Amino acid sequence characterization of mammalian vimentin, the mesenchymal intermediate filament protein

Norbert Geisler, Uwe Plessmann and Klaus Weber

Max Planck Institute for Biophysical Chemistry, D-3400 Goettingen, FRG

Received 5 September 1983

The amino-terminal 98 residues of porcine vimentin have been determined by amino acid sequence studies. Extensive overlap is seen with the corresponding region of the carboxyterminal 448 residues of hamster vimentin predicted from DNA sequence studies, which left the very amino-terminal region unknown. The combined data show that contrary to gel electrophoretic results, mammalian vimentin contains only about 467 residues, and that species-specific drift occurs mainly in the amino-terminal non-α-helical array. The results are discussed parallel to emerging concepts on intermediate filament protein diversity.

Keratin *Neurofilament* *Desmin* *GFA* *Vimentin*

1. INTRODUCTION

There are 5 classes of intermediate filament (IF) proteins. The molecular basis of their common and distinct properties has emerged only recently due to amino acid sequence data [1–8] and to a parallel dissection of the molecules into their structurally distinct domains [7]. A related α-helical middle domain able to form coiled-coils accounts for the α-type diffraction pattern but does not seem sufficient to allow filament assembly [1,7]. Thus interest has also focussed on the two non-α-helical terminal domains, which can show extreme variability both in sequence and length [1,2,4–6] and are assumed to strongly influence the biochemically distinct properties of the different IF proteins [1,5]. Therefore, it is regrettable that in spite of extensive partial sequence information on various IF proteins, a complete sequence is only available for myogenic desmin [1] and for one [5] of the probably 19 distinct epithelial cytokeratins [9].

Mesenchymal vimentin is, like desmin, a well-characterized non-epithelial IF protein. DNA sequence studies of cloned hamster vimentin have predicted the carboxyterminal 448 residues, but have left the amino-terminal region unknown [8,10]. We have determined this region by protein chemical methods in porcine vimentin for two reasons: to fully characterize a mammalian vimentin in polypeptide length and to detect how much species-specific drift can occur in this variable array.

2. MATERIALS AND METHODS

Porcine eye lens vimentin was purified [11] and fragmented with lysine-specific protease (Boehringer, Mannheim) in the presence of 6 M urea. Digests were chromatographed on CM–cellulose in order to obtain the basic amino-terminal fragment which emerges as the final peak. These procedures have been described for the neurofilament protein [4]. Digestion by trypsin and thermolysin, separation of peptides by two-dimensional paper methods as well as their characterization by amino acid composition and stepwise Edman degradation were by established procedures (see [7,11]).

3. RESULTS AND DISCUSSION

Since we were interested only in the amino-

Published by Elsevier Science Publishers B.V.

terminal array, digestion with lysine-specific protease in the presence of urea, followed by chromatography on CM–cellulose, was used. This approach selects specifically for a region devoid of lysine and rich in arginine as indicated by the DNA-deduced incomplete amino acid sequence of hamster vimentin [8]. Gel electrophoresis (not shown) indicated a major fragment (>80%) of an app. M_r of 10000, and amino acid composition pointed to a fragment whose wealth of hydroxyamino acids and arginine followed the sequence predictions. Two sets of fingerprints using trypsin and thermolysin provided the necessary information to arrive at the sequence shown in fig.1. It accounts for 98 residues. A few peptides obtained in minor yield (not shown) can extend this sequence by a further 23 residues to the next lysine residue predicted by the DNA sequence.

Comparison of the porcine sequence (fig.1) with the incomplete DNA-deduced hamster sequence, which could lack some 20–50 residues at the amino-terminus [8], leads to the following conclusions:

(i) As of residue 25 in the protein sequence there is very good alignment with the predicted sequence if allowance for a few amino acid exchanges and 3 insertions is made. We notice 7 exchanges (positions 26, 52, 53, 59, 63, 75, 84) and 3 insertions (positions 41, 61 and 62). These differences occur

```
      1                                   10                                  20
PV    (Ser,Thr) Arg Thr Val Ser Ser Ser Ser Tyr Arg Arg Met Phe Gly Gly Pro Gly Thr  Ala
HV                                                                  Pro Arg His  Leu

      21                  *               30                                  40
PV    Ser Arg  Pro Ser Ser Ser Arg Ser Tyr Val Thr Thr Ser Thr Arg Thr Tyr Ser Leu  Gly
HV    Glu Pro  Ala Gly Ser Asn Arg Ser Tyr Val Thr Thr Ser Thr Arg Thr Tyr Ser Leu  Gly

      41                                  50      *   *                       *   60
PV    Ser Ala  Leu Arg Pro Ser Thr Ser Arg Ser Leu Ser Thr Ser Ser Pro Gly Gly Val (Gly
HV    -   Ala  Leu Arg Pro Ser Thr Ser Arg Ser Leu Tyr Ser Ser Ser Pro Gly Gly Ala  Tyr
                                          •

      61        *                         70                  *                   80
PV    Tyr,Tyr) Ala Thr Arg Ser Ser Ala Val Arg Leu Arg Ser Ser Val Pro Gly Val Arg  Leu
HV    -   -    Val Thr Arg Ser Ser Ala Val Arg Leu Arg Ser Ser Met Pro Gly Val Arg  Leu

      81            *                     90                              98
PV    Leu Gln  Asp Ala Val Asp Phe Ser Leu Ala Asp Ala Ile Asn Thr Glu Phe Lys
HV    Leu Gln  Asp Ser Val Asp Phe Ser Leu Ala Asp Ala Ile Asn Thr Glu Phe Lys
```

Fig.1. Alignment of the amino acid sequence of the amino-terminal 98 residues of porcine vimentin (PV) with the corresponding sequence predicted for hamster vimentin (HV) using cloned DNA [8]. The DNA sequence does not extend to the amino-terminal end [8]. The amino-terminal residue in PV is blocked. The blocking group in IF proteins is generally the acetyl group [5]. In PV the order of residues 1 and 2 as well as 60–62, which are given in parentheses, is not yet established. Note the good alignment between PV and HV starting at residue 25. Note also the 3 deletions (dashes) and the 7 amino acid exchanges marked by a star within 60 residues. The total misfit between residues 18 and 24 (underlined in the HV sequence) could be due to a one nucleotide deletion upon cloning of hamster vimentin (see text).

exclusively prior to residue 85; i.e., the position past which myogenic desmin and mesenchymal vimentin are well aligned and show a sequence homology of about 65% [3,8,10–12]. Thus, in agreement with the hypervariable character of the headpiece region in different intermediate filament proteins [1,4–6], the two mammalian vimentins already show several amino acid exchanges. In spite of the high homology between vimentin and desmin seen along the α-helical rod domain and the carboxyterminal tailpiece [1,3,8,10–12] there is only a very limited relation in the amino-terminal headpieces, which centers mainly on residues 8–16 of porcine vimentin (fig.1) *vs* residues 11–19 in chicken desmin [1,7];

(ii) The sole ambiguity between the two sequences (fig.1) concerns the first 8 residues predicted by the DNA sequence [8]. Given the excellent alignment past these residues, which should correspond to positions 17–25 in the protein sequence, one is struck by a total misfit. We have therefore closely inspected the published nucleotide sequence in this region. We notice that an insertion of a single A (adenine) residue between nucleotides 22 and 23 would not alter the following reading frame, but would totally change the sequence prior to the hypothetical insert. Instead of reading Pro–Arg–His–Leu–Glu–Pro–Ala–Gly–Ser–Asn–Arg [8] the sequence would be Pro–Gly–Thr–Ser–Asn–Arg–Gln–Ser–Ser–Asn–Arg. This hypothetical sequence of hamster vimentin agrees in 7 out of 11 positions with residues 17–27 of the porcine protein sequence. The 4 other positions show amino acid substitutions explainable by single base changes of the corresponding codons. Thus we conclude that either the two mammalian vimentins differ distinctly over this short array, or a single nucleotide deletion close to the end of the hamster vimentin DNA has occurred during cloning.

The protein data on porcine vimentin (fig.1) taken together with the DNA data extending to the carboxyl end of hamster vimentin [8] provide a prototype mammalian vimentin. The polypeptide chain contains about 467 residues corresponding to an M_r-value close to 54000. Thus mammalian vimentin is only 4 residues longer than chicken desmin [1]. This result is unexpected considering previous gel electrophoretic data pointing to an M_r-value of 57000 for the vimentin molecule [3,9–11]. However, such deviations are well within the error of the method, since for instance a single amino acid substitution in a human actin resulted in an 8% increase in app. M_r [13]. A previous comparison of the carboxyterminal 177 residues of porcine [3,11] and hamster vimentin has only revealed two amino acid replacements [8]. This low number contrasts sharply with the 7 exchanges and 3 deletions documented in fig.1 for the region covering residues 25–84 in the amino-terminal array. Thus the headpiece region, known to be very variable among different intermediate filament proteins, is also the major region accounting for species-specific differences of the same protein.

ACKNOWLEDGEMENTS

We greatly appreciate the technical help of W. Koch and thank Dr H. Bloemendal for sending us his sequence prior to publication.

REFERENCES

[1] Geisler, N. and Weber, K. (1982) EMBO J. 1, 1649–1656.
[2] Hanukoglu, I. and Fuchs, E. (1982) Cell 31, 243–252.
[3] Geisler, N., Plessmann, U. and Weber, K. (1982) Nature 296, 448–450.
[4] Geisler, N., Kaufmann, E., Fischer, S., Plessmann, U. and Weber, K. (1983) EMBO J. 2, 1295–1302.
[5] Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.C. (1983) Nature 302, 794–800.
[6] Hanukoglu, I. and Fuchs, E. (1983) Cell 33, 915–924.
[7] Geisler, N., Kaufmann, E. and Weber, K. (1982) Cell 30, 277–286.
[8] Quax-Jeuken, Y.E.F.M., Quax, W.J. and Bloemendal, H. (1983) Proc. Natl. Acad. Sci. USA 80, 3548–3552.
[9] Moll, R., Franke, W.W., Schiller, D.L., Geiger, B. and Krepler, R. (1982) Cell 31, 11–24.
[10] Dodemont, H.J., Sariano, P., Quax, W.J., Raemakers, F., Lenstra, J.A., Groenen, M.A.M., Bernardi, G. and Bloemendal, H. (1982) EMBO J. 1, 167–171.
[11] Geisler, N. and Weber, K. (1981) Proc. Natl. Acad. Sci. USA 79, 4120–4123.
[12] Weber, K., Shaw, G., Osborn, M., Debus, E. and Geisler, N. (1983) Cold Spring Harbor Symp. Quant. Biol., in press.
[13] Vandekerckhove, J., Leavitt, J., Kanuaga, T. and Weber, K. (1980) Cell 22, 893–899.