The EMBO Journal vol.8 no.11 pp.3221 - 3227, 1989

Cytoplasmic intermediate filament proteins of invertebrates are closer to nuclear lamins than are vertebrate intermediate filament proteins; sequence characterization of two muscle proteins of a nematode

Klaus Weber, Uwe Plessmann and Werner Ulrich

Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, D-3400 Göttingen, FRG

Communicated by K.Weber

The giant body muscle cells of the nematode *Ascaris lumbricoides* show a complex three dimensional array of intermediate filaments (IFs). They contain two proteins, A (71 kd) and B (63 kd), which we now show are able to form homopolymeric filaments *in vitro*. The complete amino acid sequence of B and 80% of A have been determined. A and B are two homologous proteins with a 55% sequence identity over the rod and tail domains. Sequence comparisons with the only other invertebrate IF protein currently known (*Helix pomatia*) and with vertebrate IF proteins show that along the coiled-coil rod domain, sequence principles rather than actual sequences are conserved in evolution. Noticeable exceptions are the consensus sequences at the ends of the rod, which probably play a direct role in IF assembly. Like the *Helix* IF protein the nematode proteins have six extra heptads in the coil 1b segment. These are characteristic of nuclear lamins from vertebrates and invertebrates and are not found in vertebrate IF proteins. Unexpectedly the enhanced homology between lamins and invertebrate IF proteins continues in the tail domains, which in vertebrate IF proteins totally diverge. The sequence alignment necessitates the introduction of a 15 residue deletion in the tail domain of all three invertebrate IF proteins. Its location coincides with the position of the karyophilic signal sequence, which dictates nuclear entry of the lamins. The results provide the first molecular support for the speculation that nuclear lamins and cytoplasmic IF proteins arose in eukaryotic evolution from a common lamin-like predecessor.

Key words: intermediate filaments/keratins/lamins/nematode/vimentin

Introduction

In a higher vertebrate at least 34 distinct proteins can act as structural elements of the various cytoplasmic intermediate filaments (IFs), which are one of the three major fibrous elements (for reviews see Osborn and Weber, 1986; Steinert and Roop, 1988). Although the evolutionary origin of the complexity of vertebrate IF proteins is not understood, biochemical and immunological studies on nematodes, molluscs and annelids (Bartnik *et al.*, 1986, 1987a,b) suggest that invertebrates may display a lower IF complexity. Current evidence for invertebrates indicates two major IF prototypes: a neuronal and a non-neuronal one (Bartnik *et al.*, 1987b). An early evolutionary separation of neuronal and non-neuronal IF genes is also suggested from data on the intron positions of vertebrate IF genes (Lewis and Cowan, 1986).

Vertebrate cytoplasmic IF proteins and the nuclear lamins display related rod domains (Fisher *et al.*, 1986; McKeon *et al.*, 1986; see also Höger *et al.*, 1988) which reveal the terminal consensus sequences earlier delineated in comparisons of the various IF subtypes (Geisler and Weber, 1982, 1986; Hanugoklu and Fuchs, 1983; Steinert *et al.*, 1983). Some nuclear lamins can be forced *in vitro* to form IF structures and at least the nuclear lamina of the *Xenopus* oocyte is built from ordered meshworks of IF-related filaments (Aebi *et al.*, 1986), which seem to arise from a single lamin species (Stick, 1988). These relations between a nuclear and a cytoplasmic structure have led us and others to speculate that lamins and cytoplasmic IF proteins arose from a common lamin-like progenitor in early eukaryotic evolution (Osborn and Weber, 1986; Bartnik *et al.*, 1987b; Franke, 1987a; Myers *et al.*, 1987; Steinert and Roop, 1988). Any evaluation of this speculation requires sequence data on invertebrates. The first and still only sequence of an invertebrate lamin, that of the arthropod *Drosophila*, emphasized a clear relation to the lamins of vertebrates (Gruenbaum *et al.*, 1988; see also Höger *et al.*, 1988). The sequence of an invertebrate cytoplasmic IF protein provided a novel twist. IF isolated from the oesophagus epithelium of the mollusc *Helix pomatia* consist of two proteins, A and B (Bartnik *et al.*, 1985). They differ only at the C-terminal end where the larger A protein has an extension from the sequence common to both A and B (Weber *et al.*, 1988).

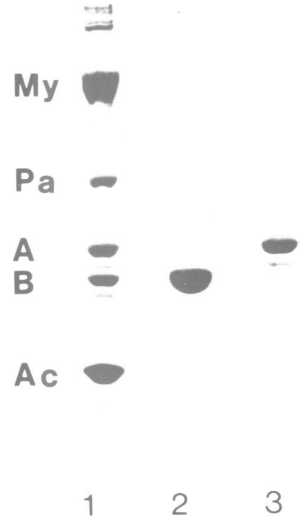

Fig. 1. Purification of the two ascaris IF proteins. **Lane 1** shows the gel electrophoretic analysis of the myofibrillar residue after low and high salt extraction. **Lanes 2** (protein B) and **3** (protein A) show the purified IF proteins. Positions of myosin heavy chain (My), paramyosin (Pa), proteins A, B and actin (Ac) are marked on the left.

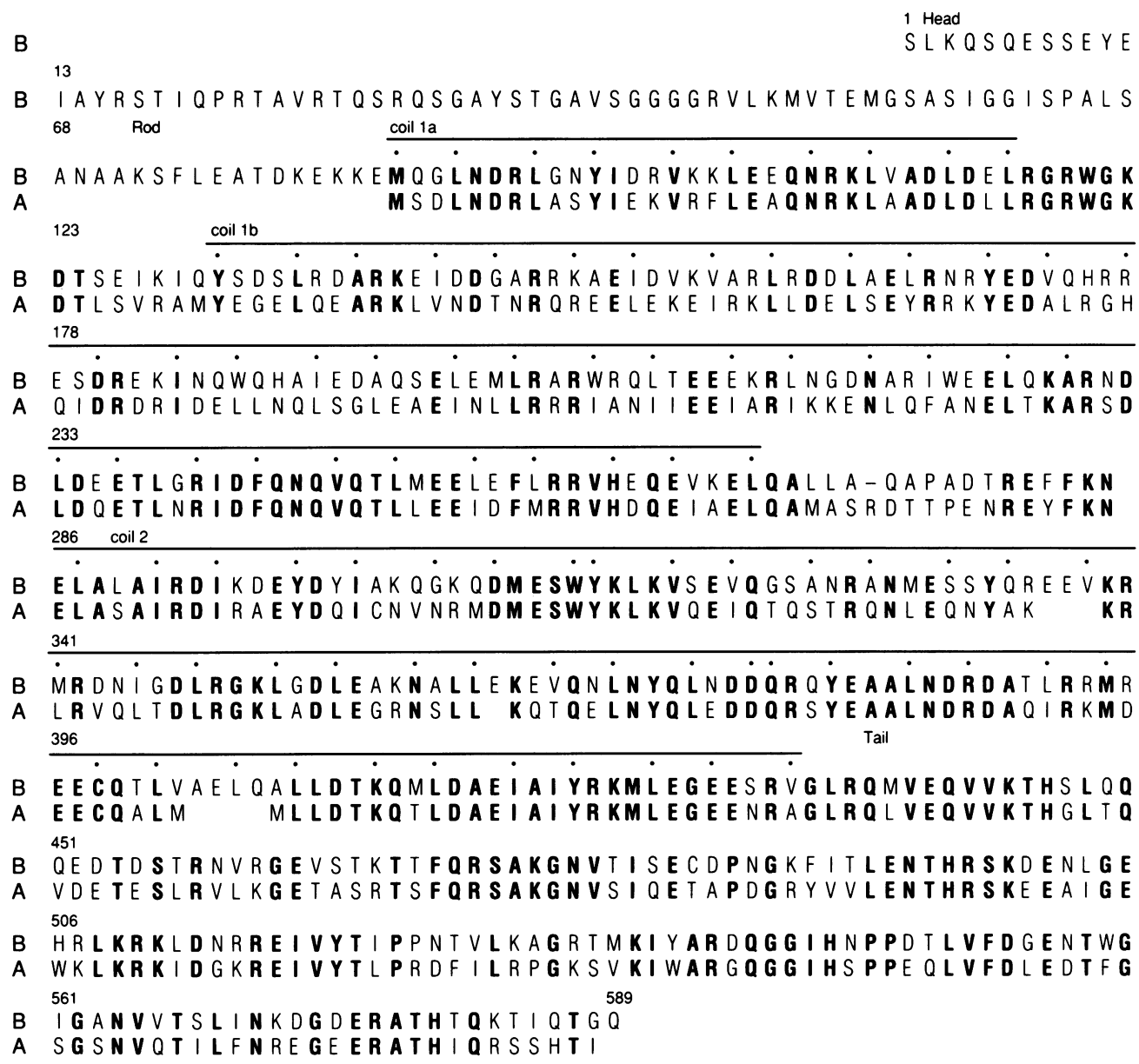

Fig. 2. The complete sequence of the *Ascaris* IF protein B and an extended partial sequence for protein A. The sequences document the three structurally distinct domains of IF proteins (head, rod and tail). Segments of the rod with coiled-coil forming ability (coils 1a, 1b and 2) are marked by lines. Dots indicate the primarily hydrophobic residues in the *a* and *d* positions of consecutive heptads (for nomenclature on IF sequences see Geisler and Weber, 1982; Geisler *et al.*, 1982a). Identical residues in the sequences of A and B are shown in bold type. Note that A and B are highly related but distinct polypeptides. They show 55% sequence identity over the rod and tail domains. The A sequence is largely complete except for the head region and a few small gaps in the rod domain. The lack of some overlapping fragments was readily overcome by the pronounced homology between A and B. A dash introduced in the spacer 2 region between coils 1b and 2 of the B sequence allows for better alignment as the sequence of A here has one additional residue. Numbers refer to the complete B sequence. For derivation of the blocked N-terminal end of B see text.

More importantly the coil 1b segment in the rod domain of the *Helix* IF proteins reveals the presence of six additional heptads (i.e. 42 residues) when compared with vertebrate IF proteins. Since all nuclear lamins show these extra heptads in their coil 1b sequences (Fisher *et al.*, 1986; Krohne *et al.*, 1987; Gruenbaum *et al.*, 1988; Höger *et al.*, 1988; Stick, 1988; Vorburger *et al.*, 1989) it was important to explore how general the results on the mollusc IF were for invertebrates. Here we show that the two muscle IF proteins of the giant nematode *Ascaris lumbricoides* (Bartnik *et al.*, 1986) also reveal the increased coil 1b length characteristic of lamins. This enhanced lamin–invertebrate IF protein relation extends unexpectedly into the large tail domains in all sequences of invertebrate IF proteins currently available.

Results

Purification of two IF proteins from Ascaris muscle and the in vitro formation of homopolymeric IF

Freshly dissected body musculature was first extracted by a low and a high salt buffer (see Materials and methods). The myofibrillar residue was solubilized in 8 M urea. Anion exchange chromatography on DE52 and Mono-Q as well as gel filtration provided both proteins in pure form (Figure 1). Western blotting of the myofibrillar residue with the murine monoclonal antibody IFA showed that A and B are the only reactive polypeptides. Their apparent mol. wts are 71 (A) and 63 kd (B) respectively (see also Bartnik *et al.*, 1986).

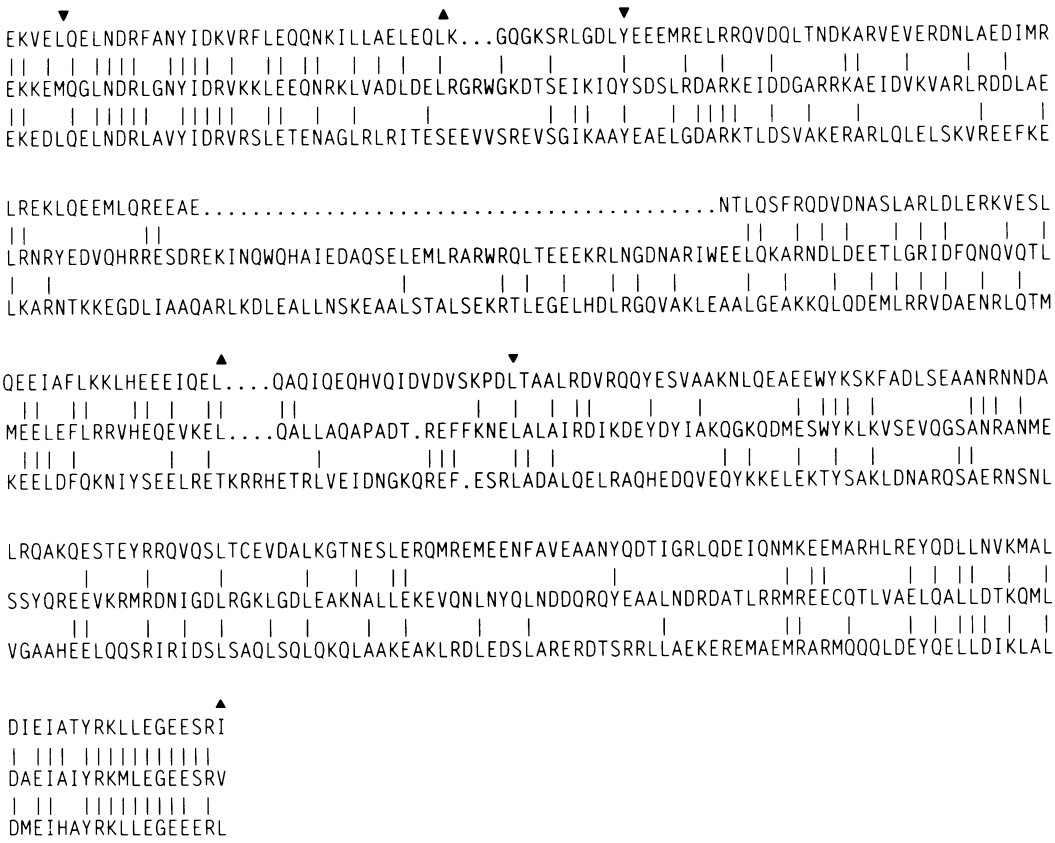

Fig. 3. Sequence alignments of rod domains. The figure shows the alignment of the rod domains of the *Ascaris* IF protein B (middle line; residues 81–433 from Figure 2), of human vimentin (upper line; Ferrari *et al.*, 1986) and of human lamin A (lower line; Fisher *et al.*, 1986; McKeon *et al.*, 1986). Arrowheads pointing down or up mark the starts and ends of coils 1a, 1b and 2 (for orientation see Figure 2). Note the six extra heptads (42 residues) in the coil 1b segment of the rod domains of the lamin and the *Ascaris* protein, which introduce a large gap in the corresponding vimentin sequence. Minor length variations are seen in spacer 1 (between coils 1a and b) for vimentin and in spacer 2 (between coil 1b and 2) for all three proteins. Only identical residues are marked (vertical lines). Note the high density of identical residues at the ends of the rods (early coil 1a and end of coil 2), where lamins and IF proteins show the same consensus type sequences (see Results). Over most of the remaining parts of the rods sequence principles rather than actual sequences are conserved. The extra six heptads found in lamin and IF-B of *Ascaris* show low homology (see text).

The complete protein sequence of Ascaris IF-B shown in Figure 2 is based on extensive automated sequencing of fragments obtained by chemical cleavage and secondary peptides derived from them (see Materials and methods). The calculated mol. wt of 67 752 daltons is in fair agreement with the previous estimate of 63 kd based on gel electrophoresis (Bartnik *et al.*, 1986). Although the sequence shows several potential cleavage sites for thrombin, this enzyme splits the native B molecule primarily at Arg 22, which is located in the head domain. The N-terminal thrombic peptide as well as secondary fragments derived from it were of great help in identification of the sequence at the blocked N-terminal end. The identification of residues 1 and 2 as serine and leucine respectively, is so far based only on the amino acid compositions of the peptides carrying the blocking group. Since *N*-acetylated hydroxyamino acids are common at the end of many IF proteins (see for instance Weber *et al.*, 1988), we have tentatively assigned serine to position 1 (Figure 2).

The sequence shows the canonical features of IF proteins. It defines the non α-helical head domains as residues 1 to ~72. The subsequent α-helical rod domain has coiled-coil forming ability and a length of 360 residues and shows the typical segmentation into subdomains (coils 1a, 1b and 2) separated by short non-helical spacers (for definition and nomenclature of IF proteins see Geisler and Weber, 1982). Based on secondary structure prediction rules the C-terminal tail domain of 157 residues could assume a globular conformation. It is relatively rich in charged residues. For instance in positions 496–517, 15 out of 22 residues are charged. A smaller cluster of 5 charges occurs about 15 residues prior to the C-terminal end.

Sequence comparisons; IF-B shares two structural features with lamins

Figure 3 shows the sequence alignment of the rod domains for the *Ascaris* IF-B protein, human vimentin and human lamin A. In contrast to vimentin and all currently known vertebrate IF proteins, the nuclear lamins have an additional six heptads (42 residues) in the coil 1b domain (for references see above). The same feature is present both in the *Ascaris* IF protein (Figures 2 and 3) and in the only other invertebrate IF protein currently known by sequence, i.e. the protein from *H.pomatia* (Weber *et al.*, 1988). The two ends of the rods display the consensus-type sequences already recognized in the first comparisons of vertebrate IF proteins (Geisler and Weber, 1982; Hanugoklu and Fuchs, 1983; Steinert *et al.*, 1983). They cover the first part of coil 1a and the end of coil 2. Over the remaining parts of the rods, sequence principles (consecutive heptads, segmentation of the helices, etc.) rather than actual sequences are conserved. When the rod of the *Ascaris* IF-B protein is compared with the rods

Fig. 4. Sequence alignment of tail domains. The upper panel (**a**) gives the alignment of the tail domain of the *Ascaris* protein IF-B with the tail domains of four different nuclear lamins. *Drosophila* lamin (LDmo; upper line; Gruenbaum *et al.*, 1988), murine lamin B (LmB; second line; Höger *et al.*, 1988), *Xenopus* lamin L1 (LXL1; third line; Krohne *et al.*, 1987) and human lamin A (LhA; fourth line; Fisher *et al.*, 1986) are presented in the consensus alignment of Höger *et al.* (1988). The sequence of the tail domain of the *Ascaris* IF B (fourth line) covers the entire tail domain (residues 434–589 from Figure 2). Note the excellent alignment of the last 118 residues (indicated by the arrow) along the lamin sequences. Bold type marks positions of amino acid identity in the *Ascaris* protein and one or more of the lamin sequences (see also Table I). Triangles mark the karyophilic signal region and asterisks indicate the cysteine containing motif CXXXM typically found at the C-terminal end of all lamins (see Discussion). The larger human lamin A has an additional domain of ~70 residues, indicated by dots, before this motif is reached. Note that the sequence alignment of the *Ascaris* IF protein introduces a deletion of ~15–20 residues around the position of the karyophilic signal. The high homology segment (arrow) starts past this position and ends at a location where the three vertebrate lamins diverge in sequence from the *Drosophila* lamin. The first 12 dots above the top sequence mark a motif detected earlier (Franke, 1987b; Höger *et al.*, 1988). It seems to occur a second time in a more degenerate form (see Discussion). The lower panel (**b**) shows the comparison of the tail domains of the *Ascaris* IF B (upper line) and the *H.pomatia* IF B (lower line). The latter sequence (Weber *et al.*, 1988) was completed in this study by the determination of the C-terminal 23 residues (see Results). Bold type marks identical residues. The arrow shows the position of the high homology segment of 118 residues defined in the comparison of the tail domains of the *Ascaris* IF-B and the nuclear lamins (panel a). Over this segment maximum homology between the two invertebrate IF proteins is only reached when several gaps are introduced which are longer than those seen in panel a. Dots above the top sequence mark a repeated motif (see panel a and Discussion).

of other IF-type molecules, maximal identity values are 115 (non-neuronal IF protein of *H.pomatia*) or around 100–110 (vimentin, desmin, neurofilament L protein from various species and the nuclear lamins). The latter numbers correspond to 28–31% identity. Noticeably lower values are found when IF-B is compared with the various vertebrate keratins present in the databank. Now only 75 (keratin I type molecules) to 85 (keratin II type molecules) identical residues are counted (21–24% identity). Figure 4a also shows that the extra six heptads of the invertebrate IF protein and the nuclear lamin are a low homology region of the rod.

In contrast to the various vertebrate IF proteins, the *Ascaris* protein shows good homology with the lamins also along the tail domain. As shown in Figure 3, the last 118 residues of the invertebrate IF-B protein can be directly aligned with the consensus presentation of four nuclear lamin sequences given by Höger *et al.* (1988). Over this segment vertebrate lamins show, depending on lamin type, between 44 and 74% sequence identity (Table I). These values drop to 37–39% when *Drosophila* lamin is compared with the three vertebrate lamins. Values of 30–35% are found when the *Ascaris* IF-B protein is compared with each of the four lamins. A direct consequence of the alignment (Figure 4a) is the necessity to introduce a deletion into the IF protein sequence. It coincides with the position of the nuclear location signal in

Table I. Sequence identity (%) along the high homology region in the tail domains of various nuclear lamins and the *Ascaris* IF protein B

	LDmo	LmB	LXL1	LhA	IFAsB
LDmo	–	37	37	39	30
LmB	37	–	74	49	35
LXL1	37	74	–	44	35
LhA	39	49	44	–	43
IFAsB	30	35	35	34	–

The high homology segment of 118 residues is defined in Figure 4a. The table compares this segment in four different nuclear lamins (LDmo, *Drosophila* lamin; LmB, murine lamin B; LhA, human lamin A; LXL1, *Xenopus* lamin L_1; for references see Figure 4) and the cytoplasmic IF protein B of *Ascaris*. Similar values are seen in a comparison with the *Ascaris* A protein (Figure 2) and the lamin sequences not used in this table (for references see Introduction).

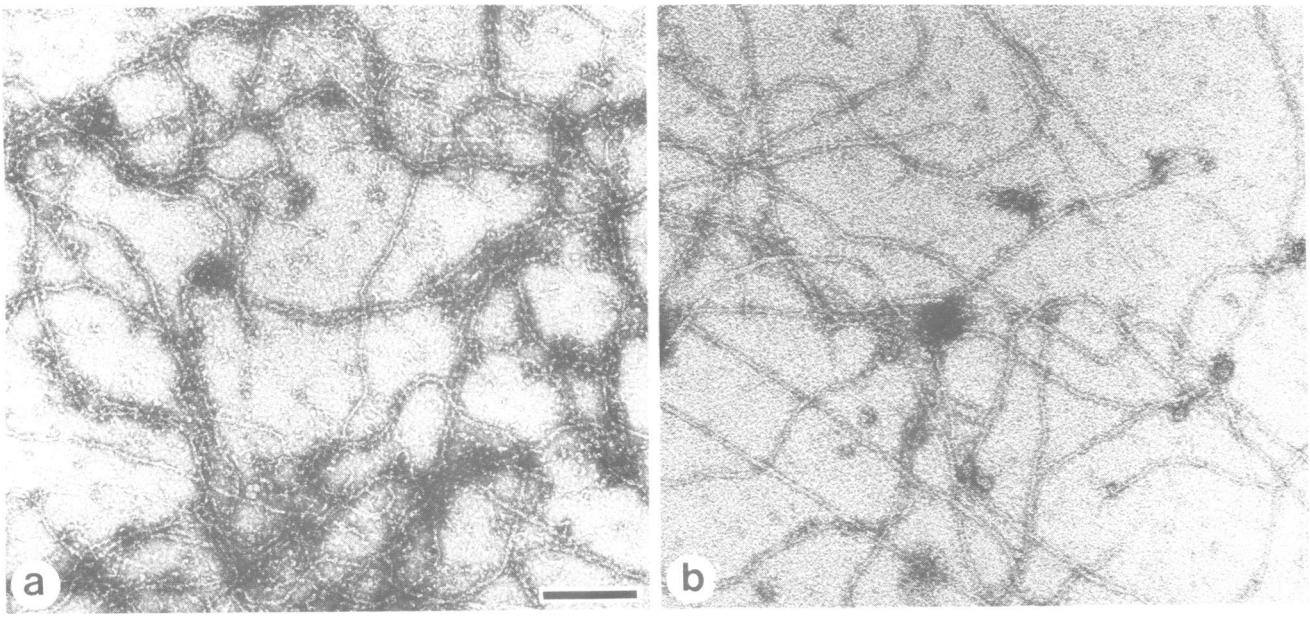

Fig. 5. *In vitro* filament formation by each of the two IF proteins of *Ascaris* muscle. When urea is removed by dialysis, protein A (**panel a**) and protein B (**panel b**) form morphologically normal IF. Micrographs are at the same magnification. The bar indicates 0.1 μm.

the lamin sequences (Loewinger and McKeon, 1988; see Discussion).

The sequences of the tail domains of *Ascaris* IF-B and *Helix* IF-A are aligned in Figure 4b. When we reported the sequence of the *Helix* protein (Weber *et al.*, 1988), we stated that it probably lacked ~20 residues at the C-terminal end. The sequence given in Figure 4b is now complete (see Materials and methods). Provided a few larger gaps are introduced, a maximum homology of 30% can be reached for the tail domains of the two invertebrate IF proteins. A similar value is also obtained along the 118 residue segment defined above. Interestingly, however, the gaps necessary in such an alignment are noticeably longer than in the alignment of *Ascaris* IF-B and the various lamins (compare Figure 4a and b). Due to these larger gaps, and the lack of the 23 C-terminal residues, we could not make a stronger case for the lamin–invertebrate IF protein relationship in our earlier study on the *Helix* protein (Weber *et al.*, 1988).

The head domain of *Ascaris* IF-B shows no striking sequence relationship when compared with the proteins present in the databank (last entry: April 1989). There is also no obvious relationship to the rather short head domains of the nuclear lamins (for sequence summaries see Höger *et al.*, 1988). There is, however, a noticeable homology between the head domains of *Ascaris* IF-B and *H.pomatia* (Weber *et al.*, 1988). A 25% sequence identity can be reached when a few gaps are introduced in the alignment (data not shown). While both proteins reveal an acidic cluster close to the N-terminal end such a feature is not seen in vertebrate IF proteins.

Ascaris IF proteins A and B are closely related but distinct proteins

Given the apparent mol. wt of 71 kd for IF-A (Bartnik *et al.*, 1986) a sequence of ~630 residues is expected. Our current analysis covers 500 residues or ~80% of the expected length. More importantly the sequence regions established include the entire tail domain and >95% of the rod (Figure 2). Over these two large domains A shares 55% sequence identity with B. Particularly high homology is seen in coil 1a, the carboxyl part of coil 1b, most of coil 2, and the entire tail region. The first part of coil 1b, which also includes the six extra heptads of invertebrate IF proteins, shows noticeably lower homology. Along these 42 residues there are only 11 identical positions (26% sequence identity). An additional but much shorter region of low homology is spacer 2, which separates coils 1b and 2. This region is known from studies of vertebrate IF proteins to tolerate both sequence and length changes (Geisler and Weber, 1982, 1986). Here proteins A and B differ in length by one residue (Figure 2). Over the entire tail domain protein A is closely related to B and shows the striking homology with the nuclear lamins described above (compare Figures 2 and 4a). The two IF proteins also show functional similarity. Each protein forms normal homopolymeric filaments when the urea is removed by dialysis against assembly buffers (Figure 5).

The head domain of A is still poorly characterized due to the fact that its N terminus is blocked. Thus we do not know whether the extra mass of A versus B arises solely from this domain and/or from an as yet unrecognized extension of the tail sequence so far established. Since the lack of this information does not interfere with the general picture emerging from the invertebrate IF sequences (see Discussion), a further characterization of A could be postponed.

Discussion

The complex three dimensional IF array documented by serial sections of the giant body muscle cells in the nematode *A.lumbricoides* (Bartnik and Weber, 1987) is built from two major structural proteins (Bartnik *et al.*, 1986). We have now shown by protein sequencing that A and B are highly homologous proteins, which share 55% sequence identity over their rod and tail domains. In contrast to the non-neuronal IF proteins of the gastropod *H.pomatia* (Weber *et al.*, 1988), the two *Ascaris* proteins must arise from two distinct genes.

The sequences of the two nematode proteins and their comparison with vertebrate IF proteins document a remarkable evolutionary drift. Over most of the coiled-coil domain, sequence principles (segmentation of the helices, consecutive heptads etc.) rather than actual sequences are observed. The noticeable exception concerns the consensus sequences present at the ends of the rod. Given their high conservation we expect that they fulfill a direct function in filament assembly. The TYRKLLEGEE sequence located at the C-terminal end of the rod contributes to the IF-A epitope present in most but not all IF proteins (Pruss *et al.*, 1981; Geisler *et al.*, 1983; Bartnik *et al.*, 1987a; Magin *et al.*, 1987). Its presence, with minor variations, in both *Ascaris* proteins (Figure 2) explains their IFA reactivity (Bartnik *et al.*, 1986).

Our results on sequences of invertebrate IF proteins (Weber *et al.*, 1988; this study) provide the first direct support for the previous speculation (Osborn and Weber, 1986; Bartnik *et al.*, 1987b; Franke, 1987a; Myers *et al.*, 1987; Steinert and Roop, 1988) that genes encoding cytoplasmic IF proteins arose in eukaryotic evolution from genes encoding nuclear lamins. All three invertebrate IF proteins currently characterized by sequence share two distinct structural features with the lamins, which are absent in all vertebrate IF proteins known to date. They have six extra heptads in the coil 1b segment of the rod and also show extensive homology with lamins in the tail domains. Here a large segment of 118 residues shows 30–35% sequence identity with all lamins, while *Drosophila* lamin shares in the same region 37–39% identity with the various vertebrate lamins (Table I).

The sequence alignment of the tail domains requires the introduction of a deletion of ~15–20 residues for the invertebrate IF proteins (Figure 4a). The corresponding stretch in the nuclear lamins displays four consecutive basic residues. They are part of the translocation signal, which is responsible for the nuclear entry of the lamins (Loewinger and McKeon, 1988). That the length of the karyophilic signal sequence exceeds 10 residues is shown by several recent reports (e.g. see Kleinschmidt and Seitner, 1987; Rihs and Peters, 1989 and references cited therein). Interestingly, the 25 residues from the end of the rod to the onset of the translocation signal are a low homology region when compared with IF proteins. In this region we note as a common feature of all lamins potential repetitive phosphorylation sites of the type RXXXS. Past the translocation signal sequence there is excellent alignment of the sequences. Thus using the consensus presentation of Höger *et al.* (1988) we could align the two *Ascaris* IF proteins without the need for additional minor gaps (Figures 2 and 4a). Interestingly, the relative position of the C-terminal end of the IF proteins coincides with the point where *Drosophila* lamin and the small lamins of the vertebrates diverge in sequence, since the latter proteins show a stretch of consecutive acidic residues. Thus the IF proteins also lack the lamin consensus motif CXXM situated at the C-terminal end (Höger *et al.*, 1988; Stick, 1988; Vorburger *et al.*, 1989). Given reports on isoprenylation (Beck *et al.*, 1988; Wolda and Glomset, 1988; see also Treston and Mulshine, 1989) and *O*-methylation (Chelsky *et al.*, 1987) of lamins, we expect in analogy with yeast a-factor (Anderegg *et al.*, 1988; Clarke *et al.*, 1988) also a proteolytic shortening up to the multiply modified cysteine of the consensus motif. The deletion of two lamin signal sequences in the invertebrate cytoplasmic IF proteins during the presumptive evolutionary process of divergence is in line with current concepts that IF proteins and lamins are compartmentalized and do not mix.

While normal computer search programs reveal no extended homology between the hypervariable tail domains of vertebrate IF proteins and the lamins, restricting the search to short motif sequences can change this view. Franke (1987b) found a moderately homologous 12 residue motif in vimentin, desmin, GFAP, keratins 8, 18 and lamins (see Höger *et al.*, 1988). While we find this motif in the published sequence of peripherin (Leonard *et al.*, 1988; Landon *et al.*, 1989), it is absent from the three neurofilament proteins. The motif occurs in the tail domains of the invertebrate IF proteins at the same relative position as in lamins (Figure 4a and b). Interestingly in these IF proteins and the lamins we also detect the motif some 75 residues later in a more degenerate form (Figure 4a and b). While Franke (1987b) drew attention to the similarity between the motif and the loop of Ca^{2+} binding proteins, we note that in nearly all cases discussed above the −Z position is not provided by an acidic residue as in the functional EF-hand proteins (Kretsinger, 1987). Thus the function of this motif, if any, remains unknown.

One of the intriguing aspects raised by the sequences of the three invertebrate IF proteins is the question whether their more lamin-like domains carry any special importance, which would separate them in functional terms from the various vertebrate IF. If not, these particular sequences could simply reflect some aspects of metazoan evolution. In this respect the six extra heptads of the coil 1b domain pose an interesting problem. They are most probably not directly required for IF formation (note their absence in all vertebrate IF proteins) and therefore their retention in all nuclear lamins could indicate a potential functional contribution in the nuclear lamina. Interestingly these special heptads mark with only 26% sequence identity a low homology region in the comparison of the two highly related *Ascaris* IF proteins (55% homology). Whether this already signals an evolutionary drift, which finally leads to the deletion of this region remains to be seen. In line with this possibility is also the observation that the gastropods, which are evolutionarily more advanced, show a lower homology versus the lamins in their IF tail domain than do the proteins of the more primitive nematode. Some of the questions raised will probably be better understood once the organization of the genes for nuclear lamins and invertebrate IF proteins is established. Here we note that the additional six heptads are inserted at a conserved intron position common to all non-neuronal IF genes in vertebrates (Parry *et al.*, 1986; Weber, 1986). It seems that the nuclear lamins are much more conserved in eukaryotic evolution than the cytoplasmic IF proteins. This conservation of the lamins most probably reflects their obligatory function, the building of a nuclear lamina. In contrast the surprising plasticity of cytoplasmic IF proteins cannot yet be related with possible functional differences of distinct IF types.

Although *Ascaris* muscle contains the two IF proteins in about equal amounts (Figure 1) they do not obey a keratin-like structure. Both A and B can form *in vitro* homopolymeric IF (Figure 5) whereas vertebrate keratin filaments are based *in vivo* and *in vitro* on obligatory heteropolymers (reviewed by Quinlan *et al.*, 1985; Steinert and Roop, 1988). Interestingly the non-neuronal IF type of gastropod, isolated from the oesophagus epithelium, also contains two proteins

able to form homopolymeric IF *in vitro* (Weber *et al.*, 1988). Therefore the few currently analysed IF proteins of invertebrates seem to resemble in structural properties the type III proteins of vertebrates (vimentin, desmin and GFAP) rather than the keratins which require keratin-pairing as additional structural information. This already seems reflected in the differential degrees of sequence homologies (see Results).

Materials and methods

Purification of Ascaris IF proteins A and B

Ascaris lumbricoides suum was collected at a pig slaughterhouse. Body musculature was dissected and then homogenized in a low salt buffer (40 mM KCl, 2 mM $MgCl_2$, 5 mM EGTA, 1 mM 2-mercaptoethanol, 1% Triton-X-100 in 20 mM PIPES pH 6.8). The insoluble residue was washed once in a high salt buffer (0.6 M KCl, 4 mM $MgCl_2$, 3 mM ATP, 3 mM EGTA, 1 mM 2-mercaptoethanol in 40 mM imidazole–HCl pH 7.1). 30 ml of buffer were used for 3 g muscle tissue. The final residue was treated with 30 ml of urea buffer (8 M urea, 1 mM DTT in 20 mM Tris–HCl pH 8.1). The soluble proteins were dialysed against urea buffer and then applied on a column of DE52 equilibrated in the same solvent. After application of a linear salt gradient (0–200 mM NaCl) the first protein-containing-fractions yielded nearly homogeneous protein B. Fractions containing predominantly protein A, some protein B and other myofibrillar polypeptides were pooled. Proteins were concentrated and subjected to gel filtration on Sepharose Cl 6B in urea buffer containing 200 mM NaCl to remove non IF proteins. A final chromatography on Mono-Q in urea buffer was used to separate A from B. Protein B obtained from the DE52 column was also subjected to Mono-Q chromatography. In earlier pilot experiments A and B were also purified by preparative gel electrophoresis in the presence of SDS. *In vitro* filament formation was induced by dialysing the individual proteins at room temperature from 8 M urea into assembly buffers. Filament formation was monitored by negative staining with uranylacetate.

Sequence analysis

After reaction of the cysteine residues with 2-vinylpyridine (Friedman *et al.*, 1970) protein B was subjected to chemical cleavage with CNBr, BNPS–skatole or hydroxylamine. Fragments obtained by HPLC were subjected to extensive gas phase sequencing on an Applied Biosystems sequenator (model A470) with an on-line PTH amino acid analyser. Secondary fragments were obtained by enzymatic cleavage with the endoproteinases Asp-N, Glu-C or Lys-C (Boehringer Mannheim). They were purified by HPLC and sequenced. Only for one region of the tail domain an additional digest with chymotrypsin was necessary to establish an overlap. Unmodified B (0.5 mg/ml in 10 mM Tris–HCl pH 7.5) was treated with 25 μl of thrombin solution (Sigma, T3010) for 30 min at 20°C. This mild treatment yielded a blocked N-terminal peptide (residues 1–22 in Figure 2) and the remainder of the molecule (residues 23–589). Treatment of the N-terminal peptide with Lys-C provided the sequence starting at position 4 and predicted a preceding lysine residue (position 3 in Figure 2). The proposal of two additional residues (serine, lysine) is based on the amino acid composition of the peptides from the blocked N-terminal region.

Chemical fragments of protein A, previously modified with 2-vinylpyridine (see above), were obtained by cleavage with CNBr or BNPS–skatole. In addition unmodified protein was treated with 2-nitro-5-thiocyanobenzoate (NTCB) for cleavage at cysteine (Geisler *et al.*, 1982b). Secondary fragments were obtained using endoproteinase Asp-N and Lys-C (see above).

IF-A from the oesophagus epithelium of *H.pomatia* was purified as before. Its C-terminal CNBr fragment (Weber *et al.*, 1988) was treated with endoproteinases Asp-N and Glu-C. Sequencing of the resulting peptides extended the previously reported sequence (Weber *et al.*, 1988) to the C-terminal end. This addition of 23 residues is shown in Figure 4b (see Results).

Acknowledgements

We thank Drs N.Geisler and M.Osborn for discussions and Dr K.Kossmagk-Stephan for observations made in an early phase of this study. W.U. was supported by a post-doctoral fellowship from the Max Planck Society.

References

Aebi,U., Cohn,J., Buhle,L. and Gerace,L. (1986) *Nature*, **323**, 560–564.
Anderegg,R.J., Betz,R., Carr,S.A., Crabb,J.W. and Duntze,W. (1988) *J. Biol. Chem.*, **263**, 18236–18240.
Bartnik,E. and Weber,K. (1987) *Eur. J. Cell Biol.*, **45**, 291–301.
Bartnik,E., Osborn,M. and Weber,K. (1985) *J. Cell. Biol.*, **101**, 427–440.
Bartnik,E., Osborn,M. and Weber,K. (1986) *J. Cell Biol.*, **102**, 2033–2041.
Bartnik,E., Kossmagk-Stephan,K., Osborn,M. and Weber,K. (1987a) *Eur. J. Cell Biol.*, **43**, 329–338.
Bartnik,E., Kossmagk-Stephan,K. and Weber,K. (1987b) *Eur. J. Cell Biol.*, **44**, 219–228.
Beck,L.A., Hosick,T.J. and Sinensky,M. (1988) *J. Cell. Biol.*, **107**, 1307–1316.
Chelsky,D., Olson,J.F. and Koshland,D.E.Jr (1987) *J. Biol. Chem.*, **262**, 4303–4309.
Clarke,S., Vogel,J.P., Deschenes,R.J. and Stock,J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4643–4647.
Ferrari,S., Battini,R., Kaczmarck,L., Rittling,S., Calabretta,B., de Riel,J.K., Philiponis,V., Wei,J.-F. and Baserga,R. (1986) *Mol. Cell Biol.*, **6**, 3614–3620.
Fisher,D., Chaudhary,N. and Blobel,G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6450–6454.
Franke,W.W. (1987a) *Cell*, **48**, 3–4.
Franke,W.W. (1987b) *Cell Biol. Int. Rep.*, **11**, 831.
Friedman,M., Krull,L.H. and Cavins,J. (1970) *J. Biol. Chem.*, **245**, 3868–3871.
Geisler,N. and Weber,K. (1982) *EMBO J.*, **1**, 1649–1656.
Geisler,N. and Weber,K. (1986) in Shay,J.W. (ed.), *Cellular and Molecular Biology of the Cytoskeleton*. Plenum Press, pp. 41–68.
Geisler,N., Kaufmann,E., Fischer,S., Plessmann,U. and Weber,K. (1983) *EMBO J.*, **2**, 1295–1302.
Geisler,N., Kaufmann,E. and Weber,K. (1982a) *Cell*, **30**, 277–286.
Geisler,N., Plessmann,U. and Weber,K. (1982b) *Nature*, **296**, 448–450.
Gruenbaum,Y., Landesman,Y., Drees,B., Bare,J.W., Saumweber,H., Paddy,M.R., Sedat,J.W., Smith,D.E., Benton,B.M. and Fisher,P.A. (1988) *J. Cell Biol.*, **106**, 585–596.
Hanugoklu,I. and Fuchs,E. (1983) *Cell*, **33**, 915–924.
Höger,T.T., Krohne,G. and Franke,W.W. (1988) *Eur. J. Cell Biol.*, **47**, 283–290.
Kleinschmidt,J.A. and Seitner,A. (1987) *EMBO J.*, **8**, 1605–1614.
Kretsinger,R.H. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 499–510.
Krohne,G., Wolin,S.L., McKeon,F.D., Franke,W.W. and Kirschner,M.W. (1987) *EMBO J.*, **6**, 3801–3808.
Landon,F., Lemonnier,M., Benarous,R., Huc,C., Fiszman,M., Gros,F. and Portier,M.M. (1989) *EMBO J.*, **8**, 1719–1722.
Leonard,D.G.B., Gorham,J.D., Cole,P., Greene,L.A. and Ziff,E.B. (1988) *J. Cell Biol.*, **106**, 181–193.
Lewis,S.A. and Cowan,N.J. (1986) *Mol. Cell. Biol.*, **6**, 1529–1934.
Loewinger,L. and McKeon,F. (1988) *EMBO J.*, **7**, 2301–2309.
Magin,T.M., Hatzfeld,M. and Franke,W.W. (1987) *EMBO J.*, **6**, 2607–2615.
McKeon,F.D., Kirschner,M.W. and Caput,D. (1986) *Nature*, **319**, 463–468.
Myers,M.W., Lazzarini,R.A., Lee,V.M.Y., Schlaepfer,W.W. and Nelson,D.L. (1987) *EMBO J.*, **6**, 1617–1626.
Osborn,M. and Weber,K. (1986) *Trends Biochem. Sci.*, **11**, 469–472.
Parry,D.A.D., Conway,J.F. and Steinert,P.M. (1986) *Biochem. J.*, **238**, 305–308.
Pruss,R.M., Mirsky,R., Raff,M.C., Thorpe,R., Dowding,A.J. and Anderton,B.H. (1981) *Cell*, **27**, 419–428.
Quinlan,R.A., Schiller,D.L., Hatzfeld,M., Achtstaetter,T., Moll,R., Jocarno,J.L., Magin,T.M. and Franke,W.W. (1985) *Ann. NY Acad. Sci.*, **455**, 282–306.
Rihs,H.P. and Peters,R. (1989) *EMBO J.*, **8**, 1479–1484.
Steinert,P.M. and Roop,D.R. (1988) *Annu. Rev. Biochem.*, **57**, 593–625.
Steinert,P.M., Rice,R.H., Roop,D.R., Trus,B.L. and Steven,A.C. (1983) *Nature*, **302**, 794–800.
Stick,R. (1988) *EMBO J.*, **7**, 3189–3197.
Treston,A.M. and Mulshine,J. (1989) *Nature*, **337**, 406.
Vorburger,K., Lehner,C.F., Kitten,G.T., Eppenberger,H.M. and Nigg,E.A. (1989) *J. Mol. Biol.*, **208**, 405–415.
Weber,K. (1986) *Nature*, **320**, 402.
Weber,K., Plessmann,U., Dodemont,H. and Kossmagk-Stephan,K. (1988) *EMBO J.*, **7**, 2995–3001.
Wolda,S.L. and Glomset,J.A. (1988) *J. Biol. Chem.*, **263**, 5997–6000.

Received on June 22, 1989; revised on July 21, 1989