Proc. Natl. Acad. Sci. USA
Vol. 78, No. 7, pp. 4120–4123, July 1981
Biochemistry

Comparison of the proteins of two immunologically distinct intermediate-sized filaments by amino acid sequence analysis: Desmin and vimentin

(cytoskeleton/differentiation/eye lens/keratin/tropomyosin)

NORBERT GEISLER AND KLAUS WEBER

Max Planck Institute for Biophysical Chemistry, D-3400 Goettingen, Federal Republic of Germany

Communicated by Manfred Eigen, March 25, 1981

ABSTRACT **Although all intermediate-sized filaments (10-nm filaments) seem to show similar morphology and share a number of biochemical properties, different cell- and tissue-specific subclasses have been distinguished by immunological experiments and by differences in apparent molecular weights and isoelectric points of the major constituent proteins. In order to understand the degree of possible homology between these proteins, we have begun amino acid sequence analysis of the polypeptides. Here we characterize a large fragment of chicken gizzard and pig stomach desmin as well as the corresponding fragment from porcine eye lens vimentin. The fragments are situated at the carboxyl end and consist of 138–140 amino acid residues—i.e., some 28% of the corresponding polypeptide chains. The results show that the two immunologically distinct porcine proteins are different gene products. They show a related amino acid sequence but differ in 36% of the residues present in the carboxy-terminal part. In contrast the two desmins, although from species as diverse as chicken and pig, are more closely related and differ in less than 9% of the residues in the corresponding carboxy-terminal region. Thus tissue specificity overrides species divergence. These results are discussed in the light of previous immunological experiments. They lend further support to the hypothesis that intermediate filaments belong to a multigene family, which is expressed in line with certain rules of differentiation during embryogenesis.**

Intermediate-sized or 7- to 10-nm filaments form one of the three major components of the cytoskeleton. Probably due to their notorious insolubility, much less is known about their proteins than about those of the two other cytoskeletal components, microfilaments and microtubules. During the last 3 years different subclasses of intermediate filaments have been distinguished on grounds of biochemical criteria such as polypeptide molecular weights and isoelectric points. In addition, and perhaps more importantly, sets of antibodies became available that differentiate at least five subclasses without crossreactivity. Interestingly, this classification parallels to some extent certain stages in cell and tissue differentiation. Currently five major classes can be described: the cytokeratins of true desmosome-expressing epithelia, the desmin fibers of certain muscle types, the glial filaments of astrocytes, the neurofilaments of neurons, and the vimentin filaments in various nonepithelial cells, in particular mesenchymal cells (for reviews see refs. 1 and 2). This cell type-specific classification of adult tissues is often more complicated in cultured cells. Thus for reasons that are currently unclear, many permanent epithelial cell lines coexpress cytokeratins and vimentin (3), certain glioma cell lines produce vimentin in addition to glial filaments (4), and hamster BHK 21 cells reveal desmin and vimentin (5, 6).

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

In spite of the differences revealed by two-dimensional gel analysis and the lack of immunological crossreactivity experienced with many antibodies, the structural proteins of these filaments share several distinct properties. The filaments show a very similar ultrastructure and morphology and belong to the highly helical k-m-e-f-type of proteins (see, for instance, ref. 7). In addition they are all insoluble in physiological buffers (for references see above) and, with the exception of certain neurofilament proteins, contain polypeptides of molecular weights in the range 45,000–68,000 (for references see refs. 1 and 2).

In order to see whether different intermediate filament proteins are related in primary structure, we have started a comparative protein-chemical study based on partial amino acid sequence analysis. We chose stomach as the source of desmin, because it is generally the tissue of choice for the purification of this type of intermediate filament (1). In addition we have used eye lens fiber tissue as the source of an intermediate filament polypeptide distinct by apparent polypeptide molecular weight and isoelectric point. This protein has been immunologically and electrophoretically characterized (8) as related to the vimentin-type filaments typically expressed in fibroblasts and other mesenchymally derived cells as well as in most cultured cells (for references see refs. 1–4).

Here we show that cleavage of desmin and vimentin with 2-nitro-5-thiocyanobenzoic acid gives rise to a 15,000 molecular weight fragment situated at the carboxy terminus. We have isolated this fragment from chicken desmin, porcine desmin, and porcine vimentin and characterized the three fragments by amino acid sequence analysis. Our results prove at the molecular level that desmin and vimentin are different but related proteins, and our findings are relevant or current models of intermediate filament diversity and organization.

MATERIALS AND METHODS

Chicken gizzard desmin and porcine stomach desmin were purified by our standard procedure (9). Porcine eye lens vimentin was purified by a related method (10). The proteins behaved homogeneously on one-dimensional $NaDodSO_4$/polyacrylamide gels and were able to polymerize into morphologically normal 7- to 10-nm-thick filaments (9, 10). All three proteins were subjected to chemical cleavage at cysteine residues by using 2-nitro-5-thiocyanobenzoic acid (11) and gave rise to three polypeptides when analyzed by $NaDodSO_4$ gel electrophoresis; 10–30% of the protein remained unchanged and the remainder was converted to two large fragments with apparent molecular weights of 15,000 and 37,000. The fragments were separated by preparative $NaDodSO_4$/polyacrylamide gel electrophoresis and eluted (12). After dialysis against 0.1% $NaDodSO_4$ and lyophilization, protein stain and $NaDodSO_4$ were removed (13) and the purified 15,000 molecular weight fragment (fragment

II) was processed for digestion with trypsin, thermolysin, chymotrypsin, *Staphylococcus aureus* V8 protease, or carboxypeptidase Y, using standard procedures. Separation of peptides was performed by the preparative two-dimensional system on Whatman 3MM paper as described (9, 14). The peptides were recovered from the paper and characterized by amino acid composition, manual dansyl Edman degradation, carboxypeptidase, and leucine aminopeptidase digestion. Amide assignments were based on the Offord plot (15) and digestion with leucine aminopeptidase and carboxypeptidase. In addition, use was made of V8 protease cleavage. This enzyme shows a highly preferential cleavage at glutamic acid residues. If necessary, secondary enzymatic cleavage was performed to complete the sequence of the peptides. Examples of this approach have previously been given for the carboxy-terminal tryptic peptide of both desmins (9).

RESULTS

Isolation of the Carboxy-Terminal 15,000 Molecular Weight Fragment. In agreement with a cysteine content of 1 per molecule (9), both chicken gizzard and porcine stomach desmin gave rise to two distinct fragments when subjected to chemical cleavage (11) with 2-nitro-5-thiocyanobenzoic acid. The large difference in apparent molecular weight—i.e., 15,000 vs. 37,000—allowed easy separation of the two fragments by preparative $NaDodSO_4$/polyacrylamide gel electrophoresis. For both desmins, we conclude that fragment II (molecular weight 15,000) is the carboxy-terminal part for the following reasons. Fragment II has a carboxy-terminal leucine residue, as does desmin from chicken and pig (9). In addition, tryptic digestion of fragment II yielded in both cases (see below) a tryptic peptide previously recognized as the carboxy-terminal peptide in both desmins (9).

When pig eye lens vimentin was cleaved at cysteine, again two fragments were obtained. These were very similar in molecular weight to those observed with desmin. Again fragment II (molecular weight 15,000) was recognized as the carboxy-terminal part. Hydrazinolysis identified glutamic acid as the carboxy-terminal residue both in vimentin (10) and in its fragment II (average yield 0.4 mol/mol). Carboxypeptidase Y digestion implicated in both cases leucine preceding the carboxy-terminal glutamic acid residue (average yield of glutamic acid 0.3 mol/mol). In addition, amino acid sequence analysis (see below) identified a tryptic peptide related to the corresponding carboxy-terminal peptide of the two desmins.

Amino Acid Sequence Analysis of Chicken and Pig Desmin Fragment II. Because the amino terminus of a cysteinyl-carrying fragment becomes cyclized due to the chemical cleavage, direct Edman degradation cannot be performed without further chemical reactions (11). However, fragmentation by trypsin,

	1									10										20*
CD	CYS	GLU	ILE	ASP	ALA	LEU	LYS	GLY	THR	ASN	ASP	SER	LEU	MET	ARG	GLN	MET	ARG	GLU	MET
PD	CYS	GLU	ILE	ASP	ALA	LEU	LYS	GLY	THR	ASN	ASP	SER	LEU	MET	ARG	GLN	MET	ARG	GLU	LEU
PV	CYS	GLU	VAL	ASP	ALA	LEU	LYS	GLY	THR	ASN	GLU	SER	LEU	GLU	ARG	GLN	MET	ARG	GLU	MET

		*				*			*	30				*						40
CD	GLU	GLU	ARG	PHE	ALA	GLY	GLU	ALA	GLY	GLY	TYR	GLN	ASP	THR	ILE	ALA	ARG	LEU	GLU	GLU
PD	GLU	ASP	ARG	PHE	ALA	SER	GLU	ALA	SER	GLY	TYR	GLN	ASP	ASN	ILE	ALA	ARG	LEU	GLU	GLU
PV	GLU	GLU	ASN	PHE	ALA	VAL	GLU	ALA	ALA	ASN	TYR	GLN	ASP	THR	ILE	GLY	ARG	LEU	GLN	ASP

										50										60
CD	GLU	ILE	ARG	HIS	LEU	LYS	ASP	GLU	MET	ALA	ARG	HIS	LEU	ARG	GLU	TYR	GLN	ASP	LEU	LEU
PD	GLU	ILE	ARG	HIS	LEU	LYS	ASP	GLU	MET	ALA	ARG	HIS	LEU	ARG	GLU	TYR	GLN	ASP	LEU	LEU
PV	GLU	ILE	GLN	ASN	MET	LYS	GLU	GLU	MET	ALA	ARG	HIS	LEU	ARG	GLU	TYR	GLN	ASP	LEU	LEU

										70										80
CD	ASN	VAL	LYS	MET	ALA	LEU	ASP	VAL	GLU	ILE	ALA	THR	TYR	ARG	LYS	LEU	LEU	GLU	GLY	GLU
PD	ASN	VAL	LYS	MET	ALA	LEU	ASP	VAL	GLU	ILE	ALA	THR	TYR	ARG	LYS	LEU	LEU	GLU	GLY	GLU
PV	ASN	VAL	LYS	MET	ALA	LEU	ASP	ILE	GLU	ILE	ALA	THR	TYR	ARG	LYS	LEU	LEU	GLU	GLY	GLU

		*			*	*		*	*	90			*							100
CD	GLU	ASN	ARG	ILE	SER	ILE	PRO	MET	HIS	GLN	THR	PHE	ALA	SER	ALA	LEU	ASN	PHE	ARG	GLU
PD	GLU	SER	ARG	ILE	ASN	LEU	PRO	ILE		GLN	THR	PHE		SER	ALA	LEU	ASN	PHE	ARG	GLU
PV	GLU	SER	ARG	ILE	SER	LEU	PRO	LEU		PRO	ASN	PHE		SER	SER	LEU	ASN	LEU	ARG	GLU

					*					110										120
CD	THR		SER	PRO	ASP	GLN	ARG	GLY	SER	GLU	VAL	HIS	THR	LYS	LYS	THR	VAL	MET	ILE	LYS
PD	THR		SER	PRO	GLU	GLN	ARG	GLY	SER	GLU	VAL	HIS	THR	LYS	LYS	THR	VAL	MET	ILE	LYS
PV	THR	ASN	LEU	GLU	SER	LEU	PRO	LEU	VAL	ASP	THR	HIS	SER	LYS	ARG	THR	LEU	LEU	ILE	LYS

										130										140
CD	THR	ILE	GLU	THR	ARG	ASP	GLY	GLU	VAL	VAL	SER	GLU	ALA	THR	GLN	GLN	GLN	HIS	GLU	VAL
PD	THR	ILE	GLU	THR	ARG	ASP	GLY	GLU	VAL	VAL	SER	GLU	ALA	THR	GLN	GLN	GLN	HIS	GLU	VAL
PV	THR	VAL	GLU	THR	ARG	ASP	GLY	GLN	VAL	ILE	ASN	GLU	THR	SER	GLN	HIS	HIS	ASN	ASP	LEU

	141
CD	LEU
PD	LEU
PV	GLU

FIG. 1. Amino acid sequence of the carboxy-terminal fragment II of chicken gizzard desmin (CD, top lines), porcine stomach desmin (PD, middle lines), and porcine eye lens vimentin (PV, bottom lines). Differences between porcine desmin and chicken desmin are marked by asterisks (top lines). Differences between porcine desmin and porcine vimentin are underlined (bottom lines).

thermolysin, chymotrypsin, and V8 protease, followed by direct dansyl Edman sequence analysis of the isolated peptides, gave sufficient information to obtain the chicken desmin sequence shown in Fig. 1. The positioning of cysteine as residue 1 was not directly confirmed by characterization of the predicted 2-iminothiazolidine-4-carboxylyl residue (11). However, the corresponding peptides showed, as expected, a blocked amino terminus that was retained in the subsequently derived secondary fragments. A full account of the sequence determination will be given elsewhere.

By using the approach outlined above, the close relatedness of porcine and chicken desmin became apparent (Fig. 1). Thus after characterization of the isolated tryptic and thermolytic peptides by amino acid composition and electrophoretic and chromatographic properties, they could be arranged in all cases according to the chicken desmin sequence. Direct dansyl Edman degradation and secondary cleavage were performed only after an amino acid substitution was indicated by the criteria given above. Two peptides (residues 75–83 and 84–92) as well as the carboxy-terminal tryptic peptide (residues 126–141) have already been characterized, when the two desmins were compared in two-dimensional separation studies (9). At the amino terminus (residues 1–7) placement was made by homology with chicken desmin, using peptide(s) with a blocked amino terminus (Fig. 1). Because of the approach taken, the possibility does exist that in a very few cases the amide assignments shown in Fig. 1 could be incorrect, because some peptides were placed by homology (for criteria see above). However, we stress that the peptides involved are relatively small (i.e., fewer than 8 amino acids), and some assignments of glutamic acid residues were checked by V8 protease-derived peptides.

Amino Acid Sequence Analysis of Pig Vimentin Fragment II. Using the same approach, we have directly analyzed the sequence of the various peptides by the dansyl Edman procedure and obtained overlapping peptides to enable us to propose the sequence given in Fig. 1. In this case we have not used homology as a sufficient argument for placement but have completely determined the sequence of all peptides obtained. Again, assignment of residues 1–7 was made from peptides with a blocked amino terminus.

DISCUSSION

The sequence information reported for the carboxy-terminal 28% of chicken desmin, porcine desmin, and porcine vimentin is of general importance for understanding the similarities and differences previously observed between intermediate filaments isolated from different tissues and cells (1, 2).

The desmin sequences determined allow assessment of species divergence for a tissue-specific protein. The porcine vimentin sequence allows us to estimate the degree of tissue specificity within the same species. Whereas pig desmin and pig vimentin fragment II differ in sequence at 50 positions (36%), the desmin fragment II from pig and chicken differ at only 12 positions (9%) (Fig. 1). This point is further illustrated by the three continuous regions without replacements covering 19, 36, and 47 residues, respectively, in the two desmins (residues 1–19, 106–141, and 35–81). In contrast, comparison of porcine desmin and vimentin reveals only one longer region covering 37 amino acids (residues 48–84) with one conservative replacement. Thus tissue specificity clearly overrides the species divergence, there being only four places (positions 82, 86, 89, and 93) where porcine desmin equals porcine vimentin rather than chicken desmin.

In spite of the high relative number of amino acid exchanges between porcine vimentin and desmin (36%), the two proteins

KERATEINE-A (52-60)	-	ARG / LYS	- TYR -	GLU -	GLU -	GLU -	VAL / ILE	- ALA -	LEU -	ARG -
KERATEINE-A (101-109)	-	LEU -	TYR -	GLU -	GLU -	GLU -	ILE -	ARG -	VAL -	LEU -
DESMIN (37-45)	-	ARG -	LEU -	GLU -	GLU -	GLU -	ILE -	ARG -	HIS -	LEU -
αB TROPOMYOSIN (220-228)	-	LYS -	TYR -	GLU -	GLU -	GLU -	ILE -	LYS -	VAL / LEU	- LEU -

FIG. 2. Related amino acid sequences found in desmin fragment II (Fig. 1), the helical fragment of merino sheep wool S-carboxymethylkerateine-A (25), and rabbit skeletal muscle tropomyosin (26). Residue numbers and sequences are from Fig. 1 and refs. 25 and 26, respectively. Two residues in the same position reflect the microheterogeneity previously established.

are closely related in structure. The substitutions are mostly conservative in nature and explainable by single-base exchanges of the corresponding codons. In addition, however, a variety of nonconservative replacements are detected, including some that necessarily change the charge patterns of the proteins involved. Thus two-dimensional analysis of tryptic peptides implied a more remote relationship than was actually revealed in the final sequences. So far we have not localized those amino acid substitutions or posttranslational modifications responsible for isoelectric variants or phosphorylated residues (see, for instance, ref. 5).

Our sequence data, showing that desmin and vimentin are related in structure, raise the question as to whether other intermediate filament proteins—i.e., glial fibrillar acidic protein, neurofilaments, and cytokeratins—are also related in sequence to vimentin and desmin. Experiments similar to those used here should yield an answer. In addition the structural relatedness of desmin and vimentin, particularly if extended to other intermediate filament proteins, might also help in explaining certain conflicting immunological results obtained in different laboratories. The antigenic determinants recognized by the various antibodies against intermediate filament proteins have not yet been elucidated. However, if all intermediate proteins are related in structure, the following types of antibodies might be expected. (*i*) Highly tissue-specific antibodies distinguishing different subclasses of intermediate filaments in a broad cross-species manner (see, for instance, refs. 3, 16, and 17). (*ii*) Tissue-specific antibodies with a much narrower cross-species specificity (see, for instance, refs. 16, 18, and 19). (*iii*) Antibodies with a lower tissue- and cell type-specificity (see, for instance, ref. 20). (*iv*) Antibodies recognizing at least one antigenic determinant common to several, or perhaps all, classes of intermediate-sized filaments (see ref. 21). This interpretation does not detract from, but indeed strongly emphasizes, the usefulness of the highly cell type-specific antibodies, which are suitable when studies in embryology, differentiation, tissue typing, and pathological disorders are at the center of interest.

The similar physical and biochemical properties of intermediate-sized filaments suggest that there may be common structural determinants that govern the three-dimensional organization of these molecules. In this connection the secondary structure of fragment II is of interest, especially because the overall similarity in amino acid composition between the fragment and the whole polypeptide chain suggests that fragment II might be diagnostic of the whole molecule. Intermediate-sized filaments are rich in α-helix and are assumed to contain extended regions of coiled-coil structure (see, for instance, refs. 7 and 22). By using the heptad convention of Stewart and

McLachlan (23–25) on fragment II from pig stomach desmin, a characteristic arrangement of amino acid residues 1–85 is apparent. The predicted positions *a* and *d* are, as required by this convention, characterized by an abundance of hydrophobic residues. This abundance is close to that noted in the α-helical segment of merino wool kerateine-A (25) and is indicative of a coiled-coil conformation. Screening the sequences of the isolated helical segment of kerateine-A (25) and rabbit muscle tropomyosin (26), which is another highly α-helical molecule with coiled-coil structure (24–28), we noticed a closely related sequence (Fig. 2). Thus two kerateine-A sequences show five identical neighboring amino acid residues and tropomyosin and kerateine share up to seven identical residues in a stretch of eight. The two desmins share with the two kerateine-A sequences either five neighboring residues or five out of six residues. If conservative amino acid exchanges are allowed for, the relatedness is even higher. This observation again suggests that desmin fragment II may be in part α-helical and involved in coiled-coil organization. This suggestion does not conform with a current model of intermediate filaments, which assumes the absence of α-helical arrangements in the carboxy-terminal third of the molecules (7). Future structural experiments on fragments II and the extension of the amino acid sequences should help to elucidate the principles of intermediate filament organization and decide if three-stranded (7) or two-stranded (24) ropes are involved.

We thank S. Lechte for technical assistance and greatly appreciate the protein chemical expertise of U. Plessmann. This work was supported by the Deutsche Forschungsgemeinschaft (Grant We 338/2) and the Max Planck Society.

1. Lazarides, E. (1980) *Nature (London)* **283**, 249–256.
2. Weber, K. & Osborn, M. (1981) *Cell Surface Rev.* **7**, in press.
3. Franke, W. W., Schmid, E., Osborn, M. & Weber, K. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5034–5038.
4. Paetau, A., Virtanen, I. & Stenman, S. (1979) *Acta Neuropathol.* **47**, 71–74.
5. Gard, D. L., Bell, P. B. & Lazarides, E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3896–3898.
6. Tuszynski, G. P., Frank, E., Damsky, C., Buck, C. A. & Warren, L. (1979) *J. Biol. Chem.* **254**, 6138–6143.
7. Steinert, P. M., Idler, W. W. & Goldman, R. D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4534–4538.
8. Ramaekers, F. C. S., Osborn, M., Schmid, E., Weber, K., Bloemendal, H. & Franke, W. W. (1980) *Exp. Cell Res.* **127**, 309–327.
9. Geisler, N. & Weber, K. (1980) *Eur. J. Biochem.* **111**, 425–433.
10. Geisler, N. & Weber, K. (1981) *FEBS Lett.* **125**, 253–256.
11. Jacobsen, G. R., Schaffer, M. H., Stark, G. R. & Vanaman, T. C. (1973) *J. Biol. Chem.* **248**, 6583–6591.
12. Lazarides, E. & Weber, K. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2268–2272.
13. Henderson, L. E., Oroszlan, S. & Konigsberg, W. (1979) *Anal. Biochem.* **93**, 153–157.
14. Vandekerckhove, J. & Weber, K. (1978) *Eur. J. Biochem.* **90**, 451–462.
15. Offord, R. E. (1966) *Nature (London)* **211**, 591–593.
16. Franke, W. W., Schmid, E., Winter, S., Osborn, M. & Weber, K. (1979) *Exp. Cell Res.* **123**, 25–46.
17. Sun, T. T., Shin, C. & Green, H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2813–2817.
18. Lazarides, E. & Baltzer, D. R. (1979) *Cell* **14**, 429–438.
19. Hynes, R. O. & Destree, A. (1978) *Cell* **13**, 151–163.
20. Campbell, G. R., Chemley-Campbell, J., Groeschel-Stewart, U., Small, I. V. & Anderson, P. (1978) *J. Cell Sci.* **37**, 303–322.
21. Pruss, R. M., Mirsky, R., Raff, M. C., Anderton, B. & Thorpe, R. (1980) *J. Cell Biol.* **87**, 178.
22. Steinert, P. M. (1978) *J. Mol. Biol.* **123**, 49–70.
23. Stewart, M. & McLachlan, A. D. (1976) *J. Mol. Biol.* **103**, 251–269.
24. McLachlan, A. D. (1978) *J. Mol. Biol.* **124**, 297–304.
25. Crewther, W. G., Inglis, A. S. & McKern, N. M. (1978) *Biochem. J.* **173**, 365–371.
26. Mak, A. S., Smillie, L. B. & Stewart, G. R. (1980) *J. Biol. Chem.* **255**, 3647–3655.
27. Stone, D. & Smillie, L. B. (1978) *J. Biol. Chem.* **253**, 1137–1148.
28. McLachlan, A. D. & Stewart, M. (1975) *J. Mol. Biol.* **98**, 293–304.