computer programs. Nucl. Acids Res. 6, 2601-2610.

- Staden R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucl. Acids Res. 10, 2951-2961.

- Steinert, P.M., Rice, R.H., Roop, D.B., Trus, B.L., & Steven, A.C. (1983). Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments. Nature 302, 794-800.

- Steinert, P.M., Parry, P.A.D., Racoosin, E.L., Idler, W.W., Steven, A.C., Trus, B.L., & Roop, D.R. (1984). The complete cDNA and deduced amino acid sequence of a type II mouse epidermal keratin of 60.000 Da: analysis of sequence differences between type I and type II keratins. Proc. Natl. Acad. Sci. USA 81, 5709-5713.

- Weiher, H., Koning, M., & Gruss, P. (1983) Multiple point mutations affecting the simian virus 40 enhancer. Science 219, 626-631.

- Ueyama, H., Hamada, H., Battula, N., & Kakunaga, T. (1984). Structure of a human smooth muscle actin gene (aortic type) with a unique intron

site. Mol. Cell. Biol. 4, 1073-1078.

- Weber, K., & Geisler, N. (1984). Intermediate filaments-from wool α-keratins to neurofilaments: a structural overview. Cold Spring Harbor Laboratory, Cancer Cells 1, The Transformed Phenotype, pp 153-159.

CHAPTER 6

HUMAN DESMIN AND VIMENTIN GENES ARE LOCATED ON DIFFERENT CHROMOSOMES

Wim Quax, P. Meera Khan*, Yvonne Quax-Jeuken
and Hans Bloemendal

Department of Biochemistry
University of Nijmegen
and
Department of Human Genetics*
University of Leiden

submitted for publication

ABSTRACT

We have used somatic cell hybrids of Chinese hamster x man and mouse x man to localize the genes encoding the intermediate filaments desmin and vimentin in the human genome. Southern blots of DNA prepared from each cell line were screened with hamster cDNA probes specific for desmin and vimentin, respectively. The single copy human desmin gene is located on chromosome 2 and the single copy human vimentin gene is assigned to chromosome 10. Partial restriction maps of the two human loci are presented. A possible correlation of the desmin locus with several reported hereditary myopathies is discussed.

INTRODUCTION

During the last decade it has been recognized that cytoskeletal structures play an important role in a number of cell processes, including maintenance of cell shape, cell motility, mitosis, intracellular transport and support for cell organelles. Structural and genetic information obtained for the subunits of these cytoskeletal structures shows that actins, which are the subunits for microfilaments, and tubulins, which constitute the microtubules, are encoded by multigene families that show strong sequence conservation during evolution (3, 13). Recent studies show that the individual members of these multigene families are unlinked and dispersed throughout the genome of many species, including man (4, 5, 17, 42, 51).

The genes that encode the third type of cytoskeletal structures, the intermediate filaments, are only lately being characterized. The sequence information obtained thus far shows that the intermediate filament genes, although still clearly related, have diverged somewhat more than those of actins and tubulins during evolution. This explains why some members exhibit the nature of a single copy gene in DNA hybridization studies,-e.g. the vimentin gene in hamster and chicken (1, 35 , 52), the desmin gene in hamster (36) and the glial filament acidic protein (GFAP) gene in mouse (28). The intermediate filament genes display a tissue-specific expression pattern that shows close parallels to embryological differentiation (11, 26). To address the question whether the expression of intermediate filament genes is regulated coordinately it is desirable to gain information on the linkage relationship between these genes in the genome. Therefore, we investigated the chromosomal location of the vimentin and desmin genes with the aid of rodent x human somatic cell hybrids. The data presented in this paper indicate that the two intermediate filament genes are not linked in the human genome.

RESULTS

Characterization of the human genes for vimentin and desmin.

It has been shown that the haploid hamster genome contains a single gene encoding vimentin (VIM) (35). This gene, which has a total length of 9 kb, conists of nine exons. From recent studies it became clear that the gene encoding desmin (DES), the muscle-specific intermediate filament, also occurs as a single copy in the hamster genome (36). Structural analysis of this gene, which comprises 6.5 kb of DNA, has revealed that its exon-intron pattern is similar to that of the vimentin gene with the same number and positions of intervening sequences (Quax et al., in preparation).

In order to determine the copy numbers of these two intermediate filament genes in the human genome we applied specific hybridization probes to Southern blots of human DNA. Probes vimA, vimB, desA and desB were isolated as M13-subclones from the corresponding hamster genes (for the content of these subclones see figure 1b). When we applied a final wash of 0.2 x SSC, 0.2% SDS at 60 ^{o}C after hybridization, probes vimB and desB both hybridized to single restriction fragments of human DNA indicating that human vimentin and desmin are encoded by single copy genes. Under the same conditions probe vimA and desA hybridized to one or more weak bands additional to a single strong band. As can be judged from figure 1A the detection of one weakly hybridizing restriction fragment is a result of cross hybridization of vimA with the human desmin gene and of desA with the human vimentin gene, respectively (figure 1a; 36, 37). The remaining weak bands may result from cross hybridization with other intermediate filament genes. Especially the GFAP gene may be recognized with these probes, since probe desA contains a region of about 100 basepairs, which shares 80% homology with mouse GFAP cDNA (28, 35).

The hybridization patterns that have been found with the different probes have allowed us to derive

a partial restriction map for the human vimentin and the human desmin gene (figure 1b). From these maps the positions of the probes and the fragments to which they hybridize can be derived.

Chromosome localization of the human vimentin and desmin gene.

For the assignment of the intermediate filament genes we used a panel of 5 mouse x human and a panel of 9 chinese hamster x human somatic cell hybrids. The cryopreserved cell lines, which were screened for the human chromosomes and chromosome-specific markers during previous studies (21, 22, 29, 30), were brought back to culture and the cell pellets of each hybrid line were divided into two portions. One part was used to verify the human chromosome content by retesting for chromosome-specific enzyme markers. The updated information is summarized in table 1. The other part was used to isolate total DNA. This DNA was digested with an appropriate restriction enzyme, electrophoresed in an agarose gel and transferred to a nitrocellulose filter. Filters were hybridized to the nick-translated insert of pVim-1 (38) or to one of the above mentioned desmin specific M13 probes, respectively. For pVim-1 it was found that the restriction enzyme EcoR1 gave a good resolution of the mouse and human bands. In the case of the hamster x human cell hybrids BamH1 gave the best results (figure 2). The 12.5 kb human EcoR1 fragment is detected in cell lines B and E. The 12.0 kb human BamH1 fragment is found in lines H, K and N. Comparison of these results with the chromosome information of the hybrids (table 1) and calculation of a discordance score indicates that the vimentin gene (VIM) is most probably in chromosome 10. The only discordant is line E, which is negative for chromosome 10, but positive for vimentin. About 10% of discordancy is frequently observed among syntenic pairs of loci, especially when they happen to be situated wide apart on the chromosome, due to an occassional occurrence of chromosomal breakage and rearrangement in the interspecific somatic cell hybrids (39). However, the use of another panel of nine cell hybrid lines (data not shown) also favoured the assignment of the vimentin gene to chromosome 10.

The specific desmin probe, desB, gave the best resolution between the human bands and those of mouse or hamster with the restriction enzyme Hind111. The human 3.7 kb Hind111 fragment formed a relatively more intense band in cell lines K and M than in lines L and P (figure 3). The data in table 1 exclude every other chromosome but chromosome 2 for the localization of the desmin gene. In the mouse x man panel no cell line was positive for human desmin (hence, not included in figure 3), which is consistant with its assignment to chromosome 2. The isocitrate dehydrogenase-1 (IDH-1) and malate dehydrogenase-1 (MDH-1) were used as chromosome 2 specific enzyme markers (21) in the present study. Both these markers were found to be absent in the mouse x man panel while their pattern of segregation and relative intensities of their expression in the individual hybrids belonging to the hamster x man panel paralleled with that of the human 3.7 kb Hind 111 band mentioned above. Such a consistent dosage relationship between the gene products expressed in a set of hybrids exhibiting the segregation of human chromosomes, has been suggested to form an additional indication that the concerned loci are situated on the same human chromosome (48). In order to further confirm the localization of the desmin gene on chromosome 2 we hybridized the hamster x human panel to another DNA probe, which is known to be located on chromosome 2, namely γ-crystallin (Den Dunnen et al., in preparation). The crystallin probe hybridized with the same lanes and in the same intensities as our desmin probe (data not shown), strengthening the assigment of the human desmin gene to chromosome 2.

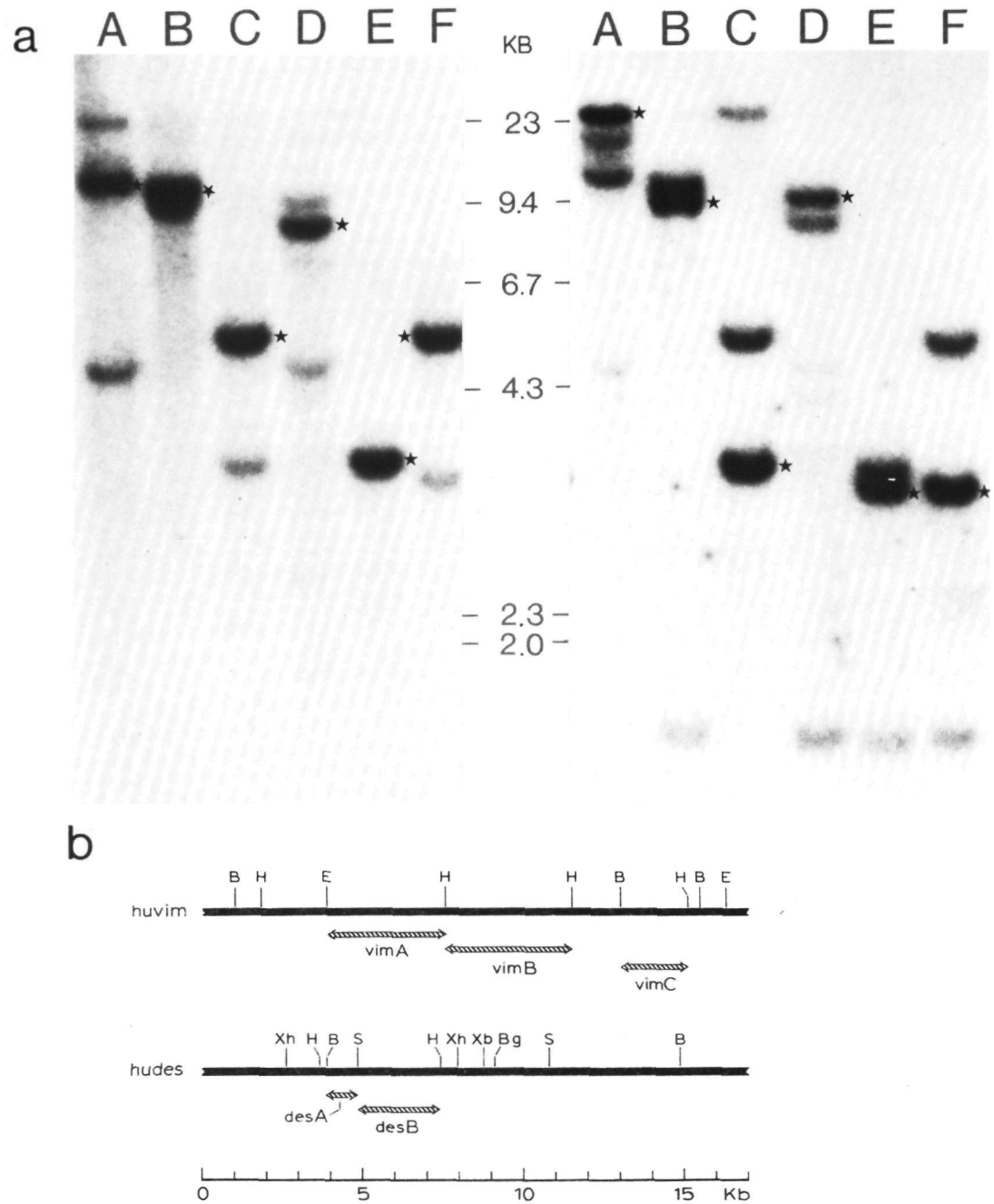
DISCUSSION

Our experiments show that in the human genome there is a single gene for vimentin and a single gene for desmin located respectively on chromosome 10 and 2. Sequence information obtained during the last years for all classes of intermediate filament subunits suggests that all of these genes evolved from a common ancestor gene (15, 18, 36, 35, 38, 43). Closely related genes that are thought to have diverged from a common ancestral gene may be linked in the genome, e.g. β-, γ- and δ-globin genes (HBB, HBG and HBD) on chromosome 11 (7); α-, β- and γ-fibrinogen (FGA, FGB and FGC) on chromosome 4 (20); the γ-crystallins on chromosome 2 (Den Dunnen et al.,in preparation); myosin heavy chain adult-1 (MYH1), adult-2 (MYH2) and embryonic-1 (MYH3) on chromosome 17 (27); two human relaxin genes on chromosome 9 (6); the α- and β-interferon genes (IFA and IFB) on chromosome 9 (31, 45). On the other hand different genes of the

Figure 1: Characterization of human vimentin and desmin genes

a: Human DNA was digested with the following set of enzymes: A) Eco R1, B) Bam H1, C) Hind111, D) Eco R1+ Bam H1, E) Eco R1+ Hind111, F) BamH1+ Hind111. Southern blots made from these digested DNAs were hybridized with probe vimA (left panel) and desA (right panel). The bands that are marked by an asterisk were identified as the true vimentin and desmin fragmens, respectively, by using specific probes (e.g. vimB and desB) and by applying more stringent washings than the 0.2x SSC, 60 °C used in this experiment. As can be judged from a comparison of the corresponding lanes of both panels one of the weaker bands in each digest of the left panel comigrates with the strongest band in the right panel and vice versa indicating that vimentin and desmin sequences cross hybridize under these conditions.

b: Partial restriction maps for the human vimentin gene (huvim) and the human desmin gene (hudes) were derived from Southern blot analysis of total human DNA digests with the aid of a number of specific probes. Enzymes used are: E= Eco R1, B= Bam H1, H= Hind111, S= Sac1, Bg= Bgl11, Xh= Xho1 and Xb= Xba1. The positions, where the different probes hybridize, are indicated by hatched bars. Probe vimA contains the sequence of amino acids 20-116 of hamster vimentin; probe vimB encodes amino acids 293-334 and vimC encodes amino acids 427-464 of hamster vimentin (35). DesA bears the information for amino acids 72-140 and desB the information for residues 276-373 of hamster desmin (Quax et al., in preparation).



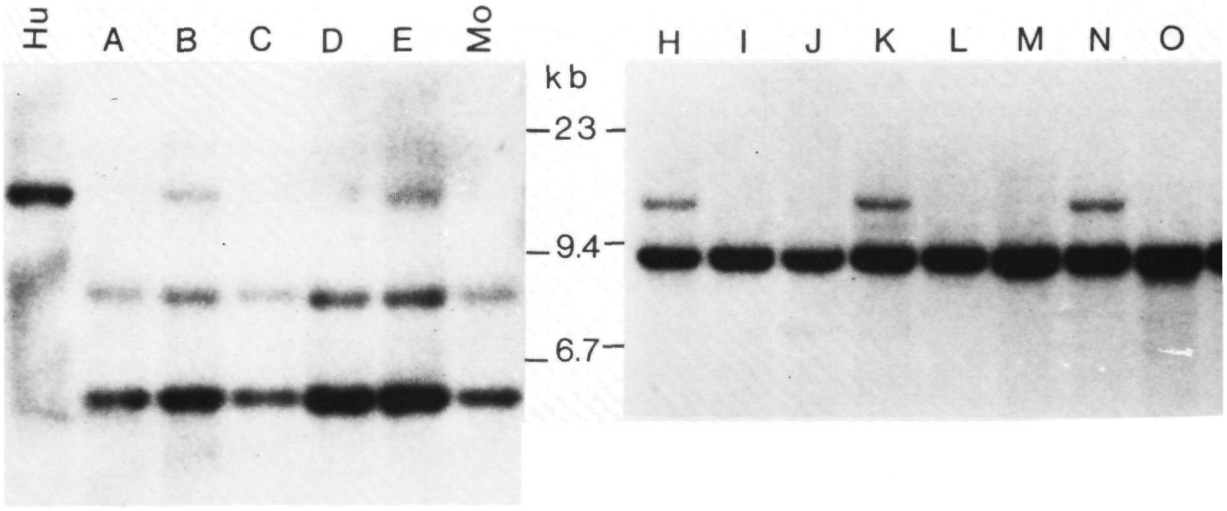


Figure 2: Assignment of the vimentin gene.

DNAs isolated from somatic cell hybrids between mouse and man were digested with Eco R1 and hybridized with the insert of pVim-1 (left panel). DNAs of cell hybrids between Chinese hamster and man, digested with Bam H1, were hybridized with the same probe (right panel). The chromosome content of each cell line, A-E and H-P (excluding that of O), is depicted in table 1. Control DNAs are Hu= human and Mo= Mouse.

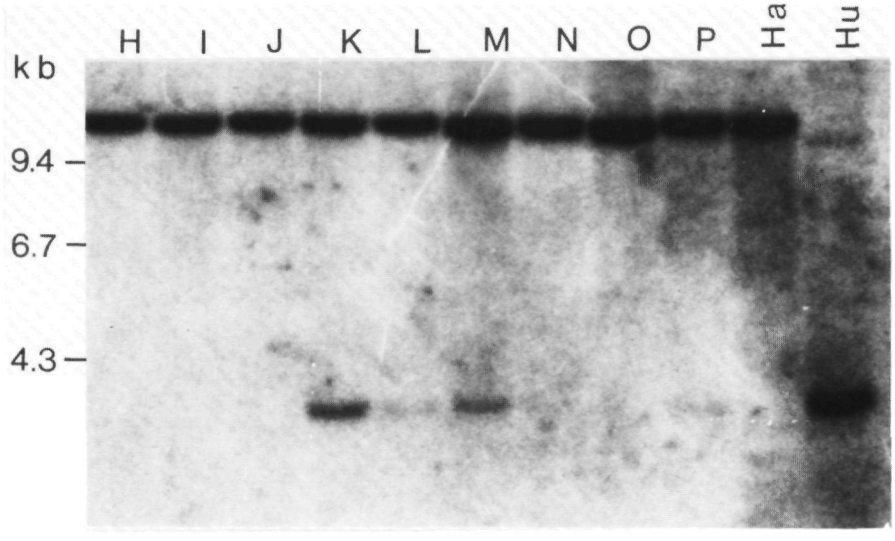


Figure 3: Assignment of the desmin gene.

DNAs isolated from somatic cell hybrids between Chinese hamster and man (H-P) were digested with Hind111, electrophoresed, blotted and hybridized with the M13 probe desB, which recognizes a 3.7 Hind111 fragment in human DNA. Cell hybrids K, L, M, and P were judged positive, the others negative for the desmin gene. Control DNAs are Ha= Chinese hamster and Hu= human.

same family can also be located on different chromosomes, e.g. the gene for pro-α2(1) collagen (COL1A2) on chromosome 7 and pro-α1(1) collagen (COL1A1) on chromosome 17 (23); the actin genes dispersed over a number of chromosomes (17, 42) including muscle α-actin (ACTA) on chromosome 1, cardiac actin (ACTC) on chromosome 15 (41) and one cytoplasmic actin gene (ACT1) on the X and another one (ACT2) on the Y chromosome (19); the α- and β-tubulin genes dispersed in the Drosophila, chicken and human genome (4, 5, 40); three estrogen inducible chicken genes- ovalbumin, ovomucoid and transferrin- on three different chicken chromosomes (24).

It is obvious from our results that the intermediate filament genes belong to the latter group. Remarkably the other genes that encode cytoskeletal filaments, the actin and tubulin genes, also belong to this group. Although it is certainly not a general rule (compare the histone genes (2)) it is possible that a disperse distribution in the genome is characteristic of structural genes that emerged early in evolution. Another contribution to this discussion has to await the chromosomal assignment of the multiple keratin genes, which should be possible by now, since cDNA probes have become available recently (12, 25, 43). The non-linked position of intermediate filament genes in the genome rules out the possibility that the spatially and temporally regulated expression of intermediate filaments is accomplished by physical linkage of the genes on the DNA.

Although the evolutionarily conserved nature of intermediate filaments suggests an important role for these structures, their precise functions remain still unknown. Therefore it is not surprising that there is at present a dearth of information on possible human diseases or aberrant conditions due to an absence or abnormal expression of intermediate filaments. However, the onset of desmin expression during muscle development and the redistribution of desmin from free cytoplasmic filaments to the Z-disk during the formation of myofibrils (14, 16) suggests a role for this gene in muscle differentiation. For this reason attention should be paid to a possible correlation between the expression of the desmin gene and various hereditary muscle disorders of unknown etiology. Several hereditary myopathies were found to exhibit aberrant accumulations of intermediate filaments in the myofibrils on electron microscopy as well as on immunofluorescence microscopy after staining with desmin specific antibodies (9, 10, 34, 44). In a case of skeletal muscle myopathy it was suggested that the occurrence of sarcomeric bodies, which were filled with insoluble filamentous material, was due to an excess in synthesis or incorrect turn-over of desmin (9). The desmin accumulation observed in muscle cells of a patient suffering from a form of congenital cardiomyopathy associated with an aberrant organization of intermediate filaments at the Z-disk, might account for the observed myocardial insufficiency (44). Finally in muscle biopsies from four related children suffering from a neuromuscular disease, peculiar desmin filaments containing inclusions were observed, which showed similarities with Mallory bodies (10), the cytokeratin containing inclusions typical of alcoholic liver disease (8). Most of the cardiomyopathies and myopathies are familial and monogenic. A linkage study between eventual Restriction Fragment Length Polymorfisms (RFLP's) in and around the desmin locus in chromosome 2 and each of the loci for unassigned myopathies should be rewarding. The results of such a study may eventually help detecting the heterozygote carriers and performing prenatal diagnosis.

MATERIALS AND METHODS

Cell hybrids and genetic analysis

The methods for production, isolation and propagation of the interspecific somatic cell hybrids were reported earlier (33, 47) Previous reports described the PG/Me hybrids (29, 49, 50), the a3y hybrids (22), the a3G hybrid (30) and the E36 series of hybrids (21) included in the present study. The hybrids were investigated for identifying the individual human chromosomes with G-banding, Giemsa-11 and/or Q-banding techniques (32). A detailed list of 37 different chromosome-specific enzyme markers employed to screen the above hybrids, including the references of the methods used, was reported elsewhere (21).

DNA preparation and Southern hybridization analysis

Total DNA of cell lines was prepared as described (46). 10 µg amounts of each DNA sample, which had been digested with restriction enzymes under conditions as recommended by the suppliers, were electrophoresed in 0.8% agarose gels, blotted onto nitrocellulose and hybridized under conditions

Lane number in figure 1	Clone	Human chromosome 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	VIM	DES
A	PG Me 1-1	+		+	+	+	+	+	+	+		+	+		+		+	+		+	+	+	+	+		
B	PG Me 4-1			+	+	+	+	+	+	+	+	+	+		+	+		+			+	+		+	+	
C	PG Me 5-1							+				+	+	+				+		+				+		
D	PG Me 6-1								+																	
E	PG Me 13-1			+				+	+	+			+	+					+	+		+	+	+	+	
H	a3Y 14-2	+		+	+	+			+		+	+	+	+	+	+	+	+		+	+	+	+		+	
I	a3Y 13-1	+		+				+	+			+	+	+	+	+	+	+		+	+	+		+		
J	a3G 5-2	+				+	+		+					+		+	+	+	+		+		+			
K	E36 33.7-2	+	+	+	+		+		+		+				+		+	+	+					+	+	+
L	E36 33.11-2	+	+					+					+	+	+	+	+	+			+	+		+		+
M	E36 10 CB 1-2	+	+	+	+			+	+			+	+	+			+	+				+		+		+
N	E36 10 CB 22B2				+				+	+	+	+					+					+		+	+	
P	E36 78-3-13-2	+	+			+						+				+	+		+	+		+		+		+
		–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–		
Discordant #	Vimentin	9	7	4	3	6	5	8	5	3	1	7	7	8	5	7	8	8	5	7	7	6	5	7		
	Desmin	4	0	7	6	7	6	7	10	8	6	8	8	7	6	6	5	7	4	8	8	7	8	6		

Table 1: Distribution of human chromosomes in cell hybrids segregating vimentin and desmin genes.

Cell line A to E are mouse x man hybrids and lines H to P are Chinese hamster x man somatic hybrids. Plus (+) indicates the presence of vimentin sequences, desmin sequences, human chromosomes and chromosome specific isozyme markers. The lowest discordance for vimentin is with chromosome 10 and for desmin with chromosome 2. The discordancy of line E -vimentin sequences detectable, but no enzyme marker present- may result from breakage and loss of a part of chromosome 10 in this line. Cell line O is omitted from this table, since the precise chromosome content of this line is not known.

as described (36). The stringency of the final washing is described in the figure legends. Plasmid inserts of pVim-1 (38) were ^{32}P-labeled by nick-translation; M13- probes were prepared by carrying out a standard T-reaction of dideoxy sequencing with [α^{32}P]-dATP of > 3000 Ci/mmole.

ACKNOWLEDGEMENTS

We thank Ms. Wilma Vree Egberts, Mr. R.G.M. van der Goes, Ms. L.M.M. Blankenstein-Wijnen and Ms. M. Monteba-van Heuvel for excellent technical assistance. Part of this study was carried out under auspicies of the Netherlands Foundation for Chemical Research (S.O.N.) with financiel aid of the Netherlands Organization for the advancement of Pure Research (Z.W.O.)

REFERENCES

1. Capetanaki, Y.G., J. Ngai, C.N. Flytzanis, and E. Lazarides. 1983. Tissue-specific expression of two mRNA species transcribed from a single vimentin gene. Cell 35: 411-420.

2. Chandler, M.E., L.H. Kedes, R.H. Cohn, and J.J. Yunis. 1979. Genes coding for histone proteins in Man are located on the distal end of the long arm of chromosome 7. Science 205: 908-910.

3. Cleveland, D.W., M.A. Lopata, R.J. MacDonald, N.J. Cowan, W.J. Rutter, and M.W. Kirschner. 1980. Number and evolutionary conservation of α- and β-tubulin and cytoplasmic β- and γ-actin genes using specific cloned cDNA probes. Cell 20: 95-105.

4. Cleveland, D.W., S.H. Hughes, E. Stubblefield, M.W. Kirschner, and H.E. Varmus. 1981. Multiple α and β tubulin genes represent unlinked and dispersed gene families. J. Biol. Chem. 256: 3130-3134.

5. Cowan, N.J., C.D. Wilde, L.T. Chow, and F.C. Wefald. 1981. Structural variation among human β-tubulin genes. Proc.. Natl. Acad. Sci. USA 78: 4877-4881.

6. Crawford, R.J., P. Hudson, J. Shine, H.D. Niall, R.L. Eddy, and T.B. Shows. 1984. Two human relaxin genes are on chromosome 9. EMBO J. 3: 2341-2345.

7. Deisseroth, A., A. Nienhuis, J. Lawrence, R. Giles, P. Turner, and F. Ruddle. 1978. Chromosomal localization of human β globin gene on human chromosome 11 in somatic cell hybrids. Proc. Natl. Acad. Sci. USA 75: 1456-1460.

8. Denk, H., W.W. Franke, R. Eckerstorfer, E. Schmidt, and D. Keraschki. 1979. Formation and involution of Mallory bodies ("alcoholic hyalin") in murine and human liver revealed by immunofluorescence microscopy with antibodies to prekeratin. Proc. Natl. Acad. Sci. USA 76: 4112-4116.

9. Edstrom, L., L.-A. Thornell, and A. Eriksson. 1980. A new type of hereditary distal myopathy with characteristic sarcoplasmic bodies and intermediate (skeletin)filaments. J. Neurol. Sci. 47: 171-190.

10. Fidzianska, A., H.H. Goebel, M. Osborn, H.D. Lenard, G. Osse, and U.Langenbeck. 1983. Mallory body-like inclusions in a hereditary congenital neuromuscular disease. Muscle & Nerve 6: 195-200.

11. Franke, W.W., E. Schmidt, D.L. Schiller, S. Winter, E.D. Jarasch, R. Moll, H. Denk, B.W. Jackson, and K. Illmensee. 1982. Differentiation-related patterns of expression of proteins of intermediate-size filaments in tissues and cultured cells. Cold Spring Harbor Symp Quant Biol 46: 431-454.

12. Fuchs, E.V., S.M. Coppock, H. Green, and D.W. Cleveland. 1981. Two distinct classes of keratin genes and their evolutionary significance. Cell 27: 75-84.

13. Fyrberg, E.A., B.J. Bond, N.D. Hershey, K.S. Mixter, and N. Davidson. 1981. The actin genes of Drosophila: Protein coding regions are highly conserved but intron positions are not. Cell 24: 107-116.

14. Gard, D.L., and E. Lazarides. 1980. The synthesis and distribution of desmin and vimentin during myogenesis in vitro. Cell 19: 263-275.

15. Geisler, N., and K. Weber. 1982. The amino acid sequences of chicken muscle desmin provides a common structural model for intermediate fila-

ment proteins. EMBO J. 1: 1649-1656.

16. Granger, B.L., and E. Lazarides. 1979. Desmin and vimentin coexist at the periphery of the myofibril Z disc. Cell 18: 1053-1063.

17. Hanauer, A., R. Heilig, M. Levin, J.P. Moisan, K.-H. Grzeschik, and J.L. Mandel. 1984. The actin gene family in man: assignment of the gene for skeletal muscle alpha-actin to chromosome 1, and presence of actin sequences on autosome 2 and 3, and on the x and y chromosomes. Cytogenet. Cell Genet. 37: 487-488.

18. Hanukoglu, I., and E. Fuchs. 1982. The cDNA sequence of a human epidermal keratin: Divergence of sequence but conservation of structure among intermediate filament proteins. Cell 31: 243-252.

19. Heilig, R., K.H. Hanauer, K.-H. Grzeschik, M.C. Hors-Cayla, and J.L. Mandel. 1984. Actin-like sequences are present on human X and Y chromosomes. EMBO J. 3: 1803-1807.

20. Henry, I., G. Uzan, D. Weil, H. Nicolas, J.C. Kaplan, G. Marguerie, A. Kahn, and C. Junien. 1984. The genes coding for the A-alpha, B-beta, and gamma chains of fibrinogen are located on chromosome 4. Cytogenet. Cell Genet. 37: 490-491.

21. Herbschleb-Voogt, E., K.-H. Grzeschik, P.L. Pearson, and P. Meera Khan. 1981a. Assignment of ADCP gene(s) to human chromosome 2 in rodent-human somatic cell hybrids. Hum. Genet. 59: 317-323.

22. Herbschleb-Voogt, E., P.L. Pearson, J.M. Vossen, and P. Meera Khan. 1981b. Basic defect in the expression of ADA in ADA SCID disease investigated through the cells of an obligate heterozygote. Hum. Genet. 56: 379-386.

23. Huerre, C., C. Junien, D. Weil, M. Chu, M. Morabito, N. van Cong, J.C. Myers, C. Foubert, M. Gross, D.J. Prockop, A. Boue, J. Kaplan, A. De La Chapelle, and R. Ramirez. 1982. Human type I procollagen genes are located on different chromosomes. Proc. Natl. Acad. Sci. USA 79: 6627-6630.

24. Hughes, S.H., E. Stubblefield, F. Payvar, J.D. Engel, J.B. Dodgson, D. Spector, B. Cordell, R.T. Schimke, and H.E. Varmus. 1979. Gene localization by chromosome fractionation: Globin genes are on at least two chromosomes and three estrogen-inducible genes are on three chromosomes. Proc. Natl. Acad. Sci. USA 76: 1348-1352.

25. Jorcano, J.L., T.M. Magin, and W.W. Franke. 1984. Cell type-specific expression of bovine keratin genes as demonstrated by the use of complementary DNA clones. J. Mol. Biol. 176: 21-37.

26. Lazarides E. 1982. Intermediate filaments: A chemically heterogeneous developmentally regulated class of proteins. Ann. Rev. Biochem. 51: 219-250.

27. Leinwand, L.A., R.E.K. Fournier, B. Nadal-Ginard, and T.B. Shows. 1984. Assignment of the sarcomeric myosin heavy chain multigene family to chromosome 17 in humans and chromosome 11 in the mouse. Cytogenet. Cell Genet. 37: 521-522.

28. Lewis, S.A., J.M. Balcarek, V. Krek, M. Shelanski, and N.J. Cowan. 1984. Sequence of a cDNA clone encoding mouse glial fibrillary acidic protein: Structural conservation of intermeidate filaments. Proc. Natl. Acad. Sci. USA 81: 2743-2746.

29. Meera Khan, P., L.M.M. Wijnen, J.Th. Wijnen, and K.-H. Grzeschik. 1983. Electrophoretic characterization and genetics of human biliverdin reductase (BLVR; EC 1.3.'.24); Assignment of BLVR to the p14 cen region of human chromosome 7 in mouse-human somatic cell hybrids. Biochem. Genet. 21: 123-133.

30. Meera Khan, P., L.M.M. Wijnen, A. Hagemeijer, and P.L. Pearson. 1984. Human formaldehyde dehydrogenase (FDH) and its assignment to chromosome 4. Cytogenet. Cell Genet. 38: 112-115.

31. Owerbach, D., W.J. Rutter, T.B. Shows, P. Gray, D.V. Goeddel, and R.M. Lawn. 1981. Leukocyte and fibroblast interferon genes are located on human chromosome 9. Proc. Natl. Acad. Sci. 78: 3123-3127.

32. Pearson, P.L., and A.M.M. van Egmond-Cowan. 1976. Further development in banding techniques or mammalian chromosome identification. In: New Techniques in Biophysics and Cell Biology (eds. R.H. Pain & B.J. Smith) Wiley, London, p213.

33. Pontecorvo, G. (1975) Production of mammalian somatic cell hybrids by means of polyethylene glycol treatment. Somatic Cell Genet. 1: 391-395.

34. Porte A., M.-E. Stoeckel, A. Sacrez, and A. Batzenschlager. 1980. Unusual familial cardiomyopathy with storage of intermediate filaments in the cardiac muscular cells. Virchows Arch (Path. Anat.) 386: 43-58.

35. Quax, W., W. Vree Egberts, W. Hendriks, Y. Quax-Jeuken, and H. Bloemendal. 1983. The structure of the vimentin gene. Cell 35: 215-223.

36. Quax, W., R. van den Heuvel, W. Vree Egberts, Y. Quax-Jeuken, and H. Bloemendal. 1984a. Intermediate filament cDNAs from BHK-21 cells: Demonstration of distinct genes for desmin and vimentin in all vertebrate classes. Proc. Natl. Acad. Sci. USA 81: 5970-5974.

37. Quax, W., Y. Quax-Jeuken, R. van den Heuvel, W. Vree Egberts, and H. Bloemendal. 1984b. Organization and sequence of the genes for desmin and vimentin. Cold Spring Harbor Laboratory, The Molecular Biology of the Cytoskeleton, in press.

38. Quax-Jeuken, Y.E.F.M., W.J. Quax, and H. Bloemendal. 1983. Primary and secondary structure of hamster vimentin prdicted from the nucleotide sequence. Proc. Natl. Acad. Sci. USA 80: 3548-3552.

39. Ruddle, F.H. 1970. Utilization of somatic cells for genetic analysis: Possibilities and problems. In: Control Mechanisms in Expression of Cellular Phenotypes (ed. M.A. Padykula) Acad. Press, New York, p233.

40. Sanchez, F., J. Natzle, D.W. Cleveland, M.W. Kirschner, and B.J. McCarthy. 1980. A dispersed multigene family encoding tubulin in Drosophila Melanogaster. Cell 22: 845-854.

41. Shows, T., R.L. Eddy, L. Haley, M. Byers, M. Henry, P. Gunning, P. Ponte, and L. Kedes. 1984. The coexpressed genes for human alpha and cardiac actin are on chromosome 1 and 15, respectively. Cytogenet. Cell Genet. 37: 583.

42. Soriano, P., P. Szabo, and G. Bernardi. 1982. The scattered distribution of actin genes in the mouse and human genomes. EMBO J. 1: 579-583.

43. 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 302: 794-800.

44. Stoeckel, M.-E., M. Osborn, A. Porte, A. Sacrez, A. Batzenschlager, and K. Weber. 1981. An unusual familial cardiomyopathy characterized by aberrant accumulations of desmin-type intermediate filaments. Virchows Arch (Path. Anat.) 393: 53-60.

45. Trent, J.M., S. Olson, and R.M. Lawn. 1982. Chromosomal localization of human leukocyte, fibroblast, and immune interferon genes by means of in situ hybridization. Proc. Natl. Acad. Sci. USA 79: 7809-7813.

46. van der Putten, H., E. Terwindt, A. Berns, and R. Jaenisch. 1979. The integration sites of endogenous and exogenous moloney murine leukemia virus. Cell 18: 109-116.

47. Westerveld, A., R.P.L.S. Visser, P. Meera Khan, and D. Bootsma. 1971. Loss of human genetic markers in man-Chinese hamster somatic cell hybrids. Nature 234: 20-24.

48. Westerveld, A., and P. Meera Khan. 1972. Evidence for linkage between human loci for 6-phosphogluconate dehydrogenase and phosphoglucomutase$_1$ in man-Chinese hamster somatic cell hybrids. Nature 236: 30-32.

49. Wijnen, L.M.M., K.-H. Grzeschik, P.L. Pearson, and P. Meera Khan. 1977a. The human PGM-2 and its chromosomal localization in man-mouse hybrids. Hum. Genet. 37: 271-278.

50. Wijnen, L.M.M., K.-H. Grzeschik, P.L. Pearson, and P. Meera Khan. 1977b. Direct assignment of citrate synthase (CS) gene to human chromosome 12 in man-mouse somatic cell hybrids. Hum. Genet. 39: 339-344.

51. Wilde, C.D., L.T. Chow, F.C. Wefald, and N.J. Cowan. 1982. Structure of two human α-tubulin genes. Proc. Natl. Acad. Sci. USA 79: 96-100.

52. 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.

SUMMARY

The cytoplasm of most, if not all, eucaryotic cells contains filamentous protein polymers, which contribute to the structural organization within the cell. This intracellular matrix has been called the cytoskeleton. On the basis of the type of protein, which constitutes the filaments, a distinction has been made between the microfilaments, which are built up by actins, the microtubules, which are formed by tubulins, and the intermediate filaments (IF) that in different cell types can contain different types of subunits. It was found that in epithelial cells the keratins represent the subunits of IF; in nerve cells the IF are built up by the neurofilament proteins; in glial cells (astrocytes) the Glial Filament Acidic Protein has been identified as the IF subunit; in muscle cells desmin is found as subunit and finally in cells of mesenchymal origin vimentin constitutes the IF filaments.

The investigations of this thesis mainly concentrate on two intermediate filament subunits: vimentin and desmin. With the aid of recombinant DNA techniques, the genetic information for both vimentin and desmin has been cloned in the Escherichia coli plasmid pBR322. Chapter 2 describes the molecular cloning of vimentin cDNA starting from hamster lens messenger RNA. Using these clones, pVim-1 and pVim-2, it has been ascertained that the hamster genome contains only one copy of the gene encoding vimentin. Chapter 4 describes how messenger RNA from Baby Hamster Kidney cells was used as starting material for the construction of the desmin cDNA clones pDes-1, -2, -3, and -4. It was found that desmin is also encoded by a single copy gene. The amino acid sequences of both vimentin and desmin have been derived from the nucleotide sequences of the corresponding cDNAs. The two different IF subunits show a strong mutual homology (65%), which forms an indication for the hypothesis that both genes originated from the same common ancestor gene. From DNA hybridization experiments it could be concluded that the divergence of the two IF genes most probably took place before the evolution of the vertebrate species (500 million years ago). Hybridization analysis of RNA isolated from various tissues shows that the transcription of RNA and the synthesis of vimentin and/or desmin takes place coordinately, which proofs that the regulation of IF expression occurs at the transcription level.

Chapter 3 describes the structure of the hamster vimentin gene. The total coding information is constituted in 9 exons which are spread over 9 kilobases of DNA. From the DNA sequence of the gene the total primary structure of hamster vimentin could be derived. Two of the nine intervening sequences precisely map at the border of two DNA fragments that encode different domains of the vimentin polypeptide. Therefore the vimentin gene forms an additional example for the hypothesis that exon-shuffling played a role in the generation of different domains in proteins. The structure of the desmin gene is discussed in chapter 5. Surprisingly, the exon-intron structure of the desmin gene is totally comparable to that of the vimentin gene. From the comparison of the total primary structure of hamster desmin, derived from the DNA sequence of the gene, with the chicken desmin sequence it emerged that the non-helical amino terminus of desmin shows great interspecies variations. On the other hand the non-helical carboxy terminal and the central α-helical domain of desmin both show an homology of more than 90% upon comparing hamster and chicken.

In chapter 5 it is also described that the cloned desmin gene was used for "gene-transfer" and subsequent expression of desmin in cells which normally do not contain desmin filaments. Transfection of a plasmid containing the transcription unit for desmin into lens cells or HeLa cells resulted in the expression of desmin and the correct assembly of a desmin IF cytoskeleton in these nonmuscle cells.

Finally chapter 6 deals with those aspects of the vimentin and desmin gene, which are of importance for the human genetics. Apart from the human restriction enzyme maps also the chromosomal localizations of the vimentin and desmin genes have been determined. Human vimentin (chromosome 10) and human desmin (chromosome 2) are encoded by two non-linked single copy genes. The possible correlation of these two chromosomal loci with hereditary determined human diseases is discussed.

SAMENVATTING

Het cytoplasma van vrijwel alle eukaryotische cellen bevat vezelachtige eitwitpolymeren, die bijdragen tot de structurele organisatie binnen de cel. Deze intracellulaire matrix wordt het cytoskelet genoemd. Op grond van het type eiwitbouwsteen heeft men onderscheid gemaakt tussen microfilamenten, welke opgebouwd zijn uit actines, microtubuli, die gevormd worden door tubulines en de intermediaire filamenten (IF), die -afhankelijk van het celtype- opgebouwd kunnen zijn uit verschillende subeenheden. Zo vormen in epitheliale cellen de keratines de bouwstenen van de IF; in zenuwcellen zijn IF opgebouwd uit neurofilament-eiwitten; in gliacellen treffen we het zure gliafilament-eiwit aan; in spiercellen is desmine de subeenheid en tenslotte in cellen van mesenchymale oorsprong is vimentine de IF-bouwsteen.

De studies beschreven in dit proefschrift hebben voornamelijk betrekking op twee van deze IF-subeenheden: vimentine en desmine. Door gebruikmaking van recombinant DNA-technieken is de genetische informatie voor zowel vimentine als desmine gekloneerd in het Escherichia coli plasmide pBR322. Hoofdstuk 2 beschrijft hoe de klonering van vimentine complementair-DNA (cDNA) uitgaande van hamster-lens messenger RNA gerealiseerd werd. Met behulp van deze cDNA klonen, pVim-1 and pVim-2, is vastgesteld, dat er in het genoom slechts een copie van het chromosomale gen coderend voor vimentine aanwezig is. In hoofdstuk 4 wordt beschreven hoe vanuit messenger RNA van baby-hamster-niercellen de desmine cDNA klonen pDes-1, -2, -3 en -4 geconstrueerd werden. Ook desmine wordt door een "single copy" gen gecodeerd. De aminozuur volgordes van vimentine en desmine konden uit de nucleotide volgorde van de cDNA's worden afgeleid. Deze twee typen IF-subeenheden vertonen een sterke onderlinge homologie (65%), waaruit kan worden afgeleid dat ze van hetzelfde voorouder-gen afstammen. Het moment, waarop deze twee genen tijdens de evolutie uit elkaar gegaan zijn ligt zeer waarschijnlijk voor de divergentie van de gewervelde dieren (500 miljoen jaar geleden), zoals kon worden afgeleid uit DNA hybridisatie-experimenten. Hybridisatie-analyse van RNA geisoleerd uit verschillende weefsels toont, dat de transcriptie van RNA en de synthese van vimentine en/of desmine met elkaar gecorreleerd zijn, waarmee is aangetoond, dat de regulatie van IF genexpressie plaatsvindt op transcriptie-niveau.

Hoofdstuk 3 beschrijft de structuur van het hamster vimentine-gen. De totale coderende informatie ligt verdeeld over 9 exonen over een afstand van 9 kilobaseparen. Vanuit de DNA-volgorde van het gen kon de volledige aminozuur-volgorde van hamster-vimentine worden afgeleid. Twee van de negen "intervening sequences" liggen juist op de grens tussen twee DNA volgordes, die verschillende eiwitdomeinen van de vimentine-subeenheid coderen. Daarmee vormt het vimentine-gen een nieuwe aanwijzing voor de veronderstelling, dat "exon-shuffling" een rol heeft gespeeld bij het ontstaan van verschillende domeinen in eiwitten. De structuur van het desmine-gen komt aan de orde in hoofdstuk 5. Heel verrassend blijkt de exon-intron structuur van het desmine-gen volledig gelijk te zijn aan die van het vimentine-gen. Uit de vergelijking van de volledige primaire structuur van hamster-desmine, afgeleid van de DNA-volgorde van het hamster gen, met de kippe-desmine volgorde blijkt, dat de "niet-α-helix" amino-terminus van desmine grote verschillen vertoont tussen verschillende diersoorten. Daarentegen vertoont het carboxy-terminale domein en het "α-helix" middengedeelte van desmine een homologie van meer dan 90% bij vergelijking van hamster en kip.

In hoofdstuk 5 staat ook beschreven hoe het gekloneerde desmine-gen gebruikt is om door middel van "gene transfer" expressie van desmine te bewerkstelligen in cellen, die normaliter geen desmine-filamenten bevatten. Transfectie van een plasmide, dat de transcriptie-unit voor desmine bevat, in lens- en HeLa-cellen resulteerde in de expressie van desmine en de correcte assemblage van desmine IF-filamenten in deze niet-spiercellen.

Tot slot worden in hoofdstuk 6 de voor de humane genetica belangrijke aspecten van het vimentine en desmine gen behandeld. Naast de restrictie-enzym kaarten zijn de chromosomale localisaties van het desmine en vimentine gen bepaald. Vimentine (chromosoom 10) en desmine (chromosoom 2) blijken door niet met elkaar verbonden "single copy" genen gecodeerd te worden. De eventuele relatie van deze twee chromosomale loci met erfelijk bepaalde ziekten wordt bediscussieerd.

LIST OF PUBLICATIONS:

- Quint, W., Quax, W., van der Putten, H. and Berns, A. (1981) Characterization of AKR Murine Leukemia Virus sequences in AKR mouse substrains and structure of integrated recombinant genomes in tumor tissues. J. Virol. 39, 1-10.

- Dodemont, H.J., Soriano, P., Quax, W.J., Ramaekers, F., Lenstra, J.A., Groenen, M.A., Bernardi, G. and Bloemendal, H. (1982) The genes coding for the cytoskeletal proteins actin and vimentin in warm-blooded vertebrates. EMBO J. 1, 167-171.

- Quax, W.J., Dodemont, H.J., Lenstra J.A., Ramaekers, F., Soriano, P., van Workum, M.E.S., Bernardi, G. and Bloemendal, H. (1982) Genes coding for vimentin and actin in mammals and birds. Adv. Exp. Medicine and Biology 158, 349-357.

- Quax-Jeuken, Y., Quax, W. and Bloemendal, H. (1983) Primary and secondary structure of hamster vimentin predicted from the nucleotide sequence. Proc. Natl. Acad. Sci. USA 80, 3548-3552.

- Quax, W., Vree Egberts, W., Hendriks, W., Quax-Jeuken, Y. and Bloemendal, H. (1983) The structure of the vimentin gene. Cell 35, 215-223.

- Bloemendal, H., Quax, W., Quax-Jeuken, Y., Dodemont, H., Ramaekers, F., Dunia, I. and Benedetti, L. (1983) Organization and expression of the vimentin gene. Mol. Biol. Rep. 9, 115-118.

- Quax, W., van den Heuvel, R., Vree Egberts, W., Quax-Jeuken, Y., & Bloemendal, H. (1984) Intermediate filament cDNAs from BHK-21 cells: demonstration of distinct genes for desmin and vimentin in all vertebrate classes. Proc. Natl. Acad. Sci. USA 81, 5970-5974.

- Quax, W., Quax-Jeuken, Y., van den Heuvel, R., Vree Egberts, W., & Bloemendal, H. (1984). Organization and sequence of the genes for desmin and vimentin. Cold Spring Harbor Laboratory. Molecular Biology of the Cytoskeleton, in press.

- Quax-Jeuken, Y., Janssen, Ch., Quax, W., van den Heuvel, R. and Bloemendal H. (1984). Bovine β-crystallin complementary DNA clones. J. Mol. Biol. 180, 457-472.

- Bloemendal, H., Quax, W., Quax-Jeuken, Y, van den Heuvel, R., Vree Egberts, W. & van den Broek, L. (1985) Vimentin and desmin cDNA clones: structural aspects of corresponding proteins and genes. Ann. N. Y. Acad. Sci.(1985), in press.

- Dodemont, H., Quax, W., Schoenmakers, J.G.G., & Bloemendal, H. (1985) Evolution of the single copy αA-crystallin gene: differently sized mRNAs of mammals and birds show homology in their 3′ non-coding regions. Mol. Biol. Rep., in press.

- Quax, W., van den Broek, L., Vree Egberts, W., Ramaekers, F., & Bloemendal, H. (1985) Characterization of the desmin gene: Expression and formation of desmin filaments in nonmuscle cells after gene transfer, submitted for publication.

- Quax, W., Quax-Jeuken, Y., Meera Khan, P., & Bloemendal, H. (1985) Characterization of the human desmin and vimentin genes: Two closely related intermediate filament genes located on different chromosomes, submitted for publication.

- van den Heuvel, R., Hendriks, W., Quax, W., & Bloemendal, H. (1985) Complete structure of the hamster αA-crystallin gene; reflection of an evolutionary history by means of exon shuffling, submitted for publication.

- Quax-Jeuken, Y., Quax, W., van Rens, G., Meera Khan, P., & Bloemendal, H. Complete structure of the αB-crystallin gene: Conservation of the exon/intron distribution in the two non-linked α-crystallin genes, submitted for publication.

CURRICULUM VITAE

Wim Quax werd op 1 februari 1956 te Geleen geboren. Hij doorliep het Gymnasium β aan de scholengemeenschap St. Michiel te Geleen, alwaar het einddiploma in 1974 behaald werd. In datzelfde jaar begon hij zijn studie biologie aan de Katholieke Universiteit te Nijmegen. Het kandidaatsexamen (B4) werd afgelegd in september 1977. In september 1980 werd het doctoraal examen biologie afgelegd (cum laude) met als hoofdvak Biochemie (Dr. A.J.M. Berns en Prof. Dr. H. Bloemendal) en als bijvakken Microbiologie (Prof. Dr. Ir. G.D. Vogels) en Aquatische Oecologie (Prof. Dr. C. Den Hartog; gestationeerd te Roscoff, Frankrijk).

Voor het verrichten van het hier gepresenteerde promotie-onderzoek was hij van oktober 1980 tot november 1984 werkzaam op het Laboratorium voor Biochemie (o.l.v. Prof. Dr. H. Bloemendal) in dienst van de Katholieke Universiteit tot april 1981 en daarna in dienst van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (Stichting S.O.N.).

Tijdens de promotietijd werd deelgenomen aan de "Eurage Lens Group workshop" te Louvain La Neuve (november 1980) en de EMBO-cursus "Expression of proteins from cloned genes (cDNA's) in eucaryotic cells" (EMBL Heidelberg, juni 1984).

Voordrachten betreffende de inhoud van dit proefschrift werden gehouden o.a.: bij de EMBO-workshop "Intermediate Filaments in Differentiation and Pathology" te Gunzburg (april 1982); bij de door de Commissie voor Biochemie en Biophysica van de Koninklijke Akademie georganiseerde bijeenkomst betreffende "Structure and Functional Aspects of Cellular Matrices" te Amsterdam (november 1983); op uitnodiging van Prof. Dr. W. W. Franke op het Deutsche Krebsforschungszentrum te Heidelberg (januari 1984); op het "Cold Spring Harbor" congres over "Molecular Biology of the Cytoskeleton", New York (april 1984).

Op 15 mei 1981 is hij gehuwd met Yvonne Jeuken.

Sinds november 1984 is hij werkzaam bij Gist-brocades N.V. te Delft. Aldaar is hij als onderzoeks-medewerker verbonden aan de afdeling "Bacteriëngenetica".

Indien U geen gebruik maakt van dit proefschrift, wordt het op prijs gesteld wanneer U het terugzendt naar het Laboratorium voor Biochemie (W.N.F.), K.U. Nijmegen.

De promovendus

1

De hybridisatie-experimenten beschreven door Capetanaki et al., vormen een onvoldoende basis voor de conclusie van de auteurs, dat het kippe-vimentine-gen veel minder homologie met andere intermediaire-filament-genen zou vertonen dan het hamster-vimentine-gen.

Capetanaki, Y.G., Ngai, J., Flytzanis, C.N., & Lazarides, E. (1983) Cell 35, 411-420.

Capetanaki, Y.G., Ngai, J., & Lazarides, E. (1984) Proc. Natl. Acad. Sci. USA 81, 6909-6913.

2

Het artikel van Steinert et al. (1980) bevat zoveel onjuistheden dat het eerder ten nadele dan ten voordele is geweest bij de opheldering van de structuur van intermediaire filament-subeenheden.

Steinert, P.M., Idler, W.W., & Goldman, R.D. (1980) Proc. Natl. Acad. Sci. USA 77, 4534-4538.

3

Uit de DNA-volgorde-gegevens van Marchuk et al. blijkt, dat Lehnert et al. de intronposities voor keratine genen ten onrechte zo nauwkeurig voorspeld hebben aan de hand van hun "R-looping" gegevens.

Marchuk, D., McCrohon, S., & Fuchs, E. (1984) Cell 39, 491-498.

Lehnert, M.E., Jorcano, J.L., Zentgraf, H., Blessing, M., Franz, J.K., & Franke, W.W. (1984) EMBO J. 3, 3279-3287.

4

De structuur van het β_{23}-crystalline (muis) zoals die voorspeld is vanuit de DNA-volgorde van het gen, is vrijwel gelijk aan het β_{A_3}-crystalline van het kalf met uitzondering van de 15 N-terminale aminozuren. Het samenvallen van de positie waar β_{23}-crystalline begint af te wijken met de plaats van een intron in verwante β,γ-crystalline-genen is op z'n minst frappant te noemen.

Inana, G., Piatigorsky, J., Norman, B., Slingsby, C., & Blundell, T. (1983) Nature 302, 310-315.

Berbers, G.A.M., Hoekman, W.A., Bloemendal, H., de Jong, W.W., Kleinschmidt, T., & Braunitzer, G. (1984) Eur. J. Biochem. 139, 467-479.

Quax-Jeuken, Y.E.F.M., Janssen, C., Quax, W., van den Heuvel, R., & Bloemendal, H. (1984) J. Mol. Biol. 180, 457-472.

5

De tegenstrijdige bevindingen van Lazarides et al. enerzijds en Holtzer et al. anderzijds betreffende het wel of niet aanwezig zijn van vimentine in volgroeid spierweefsel, kunnen wellicht verklaard worden door te veronderstellen, dat vimentine in spierweefsel een andere conformatie aanneemt en zodoende een ander antigeen karakter bezit.

Lazarides, E., Granger, B.L., Gard, D.L., O'Connor, C.M., Breckler, J., Price, M., & Danto, S.I. (1982) Cold Spring Harbor Symposia on Quantitative Biology 46, 351-378.

Holtzer, H., Bennet, G.S., Tapscott, S.J., Croop, J.M., & Toyama, Y. (1982) Cold Spring Harbor Symposia on Quantitative Biology 46, 317-330.

6

Door te stellen, dat de door hun gebruikte cDNA-kloon pAct-1 kan coderen voor α-actine geven Soriano et al. blijk van het feit, dat ze zich de bron van hun kloon niet goed realiseren.

Soriano, P., Szabo, P., & Bernardi, G., (1982) EMBO J. 1, 579-583.

7

In tegenstelling tot hetgeen de naam doet vermoeden, is het niet zozeer de methaanvormende, maar veeleer de afval-afbrekende eigenschap die methaanbacterien tot commercieel interessante organismen maakt.

8

De zeer korte "in press" tijd van sommige EMBO Journal artikelen zet de wetenschappelijke chronologie vaak op z'n kop.

Geisler, N., & Weber, K. (1982) EMBO J. 1, 1649-1656.

Lehnert, M.E., Jorcano, J.L., Zentgraf, H., Blessing, M., Franz, J.K., & Franke, W.W. (1984) EMBO J. 3, 3279-3287.

9

Gezien de rol, die wetenschappelijke publicaties spelen bij het aanvragen van gesubsidieerd onderzoek, is de term "advertisement", die door sommige tijdschriften wordt toegevoegd aan artikelen, geheel correct.

10

Bij het universitaire onderzoek wordt zo langzamerhand meer tijd besteed aan het aanvragen van subsidie dan aan het schrijven van wetenschappelijke publicaties.

11

Naarmate de mens ouder wordt, wordt zijn speelgoed duurder.

12

De suggestie van minister Deetman, dat de eventueel in te voeren "sabbatical" voor wetenschappelijk personeel in dienst van de universiteiten wellicht ook voor een zwangerschap gebruikt kan worden lijkt met het oog op de leeftijdsopbouw van het universitaire corps niet erg realistisch.

Nijmegen, 19 juni 1985
Wim Quax