Bovine Filensin Possesses Primary and Secondary Structure Similarity to Intermediate Filament Proteins

Fotini Gounari,* Andreas Merdes,* Roy Quinlan,** John Hess,§ Paul G. FitzGerald,§ Christos A. Ouzounis,* and Spyros D. Georgatos*

* Program of Cell Biology, European Molecular Biology Laboratory, 6900 Heidelberg, Federal Republic of Germany; ‡Department of Biochemistry, University of Dundee, DD1 4HN, Dundee, Scotland, United Kingdom; and §Department of Human Anatomy, School of Medicine, University of California Davis, California 95616, USA

Abstract. The cDNA coding for calf filensin, a membrane-associated protein of the lens fiber cells, has been cloned and sequenced. The predicted 755-amino acid-long open reading frame shows primary and secondary structure similarity to intermediate filament (IF) proteins. Filensin can be divided into an NH2-terminal domain (head) of 38 amino acids, a middle domain (rod) of 279 amino acids, and a COOH-terminal domain (tail) of 438 amino acids. The head domain contains a di-arginine/aromatic amino acid motif which is also found in the head domains of various intermediate filament proteins and includes a potential protein kinase A phosphorylation site. By multiple alignment to all known IF protein sequences, the filensin rod, which is the shortest among IF proteins, can be subdivided into three subdomains (coils 1a, 1b, and 2). A 29 amino acid truncation in the coil 2 region accounts for the smaller size of this domain. The filensin tail contains 6 1/2 tandem repeats which match analogous motifs of mammalian neurofilament M and H proteins. We suggest that filensin is a novel IF protein which does not conform to any of the previously described classes. Purified filensin fails to form regular filaments in vitro (Merdes, A., M. Brunkener, H. Horstmann, and S. D. Georgatos. 1991. J. Cell Biol. 115:397-410), probably due to the missing segment in the coil 2 region. Participation of filensin in a filamentous network in vivo may be facilitated by an assembly partner.

The mammalian eye lens originates from the ectoderm-derived lens placode, a layer of cells adjacent to the optic cup. In early development, the lens placode invaginates and forms the lens vesicle. Cells from the posterior surface of the lens vesicle (primary lens fibers) grow and gradually fill its cavity. In the mature lens, the anterior surface is formed by a monolayer of cuboidal epithelial cells. These cells divide around an antero-lateral (germinative) zone of the organ, elongate, and gradually differentiate into anucleate hexahedrally-shaped (secondary) lens fiber cells (LFCs).1 LFCs possess an extensive system of specialized intercellular junctions and an organized system of cytoskeletal elements which includes actin microfilaments, vimentin intermediate filaments (IFs), and some unique fibrillar structures, the “beaded filaments” (BFs) (Maisel and Perry, 1972, Bloemendal, 1981).

We have previously characterized a lens-specific, membrane and cytoskeleton-associated protein, which we termed filensin (Merdes et al., 1991). Exchange of antibodies indicated that filensin is antigenically related to the previously identified 115 kD and CP 95 proteins. These polypeptides are thought to be the structural components of the lens BFs in mammals and birds, respectively (FitzGerald, 1988; FitzGerald and Gottlieb, 1989; Ireland and Maisel, 1984b). When purified and reconstituted in isotonic salt, filensin assembled into short, 10-nm fibrils (Merdes et al., 1991; Quinlan et al., 1992) which resembled the structures formed by isolated neurofilament M and H subunits (Troncoso et al., 1989). Furthermore, isolated filensin exhibited specific binding to the IF protein vimentin and to a 47-kD lens protein (Merdes et al., 1991) and could reassociate with filensin-depleted lens membranes (Brunkener and Georgatos, 1992). These in vitro properties indicated that filensin may represent a membrane-associated IF protein (Merdes et al., 1991). Indeed, sequence information and biochemical studies reported here demonstrate that filensin shares essential structural features with intermediate filament proteins.

Materials and Methods

Reagents

Polyclonal antibodies (No. 111) were prepared by immunizing rabbits with column-purified calf filensin (resolved in SDS gels), or with a keyhole limi-
pet hemocyanine (KLH)-coupled synthetic peptide (FL 2). This peptide was modeled after a COOH-terminal sequence of calf filensin (AYEKVEVME-SIEKFSTESTI). A calf lens λgt11 cDNA library was provided by H. Bloemendal (University of Nijmegen, Nijmegen, The Netherlands).

**Screening of the λgt11 Library**

A total of 10⁶ plaques were plated. Induction of fusion proteins and antibody screening were performed exactly as described in Sambrook et al. (1989). Anti-filensin antibodies were preabsorbed against *Escherichia coli* lysates and used to screen the library at a dilution of 1/1000. Screening with DNA probes was carried out as described in Church and Gilbert (1984) using a double-stranded, 32P-labeled, 170 bp fragment excised from the lysates and used to screen the library at a dilution of 1/500. Screening with PCR on X eluates was used to select clones extending upstream to the FL-A1 clone. The 5′ primer was the 5′-CGTCAGTATCGGCGGAAITC-3′ sequence of the non-coding strand containing a SacI restriction site. The AmpliTaq polymerase (Perkin Elmer Cetus, Norwalk, CT) was used.

**PCR Screening of λgt Clones**

PCR on X eluates was used to select clones extending upstream to the FL-A1 clone. The 5′ primer was the 5′-CGTCAGTATCGGCGGAAITC-3′ sequence of the left X arm which overlapped with the EcoRI cloning site. The 3′ primer was the 5′-TCAATGAGGGCAGAGCTCA-3′ sequence of the non-coding strand containing a SacI restriction site. The AmpliTaq polymerase (Perkin Elmer Cetus, Norwalk, CT) was used.

**In Vitro Transcription, Translation, and Immunoprecipitation**

FL-A1 and FL-B were joined into Bluescript KS+ via the SacI site located 5′ of the non-coding strand containing a SacI restriction site. The AmpliTaq polymerase (Perkin Elmer Cetus, Norwalk, CT) was used.

**Sequence Analysis and Multiple Alignment**

Database searches were performed with FASTA version 1.6c (Pearson and Lipman, 1988) and BLASTP version 1.2.3 (Altschul et al., 1990). The SWISSPROT sequence database release 20 was used (Bairoch and Boeckmann, 1991). A multiple alignment of 77 complete rod domains of IF proteins with filensin was constructed using the PILEUP of the GCG sequence analysis software package version 7.1 (Deveraux et al., 1984). The default parameters for gap open (3.0) and gap elongation (0.1) were used. The amino acid comparison matrix was the normalized Dayhoff matrix (Gribskov and Burgess, 1986).

**Protein Chemical Procedures**

Calf filensin was purified as specified (Merdes et al., 1991; Quinlan et al., 1992). The NH₂-terminal 54-kD degradation product of filensin was isolated by ion-exchange and hydroxylapatite chromatography and its identity was established by microsequencing (the sequences obtained were RLD-ELAGPEDA and RAVQDITAACP). Chemical cross-linking of the entire filensin molecule was performed with 0.4 mg/ml dimethylsuberimidate (DMS) according to Geisler and Weber (1982). The 54-kD degradation product was cross-linked with 0.5 mg/ml ethylene glycol bis(succinimidylsuccinate) (EGS), exactly as described by Geisler et al. (1992). As molecular weight markers, we used a preparation of *Lethocerus* myofibrils (Lakey et al., 1990).

**Results**

**Isolation of cDNA Clones**

To identify the cDNA encoding calf lens filensin, we generated a specific antiserum against this protein which was used to screen a λgt11 calf lens library. Initial probing of 10⁶ plaques yielded 187 positive clones. 30 of these clones were further analyzed and found to contain inserts of three distinct sizes: 1.5 kb (FL-A1), 1.0 kb (FL-A2) and 0.7 kb (FL-A3). All of these clones were overlapping and possessed the same 3′-end. Sequencing of the longest one (FL-A1) revealed an open reading frame coding for 466 amino acid residues. The remaining 157 isolates were screened for clones extending upstream to the FL-A1 with a PCR approach using an antisense primer from the 5′ of FL-A1 and a sense primer overlapping with the EcoRI (cDNA cloning) site of the left X arm. The longest amplified product (FL-P) extended an additional 0.6 kb upstream to the already known sequence. For a second screening of the library, we used a 170-bp probe, containing the 5′ region of the FL-P. This yielded a 1.4-kb clone (FL-B), extending another 0.3 kb towards the 5′-end. In FL-B, we identified an ATG codon located at position 19-21 of the cDNA sequence. Immediately upstream to this codon we noticed the sequence GGAGCC, which contains the essential features of the consensus translation initiation motif GCCA+3 (Kozak, 1990, 1991). The sequence assembled from FL-A1 and FL-B comprises 2,487 nucleotides, of which 2,265 represent the coding region (Fig. 2 B).

The obtained cDNA sequence predicts a polyepitope with an M, of 83 kD, i.e., ~20 kD less than the apparent M, of calf filensin in SDS gels. To resolve this discrepancy, and to examine whether the deduced sequence comprised the entire coding region of the mRNA, FL-B and FL-A1 were joined together in a Bluescript vector (for details see Materials and Methods) and transcribed using the T3 RNA polymerase. The resulting RNA was translated in a reticulocyte lysate. As shown in Fig. 1, lanes 3, the in vitro synthesized protein has an M, of 110 kD and can be immunoprecipitated with two different antibodies recognizing filensin. Furthermore, the material obtained by immunoprecipitation of the in vitro transcribed/translated cDNA exactly comigrates (as judged by SDS PAGE in 7.5% gels) with authentic filensin immunoprecipitated after in vitro translation of total lens RNA (Fig. 1, compare lanes 2 and 3). From these data, we conclude that the cloned cDNA includes the complete amino acid sequence of calf filensin. Apparently, filensin migrates anomalously in SDS gels.

**Amino Acid Sequence and Domain Structure of Filensin**

The predicted sequence from the FL-B and FL-A1 clones has been confirmed by microsequencing of 12 tryptic and V8 protease peptides derived from purified filensin (Fig. 2 B). It is noteworthy that the peptide sequences were evenly spaced along the parent molecule and included regions near the NH₂ and COOH termini of filensin.

Computer-assisted sequence analysis reveals three distinct regions in the filensin molecule: an NH₂-terminal domain, a middle domain, and a COOH-terminal domain (Fig. 2 A). The NH₂-terminal domain (head) of 38 amino acids includes a conserved di-arginine/aromatic residue motif (YRRSY) which, with some permutations, is also found in the head domains of vimentin, desmin, peripherin, α-internexin and other IF proteins (Herrmann et al., 1992). This motif also contains a consensus site for protein kinase A phosphorylation (RRS, see Kemp and Pearson, 1990).
The middle domain of filensin (rod) is 279 amino acids long and exhibits sequence similarity to the rod domain of all intermediate filament proteins. It contains a heptad repeat pattern and is predicted to be largely \( \alpha \)-helical, as substantiated by circular dichroism (A. Merdes, E. Gounari and S. D. Georgatos, unpublished observations). Type-III IF proteins, such as human vimentin and rat peripherin, show 22\% identity to the filensin rod over an alignment length of 287 amino acids (multiple alignment, Fig. 3 A). Multiple alignment and secondary structure predictions indicate that the rod domain of filensin consists of three subdomains (coils

Figure 1. Immunoprecipitation of in vitro synthesized filensin. No Ab shows the preparations used for immunoprecipitation. aFL 2 Ab and III Ab show profiles of the material immunoprecipitated with the anti-FL 2 and the 111 polyclonal antibodies. Lane I shows the products of a mock-translation in the absence of RNA; lanes 2 and 3 represent material obtained after in vitro translation of either lens RNA, or in vitro synthesized filensin RNA, respectively (arrow depicts the position of filensin). All samples were analyzed on a 7.5\% SDS gel. \(^{14}\)C-labeled, molecular weight markers were run in lane \( M \) and the corresponding sizes are indicated on the side of the picture.

Figure 2. (A) Predicted domain structure of bovine filensin. The drawing shows the three domains of filensin (head, rod, and tail). (B) Nucleotide and predicted amino acid sequence of bovine filensin. The first methionine, as well as the peptide sequences obtained by analysis of purified bovine filensin, are underlined. In the rod domain, the \( a \) and \( d \) positions of the heptad repeats are indicated by asterisks, while the three helical domains are enclosed in shaded boxes. The region containing the tandem repeats in the COOH-terminal domain of filensin is boxed. The start of each repeat is marked by a dot. The sequence of a potential polyadenylation signal located in the 3' untranslated region of the cDNA is in oval box. "Stutter" denotes the heptad repeat stutter in the region of coil 2. These sequence data are available from EMBL under accession number X72388.
Assembly Properties of Filensin and Expression of its mRNA in Different Bovine Organs

We have previously reported that purified filensin assembles into short fibrils under isotonic salt conditions (Merdes et al., 1991). Moreover, the cDNA sequence of filensin predicts a heptad repeat pattern along the rod domain indicating that it should dimerize and oligomerize. In support to this thesis, chemical cross-linking of isolated filensin yields a ladder of high molecular weight products (Fig. 4 A). A prominent band at 214–245 kD corresponds to a dimer. A second species at ~560 kD may represent a tetramer. (The identity of the cross-linked products has been established by immunoblotting analysis [not shown].) The specificity of cross-linking is demonstrated by the absence of species corresponding to the expected size of a trimer and by the decreased yield of dimers and tetramers in the presence of urea (Fig. 4 A). Cross-linking at low and isotonic salt shows that oligomerization of filensin is promoted by increasing ionic strength (Fig. 4 B).

To further narrow down the parts involved in the dimerization and tetramerization of filensin, we performed cross-linking experiments using a 54-kD endogenous proteolytic product of this protein which comprises the entire head region and the rod domain (see below and Materials and Methods). Results in Fig. 4 C demonstrate that this peptide responds to the expected size of a trimer and by the decreased yield of dimers and tetramers in the presence of urea (Fig. 4 A). Cross-linking at low and isotonic salt shows that oligomerization of filensin is promoted by increasing ionic strength (Fig. 4 B).

The long COOH-terminal domain of filensin (tail) includes a sequence motif of 14 amino acids, tandemly repeated 6 1/2 times and predicted to form multiple β-turns (Fig. 3 B). In this region, the sequence -VKEE-XXX-PE aligns perfectly with the corresponding repeats of the COOH-terminal domains of the neurofilament M and H proteins (Myers et al., 1987; Lees et al., 1988).
Cross-linking of bovine filensin. Filensin was cross-linked as specified in Materials and Methods. The samples were analyzed in 5% polyacrylamide gels. (A) Filensin cross-linking in the presence of increasing concentrations of urea (silver stain). Bars on the left indicate molecular weight values of 800, 600, 500, 400, 205, and 117 kD from top to bottom. (B) Filensin cross-linking at 0 mM (low salt) and 150 mM (isot. salt) KCl. Arrows indicate the positions of filensin monomers. Open arrowheads (from bottom to top) indicate the positions of dimers and tetramers. Asterisks mark a trace-contaminant in the filensin preparation visible only by silver staining. (C) Cross-linking of an NH2-terminal 54-kD degradation product of filensin (arrow), comprising the filensin head and rod regions, reveals the formation of dimers (120 kD, open arrowhead), tetramers (200–220 kD) and higher oligomers.

Discussion

Molecular Features of Mammalian Filensin

There are three interesting features in the sequence of calf filensin: (a) the di-arginine/aromatic residue motif in its head domain, (b) the IF-like character and the peculiar truncation in its rod domain, and (c) the tandem repeats in its tail domain.

The di-arginine motif of calf filensin (YRRSY, position 2–6) resembles a site in the corresponding region of chicken filensin (YRRSSF, position 2–6, S. Remington, personal communication) and is similar to sequences (YRRXF) found close to the NH2 termini of vimentin, desmin, peripherin and α-internexin. In vimentin, desmin and calf filensin, the aromatic residues flanking the di-arginine motif are followed by other aromatic residues at positions 29, 27, and 15, respectively (for an alignment see Leonard et al., 1988). This arrangement favors the stacking of the phenyl rings of the aromatics (Burley and Petsko, 1985; Cohen and Parry, 1990; Conway and Parry, 1988). Therefore, the intervening (di-arginine/serine) segment may be forced to take a looping configuration. Such a charged and extended structure is expected to be surface-exposed and could play an important role in filament assembly. Indeed, mutational analysis of Xenopus laevis vimentin indicates that the di-arginine motif in this protein is essential for normal IF formation (Herrmann et al., 1992; Hofmann and Herrmann, 1992). The YRRSY motif also includes a typical kinase A phosphorylation site (RRS, Kemp and Pearson, 1990). In vitro phosphorylation experiments confirm that purified filensin can be modified by protein kinase A (M. Brunkener and S. D. Georgatos, unpublished observations). Furthermore, previous studies have shown that the avian beaded filament component, analogue of filensin is a phosphoprotein (Ireland and Maisel, 1984a).

Although shorter (279 amino acids compared to the 310 amino acid long rod of the cytoplasmic IF proteins), the rod...
domain of bovine filensin aligns well with and shows sequence homology to 77 sequenced IF proteins. In the conserved region of coil 1a, the high degree of identity indicates an evolutionary relationship between filensin and IF subunits. However, the sequence identity in other regions conserved among different IF proteins (e.g., at the COOH-end of coil 2b) is lower. cDNA sequencing of chicken filensin (S. Remington, personal communication) and analysis of a partial clone coding for rat filensin (Masaki and Watanabe, 1992) also reveal that the conservation of structural features between the various forms of filensin is more pronounced in the region of the rod domain. Considering that sequence principles rather than actual sequences are usually conserved among different IF proteins (the sequence identity between subunits of different classes can be as low as 15%), we propose that filensin represents a remote member of the IF protein family.

The filensin rod represents so far the shortest rod domain of IF proteins. The structural implications of this peculiarity can be significant. As shown in studies with IF paracrystals (Stewart et al., 1989) and in recent cross-linking experiments (Geisler et al., 1992), the most likely arrangement of IF chains in the tetrameric protofilament is the one of antiparallel, staggered dimers (for other opinions see also Steinert, 1991). Thus, in the surface lattice of IFs, dimers belonging to adjacent tetramers may overlap at their coil 2 region by 2–3 nm and interact in a longitudinal fashion (Stewart et al., 1989). This interaction may have a role in filament elongation, because short synthetic peptides modeled after the conserved coil 2b motif have been demonstrated to disassemble preformed IFs and to block elongation of IFs in vitro (Hatzfeld and Weber, 1992; Kouklis et al., 1992). For filensin, one may predict that the (shorter) coil 2 of one filensin dimer will not be long enough to reach the coil 2b of an adjacent dimer and that filensin homooligomers would therefore be unable to elongate on their own. This prediction is consistent with the fact that purified filensin polymerizes in vitro into short (80–100 nm) fibrils and not into long filaments (Merdes et al., 1991; Quinlan et al., 1992).

At a first glance, the COOH-terminal domain of calf filensin resembles the tail domains of mammalian neurofilament M and H subunits, being unusually long and containing several tandem repeats (Myers et al., 1987; Lees et al., 1988). However, unlike neurofilament proteins, the filensin tail is not particularly rich in glutamic acid residues and does not contain the characteristic KSP motif (Chin et al., 1983; Geisler et al., 1983; Julien and Mushynski, 1983; Carden et al., 1985). Interestingly, the tandem repeats are only seen in calf filensin but not in rat and chicken filensin (S. Remington, personal communication; Masaki and Watanabe, 1992). In the repeated region, the sequence conserved between calf filensin and NF-H contains the motif VKEE-XXX-PE and the tetrapeptide VKEE is variable, the PE-XXXXX. Because the intervening segment between the dipeptide PE and the tetrapeptide VKEE is variable, the similarity between neurofilament proteins and filensin seems to involve a framework rather than a linear sequence block. This framework offers the potential for multiple β-turns. Tandem repeats with a potential for turns have also been identified in secretory proteins of Chironomus tentans (Dignam et al., 1989; Galli et al., 1990) and in eggshell proteins of Ceratitis capitata (Aaggeli et al., 1991).

As mentioned previously, Masaki and Watanabe (1992) have recently reported the sequence of a partial cDNA clone coding for rat filensin (CP 94 in their nomenclature). This sequence does not include part of the rod and the NH2-terminal region of filensin. Whereas in the rod domain the sequence identity between the calf and rat filensin is high, in the tail region the two proteins seem to diverge. We believe that part of this divergence may be due to two sequencing errors in the rat sequence. Thus, addition of a single base in the rat nucleotide position 1,575 and deletion of a single base in position 1,667 will result in an amino acid sequence of this clone that matches both the calf and the chicken filensin sequences.

**In Situ Organization of Filensin and Potential Functions**

The sequence analysis presented here favors the idea that filensin belongs to the intermediate filament protein family. However, filensin does not form regular IFs on its own (Merdes et al., 1991; Quinlan et al., 1992). It may be that under in vivo conditions filensin coassembles with another component. Recent data (A. Merdes, F. Gounari, and S. D. Georgatos, unpublished observations; and P. G. FitzGerald, unpublished observations) reveal that the 47-kD filensin-binding protein (Merdes et al., 1991) also shares structural features with IF proteins. Structures assembled from filensin/47 kD may resemble more closely the typical IFs.

Filensin has been claimed to be a component of the lens BF. This is a puzzling interpretation, because the BFs are structurally distinct from IFs. However, while the majority of IF proteins are found in and reconstitute into smooth 10-nm filaments, exceptions from this structural format have been noted. Most relevant is the IF of the giant squid axon which has a beaded nature, similar, in some respects, to that of the BFs (Eagles et al., 1990).

Although further analysis is needed, we propose here that filensin is a component of a membrane-associated filament system. The binding of filensin to the plasma membrane is mediated via the COOH-terminal part of the molecule (Brunkener and Georgatos, 1992). This system may link vimentin IFs to the plasma membrane and stabilize the LFC membrane.

This work is dedicated to Stavros and Adamantia Politis.

We acknowledge S. Remington (Department of Genetics and Cell Biology, University of Minnesota at Saint Paul) for making available sequencing information of the chicken filensin before publication. We wish to thank Jeremy Lakey (European Molecular Biology Laboratory) for recording the CD spectrum of filensin. Falk Schweisinger and Marianna Brunken are acknowledged for helpful technical assistance. We thank W. W. Franke (German Cancer Research Center, Heidelberg, FRG) for communicating unpublished information generated in his laboratory by R. Ihl and U. Haus. Finally, we are indebted to J. Hofmann (German Cancer Research Center, Heidelberg, FRG) for commenting on the sequence.

Received for publication 18 November 1992 and in revised form 1 February 1993.

**References**

Aaggeli, A., S. J. Hamodrakas, K. Komitopoulos, and M. Konsolaki. 1991. Tandemly repeating peptide motifs and their secondary structure in Ceratitis capitata eggshell proteins Ces36 and Ces38. *Int. J. Biol. Macromol.* 13:307–315.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
