Sequence, Expression, and Evolutionary Conservation of a Gene Encoding a Glycine/Tyrosine-rich Keratin-associated Protein of Hair*

(Received for publication, May 22, 1992)

Antonio Fratinić, Barry C. Powell, and George E. Rogers

From the Department of Biochemistry, University of Adelaide, Adelaide 5001, South Australia

In hair differentiation several families of keratin proteins with distinctive amino acid compositions are produced. To study the role and regulation of one of these families, the glycine/tyrosine-rich keratin-associated proteins encoded by the KAP6 gene family, a partial wool follicle cDNA clone encoding a sheep KAP6 protein was sequenced and the responding gene isolated from a sheep cosmid library. The KAP6.1 gene encodes a basic protein of 82 amino acids (M_r = 8,296) with a combined glycine and tyrosine content of ~60 mol %. There are several KAP6 genes in the sheep genome, all located within a 1,050-kilobase SfiI fragment. Northern blot analysis demonstrated that at least one member of the KAP6 family is expressed in the wool follicle. A rabbit KAP6 gene was isolated and its sequence and expression patterns were compared with the sheep gene. The sheep and rabbit genes have a nucleotide sequence identity of 89.7 %, suggesting that they are equivalent genes and indicating strong selective pressure during evolution. Both genes contain several conserved sequence motifs of 7–9 nucleotides in their 5' flanking regions that may be involved in the regulation of their expression. Localization of KAP6 mRNAs in sheep wool and rabbit hair follicles by in situ hybridization suggests that the genes are expressed in the cells of the hair shaft cortex in varying expression patterns. KAP6 expression starts relatively late in hair follicle differentiation, and the proportion of hair cortical cells that express it may change from follicle to follicle.

An intriguing feature of hair structure is the multiplicity of the hair keratin proteins. The wool fiber, for example, is composed of about 50–100 keratin proteins derived from several multigene families (1–5). These proteins are broadly classified into two groups: the intermediate filament (IF) keratin proteins and the keratin IF-associated, or matrix, proteins (6). The keratin IF-associated proteins are characterized by high proportions of one or two amino acids and have been classified as ultra-high-sulfur, high-sulfur, or glycine/tyrosine-rich proteins (7). A unified nomenclature for these proteins has recently been proposed and is used in this report. The hair matrix proteins are acknowledged as keratin-associated proteins and are referred to by the prefix KRTAP, abbreviated to KAP in general usage, with a number to denote each family. The KAP6 family is the glycine/tyrosine-rich type II proteins, KAP7 is glycine/tyrosine-rich type I component C2, and KAP8 is glycine/tyrosine-rich type I component F.

Hairs have a common structure composed of a cuticle sheath, an inner cortex, and a central medulla, the relative proportions of which vary with fiber type. The cuticle and cortical cells synthesize the keratin proteins (8). Within the cortex of the wool fiber there are two major cell types, termed the paracortical and orthocortical cells (6), that contain different proportions of the keratin proteins (9). At terminal differentiation proceeds during fiber growth, the keratin IFs in the cortical cells are assembled into a filamentous scaffold that becomes embedded in a matrix of the KAPs (10). The orthocortical cells appear to have a higher IF:matrix ratio than the paracortical cells and a different ultrastructural organization (11).

The glycine/tyrosine-rich KAPs, the smallest of the hair keratins (M_r = 6,000–9,000) were originally separated into two groups on the basis of amino acid content and solubility (12), type I (KAP7 and 8), and type II (KAP6 family). These proteins are rich in glycine, tyrosine, serine, and phenylalanine, accounting for ~50 mol % of the amino acid content for KAP7 and 8 and ~77 mol % for KAP6 proteins. The glycine/tyrosine-rich KAP group is heterogeneous and may contain 15–30 components (13), although there is some speculation that the apparent heterogeneity may, in part, have arisen from the conditions of protein preparation and fractionation (14). This uncertainty can be resolved by molecular cloning of members of this family.

The glycine/tyrosine-rich KAPs vary in abundance from less than 3% in human hair and the wool of Lincoln sheep to 13% in Merino wool, and up to 30–40% in echidna quill (15). The wide range in the content of these proteins in wool and hair raises intriguing questions concerning the regulation and function of these proteins in the matrix structure of the fiber. Further, they are subject to variation through dietary, physiological, and genetic factors (7).

The determination of the amino acid sequences of KAP7 (14) and KAP8 (16) enabled the synthesis of specific oligonucleotide probes to isolate the corresponding sheep cDNAs and genes (17, 18). Recently, a partial amino acid sequence was derived from a KAP6 protein isolated from wool (7).

* This work was supported by a grant from the Wool Research Trust Fund on the recommendation of the Australian Wool Corporation. 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) M95718 and M95719.

"advertisement"

The abbreviations used are: IF, intermediate filament(s); KAP, keratin-associated protein(s); nt, nucleotide(s); bp, base pair(s); kb, kilobase pair(s).
Based on this sequence, an oligonucleotide probe was used to isolate a partial cDNA clone encoding a KAP6.1 protein from a sheep wool follicle RNA library. We report in this study the use of this clone for the isolation and sequencing of sheep and rabbit KAP6.1 genes, the investigation of the genomic organization of the sheep KAP6 family, and the expression patterns of the genes during hair follicle differentiation.

MATERIALS AND METHODS

Characterization of a Sheep KAP6.1 cDNA Clone—A sheep wool follicle poly(A)+ cDNA library (17) was screened with a 15-mer oligonucleotide mix (5'-(A)RTARTANCCRAANCC-3'; Fig. 2A, nt 232-246) complementary to the pentapeptide (Gly-Phe-Gly-Tyr-Tyr) at the carboxyl terminus of a sheep KAP6 protein (7). A strongly hybridizing cDNA clone, pKAP6.1, was isolated and purified. The cDNA insert encoded 80% of the coding sequence for a KAP6 protein, as identified by comparison with the protein sequence (7), a 3'-noncoding region of 296 bp and a poly(A) tail of 26 bp (Fig. 2A, nt 51-549). Both coding (Fig. 2A, nt 51-270) and 3'-noncoding probes (Fig. 2A, nt 295-503) were subcloned from a Bal-31 deletion clone of pKAP6.1. Unless stated otherwise, these probes were used in the library screenings, Southern and Northern blot analyses, and for the sheep in situ hybridizations.

Sheep Cosmid Library Screen—Approximately 1.5 genome equivalents of a sheep MuLV cosmid (6) were labeled with [32P]dCTP (Clontech) and screened by the colony hybridization method (19) with the sheep KAP6.1 3'-noncoding probe. From the four positive clones, one clone (KAP6 Cos1) was further purified by two additional rounds of rescreening. Cosmid DNA was prepared from 5-10 ml cultures grown overnight at 30 °C by standard methods (20).

Rabbit λ Library Screen—Approximately 5 genome equivalents of a rabbit genomic EMBL3 library (Clontech) were screened by the plaque hybridization method (21) using the sheep KAP6.1 coding probe. Of the 91 hybridizing λ clones, 12 were further purified with an additional three rounds of rescreening, and one was selected for further analysis. λ DNA was prepared from plate lysates (20).

Southern Blot Analysis—DNA restriction fragments to be used as probes were labeled with either α-32P[dATP and/or α-32P[dCTP (3,000 Ci/mmol, Bresatec, Adelaide, South Australia) by the oligolabeling method (22) using a Bresatec kit. Restriction enzyme digested DNAs, as well as HindIII-digested λ DNA and EcolI-digested SPP-1 DNA molecular weight markers (Bresatec), were electrophoresed on 0.7-1.0% agarose gels and transferred to Zeta-Probe or Zeta-Probe GT (Bio-Rad) in 0.4 M NaOH (23) by a vacuum blotting apparatus. The gels were transferred onto Zeta-Probe GT using capillary transfer (20) and 1 M NaOH, 0.02 M Na2HPO4 for 20 min. The gels were transferred onto Zeta-Probe GT using capillary transfer (20) and 1 M NaOH, 0.02 M Na2HPO4 as the transfer buffer. The filters were baked in vacuum at 80 °C for 2 h.

Tissue in Situ Hybridizations—In situ hybridizations on paraformaldehyde-fixed and sectioned sheep wool and rabbit hair follicle biopsies were performed as previously described (33). Final wash conditions were 0.1× SSPE at 55-60 °C for 30 min. The cRNA and cDNA probes were labeled with [32P]dCTP (33) and [α-32P]dUTP (34) using either T7 or SP6 RNA polymerase (34) with a kit (Bresatec). For the sheep in situ hybridizations, a pGEM-3Z(f+) clone containing the sheep KAP6.1 coding insert was linearized with HindIII and transcribed with T7 RNA polymerase to produce antisense RNA. The same clone was linearized with EcoRI and transcribed with SP6 RNA polymerase to produce sense RNA. For the rabbit in situ hybridizations, a pGEM-7Z(f+) clone containing the 1,169-bp EcoRI fragment encoding the rabbit KAP6.1 gene was linearized with HindIII and transcribed with T7 RNA polymerase to produce antisense RNA. For the sense RNA, a clone containing the same insert but in the opposite orientation was linearized with SacI (in vector polylinker), transcribed with SP6 RNA polymerase, and then subjected to partial alkaline hydrolysis (35).

Database Searches—The GenBank (36) and EMBL (37) nucleotide sequence databases were searched with the complete sheep and rabbit KAP6.1 gene sequences (Figs. 2A and 3) using the GOSTA algorithm (38).

RESULTS

Characterization and Sequence of a Sheep KAP6.1 Gene—A sheep cosmide clone was isolated with a 3'-noncoding probe from the cDNA clone pKAP6.1 (see "Materials and Methods"). Southern blot analysis of KAP6 Cos1 indicated that three EcoRI fragments of 9.9, 7.5, and 1.4 kb hybridized with the KAP6.1 coding probe, and only the 9.9-kb EcoRI fragment hybridized with the KAP6.1 3'-noncoding probe (Fig. 1). The 9.9-kb EcoRI fragment was subcloned and Southern blot analysis of this clone identified a 1.6-kb PstI fragment that hybridized with both the KAP6.1 coding and 3'-noncoding probes (data not shown). The nucleotide and deduced amino acid sequence of the KAP6.1 gene encoded by the 1.6-kb PstI fragment is shown in Fig. 2A. The nucleotide sequence of pKAP6.1 and the genomic clone were identical, indicating that the cosmide contained the genomic equivalent of the cDNA clone. The KAP6.1 gene, like the other IF-associated genes expressed in the hair follicle (6), does not contain introns. A number of general regulatory elements were identified in the noncoding and flanking regions of the gene (Fig. 2A). In addition, there were several conserved sequence motifs of 7-9 nucleotides that had been initially identified in the 5' flanking regions of the KAP7 and KAP8 genes, namely the HGT-1 and HGT-2 motifs (39).

From the nucleotide sequence, an mRNA of 610 nt is 2

E. Kuczek, unpublished experiments.

[2] R. Saint, personal communication.
1.4b

**FIG. 1.** Southern blot analysis of KAP6 Cosl. Cosmid KAP6 Cosl (250 ng) was digested with EcoRI, electrophoresed on a 0.7% agarose gel, transferred to Zeta-Probe GT and hybridized with the sheep KAP6 1.3'-noncoding probe (lane 1). The filter was given a final wash in 2 x SSC, 0.1% SDS at 65°C. The strip-washed filter was rehybridized with the sheep KAP6 coding probe (lane 2). It was given a final wash in 2 x SSC, 0.1% SDS at 65°C. The sizes of the hybridizing restriction fragments are given in kb. Note: the faint hybridization signal seen at 7.3 kb in lane 2 is most likely nonspecific since its intensity is approximately the same as the signal seen at 8.8 kb in lane 2 which is the cosmid vector, pWE15 (20).

predicted (exclusive of the poly(A) tail) encoding a basic protein of 82 amino acids (M, = 8,296) with a combined glycine (37.8 mol %) and tyrosine (22 mol %) content of ~60 mol %. The protein had a serine and cysteine content of 14.6 and 11 mol %, respectively, with leucine, arginine, phenylalanine, and asparagine accounting for the rest of the composition. Featured in the coding sequence were 9- and 12-amino acid repeats (Fig. 2B).

**Repetitive Elements Flank the Sheep KAP6.1 Gene—**A search of the nucleotide sequence databases identified two repetitive elements in the flanking regions of the sheep KAP6.1 gene (Fig. 2A). The repetitive elements had a 60/75 (80%) nucleotide match with each other. The repetitive element in the 5'-flanking region had a 185/207 (89.7%) nucleotide match with a repetitive element present in the first intron of the bovine corticotropin-1 lipotropin precursor gene (40). GenBank™ accession number J00018). The repetitive element in the 3'-flanking region contained the two split promoters that are characteristic of eukaryotic tRNA genes (41) and resembled some bovine and goat repetitive elements (42-44). Furthermore, a 44/74 (60%) match with the human glycine tRNA sequence (45) was identified (Fig. 2A, nt 610-685), and another part of the 3'-repetitive element (Fig. 2A, nt 707-767) showed identity to repetitive elements in the introns of other wool keratin genes (5, 46).

**Characterization and Sequence of a Rabbit KAP6.1 Gene—**To examine the evolutionary conservation of the KAP6 genes, a rabbit lambda genomic clone was isolated using the sheep pKAP6.1 coding probe (see "Materials and Methods"). A Southern blot filter of the lambda clone (KAP6 λ1) digested with EcoRI was hybridized with the sheep KAP6.1 coding probe, and a unique hybridizing restriction fragment of 1.2 kb was subcloned. The nucleotide and deduced amino acid sequence of the KAP6.1 gene encoded by the 1.2-kb EcoRI fragment is shown in Fig. 3. The gene does not contain introns. Many of the motifs noted in the 5'-noncoding and -flanking region of the sheep gene (Fig. 2A) were also identified in the same regions of the rabbit gene (Fig. 3).

From the nucleotide sequence an open reading frame was identified that encoded a basic protein of 79 amino acids (M, = 8,143) with a combined glycine (41.8 mol %) and tyrosine (26.6 mol %) content of ~66 mol %. The protein had a serine and cysteine content of 6.3 and 11.4 mol %, respectively, with leucine, arginine, phenylalanine, and asparagine accounting for the rest of the composition. Featured in the coding sequence were 9- and 12-amino acid repeats (Fig. 2B).

**FIG. 2.** Sequence of the sheep KAP6.1 gene. A, the nucleotide sequence of the 0.96- and 1.6-kb Psfl fragments containing, respectively, the upstream promoter elements and the sheep KAP6.1 gene is shown. The numbering begins at nt -1043 with respect to the methionine initiation codon (ATG). The derived amino acid sequence of the gene appears below the nucleotide sequence. The Kozak sequence (62) contiguous with the methionine initiation codon is underlined and the possible locations of the mRNA cap site are indicated by arrows above the appropriate nucleotide. Putative TATA and CAAT motifs, a possible general enhancer element, the CACC motif (a putative polyadenylation signal), and the T/G/T motif that are required for the efficient formation of mRNA 3' termini (64) are boxed and highlighted in reverse text. Note: there are additional putative TATA motifs at nt positions -160 to -155 (TAATTA) and -147 to -142 (CATAAA). The position of the polyadenylation site, as determined by the location of the poly(A) tail in pKAP6.1, is shown by an underlined bar. The SV40 core enhancer (65), AP-1 (66, 67), HGT-1 and HGT-2 motifs (39), and the A and B split promoters for DNA polymerase III are boxed. B, a 6-hp inverted repeat including the CACC and SV40 core enhancer homologies and capable of forming a stem-loop is underlined by apposing arrows. The position of the oligonucleotide primer (nt 232-246) that was used to isolate the pKAP6.1 is indicated by an arrow underneath the appropriate nucleotides. The repetitive elements in the 5' (nt -1043 to -838) and 3' (nt 610-816)-flanking regions are shown in lower case, and the region of identity between the two elements is indicated by lower case bold lettering. The nucleotide sequence in pKAP6.1 corresponds to nt positions 51-549 of the gene sequence. Note: The partial amino acid sequence derived from pKAP6.1 has previously been reported (68).

100% of the sequence was determined from both strands. B, the complete predicted sheep KAP6.1 protein sequence. The amino acid repeats are boxed.
Characterization of Genes Encoding Glycine/Tyrosine-rich KAPs

Fig. 3. Sequence of the rabbit KAP6.1 gene. The nucleotide sequence of the 1,169-bp EcoRI fragment containing the upstream promoter elements and the rabbit KAP6.1 gene is shown. The numbering begins at nt −867 with respect to the methionine initiation codon (+1). The derived amino acid sequence of the gene appears below the nucleotide sequence. The Kozak sequence contiguous with the methionine initiation codon is underlined and the possible locations of the mRNA cap site are indicated by arrows above the appropriate nucleotide. The putative TATA, CAAT, and CACCC motifs in the promoter region are boxed and highlighted in reverse text. Note: there are additional putative TATA motifs at nt positions −191 to −186 (CATAAA), −162 to −158 (TAATA), and −150 to −145 (CATAAA). The SV40 core enhancer, AP-1, HGT-1, and HGT-2 motifs (39) are boxed. The sequence capable of forming Z-DNA (69) is shown in bold letters. 100% of the sequence was determined from both strands. Note: from the 60 bp of 3' noncoding sequence no polyadenylation signal was found and it probably was located on an adjacent EcoRI fragment.

for the rest of the composition. The rabbit and sheep proteins are composed of the same 8 amino acids in equivalent proportions.

Sheep and Rabbit KAP6.1 Genes Are Derived from a Multigene Family—A Southern blot containing EcoRI-digested genomic DNA from three placental and three marsupial mammals was hybridized with the sheep KAP6.1 coding probe. Multiple hybridizing restriction fragments of varying intensities in all the genomic DNAs were detected (Fig. 4). The hybridization signal was much stronger in the genomic DNAs from the placental mammals, moderate in the possum (Trichosurus vulpecula) genomic DNA track and weak in the quoll (Dasyurus viverrinus) and wallaby (Macropus eugenii) genomic DNA tracks. There were 9 hybridizing EcoRI fragments in the sheep track, 8 in the human track, and more than 20 in the mouse track. A Southern blot of EcoRI-digested sheep genomic DNA hybridized with the sheep KAP6.1 3' noncoding probe showed that only the 9.9-kb EcoRI fragment hybridized to the probe, corresponding to the EcoRI fragment in the cosmid clone (data not shown). When a Southern blot of EcoRI-digested rabbit genomic DNA was hybridized with the sheep KAP6.1 coding probe, six hybridizing restriction fragments were detected (data not shown).

Pulsed Field Blot Analysis—To investigate the genomic organization of the sheep KAP6 genes, a pulsed field blot containing genomic DNA resolved in the 200–2,200-kb range was hybridized with a coding probe from the sheep KAP6.1 gene. The results showed that the KAP6 genes were contained within single SfiI (1,050 kb) and SalI, Sall, and NotI (1,800 kb) fragments (Fig. 5). No additional hybridizing restriction fragments were detected on a blot containing identically digested genomic DNAs resolved in the 32–200-kb range (data not shown).

Fig. 4. Mammalian Southern blot with the sheep KAP6.1 coding probe. EcoRI-digested genomic DNAs (4 μg) from placental mammals (sheep, human, mouse) and marsupial mammals (possum, T. vulpecula; quoll, D. viverrinus; wallaby, M. eugenii) were electrophoresed on a 0.8% agarose gel, transferred to Zeta-Probe and hybridized with the sheep KAP6.1 coding probe. The filter was given a final wash in 0.1 x SSC, 0.1% SDS at 65 °C. The sizes of the hybridizing restriction fragments are given in kb. Note: the end of the gel containing DNA <1 kb in size was accidentally lost.

Fig. 5. Pulsed field blot analysis. Approximately 15 μg of sheep genomic DNA, embedded in agarose beads, were digested with Sall, Sall, SfiI, and NotI and electrophoresed on a 1.0% agarose gel with a ramp pulse time of 50-90 s. The ethidium bromide-stained gel (A) was transferred onto Zeta-Probe GT and the filter was hybridized with a coding probe from the sheep KAP6.1 gene (Fig. 2A; 558-bp NotI fragment, nt −75 to 444 and 39 bp of the pGEM-5Zf(+) polylinker). The filter was given a final wash in 2 x SSC, 0.1% SDS at 65 °C. The sizes of the hybridizing restriction fragments (B), estimated from the migration of S. cerevisiae chromosomes (A, lane M), are given in kb.
Northern Blot Analysis of Sheep KAP6 Gene Expression in Wool Follicles—Sheep KAP6.1 coding and 3′-noncoding probes hybridized to a follicle RNA band(s) of ~650 bases (Fig. 6) in agreement with the predicted mRNA size derived from the gene. Under identical hybridization and washing conditions, and exposure times, a stronger hybridization signal was obtained with the coding probe than with the 3′-noncoding probe. Both probes were approximately the same size and labeled to approximately the same specific activity. This implied that the coding probe was detecting multiple KAP6 mRNAs with highly conserved coding regions but different 3′-noncoding regions.

Localization of Sheep and Rabbit KAP6 Gene Transcripts in Active Wool and Hair Follicles—To investigate the expression of the KAP6 family during follicle differentiation, in situ hybridizations were performed with coding region probes. These probes should detect most of the KAP6 mRNAs because the stringency conditions of the in situ hybridization were lower than those of the genomic Southern analysis (0.1 × SSPE at 60 and 65 °C, respectively) that detected multiple components (Fig. 4). KAP6 expression was first detected in differentiating hair shaft keratinocytes in sheep wool follicles and rabbit hair follicles at the same relative locations, about 200 μm above the proliferative zone of the follicle bulb (Figs. 7, A and B, and 8, A and B). In rabbit follicles two patterns of KAP6 gene expression were seen in longitudinal sections (Fig. 8, A–C). Either expression seemed to occur in all cortical cells, or it was restricted to cells in one part of the cortex. This variation was not obvious in the longitudinal sections of sheep wool follicles because of the difficulty in obtaining sufficient numbers of follicles showing a good plane of section, but examination of transverse sections confirms the presence of restricted expression patterns in sheep follicles (Fig. 7, D and F). Some of the sheep follicles show a small area of KAP6 expression in the cortex whereas in others nearly all the cells of the cortex express the genes (Fig. 7, C–F). In transverse sections of small diameter rabbit hair follicles (Fig. 8, E and G) KAP6 expression occurred in an arc of cells in the hair cortex, whereas in the larger guard hairs convoluted bands of expression were observed circling the multilobed medulla (Fig. 8, D–G). No expression was observed in the epidermis or any other hair follicle cell types (data not shown, but see “Discussion”).

Amino Acid Sequence Comparisons and Secondary Structure Analysis of the Glycine/Tyrosine-rich KAPs—The amino acid sequence derived from the sheep KAP6.1 gene (Fig. 2A) was compared with the partial sheep KAP6 amino acid sequence deduced by protein sequencing (7), hereafter referred to as KAP6.2, and with the sheep KAP7 and KAP8 gene-derived amino acid sequences (18). There were at least eight differences between the KAP6.1 and KAP6.2 sequences (Fig. 9A). The KAP6.1 amino acid sequence had 26 and 31% sequence identity with the KAP7 and KAP8 sequences, respectively (Fig. 9B). The KAP7 and KAP8 proteins shared a relatively low sequence identity of 28%. The longest stretch of identical amino acids between all three sequences was a Ser-Tyr-Gly peptide but there was a 7/8 match of amino acids at the carboxyl terminus of the KAP6.1 and the KAP7 se-
sequences (Fig. 9B). All three proteins contained a tyrosine residue at the carboxyl terminus. Despite the high content of glycine the sequence Gly-Gly is relatively uncommon, occurring twice in the KAP7 and KAP6.1 sequences (three times in the rabbit KAP6.1 sequence, see Fig. 9C) and not found in the KAP8 sequence. This is in contrast to the glycine-rich keratins of birds (47), reptiles (48), and human (49) and of mouse (50) epidermis.

The sheep and rabbit KAP6.1 proteins are highly conserved, exhibiting 11-amino acid changes and two insertions/deletions of 7 and 10 amino acids (Fig. 9C). If the 7- and 10-amino acid insertions/deletions in the rabbit and sheep proteins are excepted, the two genes had an 85 and 89% identity at the amino acid and nucleotide sequence level, respectively. An analysis of the sheep KAP6.1, KAP7, and KAP8 amino acid sequences using a suite of 10 protein secondary structure prediction programs (51) indicated that all three proteins contained a high degree of random turns, some β-sheets, and no α-helices (data not shown). The glycine-tyrosine-rich KAPs may contain the structural motif termed the glycine loop, a structure proposed to exist in glycine-rich regions of proteins (52).

**DISCUSSION**

We have described the molecular characterization and expression patterns of sheep and rabbit KAP6.1 genes during...
Characterization of Genes Encoding Glycine/Tyrosine-rich KAPs

4517

hair follicle differentiation. The sheep and rabbit genes show a high degree of sequence identity, both at the nucleotide and amino acid levels, which indicates a strong selection pressure during evolution, estimated to be ~20 million years for the sheep and rabbit genomes (53). It further suggests that the conserved regions in the two proteins have specific functions, possibly involving interactions with the hair keratin IFs of the cell cortex and the other keratin IF-associated proteins.

KAP6 Multigene Family—Both the sheep and rabbit KAP6.1 genes are derived from multigene families (Fig. 4, and data not shown). Since all the hair IF-associated genes sequenced thus far lack introns, including the two KAP6.1 genes described in this report, and the coding regions are usually less than 1 kb in size (6), it is possible that each EcoRI fragment detected in the sheep genomic Southern blot (Fig. 4) represents a KAP6 gene. All the sheep KAP6 genes appear to be contained within a 1,050-kb SfiI fragment (Fig. 5). The sheep wool keratin genes from the various families tend to occur in gene clusters (1–5), and experiments are now in progress to determine the genomic organization of these gene families by chromosomal in situ hybridization and pulsed field blot analysis.

Northern blot analysis suggests that more than one KAP6 gene is expressed in the follicle (Fig. 6). The KAP6 protein sequenced by Gillespie (7) exhibited several amino acid differences to the sheep KAP6.1 protein (Fig. 9A) and most likely represents a different member of the KAP6 family which we suggest be referred to as KAP6.2. Autoradiographs of two-dimensional polyacrylamide gel patterns of S-carboxymethylated wool keratin proteins (6) reveal multiple protein spots in the glycine/tyrosine-rich KAP region. One estimate of the complexity by protein chemical methods is 15–30 components (13), although there is some speculation of artefactual heterogeneity produced by the preparative procedures (14). Since the KAP7 and KAP8 proteins are encoded by unique genes (18) and there may be several KAP6 genes, a molecular genetic estimate of the number of sheep glycine/tyrosine-rich KAPs could be about 10.

Glycine/tyrosine-rich KAPs constitute about 19% of mouse hair protein (7). The mouse genome contains unique KAP7 and KAP8 genes (54) and up to 20 KAP6-hybridizing EcoRI restriction fragments (Fig. 4) possibly representing an equivalent number of genes (see above). In contrast, the glycine/tyrosine-rich KAP content of human hair appears to be less than 3% (15), yet the human genome contains unique KAP7 and KAP8 genes (54) and probably several KAP6 genes (Fig. 4). A number of possible explanations may account for this wide variation in the glycine/tyrosine protein content between the hairs of different species. It is well known that proteins are difficult to extract from human hair, extractions sometimes resulting in yields of less than 5% of total protein (55), and thus the low glycine/tyrosine protein content may be an underestimate. Alternatively, human hair may simply contain a greater proportion of other keratin proteins, and the glycine/tyrosine-rich KAP content is indeed low. Some of the genes may differ in transcription rates between species. The human glycine/tyrosine-rich KAP genes may be expressed at a low level, but in the mouse genome, which could contain up to twice the number of KAP6 genes, some KAP6 genes may be expressed at high levels. Another interesting finding concerns a sheep wool fiber mutant known as the felting luster mutant. It produces a wool that is virtually devoid of glycine/tyrosine-rich KAPs (56), yet the KAP7 and KAP8 genes that are abundantly expressed in Merino breed crosses (18) are present in the felting luster genome (54). If the KAP6 gene family is also present that could have interesting implications for the control of expression of those genes, particularly as the genes appear to be clustered. These two issues can now be addressed by comparative analyses of glycine/tyrosine-rich KAP gene numbers and expression by Northern blot analysis and in situ hybridization with conserved KAP6, 7, and 8 gene probes.

Expression of the Sheep and Rabbit KAP6 Genes in Active Hair Follicles—Hair shaft keratinocytes begin to express differentiation-specific hair keratins as soon as they leave the proliferative region of the follicle bulb (5, 6). The KAP6 genes are expressed in the hair shaft cortical keratinocytes (Figs. 7 and 8) at a considerable distance (200 μm) above the proliferative zone of the follicle bulb and after the transcription of the hair keratin IF genes has commenced (5). This is a relatively late stage in the differentiation of the hair shaft keratinocytes (6). It is not possible to conclude whether the KAP6 genes are also expressed in the cuticle cells because the cuticle layer of these hairs is only one cell thick and, at the position in the follicle where cortical KAP6 expression occurs, the cuticle cells are becoming compressed and are only a few microns wide. However, if expression does occur in the cuticle cells it must be restricted to those adjacent to the cortical cells that show expression. This would suggest local variation in the differentiation of cuticle cells within a follicle, but in view of the uniform ultrastructure of cuticle cells (57) this seems unlikely. As the KAP6 family has the same expression characteristics in two evolutionarily well separated species, it is likely that these data are typical of KAP6 expression in mammals in general.

An intriguing feature of KAP6 gene expression in hair follicle differentiation is the variation in spatial expression. In some follicles expression appears to be restricted to a few cortical keratinocytes whereas in adjacent follicles most cortical cells express the genes (Figs. 7, D and F, and 8, E and G). Furthermore, the observation of asymmetric expression along the hair shaft of some rabbit hair follicles (Fig. 8C) suggests that the proportion of cortical cells that express KAP6 is fixed within a follicle but variable between follicles. To establish any underlying theme it will be necessary to examine the expression of the KAP6 gene family in a series of transverse sections covering the length of many follicles. In sheep follicles there are two major types of cortical keratinocytes, orthocortical and paracortical cells, whose proportions and spatial arrangement are believed to be responsible for the crimp or curliness of wool fibers (6). KAP6 expression might be confined to one of these cell types.

Conserved Motifs in the 5'-Noncoding and -Flanking Regions—The 5'-noncoding and -flanking regions of the sheep and rabbit KAP6.1 genes are moderately conserved with ~70% identity over 394 nucleotides (compare Figs. 2A and 3). The position and sequence of several general regulatory and glycine/tyrosine-rich KAP-specific motifs are conserved in the 5'-noncoding and -flanking regions of both genes, suggesting that some of these motifs may have functional roles in the regulation of transcription or translation.

The putative TATA motif of the sheep gene (TACAAA; Fig. 2A) is identical to that present in the hamster desmin gene (58). Other putative TATA motifs were noted in the promoter region of the sheep and rabbit genes about 50–100 bp further upstream in both cases (Figs. 2A and 3). However, from an examination of the promoter regions of several hair keratin genes (1–5) these putative TATA motifs are further from the initiation codon than expected and furthermore, no CAAT motifs were identified further upstream from them. RNase protection analyses would be required to determine the functional TATA motifs.

Fisher et al. (59) have shown by immunohistochemical
means that Fos protein is present in the differentiating cells of the rat hair follicle and they suggest that Fos plays a role in keratinization. The proteins Fos and Jun form a common transcriptional activator complex that binds to AP-1 motifs (60), and since the sheep and rabbit KAP6.1 genes and a hair keratin type II IF gene (5) contain AP-1 motifs, a role for AP-1 in hair keratin gene transcription is implicated.

Among the hair keratin genes several conserved DNA sequences have been noted, and the HGT-1 and HGT-2 motifs appear to be unique to the glycine/tyrosine-rich KAP genes (39). These motifs may be involved in determining the timing and spatial pattern of KAP6.1 gene expression during hair follicle differentiation and suitable experimental strategies, such as deletion and footprint analyses and oligonucleotide binding studies (61), can now be initiated to test them.

Acknowledgments—We thank Dr. Elizabeth S. Kuczek (University of California, San Francisco) for the provision of pKAP6.1, Toni Neski for help with the in situ hybridizations, Juliana Beltrame for the pulsed field blots, Michael Calder for photographic assistance, and Dr. David A. Parry (Massey University, New Zealand) for helpful discussions.