Characterization of a First Domain of Human High Glycine-Tyrosine and High Sulfur Keratin-associated Protein (KAP) Genes on Chromosome 21q22.1*

Received for publication, June 28, 2002, and in revised form, September 9, 2002
Published, JBC Papers in Press, September 30, 2002, DOI 10.1074/jbc.M206422200

Michael A. Rogers, Lutz Langbein, Hermelinda Winter, Claudia Ehmann, Silke Praetzel, and Jürgen Schweizer

From the Section of Normal and Neoplastic Epidermal Differentiation and Division of Cell Biology, German Cancer Research Center, 69120 Heidelberg, Germany.

Analysis of the EBI/GenBank™ data base using non-human hair keratin-associated protein (KAP) cDNA sequences as a query resulted in the identification of a first domain of high glycine-tyrosine and high sulfur KAP genes located on human chromosome 21q22.1. This domain, present on the DNA accession numbers AP001078 and AP001709, was ~535 kb in size and contained 17 high glycine-tyrosine and 7 high sulfur KAP genes, as well as 9 KAP pseudogenes. Based on amino acid sequence comparisons of the encoded proteins, the KAP genes could be divided into seven high glycine-tyrosine gene families (KAP6–KAP8, and KAP19–KAP22) and four high sulfur gene families (KAP11, KAP13, KAP15, and KAP23). The high glycine-tyrosine genes described here appear to represent the complete set of this type of KAP genes present in the human genome. Both systematic cDNA isolation studies from an arrayed scalp cDNA library and in situ hybridization expression studies of all of the KAP genes identified in the 21q22.1 region revealed varying degrees and regions of expression of 11 members of the high tyrosine-glycine genes and 6 members of the high sulfur KAP genes in the hair forming compartment.

The hair follicle represents one of the few organs of the body that, throughout life, undergoes alternating cycles of growth, senescence, and rest (1). Morphologically, the hair follicle is composed of external epithelial compartments, the outer and inner root sheaths, the companion layer, and a central hair fiber-forming (trichocytic) compartment, comprising the matrix, cuticle, and cortex. Occasionally, a centrally lying medulla is present in specific hair types. Growth of the hair originates in matrix cells located in the bulb of the human follicle. This hair bulb surrounds a dermal fibroblast condensate, termed the dermal papilla, which is important for hair follicle morphogenesis (1). The main structural proteins of the hair fiber are the hair keratins and the keratin-associated proteins, KAPs, the latter being encoded by a large number of multigene families (2). Hair keratins, a subset of the large keratin family whose members are found in all cells of epithelial origin (3, 4), represent two multigene families, the type I (acidic) and type II (basic) families that comprise 15 members in humans (5, 6). They form the 8–10-nm intermediate filaments of trichocytes by co-polymerization of type I and type II members, which are differentially expressed during hair fiber development (7, 8). KAPs, which form the matrix between the hair keratin intermediate filament bundles through extensive disulfide bond cross-linking with cysteine residues in the head and tail domains of hair keratins, possess either high cysteine or high glycine-tyrosine content (2) and are further divided into three broad groups containing, to date 17 families. These are the high sulfur families KAP1–KAP3 and KAP10–KAP16, which contain less than 30 mol % cysteine (9–20), the ultra-high sulfur families KAP4, KAP5, KAP9, and KAP17, with more than 30 mol % cysteine (20–26), and the high glycine-tyrosine KAP families KAP6–KAP8 (27–30), which can be further divided into families with high (>60 mol %, KAP6 family) and lower (<60 mol %, KAP7 and KAP8 families) glycine-tyrosine content. The majority of the KAP members were initially identified in non-human species (sheep, mouse, rabbit). Recently, however, a large domain of human ultra-high/high sulfur KAP gene families (the KAP1–KAP3, KAP4, KAP9, KAP16, and KAP17 families) have been identified on chromosome 17q21.2 and their expression in the hair follicle partially characterized (20, 31). Moreover, chromosomal in situ hybridization studies using specific probes derived from human KAP5 family members have located new KAP genes to chromosomes 11p15 and 11q13 (32). In addition, the sequencing of human chromosome 21 has recently led to the identification of a further KAP gene domain on its long arm (21q23) (33). In a continuation of our previous KAP gene studies (20), we describe here the identification of a novel domain of 24 KAP genes (17 high glycine-tyrosine and 7 high sulfur KAP genes) and 9 KAP pseudogenes on chromosome 21q22.1. The genes could be grouped into 11 families, five of which are novel. Screening of a human scalp cDNA library as well as mRNA expression studies showed the expression of 17 individual KAP members, all of which were essentially localized to the hair cortex and several which also showed matrix and cuticular expression.

MATERIALS AND METHODS

Identification of Hair KAP Genes—Analysis of the EBI/GenBank™ data base with DNA sequences from regions encoding previously described

* This work was also supported in part by the Deutsche Forschungsgemeinschaft (Grant Schw 535/4-1). The costs of publication of this article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Section of Normal and Neoplastic Epidermal Differentiation and Division of Cell Biology, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. Tel.: 49-6221-423248; Fax: 49-6221-424406; E-mail: m.rogers@dkfz-heidelberg.de.

§ These authors contributed equally to this work.

1 The abbreviations used are: KAP, keratin-associated protein; ISH, in situ hybridization.
TABLE I

| KAP PCR products | PCR primer sequence | Size of product (bp) | Annealing temperature (°C) | cDNAs found | cDNA accession numbers |
|------------------|---------------------|----------------------|-----------------------------|-------------|------------------------|
| KRTAP 6.1        | ccaaccttgggaagcccaat| 146                  | 55                          | DSP         | AJ457063               |
| KRTAP 6.2        | gaattttactaat ATTGTTCTT| 133                  | 49                          | DSP         |                        |
| KRTAP 6.3        | gggtgagcactaatctat| 222                  | 54                          | DSP         | AJ457064               |
| KRTAP 7.1        | 5' AGCTTTGAGGTAATGCTTCTT 3'| 192                  | 53                          | DSP         | AK457065               |
| KRTAP 8.1        | 5' GTAATGATGGTGATGCTTCTT 3'| 200                  | 55                          | DSP         | AJ457066               |
| KRTAP 11.1       | 5' TATGGACGTCGACGGAAATCAATC 3'| 255                  | 57                          | DSP         |                      |
| KRTAP 13.1       | 5' ACTTTCTTCTCTGATGCTTCTT 3'| 149                  | 53                          | DSP         | AJ457067               |
| KRTAP 13.2       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 159                  | 53                          | DSP         | AJ457068               |
| KRTAP 13.3       | 5' ACTTTCTTCTCTGATGCTTCTT 3'| 321                  | 53                          | DSP         | AJ457069               |
| KRTAP 13.4       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 171                  | 53                          | DSP         |                      |
| KRTAP 13.5       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 154                  | 53                          | DSP         |                      |
| KRTAP 15.1       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 182                  | 54                          | DSP         |                      |
| KRTAP 15.2       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 220                  | 49                          | DSP         | AJ457067               |
| KRTAP 15.3       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 140                  | 54                          | DSP         |                      |
| KRTAP 15.4       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 143                  | 49                          | DSP         |                      |
| KRTAP 15.5       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 133                  | 55                          | DSP         |                      |
| KRTAP 16.1       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 182                  | 54                          | DSP         |                      |
| KRTAP 16.2       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 151                  | 49                          | DSP         |                      |
| KRTAP 16.3       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 139                  | 49                          | DSP         |                      |
| KRTAP 16.4       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 149                  | 54                          | DSP         |                      |
| KRTAP 17.1       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 215                  | 54                          | DSP         |                      |
| KRTAP 17.2       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 305                  | 55                          | DSP         |                      |
| KRTAP 17.3       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 174                  | 55                          | DSP         |                      |
| KRTAP 17.4       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 319                  | 53                          | DSP         |                      |
| KRTAP 17.5       | 5' TTTTCTTCTCTGATGCTTCTT 3'| 367                  | 49                          | DSP         |                      |

* Probe cross-hybridized with human KAP17.1.
* Probe cross-hybridized with a novel ultra-high sulfur KAP.
* Probe cross-hybridized with human KAP14.
RESULTS

KAP Gene Identification—Our recent detection of a high/ultra-high sulfur KAP gene domain on chromosome 17 led us to look further for human orthologs of the high glycine-tyrosine KAP gene families, KAP6–8, previously described in other species (27–30). This resulted in the discovery of a putative KAP gene domain on chromosome 21q22.1, a region not previously identified during the sequencing of this chromosome (33). We undertook a thorough DNA homology analysis and identified two contiguous genomic DNA sequences, AP001708 and AP001709, on chromosome 21q22.1, which harbored 24 high glycine-tyrosine and high sulfur KAP genes as well as 9 KAP pseudogenes (Fig. 1). The KAP gene domain covered an area of 535 kb straddling one end of both genomic sequences. All KAP genes possessed a methionine start codon and an open reading frame coding for proteins with either high glycine-tyrosine or high sulfur amino acid content. They displayed putative polyadenylation signals within a region ~400 bp downstream of their stop codon and nearly all of them contained a presumptive TATA box sequence within 100 bp upstream of the initiation codon. In addition, a variety of both high glycine-tyrosine and high sulfur KAP pseudogenes (see Table II) were identified by either the absence of an in-frame initiation codon in the region of homology or by the occurrence of frame shifts in the respective homology regions. As a rule, the genes were small (less than 1 kb in size) and comprised only one exon. The intragenic distances between the KAP genes varied widely from ~5–100 kb, and in general, the genes possessed no unique direction of transcription. The entire gene domain could be subdivided into one large group of 8 high sulfur KAP genes/pseudogene on one end, followed by a central group of 24 high glycine-tyrosine KAP genes/pseudogenes and ending thereafter with one high sulfur KAP gene (see Fig. 1). No putative KAP genes could be found on the DNA sequences immediately adjacent to this domain, but the presence of a further potential high sulfur KAP gene domain, the one immediately adjacent to this domain, but the presence of a further potential high sulfur KAP gene domain, the one immediately adjacent to this domain, but the presence of a further potential high sulfur KAP gene domain.

Table II

Genomic localization of KAP genes on DNA sequences AP001708 and AP001709, as well as the molecular weight and cysteine/glycine-tyrosine content of the derived proteins

| KAP designation | Accession no. | Open reading frame | Strand | Molecular weight of protein Da mol % | Tyrosine-glycine residues |
|-----------------|--------------|--------------------|--------|------------------------------------|--------------------------|
| KRTAP 6.1       | AP001708     | 273,712–273,930    | –      | 7279                               | 60.5                     |
| KRTAP 6.2       | AP001708     | 258,712–258,900    | –      | 6654                               | 56.5                     |
| KRTAP 6.3       | AP001708     | 253,492–253,804    | +      | 11146                              | 54.6                     |
| KRTAP 7.1       | AP001709     | 150,421–150,724    | –      | 9288                               | 34.4                     |
| KRTAP 8.1       | AP001708     | 134,054–134,245    | –      | 6825                               | 42.8                     |
| KRTAP 9.1       | AP001708     | 140,071–140,343    | –      | 9007                               | 62.2                     |
| KRTAP 9.2       | AP001708     | 147,216–147,374    | –      | 5736                               | 42.3                     |
| KRTAP 9.3       | AP001709     | 151,735–151,982    | –      | 8247                               | 66.7                     |
| KRTAP 9.4       | AP001708     | 156,881–157,135    | –      | 9106                               | 44.0                     |
| KRTAP 9.5       | AP001708     | 161,897–162,115    | –      | 7624                               | 62.5                     |
| KRTAP 9.6       | AP001708     | 201,683–201,859    | +      | 6266                               | 38.8                     |
| KRTAP 9.7       | AP001708     | 221,124–221,315    | –      | 6644                               | 53.9                     |
| KRTAP 9.7       | AP001708     | 276,481–276,651    | +      | 6201                               | 67.8                     |
| KRTAP 10.1      | AP001708     | 276,681–276,987    | +      | 6201                               | 67.8                     |
| KRTAP 10.2      | AP001708     | 295,290–295,487    | +      | 6559                               | 61.5                     |
| KRTAP 10.3      | AP001709     | 67,976–68,227      | –      | 8564                               | 50.6                     |
| KRTAP 10.4      | AP001709     | 76,184–76,403      | –      | 7035                               | 53.1                     |
| KRTAP 10.5      | AP001709     | 261,147–261,293    | +      | 5274                               | 37.5                     |

* The KAP7.1 sequence found on AP001709 contains a 1 bp insertion, which results in an incorrect reading frame. See “Discussion” for details.
Characterization of the High Glycine-Tyrosine and High Sulfur KAPs—The KAP genes listed in Fig. 1 were translated into amino acid sequences followed by a multialignment of the various KAP proteins. Included into these multiple sequence comparisons were both high glycine-tyrosine and high sulfur KAP proteins from other species, which had already been appointed to distinct KAP families or were found in the literature without special classification. This allowed the creation of an evolutionary tree in which human and non-human KAP proteins were grouped into distinct families (Fig. 2). In addition to the human orthologs of six previously described KAP families (KAP6, KAP7, KAP8, KAP11, KAP13, KAP15), we identified six novel KAP families, four high glycine-tyrosine KAP families (KAP19–KAP22) and one high sulfur KAP family (KAP23). Human orthologs segregating with the members of the murine high glycine-tyrosine KAP18 family (36) (see "Discussion") or the sheep high sulfur KAP10 (2) and mouse KAP12 proteins (10) were not detected in the evolutionary tree (Fig. 2).

The human high glycine-tyrosine KAPs can be divided into seven families, three of which are counterparts of KAP families previously described in other species: The KAP6 family with three members, KAP6.1–KAP6.3 (Figs. 2 and 3) and the hKAP7 and hKAP8 families with one member each, KAP7.1 and KAP8.1, respectively (Figs. 2, 4, and 5) (27–29). The KAP6 family members range in size from 6.6 to 11.1 kDa and exhibit a high content of glycine and tyrosine residues (54.6–60.5 mol %, see Table II). Except for KAP6.3, the completely identified

---

*a M. A. Rogers, L. Langbein, H. Winter, C. Ehmann, S. Praetzel, and J. Schweizer, unpublished data.
KAP6 proteins from all species display unique amino-terminal (MCG(S/-)YY(G/R)NY) and carboxyl-terminal (GS(G/S)F-GYY(Y/-)) sequence motifs (Fig. 3). Both the 9.3-kDa KAP7.1 and the 6.8-kDa KAP8 proteins not only exhibit a particularly high sequence homology with their non-human counterparts (Figs. 4 and 5) but also display a particularly low glycine-tyrosine content (34.4 and 42.8 mol %; Table II). All of the known KAP8 proteins from various species show a characteristic carboxyl-terminal amino acid sequence (RR(F/Y)(W/S)PFALY) (Fig. 4). However, the first nine amino acids of human KAP8.1 are significantly different from the orthologous mouse and sheep sequences (Fig. 4).

KAP19–KAP22 represent new human high glycine-tyrosine KAP families. The large KAP19 family consists of 7 members, KAP19.1–KAP19.7, whose size ranges from 5.7 to 9.1 kDa and which exhibit a particularly high glycine-tyrosine content (34.4 and 42.8 mol %; Table II). All of the known KAP8 proteins from various species show a characteristic carboxyl-terminal amino acid sequence (RR(F/Y)(W/S)PFALY) (Fig. 4). However, the first nine amino acids of human KAP1 are significantly different from the orthologous mouse and sheep sequences (Fig. 4).

KAP19–KAP22 represent new human high glycine-tyrosine KAP families. The large KAP19 family consists of 7 members, KAP19.1–KAP19.7, whose size ranges from 5.7 to 9.1 kDa and which exhibit a particularly high glycine-tyrosine content (34.4 and 42.8 mol %; Table II). All of the known KAP8 proteins from various species show a characteristic carboxyl-terminal amino acid sequence (RR(F/Y)(W/S)PFALY) (Fig. 4). However, the first nine amino acids of human KAP1 are significantly different from the orthologous mouse and sheep sequences (Fig. 4).

KAP19–KAP22 represent new human high glycine-tyrosine KAP families. The large KAP19 family consists of 7 members, KAP19.1–KAP19.7, whose size ranges from 5.7 to 9.1 kDa and which exhibit a particularly high glycine-tyrosine content (34.4 and 42.8 mol %; Table II). All of the known KAP8 proteins from various species show a characteristic carboxyl-terminal amino acid sequence (RR(F/Y)(W/S)PFALY) (Fig. 4). However, the first nine amino acids of human KAP1 are significantly different from the orthologous mouse and sheep sequences (Fig. 4).

KAP19–KAP22 represent new human high glycine-tyrosine KAP families. The large KAP19 family consists of 7 members, KAP19.1–KAP19.7, whose size ranges from 5.7 to 9.1 kDa and which exhibit a particularly high glycine-tyrosine content (34.4 and 42.8 mol %; Table II). All of the known KAP8 proteins from various species show a characteristic carboxyl-terminal amino acid sequence (RR(F/Y)(W/S)PFALY) (Fig. 4). However, the first nine amino acids of human KAP1 are significantly different from the orthologous mouse and sheep sequences (Fig. 4).
Repeat structures are and species (15 KAP13, KAP15, have been previously characterized in other and KAP22.1 contain only few repeat structures. KAP7, KAP8, KAP20, and KAP21 family members. In contrast, KAP7 which consists of four members, KAP13.1 repeat structure covering the carboxyl-terminal half of the previously described human high sulfur KAP1–KAP3 families (20). For example, KAP11.1, like mouse KAP13.1 contains a strongly conserved decameric repeat structure containing an initial CQ(L/E) motif (16), which is only partially seen in the human KAP13 family members (Fig. 7). Both, however, possess, in addition, generally less well conserved pentameric repeats in their carboxyterminal regions. A pentameric SL/G/D/C/G motif is also seen in human KAP15.1. In contrast KAP23.1 does not appear to possess an obvious repeat structure (Fig. 8).

Expression Analyses—3'-Noncoding region PCR products from all of the KAP genes presented in this study were used as probes to screen a previously described arrayed human scalp cDNA library (20). This led to the isolation of three high glycine-tyrosine (KAP7.1, KAP8.1, and KAP19.1) and two high sulfur (KAP11.1 and KAP13.1) KAP cDNA clones (see Table I). The low number of positive clones isolated, coupled with the identification of a partial KAP6.1 cDNA via 3'-RACE analysis of follicular RNA,2 led us to believe that most of the KAP genes analyzed possessed an expression level below the limits of detection in our arrayed cDNA library, which contains only 26,000 clones. We therefore subchonel all the PCR products used in the cDNA library screen (Table I) and used them as cDNA probes for in situ hybridization studies of both plucked beard hairs and scalp sections. This resulted in the demonstration of mRNA expression for a total of 17 KAP genes, 11 high glycine-tyrosine, and six high sulfur members (Fig. 9). With the exception of the KAP7.1 gene, all of the high glycine-tyrosine and high sulfur KAP genes showed similar degrees of expression in the two hair types, but the individual expression patterns varied strongly from gene to gene. The high glycine-tyrosine KAP genes, KAP7.1, KAP8.1, KAP19.1, and KAP19.2 and the high sulfur KAP gene KAP11.1 exhibited strong expression. Interestingly, the prominent expression level of KAP7.1 was limited to scalp follicles, while beard hair follicles showed a drastically reduced expression pattern (Fig. 9, B and B'). In contrast, the remaining KAP genes displayed remarkably low levels of expression (Fig. 9). The expression patterns of high glycine-tyrosine KAP genes could be further subdivided. KAP6.1 (Fig. 9A), KAP7.1 (Fig. 9, B and B'), KAP19.1 (Fig. 9D), and KAP19.2 (Fig. 9E) transcripts clearly occurred in the upper portion of the hair cortex. Of these, KAP6.1 and, in particular, KAP19.1 transcripts exhibited a vertical asymmetry of expression in the cortex. Moreover, KAP19.1 showed an additional region of expression in the hair cuticle, occurring nearly simultaneously with the onset of cortical mRNA expression (Fig. 9D). In contrast, the mRNA expression of KAP8.1 (Fig. 9C), KAP19.3 (Fig. 9F), KAP19.6 (Fig. 9H), KAP19.7 (Fig. 9I), KAP20.1 (Fig. 9K), and KAP21.2 (Fig. 9L) obviously initiated in the lower cortex or even in the matrix of the follicle. Several of
Fig. 9. Regions of mRNA synthesis of high glycine-tyrosine (A–L) and high sulfur (M–R) KAP families in plucked beard (A, B′–I, M–R) and scalp (B, K–L) hair follicles. Numbers on the lower right-hand side of each picture designate the name of each individual KAP family member. Red arrows, zone of major mRNA synthesis in the cortex; red dotted angled arrows, zone of mRNA synthesis in the hair cuticle; co, hair cortex; co*, region of asymmetric KAP expression in the cortex. Note the heterogeneity of vertical KAP-mRNA expression in the hair forming compartment. med, medulla; dp, dermal papillae. Bars: 150 μm.
these KAP genes also showed cuticular expression. For example, KAP19.6 and KAP21.2 exhibited an onset of expression in the upper portion of the cuticle, which occurred much later than their respective cortical expression. Two of the high glycine-tyrosine KAP genes (KAP8.1 and KAP19.4) displayed highly remarkable expression patterns. In the first case, the particularly strong KAP8.1 mRNA expression was essentially restricted to only one vertical half of the hair forming compartment and in beard hairs was clearly absent from the central medulla (Fig. 9C). The absence of medullar expression in beard hair sections could also be observed for the other KAP genes. In the second case, the expression of KAP19.4 was unique for it only occurred in the upper portion of the hair cuticle. Cortical expression, if present at all, was extremely weak (Fig. 9C). Finally, for six high glycine-tyrosine KAP genes (KAP6.2, KAP6.3, KAP19.5, KAP20.2, KAP21.1, and KAP22.1), neither cDNA library screening nor in situ hybridization studies resulted in the demonstration of their expression in hair follicles.

With the exception of the extraordinary strong KAP11.1 gene expression in the late matrix and the entire cortex (Fig. 9M), the expression of the remaining high sulfur KAP genes in these areas was generally very weak (Fig. 9, N–R). In a manner similar to several of the high glycine-tyrosine KAP members, a vertical asymmetry was seen for KAP23.1 transcripts (Fig. 9R). A clear-cut cuticular expression was detected for KAP13.2 (Fig. 9O), KAP15.1 (Fig. 9Q), and KAP23.1 (Fig. 9R), whose onset of expression occurred well after the initiation of the respective cortical expression. Interestingly, the cuticular expression of KAP13.2 appeared distinctly stronger than that in the cortex and was thus reminiscent of the expression pattern of the high glycine-tyrosine KAP19.4 gene (Fig. 9G). The only high sulfur KAP gene for which no expression in the human hair follicle could be demonstrated was KAP13.4.

**DISCUSSION**

The preliminary identification of a KAP gene cluster on chromosome 21q23 by the Human Genome Sequencing Project (33) led us to characterize an additional domain on chromosome 21q21.2 in a manner described recently for the elucidation of a human high/ultra-high sulfur KAP gene locus on chromosome 17q21.2 (20). Analysis of the EBI/GenBank™ data base allowed the identification of 24 high glycine-tyrosine and high sulfur KAP genes on the contiguous genomic sequences AP001708 and AP001709, which are part of chromosome 21q22.1. Like the high/ultra-high sulfur KAP genes on chromosome 17, as well as KAP genes identified in other species (2, 20), all of the KAP genes present on this locus were small in size (<1 kb) and consisted of only one exon. Remarkably, however, was the insertion of the high glycine-tyrosine KAP gene domain into the high sulfur KAP gene domain.

Homology comparison of the human high glycine-tyrosine KAP members with members from other species led to the division of these KAPs into seven families. Historically, non-human high glycine-tyrosine KAPs were divided through biochemical separation techniques into highly glycine-tyrosine-rich type II members (>60% glycine-tyrosine content) and a second type I group with a lower glycine-tyrosine content (2, 37). Subsequent sequence homology comparisons in sheep and mouse lead to the initial establishment of three high glycine-tyrosine families: the type II KAP6 family and the type I KAP7 and KAP8 families (2, 38). In the current report, three human KAP6 members and one KAP7 and KAP8 member could be identified based on their amino acid homology to sheep and mouse KAPs.

Interestingly, to date, four KAP6 members are known in the mouse (29, 36), two of which (d86420 and d86421, see Fig. 3) are completely identical in their amino acid sequences, while their mRNAs possess differing 5’- and 3’-noncoding regions. In the case of the KAP8 family, a total of four members is known in sheep, two of which (p02448 and x05639, see Fig. 4) vary from each other in only two positions, and therefore, may be polymorphisms (28, 39–41). All in all, the present composition of the KAP6 and KAP8 families in humans, mouse, and sheep clearly suggests a heterogeneity in gene number among different species. Since humans possess two KAP8 pseudogenes (see Fig. 1), the low number of functional human KAP8 members might be partially due to a species-specific inactivation of KAP8 genes during evolution. A similar type of gene loss has been described recently for the human type I keratin pseudogene hHaA, which possesses functional orthologs in the chimpanzee and gorilla (42).

Initially, the gene underlying the single human KAP7.1 protein, which shows a high sequence identity to a previously described sheep gene sequence (28), was assumed by us to represent a pseudogene due to a 1-base pair insertion into the corresponding DNA sequence on AP001709 (nucleotides 150,677), which would cause a frameshift within the region coding for the glycine-tyrosine-rich portion of KAP7.1. However, in the course of a related study in this laboratory, a KAP7.1 cDNA was isolated that contained a completely intact open reading frame (results not shown). Subsequently, cDNA/genomic analyses of four unrelated individuals by PCR clearly showed that all possessed complete open reading frames for KAP7.1. In addition, we were able to isolate a partial KAP7.1 clone from our arrayed human scalp cDNA library (see Table 1) and to successfully perform in situ hybridization studies on human hair follicles using a 3’-noncoding region of the hKAP7.1 mRNA as a probe (Fig. 9B). Therefore, either the KAP22.1 mRNA as a probe (Fig. 9B). Therefore, either this gene was inactivated in a minority of the human population, or the observed KAP7.1 gene frameshift on AP001709 was simply a DNA sequencing error.

The sequence comparison of the remaining human high glycine-tyrosine KAPs with hitherto undefined KAPs from other species allowed their classification into the new high glycine-tyrosine KAP families 19–22. In this context, the large KAP19 family appears especially interesting as its six human members co-segregated with two out of ten murine KAPs (af345294 and af4777980, see Figs. 2 and 6), whose genes were down-regulated in hair follicles of Hoxc13-overexpressing mice (36). While two of the remaining eight murine KAPs segregated either with the KAP6 (af345298; Figs. 2 and 3) or the KAP21 family (af345297; Figs. 2 and 5), remarkably, the remaining six murine KAPs clearly built up an individual, hitherto unidentified high glycine-tyrosine KAP family, termed here the KAP18 family (Fig. 2). Although a certain degree of similarity exists between the mouse KAP18 members and the human KAP19 members (data not shown), the degree of divergence seen in Fig. 2 suggests that the murine KAP18 family represents indeed a family of its own. In a manner similar to the KAP18 family, the single KAP22.1 family member (Fig. 5) has, at present, only been observed in humans.

It should be emphasized that the analysis of the current working draft of the human genome has resulted in the identification of only one further putative high glycine-tyrosine KAP gene sequence outside of the KAP gene domain characterized here. This putative pseudogene is located on chromosome 13q14 on BAC clone AL138686 (nt 63432–63717). Therefore, we believe that the entire set of human high glycine-tyrosine KAP genes are probably described here.

Remarkably, the genes encoding the seven high glycine-tyrosine rich KAP families on the contiguous DNA sequences AP001708 and AP001709 are flanked by genes that code for
High Glycine-Tyrosine/High Sulfur KAP Gene Domain

High sulfur KAP proteins (see Fig. 1). Amino acid homology comparisons of the seven high sulfur KAPs led to their division into four individual families (KAP11, KAP13, KAP15, and KAP23), three of which possess counterparts described in other species. The largest family, KAP13, consists of four members that show a high sequence homology to the murine KAP13.1, originally designated 4C32 (16). The three remaining human high sulfur KAP families comprise only single members. Based on amino acid homology, the human KAP11.1 appears to be the ortholog of a mouse KAP protein, originally termed hac1-1, identified accidentally in the search for the O'-methylguanidine-DNA methyltransferase cDNA (15). Most probably, the single human high sulfur KAP15.1 protein represents the orthology of mouse KAP15.1 (Fig. 8), one of two related KAP proteins initially termed pmg1 and pmg2 (17), but renamed recently mKAP14.2 and mKAP15.1, respectively (see Ref. 20 for explanation), with which it cosegregates in the CLUSTREE analysis (Fig. 1). It has recently been shown that the mouse KAP14.2 and KAP15.1 genes are adjacent to each other and possess opposite directions of transcription (17). Interestingly, upstream of the human KAP15.1 gene is a putative KAP pseudogene designated hKAP13A, which exhibits a fairly high homology with both mKAP14.2 and also the adjacent human KAP13 genes (Fig. 1 and results not shown). This was confirmed by a further genome wide BLASTN search with the mKAP14.2 sequence, which revealed roughly comparable sequence homologies only with the hKAP13A, hKAP15.1, and the hKAP13 genes. In addition, the human hKAP13A and KAP15.1 genes/pseudogenes exhibit, like the mKAP14.2 and mKAP15.1 genes, different directions of transcription (Ref. 17 and Fig. 1). It is therefore tempting to speculate that hKAP13A might represent an orthologous, but inactive, form of mKAP14.2, and the absence of further orthologs of the mK14.1 gene in the human genome database would point to hKAP15.1 as being the only functional human ortholog of the murine KAP14.2/KAP15.1 domain.

Chromosome 21 sequencing and its initial gene analysis by bioinformatic methods identified a surprisingly low number of 225 putative gene and 98 pseudogene loci (33). Recently, these findings have been challenged in two articles that used combinations of gene prediction and mRNA expression analysis to show, in one case, that the number of putative genes on this chromosome may be up to 10-fold higher (43, 44). Our data support this assumption, for in the chromosome 21 sequencing paper, only two exons were found by gene prediction analysis on the domain described here. As such, further biochemical and molecular biological analysis of the Human Genome Programs bioinformatically generated gene prediction data appears necessary.

In general, the expression characteristics of the human high glycine-tyrosine and high sulfur KAP genes described here correspond fairly well to what has been previously found in other species. It was known from earlier protein studies that the follicular expression of high glycine-tyrosine KAPs varies considerably, being very weak in human hairs, but strong in sheep wool and mouse hairs (45). This is in line with our finding that only cDNA clones of the most strongly expressed high glycine-tyrosine KAP members 7.1, 8.1, and 19.1 (Fig. 9) resulted from the scalp cDNA library screening (Table I). Moreover, previous in situ hybridization studies of KAP6 family members in mouse, rabbit, and sheep have shown that also their expression occurs preferentially in the upper portion of the hair and wool cortex (27, 29). This was also true for murine KAP8 family members (29), while human and sheep KAP8 expression begins much earlier in the matrix region (Ref. 2 and Fig. 9C). Whereas in sheep and humans, the cortical expression of KAP6 and KAP8 family members is vertically compartmentalized (Refs. 2 and 27 and Fig. 9, A and C)), this is not the case for mouse KAP6 and KAP8 members, which clearly exhibit an even cortical expression (29).

In mice, in situ hybridization studies of individual high glycine-tyrosine KAP19 (d86422 in Fig. 6), KAP20 (d88901 in Fig. 5) and KAP21 (d89902 in Fig. 5) members have shown that their expression was limited to the upper portion of the murine hair cortex (29) and, thus, corresponded to that of the human KAP19.1, KAP19.2, and KAP20.1 genes (Fig. 9, D, E, and K), however, deviated from human KAP21.2, whose expression initiated further down in the hair matrix (Fig. 9E). Surprisingly and at present unexplainable was the striking difference seen in the expression of KAP7.1 in scalp and beard hair follicles (Fig. 9, B and B') as well as the unique cuticular KAP19.4 expression compared with the cortical expression pattern of the other KAP19 members.

The majority of the high sulfur KAP genes analyzed in this study show a generalized mRNA expression in the upper matrix and the entire hair cortex, with KAP13.2, KAP15.1, and KAP23.1 also exhibiting cuticular expression. The strongly expressed KAP11.1 displays a nearly identical expression as its mouse ortholog (15). The demonstration of cortical KAP13.1–KAP13.3 expression in the human hair follicle was surprising, for no such expression has been previously reported in mice (16). Instead, mouse KAP13.1 was found in the periderm of embryonic mice as well as the filiform tongue papillae and the parakeratotic tail epidermis of adult mice (16). To date, the analysis of KAP13.1–KAP13.3 gene expression in human tongue and periderm epithelium has been hampered by the unavailability of the tissue. The weak expression of the human KAP15.1 gene in the matrix, cortex, and cuticle of the hair follicle (Fig. 9Q) has not yet been confirmed for its animal orthologs, although a specific mouse KAP15.1 antibody is available, which has, however, only been used in Western blots to demonstrate mKAP15.1 in mouse mammary glands and epidermis (17).

The low mRNA expression of many members of the KAP families presented here leads to questions concerning the synthesis of their respective proteins. This is especially important, for data derived from recent large scale mRNA/protein studies in yeast and human liver show only a moderate correlation between mRNA transcript and protein expression (46–48). This appears especially true for genes with low mRNA expression, for, in one study, mRNA transcripts with similar degrees of mRNA expression exhibited over 30-fold differences in protein expression (47). As such the mRNA expression data described here are no guarantee for concomitant protein expression. A detailed analysis of the KAP protein expression described here would go beyond the bounds of this paper, but will be a challenging task. For example, high glycine/tyrosine KAPs, due to their strong amino acid conservation and hydrophobic nature, are poorly amenable to the production of both pan- and individual antibodies. The small size of these proteins (<10 kDa) makes two-dimensional protein separation difficult, especially in a system that is amenable to Western blot analysis. If these handicaps can be overcome, then future analysis, possibly using mass spectrometry (matrix-assisted laser desorption ionization time-of-flight, MALDI-TOF) for protein identification, could lead to a final elucidation of KAP protein expression in the hair follicle.

Acknowledgments—We thank Dr. Herbert Spring (German Cancer Research Center) for support with confocal microscopy. The cDNA clones isolated in this study can be obtained from the German Human Genome Resource Center (RZPD) using the names provided in Table I. We also thank the RZPD members for continuing support. We also
thank Dr. Alexander Awgulewitsch (Medical University of Charleston, Charleston, SC) for providing a previously unpublished mouse KAP 18 cDNA sequence.

REFERENCES

1. Stehn, K. S., and Paus, R. (2001) Physiol. Rev. 81, 449–494
2. Powell, B. C., and Rogers, G. E. (1997) in Formation and Structure of Human Hair (Jelles, P., Zahn, H., and Hocker, H., eds) pp. 59–148, Birkhäuser Verlag, Basel, Switzerland
3. Fuchs, E., and Weber, R. (1994) in Annual Review of Biochemistry (Richardson, C. C., ed) pp. 345–382, Annual Reviews Inc., Palo Alto, CA
4. Steinitz, P. M., and Roop, D. R. (1988) Ann. Rev. Biochem. 57, 593–625
5. Rogers, M. A., Winter, H., Langbein, L., Wolf, C., and Schweizer, J. (2000) J. Invest. Dermatol. 114, 464–472
6. Rogers, M. A., Winter, H., Wolf, C., and Schweizer, J. (1998) J. Biol. Chem. 273, 26683–26691
7. Langbein, L., Rogers, M. A., Winter, H., Praetzel, S., and Schweizer, J. (2001) J. Biol. Chem. 276, 35123–35132
8. Langbein, L., Rogers, M. A., Winter, H., Praetzel, S., Beckhaus, U., Rackwitz, H.-R., and Schweizer, J. (1999) J. Biol. Chem. 274, 19874–19884
9. Zhumbaeva, B. D., Gening, L. V., and Gazaryan, K. G. (1992) Mol. Biol. 26, 550–555
10. Cole, S. K., and Reeves, R. H. (1998) Genomics 54, 437–442
11. Swart, L. S., Joubert, F. J., and Parris, D. (1976) in Proceedings of the Fifth International Wool Textile Research Conference (Ziegler, K., ed) pp. 254–264, German Wool Research Institute, Aachen, Germany
12. Mitsu, S., Ohuchi, A., Adachi-Yamada, T., Hotta, M., Tsuboi, R., and Ogawa, H. (1998) Gene (Amst.) 208, 123–129
13. Powell, B. C., Sleigh, M. J., Ward, R. A., and Rogers, G. E. (1983) Nucleic Acids Res. 11, 5327–5346
14. Frenkel, M. J., Powell, B. C., Ward, K. A., Sleigh, M. J., and Rogers, G. E. (1989) Genomics 4, 182–191
15. Huh, N., Kashigawa, M., Koniai, T., Hashimoto, Y., Kohno, Y., Nomura, S., and Kuroki, T. (1994) J. Invest. Dermatol. 102, 716–720
16. Takai, M., Takata, Y., Kuroki, T., and Huh, N.-H. (1998) J. Invest. Dermatol. 111, 128–132
17. Kuhn, F., Lassing, C., Schroeder, M., Kuroki, T., and Schweizer, J. (2001) J. Biol. Chem. 276, 19440–19451
18. Rogler, M. A., Langbein, L., Winter, H., Ehmann, C., Praetzel, S., Korn, B., and Schweizer, J. (2000) J. Biol. Chem. 275, 19440–19451
19. Powell, B. C., Arthur, J., and Nesel, A. (1995) Differentiation 58, 227–232
20. Fratini, A., Powell, B. C., Hynd, P. I., Keough, R. A., and Rogers, G. E. (1994) J. Invest. Dermatol. 102, 178–185
21. Perez, C., Auriol, J., Gerst, C., Bernard, B. A., and Egly, J. M. (1999) Gene (Amst.) 227, 137–148
22. MacKinnon, P. J., Powell, B. C., and Rogers, G. E. (1990) J. Cell Biol. 111, 2987–3000
23. McNab, A. R., Andrus, P., Wagner, T. E., Buhl, A. E., Waldon, D. J., Kawabe, T. T., Rae, T. J., Greppi, V., and Vogeli, G. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 6848–6852
24. Jenkins, B. J., and Powell, B. C. (1994) J. Invest. Dermatol. 103, 310–317
25. Pratini, A., Powell, B. C., and Rogers, G. E. (1993) J. Biol. Chem. 268, 4511–4518
26. Kuznekoff, E. S., and Rogers, G. E. (1987) Eur. J. Biochem. 166, 79–85
27. Azzi, N., and It0, M. (1997) J. Biol. Chem. 272, 20512–20518
28. Gillespie, J. M. (1991) in Physiology, Biochemistry and Molecular Biology of the Skin (Goldsmith, L. A., ed) pp. 625–659, Oxford University Press, Oxford
29. Shimomura, Y., Azzi, N., Rogers, M. A., Langbein, L., Schweizer, J., and Ito, M. (2002) J. Invest. Dermatol. 118, 236–244
30. MacKinnon, P. J., Powell, B. C., Rogers, G. E., Baker, E. G., MacKinnon, R. N., Hyland, J. Y., Cullen, D. F., and Sutherland, G. R. (1991) Mamm. Genome 1, 53–56
31. Hattori, M., Fuijyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S., Toyoda, A., Ichii, K., Totoki, Y., Choi, D. K., Suzuki, E., Okhi, M., Takagi, T., Sakaki, Y., Taudsen, S., Bleichschmidt, K., Polley, A., Menzel, U., Delahar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., and Schudy, A. (2000) Nature 405, 311–319
32. Langbein, L., Heid, H. W., Moll, I., and Franke, W. (1994) Differentiation 55, 57–71
33. Rogers, M. A., Langbein, L., Praetzel, S., Moll, I., Krieg, T., Winter, H., and Schweizer, J. (1997) Differentiation 61, 187–194
34. Tkatchenko, A. V., Visconti, R. P., Shang, L., Papenbrock, T., Pruett, N. D., Ito, T., Ogawa, M., and Awgulewitsch, A. (2001) Development (Camb.) 128, 1547–1558
35. Gillespie, J. M. (1983) in Biochemistry and Physiology of the Skin (Goldsmith, L. A., ed) pp. 475–510, Oxford University Press, Oxford
36. Rogers, G. E., and Powell, B. C. (1993) J. Invest. Dermatol. 101, 506–511
37. Gohde, T. A. (1973) Eur. J. Biochem. 44, 120–124
38. Kuznekoff, E. S., and Rogers, G. E. (1985) Eur. J. Biochem. 146, 89–93
39. Marshall, R. C., Gillespie, J. M., Inglis, A. S., and Frenkel, M. J. (1986) in Proceedings of the Sixth International Wool Textile Research Conference (Ziegler, K., ed) pp. 147–158, German Wool Research Institute, Aachen, Germany
40. Winter, H., Langbein, L., Krawczak, M., Cooper, D. N., Jave-Suarez, L. F., Rogers, M. A., Praetzel, S., Heid, P. J., and Schweizer, J. (2001) Hum. Genet. 108, 37–42
41. Reymond, A., Camargo, A. A., Deutsch, S., Stevenson, B. J., Parmigiani, R. B., Ucla, C., Bettoni, F., Lyle, R., Guipponi, M., de Souza, S., Iseli, C., Jorgensen, C. V., Bucher, F., Simpson, A. J., and Antonarakis, S. E. (2002) Genomics 79, 824–832
42. Krapavarov, P., Cawley, S. E., Drenkov, J., Bekiranov, S., Strausberg, R. L., Fedor, S. P., and Gingeras, T. R. (2002) Science 296, 916–919
43. Gillespie, J. M. (1972) Comp. Biochem. Physiol. B 41, 723–734
44. Miklos, G. L., and Malessa, R. (2001) Proteomics 1, 169–178
45. Ogawa, M., and Kuroki, T. (2001) Mol. Cell. Biol. 19, 7920–7930
46. Anderson, L., and Seilhamer, J. (1997) Electrophoresis 18, 533–537
47. Higgins, D. G., Thompson, J. D., and Gibson, T. J. (2000) Methods Enzymol. 383, 38–45
48. Gillespie, J. M. (1990) in Cellular and Molecular Biology of Intermediate Filaments (Goldman, R. D., and Steinitz, P. M., eds) pp. 95–128, Plenum Press, New York