Gene Structure and cDNA Sequence Identify the Beaded Filament Protein CP49 as a Highly Divergent Type I Intermediate Filament Protein*

(Received for publication, October 25, 1995, and in revised form, January 3, 1996)

John F. Hess, Jodi T. Casselman, and Paul G. FitzGerald‡

From the Department of Cell Biology and Human Anatomy, School of Medicine, University of California, Davis, California 95616

The fiber cell of the vertebrate ocular lens assembles a cytoskeletal structure, the beaded filament, which contains two proteins unique to the fiber cell: CP49 (phakinin) and CP115/CP95 (filensin). We report here the complete primary sequence and gene structure for human CP49. These data show that CP49 is a member of the intermediate filament family, but highly unusual in several regards. 1) CP49 primary sequence does not permit unambiguous assignment to any existing class of intermediate filament protein, but exhibits a gene structure that is identical to the Type I cytokeratins. 2) CP49 essentially lacks one of the three major domains that characterize all intermediate filament proteins, the carboxy-terminal tail domain. 3) CP49 shows substitutions at 3 of 4 residues in the otherwise highly conserved intermediate filament protein motif LNDR. Notably, this divergence includes an Arg to Cys substitution that has only been observed in the mutant human cytokeratin K14, a mutation shown to cause the skin blistering seen in the genetic disorder Dowling-Meara epidermolysis bullosa simplex.

The differentiated fiber cells of the vertebrate ocular lens assemble a cytoskeletal structure referred to as the beaded filament (1). This structure, a 5-nm diameter filament decorated by periodic beads, is morphologically distinct from both actin-containing thin filaments and vimentin-containing intermediate filaments (IFs), which are also present in these cells (2–5). Immunocytochemistry, heavy meromyosin labeling, and cell fractionation/enrichment studies have shown that beaded filaments are biochemically distinct from established cytoskeletal elements as well (6–9). Two proteins, CP49 (phakinin (10)) and CP115/CP95 (filensin (11)), have been localized to the beaded filament. Both proteins have been shown by Northern and Western blotting to be expressed only in the lens and only in the differentiated fiber cells of the lens (7, 9, 12, 13). Also tightly associated with the plasma membrane-cytoskeleton complex, including the beaded filament, is α-crystallin (14–18). α-Crystallin is the most abundant protein in the lens and exists predominantly as a soluble cytoplasmic protein. However, a small fraction of the total α-crystallin resists extraction from the plasma membrane-cytoskeleton complex and has been immunocytochemically localized to both beaded filaments and intermediate filaments (16). While the α-crystallin bound to the cytoskeleton is a small percentage of the total α-crystallin pool, it is quantitatively a major component of the cytoskeletal fraction. The role of the α-crystallin in the cytoskeletal fraction is not yet clear, but its function as a chaperonin has recently been linked to the dynamics of cytoskeleton assembly in the lens (19).

The complete cDNA sequence for bovine filensin (CP115/CP95) has been published, revealing it to be an 85-kDa protein that shows strong sequence similarity to the intermediate filament family of proteins (20). This finding was largely unexpected since antibody probes that were generally considered diagnostic for IF proteins did not react with either CP49 or filensin (21). More significantly, cytoplasmic IF proteins have not been demonstrated in structures other than the classical 8–11-nm intermediate filaments (22, 23). Thus, filensin was the first cytoplasmic IF protein to be localized to a non-IF cytoskeletal structure (9, 24). Analysis of filensin’s primary sequence revealed several other features that were atypical of IF proteins (13, 20, 25), including a primary sequence that was not clearly related to existing IF classes.

Partial sequence of mouse CP49 has been published, establishing that it, too, is an IF-like protein (26). Sequence data for bovine CP49 (11) and chicken CP49 (27, 28) have emerged as well and confirm CP49’s relationship to the IF family. However, like filensin, the primary sequence of CP49 did not show a level of identity that permitted unambiguous assignment of CP49 to any existing class of IF protein, nor was its sequence closely related to its assembly partner, filensin. Thus, the issue of whether CP49 and filensin represented novel IF classes has been left undefined. Furthermore, the reported sequences for chicken CP49 and bovine CP49 showed a dramatic divergence in the amino-terminal head domain. Since homologous IF proteins are usually well conserved, this divergence was highly unusual.

We report here the complete cDNA and predicted amino acid sequences of human CP49 as well as the structure of the human CP49 gene. This represents the first report of CP49 gene structure and the first complete and correct report of a mammalian CP49 primary sequence. These data show that 1) CP49, despite the exceptional degree of sequence divergence, is a Type I cytokeratin; 2) CP49 essentially lacks a carboxy-terminal tail domain; and 3) CP49 shows substitutions at 3 of 4 residues in the otherwise highly conserved IF protein motif LNDR, a motif considered critical to IF assembly and one that has been demonstrated to be a “hot spot” for mutations that
cause human skin blistering disorders. We also describe corrected sequence for bovine CP49 that resolves the unexpected differences seen between the human and bovine CP49 head domains. These data combine with that reported for filensinto establish that a non-intermediate filament cytoskeletal structure has been assembled from two proteins recruited from the intermediate filament family.

MATERIALS AND METHODS

Human lens total RNA was isolated from human donor lenses provided by the Lions Eye and Tissue Bank (Sacramento, CA) following the guanidine isothiocyanate/phenol extraction protocol of Chomczynski and Sacchi (29). Ten micrograms of RNA was reverse-transcribed using Superscript reverse transcriptase (Life Technologies, Inc.) following the manufacturer's instructions, using dT-adapter primers (30), CP49-specific primers, or random hexamer primers. Following cDNA synthesis, the reaction was heated to 65°C for 10 min and diluted to 500 μl with 10 mM Tris, pH 8.0, 1 mM EDTA. For PCR, 5 μl of cDNA was used as input template; cDNA pools were stored at −20°C. Amplification products were isolated from low-melting temperature agarose gels (FMC Corp. BioProducts, Rockland, ME) using Wizard preps (Promega). Purified products were subcloned into pSP72 using the Sureclone kit (Pharmacia Biotech Inc.) or the 5′9′3′ Prime PCR Cloner cloning system (5Prime3Prime, Inc., Boulder, CO). All DNA sequencing was performed using either the Sequenase or Taquence sequencing kit (U.S. Biochemical Corp.).

Following the determination of the chromosomal location of CP49 sequences (31), we purchased the following chromosome 3-specific libraries from American Type Culture Collection: 57751, 57717, and 57748. Plating and screening of the λ phage libraries were performed using standard techniques (32). Library 57751 was screened with human CP49 cDNA sequences using radioactively labeled probes synthesized using the Pharmacia oligolabeling kit. Screening of the library and purification of positive phage were performed as described (32). Inserts from positive phage were characterized by PCR. Initially, PCR was used with exon-specific primers to generate intron/exon boundary fragments for cloning and sequencing. Subsequently, λ phage genomic DNA was isolated and purified, digested with HindII, and cloned into pSP72. In some cases, oligonucleotides specific for the sequences flanking the HindII site of the λ Charon 21A phage were used (primers were generously provided by Beverly Allen, University of Florida). Amplification of the entire phage insert was performed using 35 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 3 min.

Human CP49 genomic clones were sequenced using oligonucleotide primers chemically synthesized by the University of California Davis Protein Structure Laboratory. The entire cDNA sequence presented was determined for both strands. Intron sequences presented have been determined for one strand. Comparisons of the determined sequences with data bases were performed using the University of Wisconsin Genetics Computer Group package, using fastapep.cmo with default settings.

Named primers referred to in text are as follows: primer 3953 (nucleotides 528–549), 5′-GCAGACAGAAACTATCCAGGCC-3′; primer 3815 (nucleotides 727 to 704), 3′-CTTCATAGTTTCTTGATAGAGAGC-5′; primer 4149 (nucleotides 663–686), 5′-GGACCTGGAGAGTCAAATAGAAAG-3′; and primer 4151 (nucleotides 836 to 814), 3′-CACTGAATTCTGATCGTCTCAAG-5′.

RESULTS

Human CP49 Primary Sequence—Published murine (26) and corrected bovine (GenBank accession numbers X75160 and U12016) CP49 cDNA sequences were aligned for the purpose of identifying conserved regions. Some of these regions were selected for the design of oligonucleotides that were used in PCR to amplify human lens cDNA, and amplified products were cloned and sequenced.

Collectively, these PCR-amplified products encompassed the entire open reading frame of human CP49 plus the 3′-untrans-

![Fig. 1. Complete cDNA and deduced amino acid sequences of human CP49. Introns B, C, and E–H are indicated, and the exact site of the intron is indicated by the vertical lines between bases in the nucleotide sequence. The approximate positions of these introns relative to the domain structure of CP49 are indicated in Fig. 5. Conserved IF motifs referred to in the text are shaded. Sites of oligonucleotide primers named under "Materials and Methods" are double-underlined.](http://www.jbc.org/)

Downloaded from http://www.jbc.org/ by guest on July 21, 2018
lated region. Subsequent sequencing of human genomic fragments from a genomic library confirmed the majority of the nucleotide sequences obtained from PCR products. The open reading frame was 1245 bases, encoding 415 amino acids, with a predicted molecular mass of 45,835 Da and a pl of 5.30 (Fig. 1). In previous work, Northern blotting with human CP49 probes established that the sequence was specific to human and of appropriate size (26).

The deduced amino acid sequences of the entire human CP49 protein and the human CP49 rod domain were both compared with the SWISSPROT data base. The best match was the partial mouse CP49 sequence, at 85.8% identity, a level of identity consistent with the strong conservation seen between homologous IF proteins from different species. However, the murine sequence extended only from the middle of coil 1b to the COOH terminus; thus, comparison of the amino-terminal end of the molecule was not possible. Human CP49 and the published bovine CP49 sequences (10) showed little similarity in the first 67 amino acids of the amino-terminal head domain, but aligned well from that point on. Again, because homologous IF proteins tend to be highly conserved, this dramatic divergence was surprising. To address this, we used PCR to confirm the bovine CP49 sequence and established that the published bovine cDNA sequence had omitted 2 nucleotides, resulting in a frameshift mistranslation of the first 67 residues. The corrected bovine sequence (GenBank accession number U12016) shows an 89% level of identity to the human sequence reported here.

The subsequent 35 best matches between human CP49 and the SWISSPROT data base are shown in Fig. 2. All are IF proteins, establishing a clear relationship between CP49 and the IF family.

Secondary Structural Features—Among IF proteins, the amino-terminal head and carboxyl-terminal tail domains are quite variable in both size and sequence. The central rod domains are better conserved, both in primary sequence and, to an even greater degree, predicted secondary structural features. Thus, classification of a novel sequence as a member of the IF family is based on secondary structural features as well as primary sequence identity.

To assess CP49 for the presence of secondary structural features that are characteristic of IF proteins, CP49 was aligned with the Type I cytokeratins K10 and K18. These two proteins were the closest human matches produced by the data base search, and both have been well characterized with respect to secondary structure (33). This alignment, shown in Fig. 3, permitted the identification of the central rod domain of CP49 as well as the subsequent identification of the coil and linker regions that characterize IF protein central rod domains.

The optimal alignment of CP49 and K10 permitted clear identification of CP49's variations of the highly conserved LNDR and TYRKALLEGE motifs near the ends of the central rod domain (Fig. 3, boldface) as well as essentially all of the major secondary structural features that are highly conserved among IF proteins (22, 23). Among the conserved features identified were 1) a central rod domain of appropriate overall size, 311 amino acids (34); and 2) "coil" domains within the central rod that exhibit a heptad repeat pattern of amino acids, in which positions 1 and 4 of the heptad (Fig. 3, asterisks) are dominated by, but not exclusively, apolar residues (35). This heptad repeat pattern is predictive of $\alpha$-helical secondary structure, and thus, these regions are referred to as coils. IF proteins typically exhibit three major coil domains, 1a, 1b, and 2, whose size and position are well conserved. The number, position, and length of the coil regions in CP49 are consistent with those seen among IF proteins (34) and summarized in Fig. 3. Also characteristic of many IF proteins is a "stutter" in the heptad repeats in coil 2, where the repeat pattern is interrupted (34). 3) Between coil regions are short segments that lack the heptad repeat pattern and thus the predicted $\alpha$-helicity. These are referred to as "linkers." The size, location, and number of linker regions in CP49 are consistent with those conserved among IF proteins.

The most notable departure of CP49 from the consensus domain structure of IF proteins is the virtual absence of a COOH-terminal tail domain in CP49. CP49 primary sequence terminates almost immediately after the end of the central rod domain, giving it the most abbreviated tail domain of any IF protein defined to date. Thus, with the exception of a missing/ truncated tail domain, CP49 exhibits a predicted secondary structure that is indistinguishable from the highly conserved domain structure that characterizes IF proteins.

The LNDR sequence at the beginning of coil 1a (Fig. 3, shaded) is among the most highly conserved motifs in IF proteins (22, 34, 36–38). Human CP49 shows substitutions at 3 of 4 residues, from LNDR to LGGC, the most significant divergence yet seen among the IF proteins. We have established the identical LGGC sequence in mouse CP49, and it has been reported in bovine CP49 (10) as well, but not in chicken CP49 (28).

Classification of CP49—Since the primary sequences and overall sizes of the head and tail domains of IF proteins are

| Species | IF Type | Protein | Optimal Score |
|---------|---------|---------|---------------|
| Human   | I       | K18     | 482           |
| Mouse   | I       | 59kd    | 439           |
| Mouse   | I       | K10     | 437           |
| Human   | I       | K10     | 430           |
| Xenopus | I       | endoB   | 426           |
| Human   | I       | K19     | 418           |
| Xenopus | I       | 47kd    | 413           |
| Bovine  | I       | K19     | 413           |
| Mouse   | I       | K19     | 401           |
| Human   | I       | K20     | 401           |
| Bovine  | I       | KVI B   | 398           |
| Xenopus | I       | K3      | 392           |
| Human   | I       | K15     | 391           |
| Human   | I       | K13     | 386           |
| Sheep   | I       | mfb     | 383           |
| Rat     | IV      | K21     | 380           |
| Sheep   | I       | 47.6    | 375           |
| Human   | I       | K17     | 359           |
| Human   | I       | K14     | 358           |
| Mouse   | I       | K13     | 357           |
| Human   | I       | K17     | 351           |
| Human   | I       | K9      | 343           |
| Bovine  | II      | K8      | 315           |
| Bovine  | II      | K9      | 297           |
| Human   | III     | GFAP    | 290           |
| Mouse   | III     | GFAP    | 279           |
| Rat     | IV      | a-intemex | 268         |
| Xenopus | IV      | NF-L    | 259           |
| Mouse   | II      | K8      | 257           |
| Human   | IV      | NF-L    | 250           |
| Xenopus | III     | desmin  | 245           |
| Human   | II      | K7      | 240           |
| Pig     | IV      | NF-L    | 240           |

![Fig. 2. Comparison of the full-length CP49 sequence with the SWISSPROT data base.](http://www.jbc.org/)
more variable, IF proteins have been grouped into classes or types on the basis of the degree of sequence identity among central rod domains (23). The rod domain of CP49, identified by alignment with Type I cytokeratins in Fig. 3, was compared with the SWISSPROT data base to determine whether, on the basis of sequence identity, CP49 could be assigned to any existing class of IF protein. Fig. 4 shows the results of this comparison.

When ranked on the basis of percent identity in the rod domain, the best 21 matches were Type I cytokeratins. However, the level of identity between CP49 and any Type I cytokeratin did not exceed 36.1% and ranged down to 27%. The most similar human Type II, IV, and III IF proteins were 28.2, 27.2, and 27.0%. Thus, the level of sequence identity between CP49 and the Type I cytokeratins is well below that usually seen among members of the Type I class (50–90%) and is more typical of the level of identity seen between IF classes (40%) (23, 35). On the basis of sequence identity, then, CP49 does not fit readily into any of the existing IF classes.

CP49 Gene Structure—Analysis of IF protein genes has revealed that conservation of IF protein gene structure parallels that seen in primary sequence: intron number and position have been well conserved, but type-specific variations in the number and location of introns have also emerged. Thus, gene structure constitutes an independent mechanism for the classification of an IF protein (35, 37, 39, 40).

A schematic showing the intron locations in the human CP49 gene and comparing them with those of other IF genes is presented in Fig. 5. Intron locations in the CP49 gene were defined by comparison of genomic and cDNA sequences.

As seen in Fig. 5, intron A is found only in Type II cytokeratins. Because intron A is absent from other classes of IF proteins, it is considered diagnostic for that group. PCR amplification of genomic DNA and analysis of genomic DNA isolated from a λ phage library provide no evidence for intron A in the human CP49 gene. Thus, the human CP49 gene lacks an intron that is conserved among Type II cytokeratins.

Intron B is conserved in Type I and II IF genes. PCR amplification of this region from CP49 genomic DNA was not successful. However, we have isolated genomic sequences encompassing intron B from a λ phage library and have sequenced the intron/exon boundary from that purified phage DNA. DNA sequencing shows that the human CP49 gene contains an intron between codons for Gln163 and Val164, confirming the presence of intron B.

The exact location of intron C was determined from sequencing human genomic DNA from a CP49-positive λ phage. Oligonucleotide primers 3953 and 3815 produced a 1.5-kilobase pair amplification product that was sequenced from each end, confirming the cDNA sequence and locating each end of the intron. Sequence data demonstrate that within the codon for CP49 Arg191 is a 1.2-kilobase pair intron. This amplification product also shows that no intron is present at CP49 amino acid 210, the predicted location of intron D, which is conserved among Type II and III, but not Type I, IF genes. DNA sequencing of this region was also performed on templates isolated without PCR, yielding identical results.

The location of intron E was determined from sequencing human genomic DNA from a CP49-positive λ phage. Oligonucleotide primers 3953 and 3815 produced a 1.5-kilobase pair amplification product that was sequenced from each end, confirming the cDNA sequence and locating each end of the intron. Sequence data demonstrate that within the codon for CP49 Arg191 is a 1.2-kilobase pair intron. This amplification product also shows that no intron is present at CP49 amino acid 210, the predicted location of intron D, which is conserved among Type II and III, but not Type I, IF genes. DNA sequencing of this region was also performed on templates isolated without PCR, yielding identical results.

The location of intron E was determined from a PCR product obtained during characterization of CP49 genomic DNA. Oligonucleotide primers 4149 and 4151 were used to produce a 2-kilobase pair product that was obtained only in limiting quantities with genomic input DNA. The DNA sequence of the amplified fragment again confirms the cDNA sequence and shows an 1.8-kilobase pair intron between Glu43 and Asp244. Sequence was confirmed in genomic DNA not amplified by PCR.
Introns F and G each contain HindIII sites and were isolated in two parts. Both the 5' and 3' ends of each intron have been sequenced, and the cDNA sequence was confirmed.

Intron H, important in characterizing the potential relationship between CP49 and K19, was contained in a genomic λ phage. DNA sequencing of this region of CP49 shows the presence of intron H within the codon for the last amino acid of CP49.

The phase of the triplet codon that is interrupted by an intron has been a generally well conserved feature among IF proteins and provides an additional means of verifying the authenticity of a conserved intron. Fig. 6 depicts the nucleotide sequences of the intron/exon boundaries of CP49 and identifies the phase of the triplet codon at which introns have been inserted. Introns B, E, F, and G interrupt the DNA sequence between triplet codons, while introns C and H both split the triplet codon after the second nucleotide. This pattern is in keeping with that conserved among IF genes (41, 42).

These data show that the number and precise location of introns in CP49 are identical to those seen in Type I cytokeratins, with the exception of those introns that would have been located in CP49's missing tail domain. Notably, CP49 lacks an intron seen in either Type II or III cytokeratins (intron D).

Tailless CP49 Versus Tailless K19—K19 is a Type I cytokeratin that has an abbreviated tail domain, extending 13 amino acids beyond the end of the central rod domain (43, 44). K19 is also unique among Type I cytokeratins in lacking the intron (designated here as "H") (Fig. 4) that all other Type I IF genes have at the end of the central rod domain (43). This has led to the hypothesis that K19 diverged from the rest of the Type I cytokeratins at a relatively early point (44). CP49's abbreviated tail and general low level of sequence identity to other Type I keratins suggested the possibility that CP49 might be more closely related to K19 and represent an addition to the K19 branch of Type I IF proteins. However, DNA sequence analysis of a λ phage insert containing human CP49 genomic sequences identified the presence of an intron within the last codon of CP49. The presence of this intron is consistent with Type I keratin origins, but not K19 evolutionary origins.

DISCUSSION

We report here the complete cDNA and amino acid sequences for human CP49, the first complete and correct report of a mammalian CP49 sequence, and the first description of CP49 gene structure. Comparison of CP49 sequence with the SWISSPROT database shows that the best 35 matches are IF proteins, clearly linking CP49 to the IF family. CP49's membership in the IF family is further supported by analysis of its predicted secondary structure. Alignment of CP49 with the closest Type I cytokeratin matches permits clear identification of a central rod domain in CP49 as well as all of the secondary
CP49: A Highly Divergent Type I Intermediate Filament Protein

Fig. 6. Intron/exon boundaries. The nucleotide sequence of each intron/exon boundary of the human CP49 gene is shown, as is the splice point at which the intron is excised. Introns B, C, and E–H are shown and include the first and last six nucleotides of the intron sequence. Residue positions are indicated below the nucleotide sequence.

structural features that characterize the rod domains of IF proteins and that are considered diagnostic for IF proteins (22, 23, 34–36, 45, 46). Collectively, these data constitute very strong evidence that CP49 is a bona fide member of the IF family.

While overall sequence and predicted secondary structure make a compelling case that CP49 is an IF protein, primary sequence data do not make a strong case for placement of CP49 within any of the existing classes of IF proteins. The CP49 rod domain is most similar to the Type I cytokeratins (Fig. 4), but the level of identity between CP49 and other Type I IF proteins is much lower than that usually seen among members of an IF class and is more typical of that seen between classes. Indeed, comparison of CP49 central rod domain sequence with that of other types of IF proteins showed only modestly lower levels of identity to these other types (Fig. 4). Thus, primary sequence similarity did not permit a confident assignment of CP49 to an existing class of IF protein and suggested that CP49 might constitute a novel class of IF protein. To further clarify the relationship between CP49 and the IF family, we defined the human CP49 gene structure. Within a class of IF proteins, the number and location of introns are strongly conserved, and the presence/absence of a particular intron can be diagnostic for IF class (35, 41, 42). Thus, classification based on primary sequence can be alternatively confirmed or refuted by examination of gene structure. Our data show that CP49 gene structure is identical to that of the Type I cytokeratins (35, 41). Specifically, CP49 lacks intron A (Fig. 5), which is characteristic of Type II cytokeratins, and intron D, found in Type II and III, but not Type I, IF proteins. Finally, the CP49 gene retains intron B, which is unique to Type I cytokeratins. The identical gene structure between CP49 and the Type I cytokeratins is evidence that their similarity arises from sharing a common origin rather than by convergent evolution. CP49 thus exhibits a degree of sequence divergence not previously reported among the Type I cytokeratins.

This determination of CP49 gene structure also bears on the relationship between CP49 and K19. The primary sequence of K19 extends ~13 amino acids beyond the end of the central rod domain, an unusually short carboxyl-terminal tail domain. In fact, K19 has been referred to as the “tailless” keratin because of this feature. Thus, CP49 and K19 share a common and very unusual feature among IF proteins: a highly abbreviated carboxyl-terminal tail domain. K19 is also unusual among the Type I cytokeratins in lacking an intron that is located at the end of the central rod domain and that is conserved among Type I, I1, and III IF proteins. Instead, the exon encoding the end of the central rod domain of K19 extends several dozen bases beyond the site where that intron would have occurred, encoding the abbreviated carboxyl-terminal tail. Thus, while K19 is considered a Type I IF protein, this variation in gene structure has led to the suggestion that K19 has diverged slightly from the Type I family (44). The abbreviated tail domains of K19 and CP49, combined with the low level of identity between CP49 and the Type I keratins, might be considered evidence that K19 and CP49 were closely related. Our determination that the CP49 gene contains intron H (Fig. 5), which is present in all Type I acidic cytokeratins, but not in K19, makes this evolutionary relationship unlikely.

CP49, while clearly an IF protein, also exhibits features that are unique among IF proteins. 1) CP49, as indicated, essentially lacks a carboxyl-terminal tail domain, with the amino acid sequence terminating at or near the end of coil 2. The absence of a tail domain is conserved among the human, bovine, and murine CP49 proteins. 2) CP49 shows the greatest degree of sequence divergence in one of the most highly conserved motifs among IF proteins, the LNDR motif found near the beginning of coil 1a (Fig. 3, shaded). The capacity for in vitro assembly of IF proteins into 10-nm filaments is extremely sensitive to changes in this region. Among IF proteins, 1 and very rarely 2 residues will vary from the consensus LNDR motif. CP49 shows three substitutions, LNDR to LGGC, which may be involved in the function of this motif.

The inherent human skin disorder Dowling-Meara epidermolysis bullosa simplex has been shown to be caused by a point mutation in the Type I keratin K14 that results in an Arg to Cys switch (23, 47–50). This Arg to Cys substitution is seen in inherited human skin disorder Dowling-Meara epidermolysis bullosa simplex is supported by studies on the in vitro assembly of mutant K5/K14, where the introduction of this mutation proves sufficient to disrupt assembly. Finally, engineering this mutation in trans-vitro, and murine CP49 proteins. 2) CP49 shows the greatest degree of sequence divergence in one of the most highly conserved motifs among IF proteins, the LNDR motif found near the beginning of coil 1a (Fig. 3, shaded). The capacity for in vitro assembly of IF proteins into 10-nm filaments is extremely sensitive to changes in this region. Among IF proteins, 1 and very rarely 2 residues will vary from the consensus LNDR motif. CP49 shows three substitutions, LNDR to LGGC, which may be involved in the function of this motif.

In summary, CP49 is an IF protein that represents a unique type of IF protein. Its unusual structural features and sequence divergence make it an interesting candidate for study in terms of its function and its evolutionary relationships within the IF protein family.
required to determine the relationship of filensin to existing IF classes. At this juncture, it is unclear whether CP49 and filensin represent a highly specialized keratin pair or a unique combination of IF proteins from different classes. Preliminary data on the gene structure of filensin suggest that it has similarities to Type II cytoskeletal genes.2

The accumulated data now clearly establish filensin and CP49 as IF proteins. However, sequence analyses establish both as highly unusual, a finding consistent with their presence in a nontraditional IF structure. The most provocative questions that derive from these observations are both why and how these two proteins assemble into a non-IF cytoskeletal element. How two IF proteins assemble into a non-IF structure is unknown, but the significant changes in the primary structure of these two proteins would seem a likely explanation. Interestingly, we and others have shown that in vitro, purified CP49 and filensin can assemble into classical 10-nm filaments. This suggests that the two proteins can be directed into alternative assembly pathways and, as a corollary, that some additional factor or environment is necessary to direct assembly into beaded filaments. Alternatively, the beaded filament may represent a stabilized intermediate in the process of 10-nm filament assembly. Sauk et al. (Fig. 4a of Ref. 51) have shown a beaded filament-like structure occurring in the process of in vitro filament assembly from cytotkeratin pairs.

Why a beaded filament occurs in the lens fiber cell is an equally compelling question. Intermediate filament networks composed of vimentin are present in the lens epithelium and of their evolutionary origins.

Acknowledgment—The provision of human lenses from the Lions Eye and Tissue Bank is gratefully acknowledged.

REFERENCES

1. Maisel, H., and Perry, M. (1972) Exp. Eye Res. 14, 7–12
2. Maisel, H. (1977) Experientia (Basel) 33, 525–527
3. Rafferty, N., and Esson, E. (1974) J. Ultrastruct. Res. 46, 239–253
4. Klibelar, M. A., Selten, V. A., Dunia, I., Benedetti, E. L., and Bloemenh, H. (1979) Eur. J. Biochem. 95, 543–549
5. Bradley, R. H., Ireland, M., and Maisel, H. (1979) Exp. Eye Res. 28, 441–453
6. Ireland, M., and Maisel, H. (1983) Exp. Eye Res. 36, 531–536
7. Ireland, M., and Maisel, H. (1984) Exp. Eye Res. 38, 637–645
8. FitzGerald, P. G. (1990) Curr. Eye Res. 9, 1063–1067
9. FitzGerald, P., and Gottlieb, W. (1989) Curr. Eye Res. 8, 801–810
10. Merdes, A., Gounari, F., and Georgatos, S. D. (1993) J. Cell Biol. 123, 1197–1204
11. Merdes, A., Brunkenor, M., Horstmann, H., and Georgatos, S. D. (1991) J. Cell Biol. 115, 397–410
12. FitzGerald, P. (1988) Curr. Eye Res. 7, 1243–1253
13. Masaki, S., and Watanabe, T. (1992) Biochem. Biophys. Res. Commun. 186, 190–198
14. Bloemenh, H., Hersen, T., Dunia, I., and Benedetti, E. (1982) Exp. Eye Res. 34, 63–67
15. Bloemenh, H., Berbers, G. A., De Jong, W., Ramaekers, F. C., Vermorken, A. J., Dunia, I., and Benedetti, E. L. (1984) Ciba Found. Symp. 106, 177–190
16. FitzGerald, P. G., and Graham, D. (1991) Curr. Eye Res. 10, 417–436
17. Iheanyi, F., and Takenoto, L. (1990) Curr. Eye Res. 9, 259–265
18. Gopalakrishnan, S., and Takenoto, L. (1992) Curr. Eye Res. 12, 929–933
19. Stewart, M., Hutchinson, A. M., and Quinlan, R. A. (1995) Exp. Eye Res. 60, 181–192
20. Gounari, F., Merdes, A., Quinlan, R., Hess, J., FitzGerald, P. G., Ouzounis, C. A., and Georgatos, S. D. (1993) J. Cell Biol. 121, 847–853
21. Ramaekers, F. C. S., Osborn, M., Schmid, E., Weber, K., Bloemenh, H., and Franke, W. W. (1980) Exp. Cell Res. 127, 309–327
22. Albers, K., and Fuchs, E. (1992) Int. Rev. Cytol. 134, 243–279
23. Fuchs, E., and Weber, K. (1994) Annu. Rev. Biochem. 63, 345–382
24. Ireland, M., and Maisel, H. (1989) Lens Toxic. Res. 6, 623–638
25. Remington, S. G. (1993) J. Cell Sci. 105, 1057–1068
26. Hess, J. F., Casselman, J. T., and FitzGerald, P. G. (1993) Curr. Eye Res. 12, 77–88
27. Orii, H., Agata, K., Sawada, K., Eguchi, G., and Maisel, H. (1993) Curr. Eye Res. 12, 583–588
28. Sawada, K., Agata, J., Eguchi, G., Quinlan, R., and Maisel, H. (1995) Curr. Eye Res. 14, 545–549
29. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
30. Frohman, M., Dush, M., and Martin, G. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 8998–9004
31. Hess, J., Casselman, J., and FitzGerald, J. (1995) Curr. Eye Res. 14, 11–16
32. Sambrook, J., Fritsch, E., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
33. Reiger, M., and Franke, W. (1988) J. Mol. Biol. 204, 841–849
34. Parry, D. (1990) Cell and Molecular Biology of Intermediate Filaments (Goldman, R., and Steinert, P., eds) pp. 175–204, Plenum Press, New York
35. Steinert, P., and Roop, D. R. (1988) Annu. Rev. Biochem. 57, 593–625
36. Parry, D., and Steinert, P. (1992) Curr. Opin. Cell Biol. 4, 94–98
37. Stewart, M. (1990) Curr. Opin. Cell Biol. 2, 90–100
38. Stewart, M. (1993) Curr. Opin. Cell Biol. 5, 3–11
39. Weber, K., Riemer, D., and Dodemont, H. (1991) Biochem. Soc. Trans. 19, 1021–1023
40. Dodemont, H., Riemer, D., and Weber, K. (1990) EMBO J. 9, 4083–4094
41. Krieg, T. M., Schafer, M. P., Cheng, C. K., Filippa, D., Flaherty, P., Steinert, P. M., and Roop, D. R. (1986) J. Biol. Chem. 261, 5876–5870
42. Marchuk, D., McCrohon, S., and Fuchs, E. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 1609–1613
43. Bader, B. L., Magin, T. M., Hatzfeld, M., and Franke, W. (1986) EMBO J. 5, 1065–1075
44. Stasiak, P. C., Purkis, P. E., Leigh, I. M., and Lane, E. B. (1989) Invest. Dermatol. 92, 707–716
45. Fuchs, E., and Hanukoglu, I. (1983) Cell 34, 332–334
46. Weber, K., and Geisler, N. (1985) Annu. Rev. Biochem. 54, 126–143
47. Coulombe, P. A., Hutton, M. E., Letal, A., Hebart, A., Paller, A. S., and Fuchs, E. (1991) Cell 66, 1301–1311
48. Fuchs, E., and Coulombe, P. A. (1992) Cell 69, 899–902
49. Fuchs, E. (1991) Biochem. Soc. Trans. 19, 1112–1115
50. Fuchs, E., Coulombe, P., Chen, J., Chan, Y., Hutton, E., Syder, A., Degenstein, L., Yu, Q. C., Letal, A., and Vassar, R. (1994) Invest. Dermatol. 103, 255–305
51. Saku, K., Krumweide, M., Cocking-Johnson, D., and White, J. (1984) J. Cell Biol. 99, 1590–1597
52. Ellis, M., Alousi, A., Lawiczak, J., Maisel, H., and Welsh, M. (1984) Exp. Eye Res. 38, 195–202
53. Maisel, H., and Ellis, M. (1984) Curr. Eye Res. 3, 369–377
54. Sandilands, A., Prescott, A., Carter, J., Hutcheson, A., Quinlan, R., Richards, J., and FitzGerald, P. (1995) J. Cell Sci. 108, 1397–1406
Gene Structure and cDNA Sequence Identify the Beaded Filament Protein CP49 as a Highly Divergent Type I Intermediate Filament Protein
John F. Hess, Jodi T. Casselman and Paul G. FitzGerald

J. Biol. Chem. 1996, 271:6729-6735.
doi: 10.1074/jbc.271.12.6729

Access the most updated version of this article at http://www.jbc.org/content/271/12/6729

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 52 references, 11 of which can be accessed free at http://www.jbc.org/content/271/12/6729.full.html#ref-list-1