Epiplakin Gene Analysis in Mouse Reveals a Single Exon Encoding a 725-kDa Protein with Expression Restricted to Epithelial Tissues*

Daniel Spazierer‡§, Peter Fuchs‡§, Verena Pröll‡, Lubomir Janda‡, Susanne Oehler‡, Irmgard Fischer‡, Rudolf Hauptmann‡, and Gerhard Wiche‡**

From the ‡Vienna BioCenter, Institute of Biochemistry and Molecular Cell Biology, University of Vienna, A-1030 Vienna, Austria and the §Department of Exploratory Research, Boehringer Ingelheim Austria, A-1121 Vienna, Austria

Based on cDNA cloning and sequencing, human epiplakin has been classified as a member of the plakin protein family of cytolinkers. We report here the characterization of the mouse epiplakin gene locus and the isolation of full-length mouse epiplakin cDNA using BAC vectors. We found that the protein is encoded by a single remarkably large exon (>20 kb) that consists of a series of 0.8–1.5-kb-long DNA repeats, eight of which are virtually identical. Consequently, mouse epiplakin contains 16 plakin repeat domains, three more than reported for the human protein and eight more than predicted for the mouse protein based on the contig characterized by the Mouse Genome Sequencing Consortium. Using antibodies raised to a highly conserved repeating epiplakin sequence domain, we show that the protein in cells is expressed in its full length (725 kDa), and we provide evidence that the size of human epiplakin previously may have been underestimated. In addition we show on transcript and protein levels that epiplakin is restricted to epithelial tissues and that its gene maps to mouse chromosome 15 (human chromosome 8). This study lays the groundwork for future genetic approaches aimed at defining the biological role of this unique protein.

Epiplakin was originally identified as a 450-kDa epidermal autotigien showing immunoreactivity with the serum of a patient suffering from a subepidermal blistering disease (1, 2). The recent isolation and sequence analysis of epiplakin cDNA from humans (3) classified the protein as a member of the plakin or cytolinker protein family (for review see Refs. 4 and 5). This family comprises large multidomain proteins that serve as bridging elements between cytoskeletal filaments and as filament anchoring structures of membrane-associated adhesive junctions. Concurrent with their role as cytoskeletal linker proteins, the functional impairment of these proteins leads to diseases accompanied by skin blistering and other structural motifs characteristic of plakins must be considered an atypical family member. However, the redundance of its PRDs makes it a very attractive model for studying the specific function(s) of this domain and of plakin cytolinker proteins in general.

PRDs comprise a highly conserved ~20-kDa core region, called a module (8, 9), and a less conserved linker region of variable length. Modules are composed of four complete and one incomplete tandem copies of a 38-residue-long sequence motif, referred to in data bases as the plectin (PLEC) repeat (9). As predicted on the basis of computer modeling (9), x-ray crystallography showed each of the first four PLEC repeats to consist of a β-hairpin followed by two anti-parallel α-helices, whereas the last PLEC repeat forms a hairpin-helix-loop-hairpin-type motif (10). Although little is known about the function of these modules, the linkers connecting them are thought to serve as binding sites for interacting proteins. For example in plectin, the best studied plakin to date, the binding site for intermediate filaments (IFs) of various types (vimentin, cytokeratins, and desmin) has been mapped to a stretch of about 50 amino acid residues in the linker region between PRD modules 5 and 6 (11). For desmoplakin, too, it could be shown that residues downstream of its third PRD module were important for keratin binding, whereas the linker between the second and third module as well as the second PRD module itself have been reported to be required for vimentin interaction (12–15) (note that the PRD modules of desmoplakin are also known as A, B, and C domains). Binding to keratins and neuronal IFs has also been demonstrated for the PRDs of BPAG1e and BPAG1n (16).

To lay a basis for future genetic studies, we characterized the mouse epiplakin gene locus and determined the nucleotide sequence of the mouse epiplakin cDNA in full. We found that the entire coding sequence of mouse epiplakin is contained within a very large (~20 kb) single exon encoding a series of

* This work was supported by Grant SFB6-11 from the Austrian Science Research Fund. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY312170.

† Fellow of the International Ph.D. Program at the Vienna Biocenter, supported by Grant W1 from the Austrian Science Research Fund.

‡ To whom correspondence should be addressed. E-mail: wiche@ubc.univie.ac.at.

The abbreviations used are: BPAG1, bullous pemphigoid antigen 1; PRD, plakin repeat domain; IF, intermediate filament; PIPES, 1,4-piperazinediethanesulfonic acid; GST, glutathione S-transferase; ORF, open reading frame.

1 The abbreviations used are: BPAG1, bullous pemphigoid antigen 1; PRD, plakin repeat domain; IF, intermediate filament; PIPES, 1,4-piperazinediethanesulfonic acid; GST, glutathione S-transferase; ORF, open reading frame.
virtually identical DNA repeats. Only by using BAC technology, but not standard DNA cloning methods, were we able to clone the gene and establish its nucleotide sequence. Our analysis shows that mouse epilakin contains 16 PRDs, and its size is therefore considerably larger than that reported for the human ortholog. Because we could confirm a similar large size of mouse and human epilakin by immunoblotting, it is possible that the size of human epilakin in former studies has been underestimated. In addition, we show the immunolocalization of epilakin on epithelial sections of mouse tissue using a newly generated mouse hybridoma, using the DyeDector system, and we document the tissue distribution of epilakin transcripts using RNase protection assays. The predicted structure of epilakin is discussed in comparison with that of other plakin family members.

EXPERIMENTAL PROCEDURES

Isolation of cDNA and DNA Sequencing—cDNA clones were isolated from a mouse skin cDNA library (strain C57Bl/6; Stratagene) using a 830-bp-long rat epilakin cDNA fragment as probe. 5'- and 3'-rapid amplification of cDNA ends as well as PCR analyses of Marathon-Ready™ cDNA derived from mouse kidney and 11-day-old mouse em- bryo (Clontech; Palo Alto, CA) were performed using Advantage DNA polymerase (Clontech) in a Perkin-Elmer GeneAmp 9700 thermal cy- cler, following the protocols supplied by the manufacturers. Nested epilakin-specific primers were designed with a melting temperature higher than 60 °C using the Oligo 4.0 program. Optimized PCR con- ditions for the first PCR consisted of five cycles of 94 °C for 5 s and 72 °C for 3 min, 5 cycles of 94 °C for 5 s, and 70 °C for 3 min and 30 cycles of 94 °C for 5 s and 68 °C for 3 min. 2 μl of a 1:50 dilution of the initial PCR was used in a second round of PCR with nested primers (40 cycles of 94 °C for 5 s, 64 °C for 30 s, and 72 °C for 3 min). PCR products were cloned into plasmid pCR2.1 (Invitrogen) for further analyses and se- quencing. The nucleotide sequences were determined by the chain ter- minal nuclease method using the DyeDector terminator cycle sequencing kit (Applied Biosystems, Foster City, CA).

Isolation of Genomic Clones and Analysis of Exon-Intron Organiza- tion—To isolate genomic clones, a mouse genomic library (strain 129; Stratagene, La Jolla, CA) was screened using mouse epilakin cDNA clone pDS 104 and a 900-bp Sau3A/BalI fragment of lambda clone EPI1 as probes. Exon-intron boundaries were identified by comparison of geno- mic DNA and cDNA sequences. The intron was sequenced in its entirety.

Data Base Search and Sequence Alignments—Data base searches were performed using the BLAST program (17). All of the sequence alignments were generated with the LALIGN program (www.ch. embnet.org/software/LALIGN_form.html) using the algorithm of Huang and Miller (18). Secondary structure predictions were made using the data base search and sequence alignments. The predicted structure of epiplakin is discussed under "Results.

RESULTS

The Mouse Epilakin Gene: Eight of Sixteen Sequence Re-peats Are Nearly Identical—When we screened a rat genomic library with a rat plectin cDNA probe, covering the rod domain and one and a half plectin modules, two overlapping lambda clones were isolated that carried sequences differing from plectin. Analysis of these clones showed that parts of them were

S. Oehler, unpublished data.
highly homologous to published partial sequences of a human epidermal 450-kDa autoantigen (2), later identified as epiplakin (3). To isolate the gene and entire cDNA, a mouse skin cDNA library was screened using a PCR-amplified part of the isolated rat sequence as a probe. The screening yielded a 2420-bp cDNA clone (pDS 104), which contained a stop codon in a highly homologous to published partial sequences of a human epidermal 450-kDa autoantigen (2), later identified as epiplakin (3). To isolate the gene and entire cDNA, a mouse skin cDNA library was screened using a PCR-amplified part of the isolated rat sequence as a probe. The screening yielded a 2420-bp cDNA clone (pDS 104), which contained a stop codon in a

The restriction enzyme sites were K, KpnI, B, BamHI, BgII, and S, SpeI. HindIII sites used for cloning of the entire 1.5-kb BamHI/BglII repeats-containing region are in bold type. The lower part of the figure shows positions of isolated lambda, BAC, and cDNA clones. The open boxes in the clones represent 1.5-kb DNA repeats. The dashed line indicates an uncertain number of 1.5-kb DNA repeats contained in lambda clone EP 10. The KpnI sites used for determination of the exact number of DNA repeats in pDS 166 are in bold type.

To isolate and characterize the genomic region 3' of clone EP1, we used a BAC clone shown by BLAST search to harbor epiplakin genomic sequences (RPC123, clone 208H22). The characterization of this clone by restriction enzyme digestion and Southern blotting enabled us to identify a large genomic region (>12 kb), starting near the 3' end of clone EP1 and consisting of an unknown number of ~1.5 kb repeats (Fig. 1). The length of the repeats was determined by digestion of the clone with either BamHI or BglII and Southern blotting of the fragments using a carboxyl-terminal epiplakin fragment as a probe (data not shown). Because no suitable singular restriction sites could be found in this region, the entire gene segment containing the repeats was subcloned into a bacterial artificial (BAC) vector (pBeloBac 11) via two flanking HindIII restriction enzyme sites to yield clone pDS 166 (Fig. 1). KpnI digestion of clone pDS 166 and gel electrophoresis of the fragments showed that the region containing the repeats comprised ~13.8 kb (Fig. 2A). Knowing the exact length of the sequences flanking the repeat region (5' KpnI-BglII fragment, 1140 bp; 3' BglII-KpnI fragment, 345 bp), the number of 1.5 kb repeats could be defined as eight (1140 + 8 × 1545 + 345 = 13,845 bp). Because these repeats may be subject to recombination events in meiosis because of their sequence identity, we examined the genomic stability of the epiplakin locus in three different mouse strains. Southern blot analysis of KpnI-digested DNA from strains B6, CBA, and 129 revealed no differences in the length of fragments (Fig. 2B), confirming genomic integrity of all repeats in different genetic backgrounds.

In the absence of singular restriction sites in clone pDS 166 and because of its sequence redundancy, the analysis of its exact nucleotide sequence posed a major challenge. Following a shotgun cloning strategy, clone pDS 166 was digested with either BglII or BamHI, and the 1.5-kb fragments generated were subcloned (Fig. 3A). Because of the lack of potentially recombining similar sequences, single 1.5-kb clones were stably replicating and therefore could be subjected to sequence analysis. In total 21 BamHI and 30 BglII clones were analyzed. Additional sequence information about the 5' and 3' ends of this highly homologous repeat region was obtained from clones pDS 167 and pDS 170 that were isolated using singular upstream or downstream restriction sites. Based on this analysis we could identify four different classes of BamHI and five different classes of BglII subclones (Fig. 3B). Sequence alignment of a 5' SpeI-BamHI fragment (pDS 167) with BglII fragments enabled the extension of the exact sequence until the 3' end of the first BglII fragment, because only one class of BglII fragments (class I) revealed the same nucleotides (T, T, and G) at variable positions 1–3 (Fig. 3, B and C). However, gene walking in the 3' direction of clone pDS 167 by searching for a unique class of overlapping BamHI clones was not possible, because more than one class (classes I and III) matched nucleotides A and C at the variable positions 4 and 5 of the BglII class I clone. Performing BglII partial digestion, we were also able to clone the very 3' 1.5-kb BglII fragment as part of a BglII-KpnI subclone (pDS 170) of pDS 166 (Fig. 3A). Sequence analysis of this clone revealed the nucleotides C, C, A, T, C, and T at variable positions 1–5, and 1, respectively (Fig. 3B). Again, however, it was impossible to extend the sequence of pDS 170 in the 5' direction, because no unique classes were found, but two BamHI subclone classes (classes I and IV) were identified with nucleotides C, C, and A at variable positions 1–3 of the overlapping region. Consequently, the sequences of only two (the first and last) of the eight 1.5-kb BglII fragments identified could unambiguously be determined, and the order of the other six repeats remained undefined. Because all of the isolated BamHI fragments carried a cytosine at the variable positions 1 and 2, these positions could be specified in all repeats, leaving just three positions (positions 3–5) in repeats 2–7 undetermined (Fig. 3C). Consequently, a sequence could be deduced that lacked exact nucleotide specifications only at three positions within each of the six interior BglII 1.5-kb fragments. The two alternative nucleotides found at each of these positions

Downloaded from http://www.jbc.org/ by guest on July 26, 2018
gave rise to different amino acid residues only at positions 3 (Gly or Glu) and 4 (Glu or Val) (Fig. 3C). At position 5 the alternative nucleotides were at the wobble position of lysine codons. In summary, our analysis revealed that the mouse epiplakin gene comprises two exons, a noncoding exon 1 and a large (~20 kb) coding exon 2 (Fig. 1). Overall, the isolated lambda clones and BAC subclones resulted in a continuous span of 42 kb of genomic sequence.

**Exceptionally Large Size of Epiplakin Confirmed on Transcript and Protein Levels**

— A Northern blot analysis of RNA isolated from mouse salivary gland using a 802-bp carboxyl-terminal cDNA fragment of epiplakin as a probe for hybridization (Fig. 2).

**FIG. 2.** Numerical analysis of 1.5 kb DNA repeats. A, agarose gel electrophoresis of plasmid pDS 166 and BAC 208H22 cut with KpnI. The arrows mark the calculated lengths of KpnI fragments containing either 7, 8, or 9 BglII repeats of 1545 bp; the bold arrow points at the band generated. Size markers were run in parallel. B, Southern blot analysis. Genomic DNA from mouse strains B6, 129, and B6/CBA and from BAC 208H22 were cut with KpnI, and the fragments generated were subjected to Southern blotting using a 1.5-kb BamHI repeat as a probe; the positions of size markers are indicated.

**FIG. 3.** Nucleotide and amino acid sequences of highly homologous epiplakin repeat domains. A, schematic drawing of genomic repeat-containing region (clone pDS 166). The restriction sites were: Ba, BamHI; Bg, BglII; H, HindIII; K, KpnI; and S, SpeI. The white boxes represent 1.5-kb BglII and BamHI fragments. The 5′ end of the repeat region was cloned as part of a SpeI-BamHI fragment (pDS 167), the 3′ end was cloned as a BglII-KpnI subclone (pDS 170) via BglII partial digestion (asterisk). One BglII and one BamHI subclone are drawn enlarged (representative for many analyzed BglII and BamHI fragments) with variable positions marked by numbers. The clones shown overlap in the region harboring variable positions 4 and 5. Epiplakin modules are indicated in gray. B, nucleotides at the variable positions of isolated BglII and BamHI subclone classes and of the clones harboring the 5′ and the 3′ ends of the repeat region (pDS 167 and pDS 170). C, nucleotide sequence specifying nucleotides at any variable positions of repeat region and deduced amino acid sequence. Numbers of nucleotides (n) and amino acid residues (a) between variable positions are indicated. Bg, BglII restriction sites. Positions of BglII repeats 1–8 and of the repeat domains 9–16 are marked in nucleotide and amino acid sequences, respectively.
tion revealed the presence of a very large transcript (Fig. 4A). Based on the size of plectin transcripts (34), the size of epiplakin transcripts was estimated as ~22 kb. This correlated very well with the transcript size predicted: 1.5-kb 5′-untranslated region, 20-kb ORF, and 1-kb 3′-untranslated region (Fig. 1). In the absence of other detectable signals, the expression of other major isoforms caused by splice events within exon 2 of epiplakin could be ruled out.

The 19,644-bp ORF of epiplakin potentially encodes a protein of 725 kDa. To confirm this size we raised antibodies to a GST fusion protein containing a fragment of epiplakin (residues 6041–6364) without any significant sequence homologies to plectin, desmoplakin, or other proteins. The antibodies obtained were affinity-purified and used for Western blotting. Analyzing proteins extracted from a newborn mouse, we could detect a single band of molecular weight considerably higher than that of plectin (>500,000), fitting very well the expected molecular mass of 725 kDa (Fig. 4B). An immunoreactive protein band of similar size was observed when protein extracts of human HaCaT cells were analyzed, indicating that human epiplakin had the same apparent molecular weight as the mouse protein (Fig. 4C). This deviated from an earlier report describing a 552-kDa protein (3). This deviated from an earlier report describing a 552-kDa protein (3). To determine the chromosomal localization of the mouse epiplakin, we used DNA progeny derived from mating of ((C57BL/6J × Ei) F1) mice (Fig. 3). No detectable hybridization was found in either mouse or human tissue sections, thereby excluding the presence of a false positive reaction. In conclusion, we were able to identify the chromosomal localization of the mouse epiplakin gene.

**Epiplakin Structure Predictions**—The predicted amino acid sequence comprised 16 PRDs, each containing a linker followed by a module (Fig. 5A). The modules were homologous to the B-type modules of plectin, desmoplakin, and BPAG1 (9, 35). The number of modules identified (16 modules) differs from the 13 domains previously reported for human epiplakin (3). This discrepancy is due to the different numbers of almost identical repeat domains found in the carboxyl-terminal part of the protein. Only five such repeats were reported to be expressed in humans, whereas eight were identified in this study. Contrary to the linker regions, those parts of epiplakin repeat domains corresponding to modules were highly conserved compared with other plakin family members. Furthermore, similarities between human and mouse epiplakin were more pronounced among modules (e.g. ~90% sequence identity of modules 9) compared with linkers (~63% of the corresponding linkers). A closer look identified two groups of B-type domains: group I with ~70% identity (modules 3, 6, and 8–16) and group II with ~50% or less identity to the first B-type domain of plectin (modules 1, 2, 4, 5, and 7) (Table I). A similar differentiation has been made recently in an analysis of the crystal structure of B- and C-type domains of desmoplakin (10). The proposed characteristics of the type B module of desmoplakin are found only for group I but not group II modules of epiplakin.

In defining the start of each module, the amino acids alanine and glycine at positions 3 and 4 were used (Fig. 5B, lower panel). This definition applies to all modules found in any of the plakins known, except for module 4 of epiplakin (Gly-Gly) and the third domain of desmoplakin (Ala-Ala). Each module of epiplakin is composed of five tandem copies of a repeat motif, in data bases (SMART, smart.embl-heidelberg.de, and Pfam, www.sanger.ac.uk/cgi-bin/Pfam, respectively) defined as the PLEC repeat (Fig. 5B). The first PLEC repeats (comprising 42 amino acids in modules 1–7 and 47 amino acids in modules 8–16, respectively) form a hairpin-helix-loop-helix (β2α2ο) motif. Repeats 2, 3, and 4 (composed of 38 amino acid residues) exhibit a similar structure. In contrast, the fifth and last repeat in each module (comprising only 32 residues) exhibits a hairpin-helix-loop-hairpin-type structure (β2ο2β). The linker regions between the modules can be divided into four different types: Linker 1 (type I) is unique and shows no homology to the rest of the protein. Type II, III, and IV linkers differ in their length but are at least partially homologous to each other (Fig. 6). Linker 2 (type II), the shortest, shows homology to the carboxyl-terminal part of linker types III and IV. Linkers 3–8 belong to type III, linkers 9–16 belong to type IV. Type IV linkers, which are virtually identical in primary structure, are the longest linkers and can be subdivided into three homology regions. Regions 1 and 2 are homologous to each other and to the amino-terminal part of type III linkers, whereas region 3 shows homology to type III linkers along their entire length (Fig. 6). Interestingly, the only two positions in the 1.5-kb sequence repeats, where our sequence analysis did not allow the assignment of one of two possible amino acid residues (Fig. 3C), were located within the linker domains. The other nucleotide differences detected, which did not affect amino acid exchanges, were located in regions encoding epiplakin modules. It follows that modules 8–15 were fully identical in sequence (Fig. 3).

The Epiplakin Gene Is a Close Neighbor of the Plectin Gene—To determine the chromosomal localization of the murine epiplakin gene, interspecific backcross analysis using DNA progeny derived from mating of (C57BL/6J × SPRET/Ei) mice was carried out. To identify polymor-
Fig. 5. Analysis of epiplakin amino acid sequence. A, domain structure of epiplakin. 16 repeat domains, each comprising a linker (white bar) and a module (gray ellipse), are indicated. Almost identical modules are shown in light gray, and modules with lower homology are in dark gray. B, schematic drawing of epiplakin modules and linkers (upper part) and amino acid sequence of the 16 epiplakin modules (lower part). In the...
TABLE I
Comparison (percentage of sequence identity) of epiplakin modules, plectin modules 1–6, and desmoplakin module B

| Epiplakin module | Plectin module | Desmoplakin module B |
|------------------|----------------|----------------------|
|                  | 1   | 2   | 3   | 4   | 5   | 6   |                  | 1   | 2   | 3   | 4   | 5   | 6   |
| 1 (49–238)*      | 52.6| 51.1| 52.6| 52.6| 51.1| 33.2| 52.6              | 52.6|     |     |     |     |     |
| 2 (193–476)      | 50.5| 47.9| 53.2| 49.5| 47.3| 33.5| 46.3              |     |     |     |     |     |     |
| 3 (619–806)      | 76.1| 56.9| 72.3| 50.0| 67.6| 35.6| 67.0              |     |     |     |     |     |     |
| 4 (939–1123)     | 48.4| 48.4| 50.0| 48.3| 33.2| 42.4| 45.2              |     |     |     |     |     |     |
| 5 (1255–1442)    | 50.0| 47.9| 54.8| 46.3| 47.3| 30.3| 47.3              |     |     |     |     |     |     |
| 6 (1580–1767)    | 78.2| 54.3| 72.3| 48.4| 68.1| 36.2| 68.6              |     |     |     |     |     |     |
| 7 (1906–2093)    | 45.7| 43.6| 48.4| 44.7| 42.0| 34.0| 44.1              |     |     |     |     |     |     |
| 8* (2233–2425)   | 74.1| 53.9| 71.0| 50.8| 67.4| 34.4| 67.4              |     |     |     |     |     |     |
| 16 (6353–6545)   | 72.5| 54.4| 69.4| 49.2| 67.4| 34.9| 67.9              |     |     |     |     |     |     |

* Numbers in parentheses refer to amino acid positions.

Epiplakin Is Predominantly Expressed in Epithelial Tissues—To quantify epiplakin transcripts in tissue extracts, RNase protection assays were carried out using two different epiplakin-antisense riboprobes (specific for the 5′-untranslated region and the amino terminus of epiplakin) and a ribosomal protein S16-specific probe for standardization (Fig. 8). Both epiplakin-specific probes revealed high levels of expression in skin, small intestine, and salivary gland, comparatively lower levels in lung, uterus and liver, and no detectable expression in brain, kidney, muscle, heart, and spleen. At large, this pattern was consistent with the immunolocalization data of epiplakin on cryosections of various mouse tissues (Fig. 9). Strong epiplakin-specific signals were found in all cell layers of the epidermis, whereas no signals could be detected in the subjacent dermis (Fig. 9A). In small intestine (Fig. 9B), epiplakin was exclusively expressed in the epithelial cell layer of the villi, whereas the inner, connective tissue showed no staining. This correlated well with negative immunoblotting results obtained with protein extracts from mouse fibroblasts (data not shown). In liver (Fig. 9C), epiplakin was prominently expressed at the margins of hepatocytes, with additional less pronounced staining of bile canaliculi and of patchy or partly filamentous structures distributed throughout the cytoplasm. In salivary gland and pancreas (Fig. 9, D and E), epiplakin-specific staining was found in the cubic epithelium of the ducts and in myoepithelial cells. No signals could be detected in skeletal muscles, brain, and kidney (Fig. 9F and data not shown). Controls using rabbit nonimmune serum were negative (data not shown). Immunofluorescence microscopy of mouse keratinocytes revealed no filamentous cytoplasmic staining, contrary to expectations considering the subcellular localization of other plakin family members (data not shown).

DISCUSSION
Our detailed analyses of the mouse epiplakin gene locus and of epiplakin cDNA revealed that the protein is a 725-kDa translation product of a ~22-kb mRNA. Of the two exons constituting its gene, only one (with a remarkable size of ~20 kb) was found to be coding, whereas the preceding second one was noncoding. The size of the single coding exon of epiplakin exceeds that of the longest single exon reported by the International Human Genome Sequencing Consortium, a 17.1-kb exon of the titin gene (36). An even longer tandem repeat domains-containing coding exon (~34 kb) was reported for submaxillary mucine in pig (57). Another intriguing feature of the epiplakin gene is the existence of eight virtually identical 1.5-kb DNA repeats, arranged in tandem at the 3′ end of the coding sequence. Sequence analysis revealed six different types of such 1.5-kb DNA repeats. Their alignment disclosed differences in nucleotides at only five positions. We were able to determine the type and therefore the exact sequence of the first and the last of the eight repeats; the order of the remaining six could not be defined. As a result the nearly 20-kb-long coding sequence of the mouse epiplakin gene could be determined for all but 18 nucleotides residing within a span of ~9 kb. For each of these 18 positions only two alternative bases remained an option. Despite the near-identity of some of its DNA repeats, the epiplakin gene seems to be genomically stable as far as the number of transmitted repeats is concerned, as shown by our analysis of three different inbred mouse strains.

The isolation and analysis of the genomic locus of epiplakin, especially of the part comprising the 1.5-kb DNA repeats, was a technical challenge. In standard cloning approaches the number of DNA repeats was not stably maintained because of recombination events in E. coli. Only the use of the BAC cloning system enabled the isolation of these repeats in full and their subsequent sequence analysis. Furthermore, because all PCR-based methods and cDNA isolation techniques meet the same problems, the technical approach taken would seem to be the only reliable way to identify the exact length, sequence, and, ultimately, structure of epiplakin, or similarly repetitive
protein structures. In this context it is not surprising that the shotgun sequencing strategy used by the mouse genome sequencing project was not successful in determining the whole sequence of the epiplakin gene. The detailed analysis of the epiplakin gene locus reported here will help to bridge a gap in the draft sequence of the mouse genome, which probably would not have been closed by conventional sequencing approaches in the near future.

The single epiplakin-specific signal detected by Northern blotting corresponded well to the expected size of the full-length transcript (91/11011 kb), ruling out any major splice events. Furthermore, only one major immunoreactive protein band of 725 kDa could be detected by immunoblotting analysis of newborn mice lysates, suggesting that proteolytic processing of full-length expressed protein was not a major issue. The different lengths of the mouse (this report) and the human (3) epiplakin cDNA ORFs (19644 versus 15195 bp) were due to different numbers of 1.5-kb DNA repeats identified in each case (eight in mouse versus five in human), whereas the remaining parts of the sequences were found homologous to each other. Because only classical screening and PCR-based methods were used for the isolation of human epiplakin cDNA, one or more of the 1.5-kb repeats could have easily been missed. This was supported by our immunoblotting data showing that human

sequence of the epiplakin gene. The detailed analysis of the epiplakin gene locus reported here will help to bridge a gap in the draft sequence of the mouse genome, which probably would not have been closed by conventional sequencing approaches in the near future.

The single epiplakin-specific signal detected by Northern blotting corresponded well to the expected size of the full-length transcript (~22 kb), ruling out any major splice events. Furthermore, only one major immunoreactive protein band of ~725 kDa could be detected by immunoblotting analysis of newborn mice lysates, suggesting that proteolytic processing of full-length expressed protein was not a major issue. The different lengths of the mouse (this report) and the human (3) epiplakin cDNA ORFs (19644 versus 15195 bp) were due to different numbers of 1.5-kb DNA repeats identified in each case (eight in mouse versus five in human), whereas the remaining parts of the sequences were found homologous to each other. Because only classical screening and PCR-based methods were used for the isolation of human epiplakin cDNA, one or more of the 1.5-kb repeats could have easily been missed. This was supported by our immunoblotting data showing that human
and mouse keratinocyte epiplakins have a similar apparent high molecular mass of >700 kDa, clearly above the 552 kDa reported for the human species (3). However, the exact size of human epiplakin remains to be identified using a methodological approach similar to the one presented in this study.

RNase protection assays revealed high level expression of epiplakin transcripts in mouse skin, small intestine, and salivary gland and lower expression in liver, lung, and uterus. No expression could be detected in brain, muscle, heart, kidney, or spleen. Considering that epiplakin fragments were successfully amplified from a kidney cDNA library, the lack of any signal for kidney in these assays suggested that only trace amounts of epiplakin mRNA were present in this tissue. Thus, in mouse we found expression of epiplakin transcripts to be largely restricted to epithelial tissues, at variance with human epiplakin, which was reported to be widely distributed in a variety of tissues (3). The expression pattern of mouse epiplakin mRNA was in agreement with immunolocalization data of the protein on cryosections of tissues, showing its expression in all layers of the epidermis, in the epithelial layer of the small intestine, in the cubic epithelium of pancreas and salivary gland, and in liver. The more prominent expression of epiplakin observed in suprabasal compared with basal keratinocytes of the epidermis may indicate a role of epiplakin in skin barrier function, as suggested for other plakins (38, 39).

Mouse epiplakin can be considered as a highly ordered protein structure consisting of 16 homologous parts (PRDs), each one of them composed of a module and a linker domain. All modules and, with the exception of the first, all linkers show similarity among each other. The 16 epiplakin modules each are composed of five structural motifs known as PLEC repeats (9), based on which epiplakin counts as a member of the plakin protein family (4, 5). However, in the absence of other molecular domains shared with plakins, epiplakin is atypical for this protein family. Its closest relative with regard to sequence homology of PRDs would be plectin. This close relationship is further strengthened by the fact that plectin contains the highest number of PRDs of all plakins besides epiplakin and that the epiplakin and plectin genes are just 60 and 45 kb apart on mouse chromosome 15 (40) and human chromosome 8 (34), respectively. On these grounds it can be assumed that epiplakin is a relatively young protein in vertebrate evolution and probably emerged through a duplication and subsequent amplification of a plectin PRD.

Several plakins have been shown to harbor binding sites for various types of IFs in their PRDs (11, 13–16, 41–47). Epiplakin, containing a multitude of such domains, might therefore be expected to contain one or more such sites. However, neither the human nor the mouse protein species harbors sequences characteristic for previously well characterized essential IF-binding sites of plakins. In particular they lack the versatile IF interaction domain, originally identified in the linker domain between PRD modules 5 and 6 of plectin (11), which is found also in other plakins, including desmoplakin and module-less periplakin. Because epiplakin, in addition, lacks an actin-binding domain found at the amino terminus of several other plakins, it remains to be shown whether it qualifies as a true cytolinker. Its unique, highly ordered structure and especially the eight almost identical carboxyl-terminal repeat domains may be perfect preconditions for its putative role as a scaffolding platform providing multiple docking sites for complex protein machineries such as those involved in signaling. This study lays the groundwork for future genetic approaches aimed at establishing the biological role of this unique protein, especially because it opens the door for generating genetically altered mice.

Acknowledgment—We thank Norbert E. Fusenig (German Cancer Research Center) for the donation of HaCaT cells.

REFERENCES
1. Fujiwara, S., Kohno, K., Iwamatsu, A., and Shinkai, H. (1994) Dermatology 189, (Suppl. 1) 120–122
2. Fujiwara, S., Kohno, K., Iwamatsu, A., Naito, I., and Shinkai, H. (1996) J. Invest. Dermatol. 106, 1125–1130
3. Fujiwara, S., Takeo, N., Otani, Y., Parry, D. A., Kunitzma, M., Lu, R., Sasaki, M., Matsuo, N., Khaleeduzman, M., and Yoshikawa, H. (2001) J. Biol. Chem. 276, 13540–13547
4. Leung, C. L., Green, K. J., and Liem, R. K. (2002) Trends Cell Biol. 12, 37–45
5. Roper, K., Gregory, S. L., and Brown, N. H. (2002) J. Cell Sci. 115, 4215–4225
6. Pulkkinen, L., Smith, F. J., Shimizu, H., Murata, S., Yonita, H., Hashisuka, H., Nishikawa, T., McLean, W. H., and Uitto, J. (1996) Hum. Mol. Genet 5, 1539–1546
7. Armstrong, D. K., McKenna, K. E., Purkis, P. E., Green, K. J., Eady, R. A., Leigh, I. M., and Hughes, A. E. (1999) Hum. Mol. Genet 8, 143–148
8. Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997) Proteins 28, 405–420
9. Janda, L., Damborsky, J., Rezniczek, G. A., and Wiche, G. (2001) Bioessays 23, 1064–1069
10. Choi, H. J., Park-Snyder, S., Pascoe, L. T., Green, K. J., and Weis, W. I. (2002) Nat. Struct. Biol. 9, 612–620
11. Nikolic, B., Mac Nulty, E., Mir, B., and Wiche, G. (1996) J. Cell Biol. 134, 1455–1467
12. Stappenbeck, T. S., and Green, K. J. (1992) J. Cell Biol. 116, 1197–1209
13. Stappenbeck, T. S., Bornsalager, E. A., Corcoran, C. M., Luu, H. H., Virata, M. L., and Green, K. J. (1995) J. Cell Biol. 123, 123
14. Kouklis, P., Hutton, E., and Fuchs, E. (1994) J. Cell Biol. 127, 1049–1060
15. Meng, J. J., Bornsalager, E. A., Green, K. J., Steinert, P. M., and Ip, W. (1997) J. Biol. Chem. 272, 21495–21503
16. Yang, Y., Dowling, J., Yu, Q. C., Kouklis, P., Cleveland, D. W., and Fuchs, E. (1990) Cell 66, 655–663
17. Aleschfeld, S. F., Gieh, W., Miller, W. E., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
18. Huang, X., and Miller, W. (1991) Adv. Appl. Math. 12, 357–357
19. Garnier, J., Gibrat, J. F., and Robson, B. (1996) Methods Enzymol. 266, 540–553
20. Guermeur, Y. (1997) Combinaison de Classifieurs Statistiques, Application à la Prédiction de la Structure Secondaire des Protéines. Ph.D. thesis, Université Paris 6
21. Jones, D. T. (1996) J. Mol. Biol. 266, 540–553
22. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998) Bioinformatics 14, 892–893
23. Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S., Thurston, S. J., and Birkenmeier, E. H. (1994) Mamm. Genome 5, 253–274
24. Barnes, W. M. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 2216–2220
25. Manly, K. F. (1993) Mamm. Genome 4, 303–313
26. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
27. Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R., and Berns, A. (1991) Nucleic Acids Res. 19, 4293
28. Church, G. M., and Gilbert, W. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 1991–1995
29. Eger, A., Stockinger, A., Wiche, G., and Foisner, R. (1997) J. Cell Sci. 110, 1307–1318
30. Andra, K., Kornacker, I., Jorgl, A., Zorer, M., Spazierer, D., Fuchs, P., Fischer, I., and Wiche, G. (2003) J. Invest. Dermatol. 120, 189–197
31. Boukamp, P., Petrussevska, R. T., Breitkreutz, D., Hornung, J., Markham, A., and Fusenig, N. E. (1988) J. Cell Biol. 106, 761–771
32. Laemmli, U. K. (1970) Nature 227, 680–685
33. Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8148
34. Liu, C. G., Maercker, C., Castaño, M. J., Hauptmann, R., and Wiche, G. (1996) Proc. Natl. Acad. U. S. A. 93, 4278–4283
35. Green, K. J., Parry, D. A., Steinert, P. M., Virata, M. L., Wagner, R. M., Angst, B. D., and Nilles, L. A. (1990) J. Biol. Chem. 265, 11406–11407
36. Lander, E. S. (2001) Nature 409, 860–921
37. Eckhardt, A. E., Timpte, C. S., DeLuca, A. W., and Hill, R. L. (1997) J. Biol. Chem. 272, 33204–33210
38. Steinert, P. M. (2000) J. Cell Biol. 151, F5–F8
39. Maïtta, A., DiColandrea, T., Groot, K., and Watt, F. M. (2001) Mol. Cell. Biol. 21, 7047–7055
40. Fuchs, P., Zorer, M., Reznicek, G. A., Spazierer, D., Oehler, S., Castaño, M. J., Hauptmann, R., and Wiche, G. (1999) Hum. Mol. Genet. 8, 2461–2472
41. Wiche, G., Gronow, D., Donovan, A., Castaño, M. J., and Fuchs, E. (1993) J. Cell Biol. 121, 607–619
42. Steinbock, F. A., Nikolic, B., Coulombe, P. A., Fuchs, E., Traub, P., and Wiche, G. (2000) J. Cell Sci. 113, 483–491
43. Leung, C. L., Sun, D., and Liem, R. H. (1999) J. Cell Biol. 144, 435–446
44. Guo, L., Degenstein, L., Dowling, J., Yu, Q. C., Wollmann, R., Perman, B., and Fuchs, E. (1995) Cell 81, 253–263
45. Kazerounian, S., Uitto, J., and Aho, S. (2002) Exp. Dermatol. 11, 428–438
46. DiColandrea, T., Karashima, T., Mainz, A., and Watt, F. M. (2000) J. Cell Biol. 151, 573–586
47. Karashima, T., and Watt, F. M. (2002) J. Cell Sci. 115, 5027–5037
48. Cartwright, E., Dawson, C., Frances, S., McGarvey, D., Phillips, J., and Williams, J. (2000) QIAGEN News, Issue 3, pp. 18–21
49. Cartwright, E., Dawson, C., Frances, S., McGarvey, D., Phillips, J., and Williams, J. (2000) QIAGEN News, Issue 4, pp. 22–25
Epilakin Gene Analysis in Mouse Reveals a Single Exon Encoding a 725-kDa Protein with Expression Restricted to Epithelial Tissues
Daniel Spazierer, Peter Fuchs, Verena Pröll, Lubomir Janda, Susanne Oehler, Irmgard Fischer, Rudolf Hauptmann and Gerhard Wiche

J. Biol. Chem. 2003, 278:31657-31666.
doi: 10.1074/jbc.M303055200 originally published online June 5, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M303055200

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 45 references, 20 of which can be accessed free at http://www.jbc.org/content/278/34/31657.full.html#ref-list-1