

Cloning and Characterization of Multiple Human Genes and cDNAs Encoding Highly Related Type II Keratin 6 Isoforms*

(Received for publication, March 2, 1995, and in revised form, May 3, 1995)

Kenzo Takahashi‡, Rudolph D. Paladini, and Pierre A. Coulombe§

From the Departments of Biological Chemistry and Dermatology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

The human type II keratin 6 (K6; 56 kDa) is expressed in a heterogeneous array of epithelial tissues under normal conditions, but is better known for its strong induction in stratified epithelia that feature an enhanced cell proliferation rate or abnormal differentiation. Previous work has established the existence of two functional genes encoding K6 protein isoforms in the human genome, although only a partial cDNA clone is available for K6a, the dominant human K6 isoform in skin epithelial tissues (Tyner, A., and Fuchs, E. (1986) *J. Cell Biol.* 103, 1945–1955). We screened human genomic and skin cDNA libraries with probes derived from the K6b gene, and isolated clones containing the full-length gene and cDNA predicted to encode K6a. A thorough characterization of a large number of genomic (57) as well as cDNA (64) clones further revealed the existence of as many as six different human K6 protein isoforms that are highly related at the gene structure, nucleotide sequence, and predicted amino acid sequence levels. Based on the information accumulated to date we propose an evolutionary model in which the multiplicity of human K6 genes is explained by successive gene duplication events. We further demonstrate that K6a is clearly the dominant K6 isoform in skin tissue samples and cultured epithelial cell lines and that the various isoforms are differentially regulated within and between epithelial tissue types. Our findings have direct implications for an understanding of the regulation and function of K6 during hyperproliferation in stratified epithelia and the search for disease-causing mutations in K6 sequences in the human population.

Keratins are epithelial-specific intermediate filament (IF)¹ proteins encoded by a large multigene family. The ~25 keratins (molecular mass 40–70 kDa) expressed in “soft” epithelial tissues (excluding hair and nail) have been subdivided into type I (K9–K20) and type II (K1–K8) IF sequences (1, 2). As keratin filament assembly begins with the formation of a type I-type II heterodimer (3), epithelial cells express at least one

member of each subtype. Pairwise keratin gene expression is regulated in an epithelial tissue-type and differentiation-specific manner, creating patterns that have been well conserved among mammalian species (1, 4). In stratified epithelia, the type II K5 and type I K14 genes are transcriptionally active in mitotically active basal cells (5, 6), while other pairs of keratin genes are transcribed in the differentiating cell layers. In epidermis, the main differentiation-specific keratins are K1 and K10, while in esophagus and cornea, they are K4 and K13, and K3 and K12, respectively (4–7). These keratin pairs appear to be specific for the program of terminal differentiation executed in these tissues (4). In the cytoplasm of epidermal cells, the primary function of keratin filaments is to provide the strength necessary to maintain integrity when skin is subjected to mechanical stress. Alterations in the structure of keratin filaments at any level within the epidermis causes it to rupture within the cell layer(s) affected upon mild mechanical trauma (3, 8). The production of such phenotypes through the directed expression of mutant keratins in the skin of transgenic mice paved the way for the discovery of mutations affecting specific keratins in individuals suffering from a variety of genodermatoses featuring trauma-induced blistering of the epidermis (8–10).

The type II keratin 6 (K6; 56 kDa) is remarkable by several criteria. In contrast to many other keratins, the pairwise expression of K6 and its type I partners K16 and/or K17 is not linked with a well defined program of terminal differentiation (1, 4). Thus, K6 is constitutively expressed in distinct types of epithelia, such as filiform papillae of tongue, several “wet” stratified epithelia lining the oral mucosa and esophagus, the outer root sheath of hair follicles, and in glandular epithelia (11–14). With the exception of specific body sites, *e.g.* palm and sole, K6 is not expressed in epidermis unless it undergoes enhanced proliferation or abnormal differentiation (15–17). K6 and K16 are in fact best known for their induction in epidermis and other stratified epithelia undergoing hyperproliferation, *e.g.* during wound healing, in several diseases (*e.g.* psoriasis, actinic keratosis) and in cancer (12, 15, 17). Likewise, K6 induction also occurs when epidermal, corneal, and tracheal cells are seeded in primary culture *in vitro* (4, 7, 15, 18). The association between a faster cell turnover rate and K6 expression in stratified epithelia is intriguing, given that expression occurs post-mitotically and can be uncoupled from mitosis in cultured keratinocytes (7, 18). The function(s) that K6 may play in stratified epithelia displaying an enhanced mitotic activity and during wound healing remain to be defined.

Another characteristic of K6 is the existence of two functional genes encoding highly related protein isoforms in the human (16), bovine (19), and mouse (20) genomes. Of the two known human K6 isoform genes, K6a is more abundantly expressed than K6b at the mRNA level in skin explant cultures (21). While the human K6b gene has been characterized (21), only a partial cDNA is available for human K6a (22). As part of

* This work was supported in part by the National Institutes of Health Grant R55-AR42047. 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 nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L42575–L42612.

‡ Paul Janssen Postdoctoral Fellow of the Dermatology Foundation.

§ Supported by a Junior Faculty award from the American Cancer Society. To whom correspondence should be addressed: Dept. of Biological Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205. Tel.: 410-614-0510; Fax: 410-955-5759.

¹ The abbreviations used are: IF, intermediate filament; K, keratin; bp, base pair; kbp, kilobase pair; PCR, polymerase chain reaction.

our efforts to understand the role(s) of K6 in hyperproliferative stratified epithelia, we cloned and characterized the human K6a gene and cDNA. While doing so, we discovered the existence of hitherto unknown functional K6 genes in the human genome. The K6 genes and cDNAs we isolated are predicted to encode five or six highly related K6 isoforms, and are differentially regulated in several epithelial cells and tissues examined. We are also proposing a model for the evolution of human K6 genes. While the functional consequences of this remarkable sequence multiplicity have yet to be determined, our findings have direct implications for an understanding of the function of K6 during wound healing as well as the search for point mutations in K6 sequences in the human population.

EXPERIMENTAL PROCEDURES

Materials—Materials were obtained from the following sources: λ DashII, λ ZapII, pBluescript vector, and Gigapack Gold II from Stratagene; restriction endonucleases and DNA modifying enzymes from New England Biolabs; Moloney murine leukemia virus, and Superscript II reverse transcriptase from Life Technologies, Inc.; oligo(dT)-Latex from Roche; Nytran from Schleicher & Schuell; Biotyne nylon membrane from Pall Ultrafine Filtration Corp.; GeneScreen Plus from Dupont. Cell culture medium, fetal bovine serum, glutamine, and antibiotics were purchased from BioWhittaker. All other chemicals were of reagent grade.

Human Tissues and Construction of Libraries—DNA was extracted from human placenta and used for the construction of a genomic library. A skin cDNA library was constructed from poly(A) mRNAs extracted from a squamous cell carcinoma of the lower leg (with adjacent normal tissue) obtained from a patient (excision surgery). Other human tissues were obtained as discarded material in the course of surgery or at autopsy.

Cell Culture—SCC-13, a human skin squamous cell carcinoma line (23), was grown on a NIH 3T3 fibroblast feeder layer in Dulbecco's modified Eagle's medium supplemented with 20% fetal bovine serum, 0.4 μ g/ml hydrocortisone, and 10 ng/ml epidermal growth factor. SCC-4 and SCC-9, two human tongue squamous cell carcinoma lines (23), were grown in 1:1 mixture of Ham's F-12 and Dulbecco's modified Eagle's media supplemented with 10% fetal bovine serum and 0.4 μ g/ml hydrocortisone. PtK-2, a kangaroo rat kidney cell line (24), was grown in Eagle's minimum essential medium supplemented with 10% fetal bovine serum, non-essential amino acids, and sodium pyruvate. All cells were cultured at 37 °C in a humidified atmosphere containing 5% CO₂.

Human Genomic DNA Library Screening—We used a human genomic library constructed in the λ DashII cloning vector (25) to screen for human K6 genes. Approximately 1.5×10^6 phage clones were screened with DNA probes radiolabeled with [α -³²P]dCTP. The initial screening was done using three probes on replicated filters: a 483-bp *Apa*I fragment derived from exon 1 and a 286-bp *Rsa*I-*Sac*I fragment from the 3' non-coding region of the cloned human K6b gene (21); and a 216-bp *Alu*I-*Spe*I fragment from the 3' non-coding region of the cloned human partial K6a cDNA (22). Hybridization was carried out under stringent conditions at 65 °C in 1 M NaCl, 10% dextran sulfate, and 1% SDS. Filter washing was performed at 65 °C in 0.1 $\times$ saline sodium citrate (SSC; 150 mM NaCl and 15 mM sodium citrate). Hybridization-positive clones were isolated by repeated plaque purification using standard procedures (26). DNAs extracted from the purified phage clones were analyzed by restriction digests and Southern blotting using probes corresponding to one or several exons of the human K6b gene. DNA fragments expected to contain an exon or exons on the basis of blot-hybridization analysis were isolated and subcloned into pBluescript SK(+). DNA sequencing was carried out as described (27). Nucleotide and amino acid sequences were compared using the DNASIS-Mac 2.0 software using the simple homology routine (Hitachi Software Engineering Co.).

Human Skin cDNA Library Screening—To obtain human K6 cDNA clones, we screened a human cDNA library constructed using poly(A) RNA extracted from a leg skin squamous cell carcinoma. After oligo(dT)-driven reverse transcription and second-strand cDNA synthesis, *Eco*RI adapters were ligated, and fractions containing cDNAs of 2–5 kb were inserted into a λ Zap II vector. A total of 1.4×10^6 cDNA clones were screened at high stringency (0.1 $\times$ SSC, 65 °C) with two different probes derived from exon 1 of the human K6a gene isolated in this study (a ~450-bp *Xho*I-*Nar*I and a ~500-bp *Apa*I fragment). Positive cDNA clones were further analyzed by oligonucleotide hybridization, PCR, and DNA sequencing. Oligonucleotide hybridization was used to dis-

criminate among K6 isoform cDNAs on the basis of codon 155 sequence (encoding either Ala or Thr, depending on the isoform). The antisense oligonucleotides used were: probe A (Table I), to detect the K6a isoform (Thr¹⁵⁵); and probe B (Table I), to detect all other K6 isoforms (Ala¹⁵⁵; see "Results"). Isolated K6 cDNA clones were blotted onto nylon membrane, and hybridization with probes A or B was carried out at 37 °C in 5 $\times$ SSC, 30% formamide, and washed in a 0.5 $\times$ SSC, 0.5% SDS solution at 60 °C. cDNA clones hybridizing positively with probe A or B were purified and rescued into pBluescript SK(-) for analysis.

Northern Blot Analysis and Primer-extension Analyses—RNAs were isolated from primary cultures of human skin keratinocytes and several human epithelial carcinoma cell lines by the acid-phenol extraction method (28). These RNA samples were subjected to Northern analysis using genomic DNA fragments from the 3' non-coding sequence of the K6a and K6b genes as probes. For primer-extension analysis of the K6a mRNA, we used an oligodeoxynucleotide primer specific for the 5' leader sequence of the K6a mRNA (K6a in Table I). The primer was labeled with ³²P at its 5' end (2.0–5.0 $\times 10^4$ cpm), hybridized to total RNA (20 μ g) for 4 h at 30 °C, and the mixture was subjected to reverse transcription as described previously (25). The primer-extended products were electrophoresed along with the corresponding sequencing reaction on a 7 M urea, 6% polyacrylamide gel.

Quantitation of K6 Isoform mRNA Levels by a Colony Hybridization Assay—mRNAs were extracted from various human tissues and epithelial cell lines (see above), primed with oligo(dT), and cDNA synthesis was carried out *in vitro* (26). Small aliquots of cDNA products were used for PCR reactions (94 °C for 40 s; 54 °C for 40 s; 72 °C for 60 s; total of 30 cycles) with a set of universal K6 primers, 5' K6-primer, and 3' K6-primer (Table I). The target sequences of these two primers are perfectly conserved among the K6 isoform genes isolated (K6 a, b, c, d) and cDNAs (K6 a, b, e, f), and amplify a 636-bp long fragment encompassing exons 1 and 2. PCR products were subcloned, and transformants were transferred on several duplicate nylon membranes and grown on LB plates for use in a colony hybridization assay. Membranes were hybridized with oligonucleotide probes A and B (Table I). Selective hybridization with probe A indicated the K6a isoform, while hybridization with probe B indicated the K6 b, c, d, e, or f isoform. These isoforms were discriminated by hybridization of duplicate filters with individual oligonucleotide probes C, D, E, F, and G (see Table I). Optimal washing conditions for all the oligoprobes were determined using appropriate control DNAs. Under 0.5 $\times$ SSC and 42 °C conditions, each probe was found to hybridize specifically with the expected purified isoform cDNA(s). Following autoradiography, each clone was scored for positive hybridization with oligonucleotide probes A–G. Hybridization with probe A indicated the K6a isoform; hybridization with probes D and F indicated either the K6b or K6f isoform; hybridization with probes C, E, and G indicated the K6c isoform; and hybridization with probes F and G indicated either the K6d or K6e isoform. Because our K6d genomic clone lacked 198 bp at the 5' end (see "Results"), we could not use probe C to distinguish it from the K6e isoform.

Transient Expression of K6 Isoforms—Genomic DNA fragments containing the entire coding sequence of the human K6a gene (~8-kbp *Eco*RI fragment), K6b (~18-kbp *Sal*I fragment), and the K6c gene (~20-kbp *Sal*I fragment) were subcloned into a cytomegalovirus vector (29) containing a cytomegalovirus promoter-enhancer and a SV40 polyadenylation signal. Transient transfection assays were done in PtK2 epithelial cells (24) cultured on glass coverslips, using the calcium-phosphate precipitation method (30). At 72 h post-transfection, cells were fixed in absolute methanol (–20 °C, 15 min) and processed for indirect immunofluorescence (30). The primary antisera used were a rabbit polyclonal anti-K6 (17) and L₂A₁, a mouse monoclonal anti-K8/K18 (31). Bound primary antibodies were detected with fluorescein isothiocyanate-conjugated goat anti-rabbit IgG (Vector Labs), and a biotin-conjugated goat anti-mouse IgG (Kirkegaard and Perry Labs) followed by streptavidin-Texas Red conjugate (Vector Labs).

RESULTS

Isolation of Multiple Human Genes Encoding Keratin 6—We first performed Southern blot analyses of human genomic DNA using probes derived from the previously characterized human K6b gene (21). When digested human DNAs were probed with a ~600-bp *Nco*I fragment corresponding to the coding portion of the human K6b gene exon 1, nearly 10 hybridization bands were apparent (data not shown). This *Nco*I fragment, however, also hybridized with the human K5 gene, even under stringent washing conditions. When a smaller (283 bp) *Nco*I-*Nar*I frag-

ment derived from the K6b gene exon 1 (which does not hybridize with the K5 gene) was used under stringent conditions ($0.1 \times \text{SSC}$, 70°C), between three and five strong hybridization signals were detected in digested genomic DNAs from several randomly selected individuals (Fig. 1). These results suggested the existence of additional K6 isoform genes or K6-related gene(s) in the human genome.

We screened a human genomic DNA library to isolate the human K6a gene as well as any other K6-encoding genes. A total of 1.5×10^6 clones were screened with various ^{32}P -labeled cDNA probes derived from the coding sequence of the cloned human K6b gene (21). Over 100 clones were found to hybridize strongly under stringent conditions with a probe derived from K6b exon 1. Of these, 57 independent clones were isolated by repeated plaque purification, and analyzed by hybridization under stringent conditions with probes derived from the 3' non-coding regions of the human K6b gene and K6a partial cDNA (21, 22). This revealed that 30 clones hybridized with the K6a 3' non-coding probe (group 1), 7 clones hybridized with the K6b 3' non-coding probe (group 2), while the remaining 20 clones did not hybridize with either probe (group 3). Each group of genomic clones was further analyzed by digestion of purified phage DNAs and their hybridization with specific fragments from the K6b gene. Based on this analysis, the genomic clones in groups 1 and 3 could be further divided into three and four subgroup(s), respectively. Subsequently, hybridization-positive restriction fragments were isolated from representative clones in each of the 8 subgroups, subcloned into pBluescript SK(+), and analyzed by DNA sequencing. The sequence data obtained was compared with those reported for the human K6a partial cDNA and K6b gene (21, 22).

Group 1 clones, which hybridized to the K6a 3' non-coding probe and included three restriction mapping subgroups, consisted of the human K6a gene and two novel K6-encoding genes, named K6c and K6d. Subsequent analyses indicated that these three genes shared virtually identical 3' non-coding sequences (see below). Group 2 clones, which hybridized to the K6b 3' non-coding probe, were independent isolates of the human K6b gene. The four subgroups identified among group 3 clones were found to include: (i) another potential K6-encoding gene (albeit partial); (ii) the K5 gene; and (iii) two full-length genes whose sequence display features characteristic of both the human K5 and K6 genes, designated K5/6- α and K5/6- β .

Structural Organization of Keratin 6-Encoding Human Genes and Characterization of Their mRNAs—We completed the sequencing of the coding region, 5'- and 3'-flanking sequences and intron-exon junctions of the genomic clones potentially encoding K6-like proteins. Comparison of these sequences with the previously reported human K6b gene sequence (21) enabled us to locate all exons and define the 5' regulatory sequences. Overlapping phage clones yielded full-length genes for K6a, K6b, and K6c, and restriction digestion/Southern blotting analyses of suitable phage clones allowed us to assign each of them to a specific hybridization product in human genomic DNAs processed in parallel (Fig. 1). The K6d genomic clone, on the other hand, lacked 198 bp at the 5' end of exon 1, thus preventing assignment to a specific product in digested human genomic DNA. All K6 isoform genes except one are ~6–7 kbp long (Fig. 2) and thus are analogous to other known human type II keratin genes (21, 32–34). In contrast, the K6c gene is remarkably long for a type II keratin gene, extending over 17 kbp. The K6 genes all contain nine exons interrupted by eight introns. The position of all eight introns (A–H) is identical in the K6a, K6c, and K6d genes and in the previously characterized human K6b, K7, and K5 genes (21, 32, 35). The introns are located within the protein-coding regions,

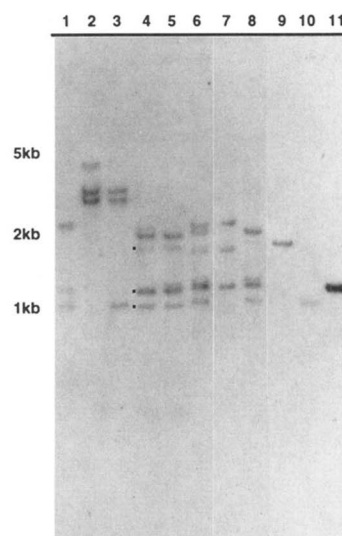


FIG. 1. Southern blot analysis of human genomic DNAs and cloned phage DNAs. Restricted DNAs were resolved by 0.8% agarose gel electrophoresis, blotted onto nylon membrane, and hybridized with a radiolabeled 283-bp probe derived from exon 1 of the cloned human K6b gene. The probe used did not contain the recognition motif for any of the enzymes used in these analyses, nor did it hybridize with the human K5 gene (not shown). The position of size markers in kilobase (kb) is indicated at left. Lanes 1–4, 10 μg of genomic DNA from a normal human subject were digested with the *HincII* (lane 1), *PstI* (lane 2), *PstI* and *SacI* (lane 3), *PstI* and *HincII* (lane 4). Lanes 5–8, 10 μg of genomic DNA from four randomly selected human subjects were digested as in lane 4. Lanes 9–11, 10 pg of the cloned K6a (lane 9), K6c (lane 10), and K6b (lane 11) phage DNAs were digested as in lanes 4–8. Between three and five hybridizing products were detected in the human genomic DNAs tested. After *PstI*-*HincII* double-digestion, fragments of corresponding sizes can be found in genomic DNAs (lanes 4–8) and in the cloned DNAs for K6b, K6c, and K6a (polymorphisms are obvious in this latter case). These fragments are identified by dots in lane 4.

and the sequences of the exon-intron boundaries conform to the consensus splicing signal (36). From restriction mapping and Southern blotting analyses of purified genomic DNAs, we deduced that the K6d and K6c genes exist in tandem with the same transcriptional orientation and separated by approximately 12 kbp. Likewise, we found that the K6-related gene K5/6- α (see below) is positioned immediately 3' downstream from the K6b gene (Fig. 2).

We estimated the size of all introns in the K6 isoform genes by a combination of sequencing and PCR amplification between neighboring exons in cloned DNAs, and found significant differences only for introns A and F (Fig. 2). The size of intron A, located between exon 1 and 2, is ~11 kbp in the K6c gene, while it is ~1.7 kbp in all other human K6 genes. Southern blot analysis of genomic DNA from randomly selected individuals with an intron A-specific probe confirmed the presence of this unusually long intron A in the general human population (data not shown). The size of intron F, located between exon 6 and 7, also varied among human K6 genes, and is ~0.6 kbp in K6d, ~0.8 kbp in K6b, and ~1.5 kbp in both K6a and K6c (Fig. 2). The size of introns B, C, D, E, G, and H appears very similar among K6 genes, and are of approximately 360, 90, 200, 500, 220, and 200 bp, respectively. These size differences in intron F were exploited in the context of a PCR assay to further confirm that the various K6 genes cloned from the human genomic DNA library used also occur in the general population. A set of oligonucleotide primers corresponding to segments of exon 6 (5' exon 6; Table I) and exon 7 (3' exon 7; Table I) that are perfectly conserved among K6 a, b, c, and d genes were used to amplify the entire intron F and flanking exon (~200 bp) sequences from the genomic DNA of several individuals, followed by electro-

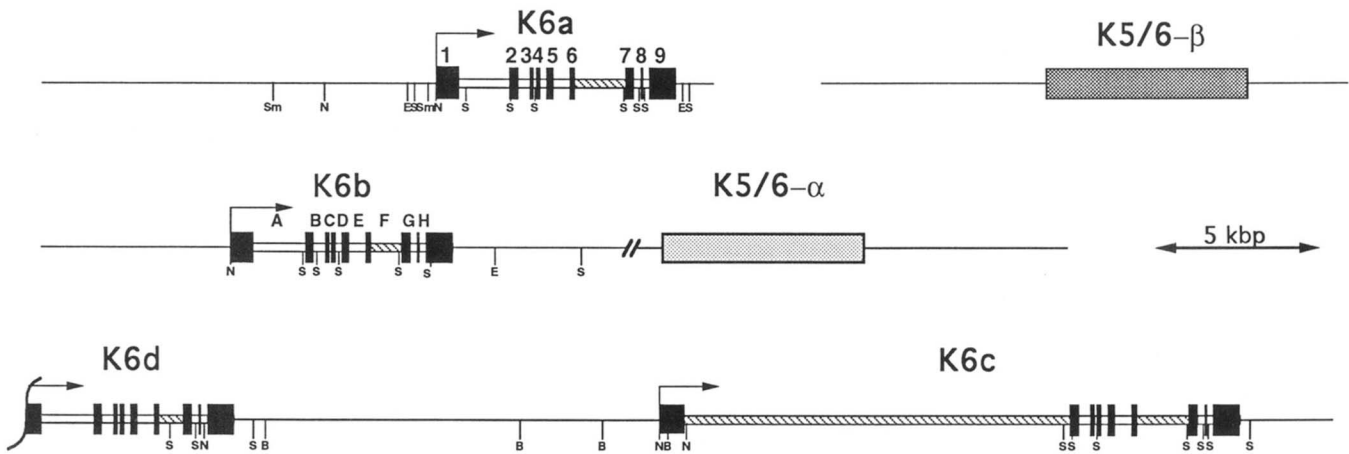


FIG. 2. Genomic organization of human keratin K6 and K6-related genes. The human K6a, K6b, K6c, and K6d genes contain 9 exons (black boxes, labeled 1–9 for K6a) separated by 8 introns (white boxes, labeled A–H for K6b). Intron F, whose length varies between 0.6 and 1.5 kb among K6 genes, and the unusually long intron A (11 kb) in the K6c gene are marked by diagonal lines. Direction of transcription in K6 genes is indicated by an arrow at the beginning of exon 1. Two genes with high sequence homology to both K5 and K6, K5/6- α and K5/6- β , have been only partially characterized. The location of the recognition sequence for some of the restriction enzymes used for the organization of phage library clones into subgroups is indicated for the K6 genes (B, *Bam*HI; E, *Eco*RI; N, *Nco*I; S, *Sac*I; Sm, *Sma*I; with the exception of *Sac*I, not all sites are depicted). Restriction mapping showed that the K6d and K6c genes, and the K6b and K5/6- α genes, are located in tandem in the same locus. A scale is given in kilobase pairs.

TABLE I
List of oligonucleotide primers and probes

Oligonucleotide probe/primer	Nucleotide sequence	Target sequence	Use
K6a	5'-GTGCTGGCCATGGTTCCAGGAGATGAGAGA-3'	+11 to -19	Primer extension of K6a mRNA
5' K6-primer	5'-CCGGAATTCAGGGGCAGTGGTGGCCTG-3'	+121 to +141	PCR amplification of K6 cDNAs
3' K6-primer	5'-CCGGAATTCGCTCCACCAGGTCCTGCAT-3'	+745 to +724	PCR amplification of K6 cDNAs
5' exon 6-primer	5'-CCTGCGCAACACCAAGCAGG-3'	+1119 to +1138	PCR amplification of K6 genes
3' exon 7-primer	5'-CCTCCAGCAGCTTGCCTGAG-3'	+1411 to +1392	PCR amplification of K6 genes
Probe A	5'-TGGATGGTGGGATCGATTTCG-3'	+470 to +450	Identify K6a
Probe B	5'-TGGATGGCGGGTCAATTTCG-3'	+470 to +450	Identify K6b, K6c, K6d, K6e and K6f
Probe C	5'-GCCGAGCTTATATG-3'	+173 to +187	Identify K6c and K6e
Probe D	5'-GCCGAGCTCTGTATG-3'	+173 to +187	Identify K6a, K6b, and K6f
Probe E	5'-TCTGGAACAAAGTG-3'	+570 to +584	Identify K6a and K6c
Probe F	5'-TCTGGACACCAAGTG-3'	+570 to +584	Identify K6b, K6d, K6e and K6f
Probe G	5'-ACGGGCGCGCTGGA-3'	+693 to +717	Identify K6a, K6c, K6d and K6e
Probe H	5'-CTGCTCTTGACGTGGTCTGA-3'	+1203 to +1184	Identify all K6 isoform genes

phoresis and Southern blotting with a coding sequence-specific oligonucleotide probe (probe H; Table I). Hybridization-positive products of 1.7 kb (K6a and c), 1.0 kb (K6b), and 0.8 kb (K6d) were detected within each individual genomic DNA tested (Fig. 3), further supporting the notion that the K6 genes discovered in the genomic DNA library are present in the general human population.

The nucleotide and predicted amino acid sequences of the K6a gene, whose mRNA predominates among K6 isoforms in human skin tissues and cell lines, are shown in Fig. 4. The relationship between gene structure and protein structure in K6a is identical to that of K6b and several other human type II keratin genes (21, 32, 35). Exon 1 (588 bp) encodes the entire 5' non-coding region and the protein-coding region extending from the amino-terminal head domain (160 amino acids) to the midpoint of the 1A segment in the α -helical (rod) domain. Exons 2–7 (215, 61, 96, 165, 126, and 221 bp, respectively) encode the bulk of the central rod domain sequence, consisting of four α -helical segments (1A, 41 amino acids; 1B, 101 amino acids; 2A, 40 amino acids; 2B, 99 amino acids) featuring heptad repeats of hydrophobic residues (37), separated by short non-helical linker segments (L1, 11 amino acids; L12, 14 amino acids; L2, 10 amino acids). The small exon 8 (39 bp) encodes the last two residues of the rod and the first 11 residues of the non-helical tail domain. The remainder of the tail domain (78 amino acids) and all of the 3' non-coding region are

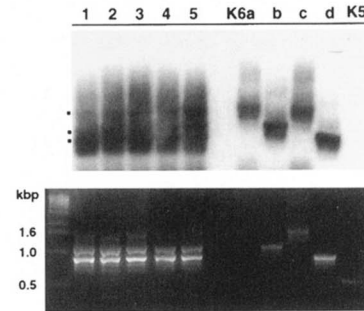


FIG. 3. PCR amplification experiment from human genomic DNAs and cloned phage DNAs. A set of universal K6 primers was used to amplify the intron F and flanking exon sequences from all known K6 genes in genomic DNA from randomly selected individuals (lanes 1–5) and from the cloned K6a, K6b, K6c, K6d, and K5 genes. PCR products were resolved by agarose gel electrophoresis (bottom panel, ethidium bromide stain), transferred onto nylon membrane, and probed with a radiolabeled (universal) K6 probe H (Table I) (upper panel). Positively hybridizing products of ~0.8 kb (K6d), ~1.0 kb (K6b), and ~1.7 kb (K6a and K6c) were detected in the five human genomic DNAs tested (marked by dots in upper panel). The corresponding amplified product from the human K5 gene (~0.5 kb) did not react with the K6 oligoprobe, indicating the specificity of the assay.

encoded by exon 9 (approximately 780 bp).

The nucleotide sequence of the coding region, 5'- and 3'-flanking sequences, and intron-exon junctions of the human

FIG. 4. **Nucleotide sequence of exons and flanking regions for the human K6a gene.** Sequences corresponding to the coding strand of exons 1-9 are shown in *upper case letters*, and the predicted amino acid encoded (one-letter code) is indicated directly *above*. Amino acid residues are numbered 1-564 starting with the initial methionine, and positions which differ among K6 isoforms are identified with *asterisks*. The transcription initiation site was mapped and defines the 5'-end of of exon 1 (+1). Sequences corresponding to the 5'-upstream region and to intron boundaries are shown as *lower case letters*. Intron sequences were not completely determined; the approximate size of each intron is indicated between parentheses. As the precise 3'-end of the K6a mRNA remains to be determined, the sequence beyond the canonical poly(A) signal sequence (*underlined* AATAAA) is shown as *upper case letters*. In the 5'-upstream sequence, the canonical TATA sequence is *underlined*, as are potential binding sites for known transcription factors.

TABLE II
Nucleotide sequence identity among human K6 isoform genes

Isoform compared	Segments compared				
	Exon 1-exon 9	5' Sequence ^a	Exon 1 only ^b	Exon 2-exon 8	Exon 9 only ^c
	%	%	%	%	%
K6a versus K6b	84.6	84.7	98.3	97.4	60.2
K6a versus K6c	97.9	84.7	97.6	100	99.9
K6a versus K6d ^d	98.1	ND ^d	97.4	97.7	99.9
K6b versus K6c	84.9	87.6	98.1	97.3	60.2
K6b versus K6d ^d	84.4	ND ^d	99.1	98.9	60.2
K6c versus K6d ^d	98.9	ND ^d	100	97.7	100

^a The 5' leader sequence extends from the TATA box to the translation initiation codon.

^b The exon 1 segment analyzed does not include the 5' leader (untranslated) sequence.

^c The exon 9 nucleotide sequence includes the 3' non-coding segment up the poly(A) (AATAAA) signal sequence.

^d The K6d genomic clone lacks 198 bp at the 5' end of protein-coding exon 1.

K6b, K6c, and K6d genes have also been determined. The gene and corresponding protein domain structure of the K6c and K6d isoforms are identical to those found in K6a (this study) and K6b (21) (data not shown). We calculated the nucleotide sequence identity between all pair combinations of human K6 genomic clones for the entire coding sequence (exon 1-exon 9) as well as specific segments of the coding and 5'-upstream sequences (Table II). This revealed remarkable trends in the conservation and divergence of particular exons among human K6 isoform genes. The K6a and K6c genes display a completely identical nucleotide sequence in the protein coding region covered by exons 2–9, while a single nucleotide difference was found in the 3' non-coding sequence of exon 9. In contrast, the coding segment of their exon 1 and especially their proximal 5'-upstream sequence are more different (Table II). The K6d gene, on the other hand, is identical to the K6c gene over exon 1 (our K6d genomic clone lacks 198 bp coding sequence at its 5' end) as well as exon 9, while it is 97.7% identical over exons 2–8. The K6b gene sequence is clearly different over all the segments analyzed, although its exon 1-exon 8 segment is related to K6d. These data have significant implications for the evolution of human K6 genes, as discussed below.

To examine the expression of these K6 isoform genes, we performed Northern analysis on RNA extracted from cultured epithelial cells. Hybridization of normal human epidermal keratinocyte RNA with probes derived from the 3' non-coding portion of the K6a or K6b genes each gave rise to a single band of ~2.3 kb (Fig. 5A). The signal obtained using the K6a 3' non-coding probe is a mixed one with contributions by the K6a, K6c, and K6d mRNAs. Identical size mRNAs were detected when these probes were used on RNA prepared from SCC-4, a tongue squamous cell carcinoma line (Fig. 5A). These results indicate that conventional Northern analysis allows only partial discrimination among K6 isoform mRNAs.

Primer-extension analysis was carried out to define the 5' termini of the K6a mRNA. The 5' termini of the K6b, K6c, and K6d mRNAs could not be mapped using primer extension, either because the genomic clone lacked the required 5' end sequence (K6d) or the existence of additional isoform-encoding mRNAs with highly related sequences (see cDNA cloning section below). In contrast, the 3'-nucleotide of the K6a primer used for this analysis (Table I) was completely specific for the K6a isoform sequence. One major primer-extended product was seen for K6a in total RNA extracted from human epidermal cells in primary culture (Fig. 5B). On the basis of the gene sequence data, the 5' terminus of the K6a mRNA was assigned at 48 bp upstream from the translation initiation site. Inspection of the 5' sequence in K6a revealed the presence of a "TATAA" motif located 50 bp upstream from the putative transcription initiation site (Fig. 4). Also, the nucleotide sequence surrounding the ATG start codon in the K6a gene agrees well with the consensus sequence (38). Based on this and our North-

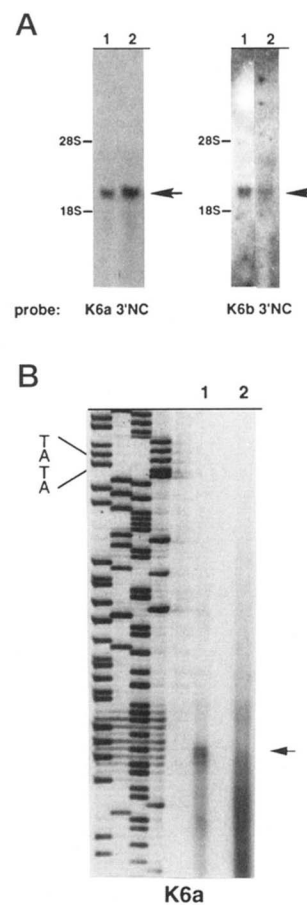


FIG. 5. RNA blot-hybridization and primer-extension analysis of K6 isoform mRNAs. A, total RNAs (5 μ g) extracted from primary cultured human skin keratinocytes (lane 1) and the tongue squamous cell carcinoma line SCC-4 (lane 2) were electrophoresed, transferred onto nylon membrane, and probed with genomic fragments corresponding to the 3' non-coding sequence of the human K6a (K6a 3'NC) and K6b (K6b 3'NC) genes. Migration of ribosomal RNA (28 S, 18 S) is shown at left. Both these probes detect a single band of ~2.3 kb (arrow) in both RNA extracts. B, for primer extension analysis, a radiolabeled oligonucleotide complementary to the 5' portion and specific for the K6a gene was hybridized to total RNA extracted from skin keratinocytes (30 μ g, lane 1) and reverse-transcribed. The same oligonucleotide primer was used for the sequencing of the corresponding genomic DNA by the chain termination method and the four reactions (loaded as A, C, G, T from left to right) were ran in parallel with primer-extended product(s). Yeast tRNA was used as a negative control (lane 2). A major primer extended product was present only in lane 1, as indicated by the arrow. The position of the TATA box is indicated at left.

ern blot data, we estimate the length of the 3'-untranslated region of the K6a mRNA to be ~500–600 nucleotides. Consistent with this, there is a single potential polyadenylation/processing signal, AAUAAA (39), located ~520–530 nucleotides down-

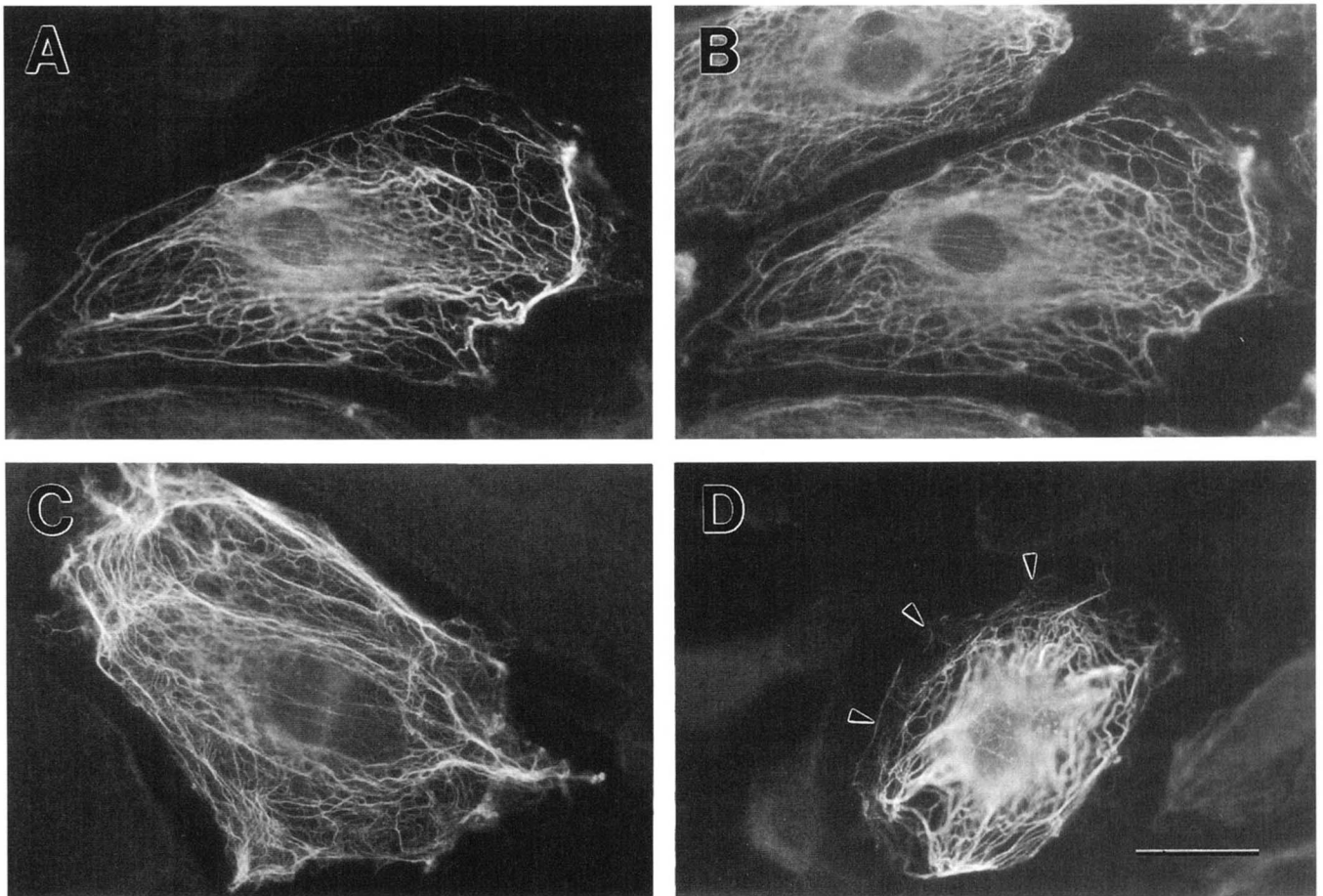


FIG. 6. Expression of full-length human K6 genes in cultured epithelial cells. The genomic inserts of K6a (8 kbp), K6b (18 kbp), and K6c (20 kbp) were subcloned in a cytomegalovirus promoter-based expression vector and transiently transfected into the kidney epithelial cell line PtK2. At 72 h post-transfection, cells were fixed and double-labeled with a rabbit anti-K6 antiserum followed by fluorescein isothiocyanate-conjugated goat anti-rabbit IgG, and a mouse anti-K8-K18 antibody followed by a biotin-conjugated goat anti-mouse IgG and streptavidin-Texas Red. **Frames A and B**, double-immunofluorescence labeling of a human K6a-transfected cell, with the anti-K6 signal shown in **A** and the anti-K8-K18 signal shown in **B**. The signals for both K6a as well as the endogenous PtK2 K8/K18 filament network co-localize perfectly at this level of resolution. Similar results were obtained when the K6b (**frame C**) and K6c (**frame D**) genes were transfected (only the anti-K6 stainings are shown). For all three genes, a subset of transfected cells showed a slightly altered organization of their endogenous filament network (see **arrowheads** in **D**). **Bar** = 25 μ m.

stream from the translation stop codon in the K6a gene (Fig. 4).

We analyzed the proximal ~200 bp of 5'-upstream sequence of the K6a gene for the presence of potential binding sites for transcription factors involved in keratin gene expression or in the response to wounding (Fig. 4). Upstream from the TATA box, we found two potential binding sites for AP2, a transcription factor of neural and epidermal lineages playing a crucial role in the regulation of epidermal keratin gene expression (40). A site known to be involved in the up-regulation of the K6b gene upon epidermal growth factor treatment of cultured human epidermal keratinocytes (41) is conserved in sequence and location in the K6a gene upstream sequence. There are several additional potential regulatory sequence elements in the 5'-flanking region analyzed, including AP-1 (response to phorbol esters, cAMP, transforming growth factor- β , retinoic acid, interleukin-2; Ref. 42), PEA3 (response to epidermal growth factor, phorbol esters, serum; Ref. 42), GAS (response to interferon- γ ; Ref. 42), and NF-1 (response to transforming growth- β ; Ref. 42). These elements are of interest in the context of a gene whose expression is induced by wounding. Additional studies will be required to determine whether these sequence elements and their cognate transcription factors play a role in the regulation of K6a gene expression.

The K6a, K6b, and K6c Genomic Clones Give Rise to a K6 Protein When Expressed in Cultured Cells—The genomic in-

serts of K6a, K6b, and K6c were subcloned in a cytomegalovirus promoter-based expression vector and transiently transfected into PtK2 cells (K6-, K7+, K8+, K18+, K19+, vimentin+; Ref. 24). At 72 h post-transfection, cells were fixed and processed for immunofluorescence microscopy (30). Mock-transfected PtK2 cells showed no reactivity with the anti-K6 antiserum (not shown). In contrast, each of these clones gave rise to an antigen immunoreactive with the anti-human K6 antiserum in transfected cells. In a majority of cells, the signal was filamentous and co-localized with that for the endogenous K8-K18 filaments (Fig. 6, A, B, and D). In each case, however, a subset of cells featured an abnormal keratin filament network (note the retracted IF network in the K6c-transfected cell in Fig. 6D). Additional studies will be required to characterize this phenomenon, and compare the assembly properties of these human K6 isoforms. These data further support the notion that the K6a, K6b, and K6c genomic clones contain functional genes.

Isolation of Other K6-related Genomic Clones—We isolated an additional phage clone containing a partial K6-encoding gene, whose sequence extended from the 5'-upstream region to intron C. Over exons 1–3, the sequence of this clone is identical to that of K6b; however, the intron B nucleotide sequence slightly differs between the two clones (data not shown). In the absence of a full-length clone for this gene, we cannot deter-

TABLE III
Expression of human K6 isoforms

	No. of clones examined ^a	Isoform profile ^b					
		K6a	K6b	K6c	K6d	K6e	K6f
Skin cDNA library ^c	64	77%	13%	0%	0%	5%	5%
Tissue samples ^d		K6a	K6b + K6f	K6c	K6d + K6e		
Scalp skin	155	66%	31%	0%	3%		
Sole skin	86	84%	14%	1%	1%		
SCC skin	115	69%	27%	1%	3%		
Cultured samples ^e							
Foreskin	88	99%	1%	0%	0%		
NHSK	51	80%	14%	2%	4%		
SCC-13	28	96%	0%	4%	0%		
SCC-9	53	83%	7%	4%	6%		

^a This column refers to the total number of independent K6 cDNA clones examined in each sample.

^b Different experimental strategies were used to discriminate among K6 isoforms in the cDNA library and in tissue and cell samples. See text for details.

^c The skin cDNA library was prepared using poly(A) mRNA extracted from a squamous cell carcinoma of the leg.

^d Human tissue samples were obtained as discarded materials in the course of surgery, and were immediately frozen and stored at -80°C until use.

^e NHSK refers to primary cultured normal human skin keratinocytes. SCC-13 and SCC-9 are cell lines derived from squamous cell carcinomas of the skin and tongue, respectively.

mine with certainty whether it corresponds to a distinct K6 isoform, a different allele of the K6b isoform, or a pseudogene derived from the functional K6b gene. A partial sequence analysis of the two K6-related clones K5/6- α and K5/6- β obtained in our genomic library screen (Fig. 2) revealed that if functional, these genes would encode novel type II keratin-like sequences. The nucleotide sequence of the K5/6- α clone shows $\sim 60\%$ homology with the head domain of K6, while the K5/6- β clone shows $\sim 45\%$ homology with that of K6 (data not shown). In addition to an exon 1 probe, these two clones hybridize with several other exon probes derived from the K6a cDNA under stringent washing conditions. Additional analyses will be necessary to determine the identity and functional status of these K6-related genes.

cDNA Cloning and Comparison of Predicted Amino Acid Sequences of Human K6 Isoforms—We screened a human skin cDNA library constructed from a patient with squamous cell carcinoma of the lower leg, utilizing an exon 1 probe derived from the K6a genomic clone. Positive clones were grouped on the basis of hybridization with the 3' non-coding probes used in the Northern blot analysis, and analyzed by hybridization with oligonucleotide probes specific for either a single or a subset of K6 isoforms (Table I) under optimized stringency conditions (see "Experimental Procedures"). 64 independent K6 cDNA clones were isolated and analyzed by oligonucleotide hybridization and DNA sequencing. Among these, we discovered several clones with cDNA inserts corresponding to either the K6a or K6b genes. In each case, one clone was selected and its complete sequence determined. We did not find cDNA clones corresponding to either the K6c or K6d gene in this particular library (Table III). On the other hand, we found six clones (out of 64) whose insert sequence did not exactly correspond to any of the four K6 genomic clones isolated and characterized. These six clones could be clearly partitioned into two distinct groups by sequencing, and the nucleotide sequences were identical within each subgroup. At the nucleotide sequence level these two novel cDNAs, designated as K6e and K6f, were, respectively, ~ 96 – 98 and ~ 97 – 98% identical to the other K6 isoforms characterized in our genomic cloning effort.

The amino acid sequence predicted from the K6a cDNA clone is in complete agreement with that predicted from the corresponding genomic clone shown in Fig. 4 (data not shown). While the amino acid sequence predicted from both our K6b genomic and cDNA clones are in complete agreement (data not shown), they differ slightly from the one previously reported

(see Ref. 32).² Likewise, the protein sequences predicted from the K6c (complete) and K6d (starting in the middle of exon 1) genomic clones, and the K6e and K6f cDNA clones, were determined (data not shown). The K6a, K6b, K6c, K6e, and K6f isoforms are thus all predicted to consist of 564 amino acids, with calculated molecular weights (M_r) of 60,042, 59,996, 60,183, 60,220, and 60,064, respectively. The virtually identical predicted M_r values for the human K6 isoforms provide a good explanation for the inability to discriminate among them using regular polyacrylamide gel electrophoresis (16). These calculated values are slightly larger than the experimentally measured one, 56 kDa (1; this has been repeatedly described for human keratins; e.g. Refs. 32, 34, and 35). Each K6 isoform protein is predicted to show at least 97.6% identity to other isoforms, and substitutions occur at a total of 16 positions among K6 a-f. The amino acid residues predicted to occur at these variable positions in each of the six human K6 isoforms, and in the corresponding position of a mouse K6 (43) and the human K5 (32) sequences, are compared on Fig. 7. Substitutions are relatively more concentrated in the non-helical head and tail domains, with each containing 5 variable positions, compared to the substantially longer rod domain, which shows only 6 variable positions (Fig. 7). Within the rod, the substitutions only affect residues located in the α -helical segments. Remarkably, a maximum of two different amino acid residues may occur at each of the 16 variable positions among human K6 isoforms, and in fact, many of these substitutions are very conservative. At several of these positions, interestingly, the corresponding amino acid in the human K5 sequence is predicted to be identical to a subset of K6 isoforms. The two human K6 isoforms predicted to differ the most are K6a and K6e, with 13 amino acid substitutions. The mouse K6 amino acid sequence (43; see Ref. 6) shows ~ 80 – 85% sequence identity with human sequences. However, among the 15 "variable" amino

² The differences between the human K6b amino acid sequence predicted from the published genomic clone (see Refs. 21 and 32) and that predicted from the genomic and cDNA clones characterized in this study are as follows: A⁸⁹G⁹⁰→G⁸⁹S⁹⁰; C¹⁰⁷G^{→107}; P¹¹⁶A¹¹⁷L¹¹⁸L¹¹⁹C¹²⁰; A^{→116}G¹¹⁷L¹¹⁸A¹¹⁹G¹²⁰G¹²¹; A³⁵³S^{→354}; Y⁴⁰¹C^{→402}. Note that the amino acids predicted to occur at these positions are identical in all K6 isoform genes and cDNAs cloned in this study. Note also that at codon positions 89, 107, 116, 117, 120, 121, 354, and 402 (i.e. 8 out of 10), these differences make our K6b clone identical to human K5 (32) and all other human K6 isoforms. Additional studies will be required to ascertain the reason(s) for these differences.

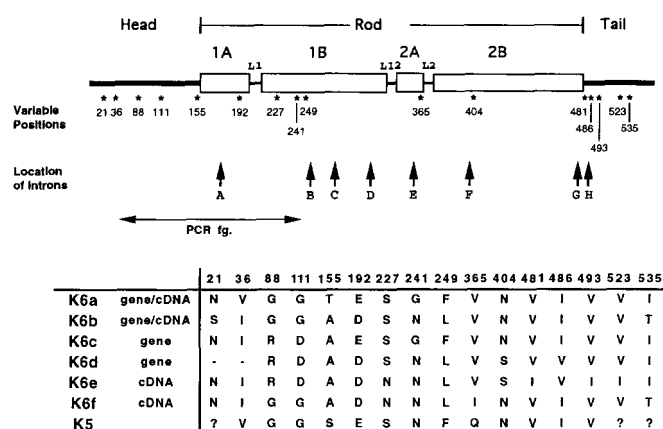


FIG. 7. Schematic representation of the secondary structure and variable amino acid positions among human K6 isoforms. The tripartite domain organization typical of cytoplasmic IF proteins and shared by human K6 isoforms is illustrated. A central domain (rod) contains four sequence segments featuring a heptad repeat of hydrophobic residues and predicted to be α -helical (coils 1A, 1B, 2A, and 2B). This repeat is interrupted at three conserved locations within the rod, generating the non-helical linker segments L1, L12, and L2. The central rod domain is flanked by head and tail domains, two nonhelical sequences located at the amino- and carboxyl-terminal portion of the protein, respectively. The location of introns A–H along the protein coding sequence, and of the PCR fragment used for the determination of the K6 mRNA profile in human samples, are depicted. Amino acid differences occur at a total of 16 positions among human K6 isoforms: their location is depicted with an asterisk along with the amino acid residue number (see Fig. 4) under the protein structure representation. The identity of the amino acid residues occurring at each of these variable positions in the human K6 isoforms a–f, and in the previously characterized mouse K6 (mK6; Ref. 43) and human K5 (32) are listed below. In a few cases, it was not possible to determine with certainty the identity of the amino acid in the mK6 and K5 sequences that corresponded to the same position in human K6 isoform sequences (these are identified with a “?”). The source of these sequences (gene, cDNA, or both) is also indicated.

acid positions among human K6 isoforms that could be directly aligned with the mouse K6 sequence, 11 are identical to K6a, including Thr¹⁵⁵ (Fig. 7).

Expression of K6 Isoforms in Selected Human Epithelial Tissues and Cell Lines—To investigate whether the human K6 genes are differentially regulated, particularly in skin, we determined the K6 isoform profile in a variety of human skin tissue samples and cell lines. First, a combination of colony hybridization assay (see “Experimental Procedures”) and DNA sequencing was applied to the 64 independent clones isolated during our cDNA library screening effort. Given that the cloning strategy was based upon hybridization with an exon 1 probe, we had access to near full-length coding sequences to ascertain the identity of the K6 isoform encoded. The K6a isoform was clearly the dominant species in this cDNA library, constituting 77% (49 out of 64) of the isolated K6 clones (Table III). The K6b, K6e, and K6f isoforms, respectively, accounted for 13, 5, and 5% of the K6 clones. As mentioned above, no clone corresponding to either the K6c or K6d isoform was found in this particular library (Table III). Note that the surgically excised squamous cell carcinoma of the skin used to construct the cDNA library included a significant amount of surrounding normal tissue.

Second, expression of K6 isoform mRNAs was analyzed in human foreskin, scalp skin, sole skin, and a squamous cell carcinoma, in cultured primary human skin keratinocytes, and in human cancer cell lines known to express K6, such as SCC-13 and SCC-9. The foreskin sample obtained had been incubated in cultured media, so that K6 induction likely had occurred by the time mRNA was extracted (16). Reverse-trans-

cribed cDNAs were subjected to PCR amplification with two “universal” K6 oligonucleotide primers, generating a 626-bp fragment covering parts of exons 1 and 2 (Fig. 7). After subcloning, independent bacterial clones were spotted on duplicate filters and subjected to colony hybridization as described under “Experimental Procedures.” The oligonucleotide probes used (Table I) could not unequivocally “resolve” K6b from K6f as well as K6d from K6e, partly because only one-third of the coding sequence was available to discriminate among K6 isoforms. The K6a and K6c isoforms, on the other hand, could be discriminated in this assay. The K6a isoform is the dominant K6 mRNA in all human tissues and cultured cell lines tested (see Ref. 21), although the extent of its expression varied appreciably among the samples tested (Table III). In normal human scalp skin, where K6 is expressed in hair follicles (14) but not in epidermis, 66% of K6 mRNAs encode K6a, while 31% encode either K6b or K6f. A similar partitioning of K6 mRNAs occurs in a well differentiated squamous cell carcinoma of skin (Table III). In sole skin, where K6 is constitutively expressed in the differentiating suprabasal layers of epidermis (1, 12), K6a constitutes 84% of K6 mRNAs, while that of K6b+K6f is lower (14%) compared to scalp skin. In all samples subjected to culture *in vitro*, the K6a mRNA is even more dominant, with proportions ranging from 80% to as much as 99%. This is especially true in cultured epidermal keratinocytes, *e.g.* the foreskin explant and SCC-13 cells (Table III). While the K6c and K6d+K6e mRNAs are present in the majority of samples tested, they constitute a minor fraction of the total K6 mRNA pool (Table III). Collectively these data provide direct evidence that the K6a and K6c genes cloned in this study are functional, and that the human genome is likely to contain additional K6 genes, as per our discovery of the K6e and K6f sequences in a skin cDNA library and in several skin tissue and cell samples. They also denote considerable heterogeneity in the expression of K6 genes among the tissues surveyed, although in all cases K6a is the most abundantly expressed isoform at the mRNA level.

DISCUSSION

Discovery of Multiple Genes and cDNAs Encoding Highly Related K6 Isoforms in the Human Genome—We cloned and characterized multiple human type II keratin genes and cDNAs predicted to encode highly related K6 isoform proteins. In keeping with the previous human K6 cloning efforts (16) and the keratin nomenclature (1), we designated the genes cloned in this study as K6a, K6b, K6c, and K6d. That these genomic clones represent separate K6 genes is supported by several lines of evidence. First, the K6b, K6c, and K6d genes map to distinct genomic loci. Second, the K6a, K6b, K6c, and K6d genes each display a unique genomic structure, with introns A and F showing varying lengths. We exploited the variations in intron F length to show that the K6a/K6c, K6b, and K6d genes are present in the general population. Third, while the coding and in some cases, the 3′ non-coding sequences are very similar among human K6 isoforms, significant differences occur in their proximal 5′-upstream sequences. Finally, we showed that the K6a, K6b, and K6c genes are differentially regulated in various skin samples and cell lines (we did not obtain direct evidence of expression for the K6d genomic clone). The four human K6 isoform genes cloned show a structure similar to several other human type II keratin genes (21, 32, 33, 35). The K6a, K6b, and K6d genes are each contained within 7 kb, and thus are typical of human type II keratin genes. In contrast, the human K6c gene appears unique as it extends over 17 kb. The large size of this gene is due exclusively to the presence of an unusually large intron A (~11 kb). Only one other human type II keratin gene, K7, is nearly as long at ~15 kb long (35). We

further demonstrate that the K6c and K6d genes, and the K6b and K5/6 α ones, are located in tandem in the human genome. Applying probes derived from 3' non-coding sequences to a library of somatic cell hybrid DNAs, Rosenberg *et al.* (44) previously localized the human K6b gene and the gene encoding the human K6a cDNA to chromosome 12. Given that the K6a probe they used reacts with the 3' non-coding region of the K6a, K6c, and K6d genes, this implies that these three genes, along with the K6b and K5/6 α ones, reside on human chromosome 12. This has obvious implication for the search for K6 mutations in the human population, as discussed below.

Clones corresponding to the mRNAs encoded by the K6a and K6b genes were easily retrieved in a cDNA library prepared from a squamous cell carcinoma of skin. Among a population of 64 K6 cDNA clones, however, none corresponded to the coding sequences for the K6c and K6d genes. The significance of this absence is doubtful, as only one among 115 partial K6 cDNA clones found by PCR in another skin squamous cell carcinoma corresponded to K6c (a similar argument applies for K6d; Table III). On the other hand, we discovered two novel K6-encoding sequences in this skin cDNA library. The extent of nucleotide sequence divergence between these two clones and other K6 isoforms over their coding segment is ~2–3%, exceeding the known frequency of polymorphism in the human genome (0.2%; see Ref. 45). It appears likely that the cDNA sequence designated as K6e is the product of a yet another K6 gene. Compared to K6a and K6b, K6e shows a unique 3' non-coding sequence as well as nucleotide differences leading to several amino acid substitutions (>10; Fig. 7). In contrast, the other cDNA clone, K6f, shows only three differences with K6b at the protein sequence level, and the high homology with the K6b sequence extends to the 3' non-coding sequence (not shown). At present, therefore, we cannot be certain that the K6f isoform is the product of a distinct gene. It is likely that the genes coding for the K6e and K6f mRNAs were among the population of 57 genomic clones isolated, but were missed because they share a similar restriction map with one of the representative clones selected for complete characterization. Further studies will be required to address this possibility and identify the genes encoding these K6 cDNAs.

The nucleotide and amino acid sequence identity among human K6 isoforms is remarkably high, and certainly accounts for the underestimation of the multiplicity of K6 sequences in previous efforts. The occurrence of so many isoform-encoding genes is unparalleled among keratins and IF proteins in general, and is not restricted to the human genome. Two K6 isoform-encoding genes have been found so far in the mouse genome (20), while three of them have been documented in the bovine genome (19)³ The other known instance of human keratin multiplicity involves K2, a type II keratin, although there are important differences with K6. While K2e and K2p cannot be resolved by standard protein electrophoresis techniques, they show different sizes for the non-helical head and tail domains and share only 72% amino acid sequence identity in the rod domain (46). The two genes are also differentially expressed, with K2e found in epidermis and K2p found predominantly in the palate and other oral epithelia. The two human K2 sequences thus probably represent distinct keratins which originally received a common designation because of their related size and charge properties (46).

Evolution of Human K6 Genes—The K5 and K6 genes probably arose from a common ancestral gene by gene duplication (47). Indeed, not only do the K5 and K6 genes show the highest sequence homology among type II keratins (32), they are also

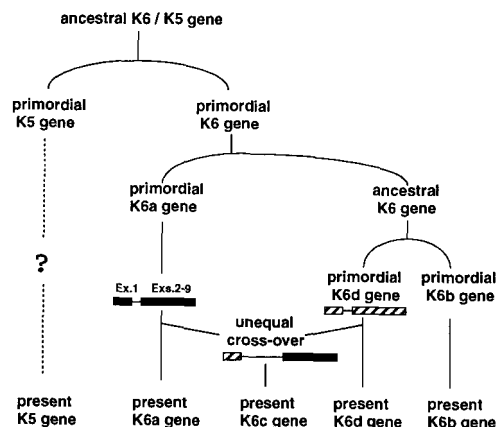


FIG. 8. A model for the evolution of human K6 isoform-encoding genes. This model is based upon the comparison of genomic structure and nucleotide sequences among the K6 genes cloned in this study. The K6 and K5 genes are thought to have arisen from a common ancestral gene. Subsequently, sequential gene duplication events resulted in the generation of at least three distinct primordial K6 genes, corresponding to the current K6a, K6b, and K6d genes. More recently, an unequal crossing-over event took place at the level of intron A between the postulated primordial K6a and K6d genes, generating a hybrid gene designated as K6c. See "Discussion" for additional details.

both expressed in stratified epithelia. We propose that, subsequent to the event that gave rise to the primordial K6 gene, a series of duplication events generated the multiple K6 isoform genes. This notion is supported by the perfect conservation of intron positions among K6 isoform-encoding genes, together with their highly homologous nucleotide sequences.

Our model of K6 gene evolution, presented in Fig. 8, is based on the assumption that the first duplication of the primordial K6 gene gave rise to the primordial K6a gene and the ancestor of the K6b and K6d genes. There are several lines of evidence supporting the notion that K6a is the most ancient among K6 isoforms genes. The nucleotide sequence of K6a is the one most related to K5, and furthermore, the sequences of K6b and K6d are clearly more related to one another than to K6a, suggesting that the latter evolved independently. The notion that the current K6b and K6d genes evolved from a common ancestor gene is favored by the very high nucleotide identity between their exons 2–8, and the relatedness of their intron F. After a postulated duplication event that led to the creation of the primordial K6b and K6d genes, exon 9 of the K6b gene diverged considerably, and became unique. Subsequent to the appearance of the primordial K6b and K6d genes, we propose that an unequal crossing-over event took place at the level of the intron A between K6a and K6d. The result was a novel hybrid gene featuring the 5' end (exon 1) of the ancestral K6d gene and the 3' end (exons 2–9) of the K6a gene (Table II), corresponding to the K6c gene discovered in this study. The extensive homology and the close physical proximity of the K6 genes is expected to predispose this locus to unequal recombination events, as suggested for the human visual pigment genes (48). An unequal crossing-over mechanism is further supported by the fact that the 5' end of intron A sequence is highly similar in K6c and K6d, while its 3' end is highly similar in K6c and K6a (also, introns B–H sequences are highly similar in K6a and K6c). This mechanism could also account for the unusually long size of intron A in the K6c gene (Fig. 8). We further surmise that the K6c gene appeared recently during human genome evolution, since its nucleotide sequence is perfectly conserved with distinct parts of the K6a and K6d genes. The human K6 genes are thus probably still evolving, and it is possible that some of the poorly expressed K6 genes, such as K6c and K6d (assuming that their expression in skin reflects that in other epithelia

³ Dr. Jose Jorcano, personal communication.

tissues), are destined to be inactivated in the future.

Functional Significance of the Existence of Multiple Human K6 Genes and Relevance to Disease—Understanding the regulation and function of the K6 genes is of considerable biological and clinical interest, as they are expressed postmitotically in stratified epithelia showing an enhanced turnover rate. A physiological example of increased mitotic activity is epithelial wound healing. Injury to human epidermis triggers an induction of K6 and K16 expression at the wound edge within hours (16, 49, 50), preceding the onset of enhanced mitotic activity, which occurs at ~24 h (50, 51). As regeneration is completed, the K6 and K16 proteins disappear from epidermal tissue (49). Similar events take place in injured mouse skin.⁴ Induction of K6-K16 is therefore part of the early transcriptional response of epidermis to injury, although the mechanisms involved and their contribution to regeneration remain unknown. On the other hand, K6 expression appears "constitutive" in skin disorders featuring epidermal hyperproliferation, such as psoriasis, viral infections, and various carcinomas (15, 17). In those disorders as in regenerating epidermis, abundant expression of K6 and K16 can be associated with abnormal terminal differentiation (see Ref. 52). Clearly, an understanding of the regulation of the human K6 genes and of the assembly properties of their products will provide significant insights into the biology of wound healing and hyperproliferative disorders. The isolation of the human K6a gene and cDNA provides us with useful tools to examine these important issues.

Amino acid substitutions are predicted to occur at 16 positions among the human K6 isoforms. Whether any of them modifies the assembly properties or regulation of the K6 isoform(s) concerned is not known. Functional differences could occur at three distinct levels: (i) structure and interaction between 10-nm filaments; (ii) regulation of the assembly/disassembly processes; and (iii) interaction with associated proteins. Amino acid substitutions with potential structural significance occur at positions 88 and 111 of the head domain: in K6c, K6d, and K6e, these codons encode oppositely charged residues, *i.e.* Arg⁸⁸ and Asp¹¹¹, while in the other isoforms and in K5, they encode nonpolar Gly residues (Fig. 7). This region of the head domain otherwise contains few charged residues, and is rich in glycine and hydrophobic amino acids (Fig. 4). K6 is phosphorylated *in vivo* (53, 54), and substitutions in specific K6 isoforms could alter these modifications. Substitutions at positions 21 and 155 in the head domain, and at position 535 in the tail, involve the appearance of either a Ser or a Thr residue in specific isoforms (Fig. 7). In the nonhelical end domains of many IF proteins including keratins, specific Ser and Thr residues are phosphorylated (55) or O-glycosylated (56) in a cell cycle-dependent fashion. In lamins, vimentin, and K18, phosphorylation of specific serine residues located in the end domains mediates a major reorganization of filaments during and after mitosis (55, 56). Obviously, experimental testing will be required to establish whether the various human K6 isoforms are differentially phosphorylated *in vivo*.

Mutations in many of the keratin genes constitutively expressed in skin underlie several dominantly inherited genetic skin diseases that share trauma-induced blistering of the skin as their predominant clinical manifestation (reviewed in Refs. 2, 3, and 8–10). These diseases include epidermolysis bullosa simplex (involving mutations in K5 or K14; Refs. 8–10), epidermolytic hyperkeratosis (K1 or K10 mutations; Refs. 8–10), ichthyosis bullosa of Siemens (K2e mutations; Refs. 8 and 9), the epidermolytic (K9 mutations; Refs. 8 and 9), and non-epidermolytic (K1 mutations; see Ref. 57) variants of palmo-

plantar keratoderma, and more recently, pachyonychia congenita (K16 or K17 mutations; see Ref. 58). The spectrum of clinical manifestations associated with these diseases depends on several factors, including the pattern of expression of the mutated keratin gene as it relates to the anatomical site(s) subjected to trauma, the position of the amino acid affected along the keratin polypeptide chain, and the nature of the substitution (3, 8). In addition to demonstrating a crucial role for keratin IFs in maintaining the integrity of the epidermis, these findings raise the important question of whether other keratin genes are mutated in the human population, and if so, which disease(s) result from such mutations.

Mutations are thus likely to occur in the human K6 isoforms, and could lead to a perturbation of normal skin physiology and resistance to mechanical stress in tissues such as palmar and plantar epidermis, the outer root sheath or hair follicles, sebaceous and sweat glands, nails, oral mucosa, tongue, and perhaps even in regenerating epidermis. In addition to the determining factors cited above, however, the clinical manifestations associated with the a K6 mutation would depend on the stoichiometry between the various isoforms expressed in the affected epithelial tissue(s). Quite possibly, previously discovered point mutations in the human K5, K1, and K2e genes affecting residues that are well conserved among type II keratins (*e.g.* Ref. 8) may not generate a clinically visible phenotype if they affect a minor K6 isoform. At present, on the basis of their co-regulation with the K16 and K17 genes, it appears likely that mutations in a K6 gene account for a subset of pachyonychia congenita disease-causing mutations (see Ref. 58). Future efforts should reveal the identity of the diseases caused by function-disrupting mutations in K6 sequences.

Acknowledgments—We are very grateful to Dr. Elaine Fuchs (HHMI, University of Chicago) for making available the human K6b genomic and partial K6a cDNA clones, the anti-K6 antiserum, and SCC-13 cells used in this study. We thank Dr. Lorne Taichman (SUNY, Stony Brook, NY) for his gift of SCC-4 and SCC-9 cells. We thank Drs. Jeremy Nathans, Carole Parent, and Ormond MacDougall (Johns Hopkins University School of Medicine) for their comments.

REFERENCES

- Moll, R., Franke, W. W., Schiller, D. L., Geiger, B., and Krepler, R. (1982) *Cell* **31**, 11–24
- Fuchs, E., and Weber, K. (1994) *Annu. Rev. Biochem.* **63**, 345–382
- Coulombe, P. A. (1993) *Curr. Opin. Cell Biol.* **5**, 17–29
- O'Guin, W. M., Schermer, A., Lynch, M., and Sun, T.-T. (1990) in *Cellular and Molecular Biology of Intermediate Filaments* (Goldman, R. D., and Steinert, P. M., eds) pp. 301–334, Plenum Publishing Corp., New York
- Fuchs, E. (1993) *J. Cell Sci. Suppl.* **17**, 197–208
- Schweizer, J. (1993) *Molecular Biology of the Skin: The Keratinocyte*, pp. 33–78, Academic Press, New York
- Schermer, A., Jester, J. V., Hardy, C., Milano, D., and Sun, T. T. (1989) *Differentiation* **42**, 103–110
- Fuchs, E., Coulombe, P. A., Cheng, J., Chan, T.-M., Hutton, E., Syder, A., Degenstein, L., Yu, Q.-C., Letai, A., and Vassar, R. (1994) *J. Invest. Dermatol.* **103**, 25S–30S
- Lane, E. B. (1994) *Curr. Opin. Genet. Dev.* **4**, 412–418
- Steinert, P. M., and Bale, S. J. (1993) *Trends Genet.* **9**, 280–284
- Moll, R., Moll, I., and Wiest, W. (1982) *Differentiation* **23**, 170–178
- Moll, R., Krepler, R., and Franke, W. W. (1983) *Differentiation* **23**, 256–269
- Ouhayoun, J., Gosselin, F., Forest, N., Winter, S., and Franke, W. W. (1985) *Differentiation* **30**, 123–129
- Stark, H. J., Breikreutz, D., Limat, A., Bowden, P., and Fusenig, N. E. (1987) *Differentiation* **35**, 236–248
- Weiss, R. A., Eichner, R., and Sun, T. T. (1984) *J. Cell Biol.* **98**, 1397–1406
- Tyner, A. L., and Fuchs, E. (1986) *J. Cell Biol.* **103**, 1945–1955
- Stoler, A. L., Kopan, R., Duvic, M., and Fuchs, E. (1988) *J. Cell Biol.* **107**, 427–446
- Kopan, R., and Fuchs, E. (1989) *J. Cell Biol.* **109**, 295–307
- Blessing, M., Zentgraf, H., and Jorcano, J. L. (1987) *EMBO J.* **6**, 567–575
- Rothnagel, J. A., Seki, T., Longley, M. A., Holder, R., Bundman, D., Adachi, K., and Roop, D. R. (1993) *J. Invest. Dermatol.* **101**, 563 (Abstr. 235)
- Tyner, A. L., Eichman, M. J., and Fuchs, E. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 4683–4687
- Hanukoglu, I., and Fuchs, E. (1983) *Cell* **33**, 915–924
- Wu, Y. J., Parker, L. M., Binder, N. E., Beckett, M. A., Sinard, J. H., Griffiths, C. T., and Rheinwald, J. G. (1982) *Cell* **31**, 693–703
- Franke, W. W., Schmid, E., Osborn, M., and Weber, K. (1978) *Cytobiologie* **17**, 391–411
- Takahashi, K., Tanaka, A., Hara, M., and Nakanishi, S. (1992) *Eur. J.*

⁴ K. Takahashi and P. A. Coulombe, unpublished data.

- Biochem.* **204**, 1025–1033
26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
 27. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
 28. Bessho, Y., Nakanishi, S., and Nawa, H. (1993) *Brain Res. Mol. Brain Res.* **18**, 201–208
 29. Compton, D. A., Cleveland, D. W. (1993) *J. Cell Biol.* **120**, 947–957
 30. Letai, A., Coulombe, P. A., and Fuchs, E. (1992) *J. Cell Biol.* **116**, 1181–1195
 31. Chou, C. F., and Omary, M. B. (1993) *J. Biol. Chem.* **268**, 4465–4472
 32. Lersch, R., Stellmach, V., Stocks, C., Giudice, G., and Fuchs, E. (1989) *Mol. Cell. Biol.* **9**, 3685–3697
 33. Johnson L. D., Idler, W. W., Zhou, X.-M., Roop, D. R., and Steinert, P. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1896–1900
 34. Krauss, S., and Franke, W. W. (1990) *Gene (Amst.)* **86**, 241–249
 35. Glass, C., and Fuchs, E. (1988) *J. Cell Biol.* **107**, 1337–1350
 36. Shapiro, M. B., and Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174
 37. Parry, D. A. D. (1990) in *Cellular and Molecular Biology of Intermediate Filaments* (Goldman, R. D., and Steinert, P. M., eds) pp. 175–204, Plenum Press Corp., New York
 38. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8132
 39. Breatnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383
 40. Byrne, C., Tainsky, M., and Fuchs, E. (1994) *Development* **120**, 2369–2383
 41. Jiang, C.-K., Magnaldo, M., Ohtsuki, I. M., Freedberg, F., and Blumenberg, M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6786–6790
 42. Faisst, S., and Meyer, S. (1992) *Nucleic Acids Res.* **20**, 1–26
 43. Steinert, P. M., Parry, D. D., Racoosin, E. L., Idler, W. W., Steven, A. C., Trus, B. L., and Roop, D. R. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5709–5713
 44. Rosenberg, M., Fuchs, E., LeBeau, M., Eddy, R. L., and Shows, T. B. (1991) *Cytogenet. Cell Genet.* **57**, 33–38
 45. Jeffreys, A. J. (1979) *Cell* **18**, 1–10
 46. Collin, C., Ouhayoun, J. P., Grund, C., and Franke, W. W. (1992) *Differentiation* **51**, 137–148
 47. Blumenberg, M. (1988) *J. Mol. Evol.* **27**, 203–211
 48. Nathans, J. (1994) *Cell* **78**, 357–360
 49. Mansbridge J. N., and Knapp, M. (1987) *J. Invest. Dermatol.* **89**, 253–263
 50. deMare, S., vanErp, P. E. J., Ramaekers, F. C. S., and van de Kerkhof, P. C. M. (1990) *Arch. Dermatol. Res.* **282**, 126–130
 51. Bereither-Hahn, J. (1986) *Biology of the Integument*, Vol. 2, pp. 443–471, Springer Verlag, Berlin
 52. Takahashi, K., Folmer, J., and Coulombe, P. A. (1994) *J. Cell Biol.* **127**, 505–520
 53. Gilmartin, M. E., Culbertson, V. B., and Freedberg, I. M. (1980) *J. Invest. Dermatol.* **75**, 211–216
 54. Yatsunami, J., Komori, A., Ohta, T., Suganuma, M., Yuspa, S. H., and Fujita, H. (1993) *Cancer Res.* **53**, 992–996
 55. Eriksson, J. E., Opal, P., and Goldman, R. D. (1992) *Current Opin. Biol.* **4**, 99–104
 56. Ku, N.-O., and Omary, M. B. (1994) *J. Cell Biol.* **127**, 161–171
 57. Kimonis, V., DiGiovanna, J. J., Yang, J.-M., Doyle, S. Z., Bale, S. J., Compton, J. G. (1994) *J. Invest. Dermatol.* **103**, 764–769
 58. McLean, W. H. I., Rugg, E. L., Lunny, D. P., Morley, S. M., Lane, E. B. Swensson, O., Dopping-Hepenstal, P. J. C., Griffiths, W. A. D., Eady, R. A. J., Higgins, C., Navsaria, H. A., Leigh, I. M., Strachan, T., Kunkeler, L., and Munro, C. S. (1995) *Nature Genetics* **9**, 273–278