An Ultra-high Sulfur Keratin Gene is Expressed Specifically During Hair Growth

Alistair R. McNab, Linda Wood, M.S., Nicole Theriault, Ph.D., Todd Gierman, B.S., and Gabriel Vogeli, Ph.D.

Molecular Biology Research (ARM, LW, GV), Biopolymer Chemistry (NT), and Cardiovascular Diseases Research (TG), The Upjohn Company, Kalamazoo, Michigan, U.S.A.

To study the regulation of the hair cycle in the mouse, we have isolated and characterized a gene for ultra high sulfur keratin that is expressed specifically during the active hair growth cycle. The gene (gUHSK-704Eco) was isolated as a member of a gene cluster on a recombinant phage with a DNA insert of 18 kb that was isolated by screening a murine genomic library at low stringency with a synthetic oligonucleotide derived from a sheep high sulfur keratin gene (Powell, Nucleic Acids Res. 1983 11, 5327). The murine ultra-high sulfur keratin gene has no intervening sequence; the 558 nucleotide of the coding region specify 186 amino acids, of which 70 (37%) are cysteine. A Cys-Cys-Gln-Pro repeat is found 12 times within the coding region. RNA dot blots show that the ultra-high sulfur keratin gene is expressed during the hair cycle concomitant with the anterior-posterior temporal pattern of the normal murine hair cycle. *J Invest Dermatol 92:263–266, 1989*

Mammalian hair is formed from epithelial cells that undergo terminal differentiation including complete cornification [1]. The ultra-high sulfur keratin proteins are related to the high sulfur keratin proteins both in sequence and proposed function [for review see Refs 2 and 3]. These high sulfur keratins accumulate together with other hair specific proteins within the meshwork of the cytokeratins that form the intermediate filaments of epithelial cells [4,5]. The ultra-high sulfur proteins are predominantly found in the cuticle with some also found in the cortex, whereas the high sulfur proteins are found in the hair cortex [6]. The ultra-high sulfur and the high sulfur keratins provide the cysteines for the cross-linkage of the hair proteins to form intracellularly the physical and chemical inert structure of the hair.

In mice, waves of hair growth progress from head to tail along the body [7]; when the mice grow older these waves give way to patchy hair growth. During hair growth, epidermal cells in the hair follicle shift their biosynthesis from skin specific cytokeratins and basement membrane components to hair specific proteins, under the inductive influence of the dermal papilla [8,9]. To study this inductive influencing, we are isolating and characterizing molecular probes involved in these changes. Because we already have molecular probes for the analysis of basement membrane biosynthesis [10,11], we have isolated a molecular probe specific for a hair protein in order to analyze at the molecular level mechanisms regulating the biosynthesis of hair. Here we report the isolation and characterization of an ultra-high sulfur keratin gene and demonstrate that it is specifically expressed during hair growth.

Manuscript received April 20, 1988; accepted for publication August 19, 1988.

Reprint requests to: Alistair McNab, Molecular Biology Research, 7242-267-5, The Upjohn Company, Kalamazoo, MI 49001.

Abbreviations:
aa: amino acid(s)
C: carboxy terminal
cDNA: DNA complementary to mRNA
EDTA: N,N'-1, 2-Ethanediylbis[N-(carboxymethyl)glycine]
N: amino terminal
SDS: sodium dodecyl sulfate

MATERIALS AND METHODS

Screening of the Genomic Mouse DNA Library The oligonucleotide AMC-16 (5'-GCAGGTGGGCTGGCAGCAGCAGGCTGGGCGGCAGCA) used to isolate clones from a mouse genomic library was derived from nucleotide 854 to 890 of the published sequence B2A from the sheep [12]. The methods used for plating and screening of the genomic DNA library (a gift of Dr. Seidman, NIH), and isolation and purification of phage DNA have been published [13,14]. In short, 10,000 λ Charon 4A recombinant clones were plated onto 15 cm dishes and transferred to NEN colony hybridization membrane (Dupont). The filters were treated following the manufacturer's recommendations. They were prehybridized and hybridized at low stringency in 0.9 M NaCl, 0.2 M Tris-HCl pH 7.4, 20 mM EDTA pH 7.5, 30% formamide, 20 mM Na-phosphate buffer pH 7.4, 1% SDS, 50 μg/ml each of poly C and poly A (Collaborative Research), 1 mM ATP (Sigma), 50 μg/ml denatured Salmon testis DNA (Sigma) and 100 μg/ml denatured E. coli DNA (Sigma, EC-DNA). The hybridization mixtures also contained 10% Dextran sulfate. A maximum of 20 filters were transferred singly into the pre-hybridization and the hybridization mixture before being sealed into plastic bags. Hybridization with 5' end-labeled oligonucleotide probe (γATP crude from ICN: 10^8–10^9 cpm/ug, 500,000 cpm [32P]/ml) was at 42°C for 16 h. All filters were washed in 0.1 × SSC (15 mM NaCl, 1.5 mM Na-Citrate, pH 7.2), 0.1% SDS at room temperature for low stringency. High stringency washes were in 0.1 × SSC, 0.1% SDS at 51°C. The filters were exposed at −70°C with Lightening Plus Screens (Dupont) to XAR-5 x-ray film (Kodak) that had been pre-flashed. The areas of the positive colonies were cut out from the filter, eluted with 1 ml LB broth and dilutions were plated again to purify the positive clones. We found two genomic clones, only one (gUHSK-704Eco) was further analyzed and sequenced.

DNA Isolation, Subcloning, and Sequencing The λ phage gUHSK-704Eco was grown in E. coli strain LE392 and the phage DNA was purified by standard protocols [13,14]. The subclone gUHSK-704Eco-pUC (-M13) was made by ligating a 2 Kb Eco R1 fragment into the plasmid pUC 13 (BRL) and the sequencing phage M13mp18 (BRL), respectively. The bacteria infected with the re-

0022-202X/89/$03.50

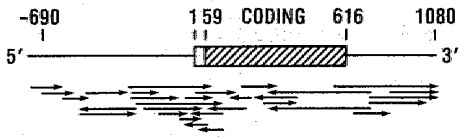

Figure 1. Sequencing strategy and the position of the gene within the subclone gUHSK-704Eco-pUC. The numbering of the nucleotides starts with the prospective RNA transcription site. The 5′ untranslated region is *boxed;* the translated region is boxed and *shaded.*

combinant plasmid was grown to stationary phase in modified super-broth containing per liter 24 g Yeast extract, 12 g tryptone, 5 ml glycerol, 2 g uridine, 25 mg of ampicillin, and 0.1 M K-phosphate buffer pH 7.5. No chloramphenicol amplification was done. The plasmid DNA was isolated using a modified protocol of Marko et al [15] followed by two CsCl density gradients. Sequencing of gUHSK-704Eco (Fig 1) was done with the chemical degradation method [16] and the dideoxy sequencing method on single stranded and double stranded DNA [17,18]. The oligonucleotide used as primers were used directly after de-protection and lyophilization without further purification.

RNA Isolation and Dot Blot Analysis Bald/c mice of different ages were killed, the skin removed, and the RNA from the skin was isolated following the published procedure [19]. The skin of 3-d-old mice was divided into three regions before RNA extraction. The RNA was denatured with formaldehyde and formamide and was applied as dot blots onto NEN gene-screen plus (Dupont) following the manufacturer's protocol and hybridized as described above. The oligonucleotide AMC-12 (5′-GTGCGGCAGCAGCTG-CTCACACAGCATGGCTGGAAGCAGGTC) and AMC-48 (5′-CAGCAGGTCTGAGACTAAATTATTCTATTCCAGTA-GAATAAT) that was used to hybridize after end-labeling with the dot blots was derived from position 232–273 and from position 626–664, respectively (Fig 2).

RESULTS

Nucleotide and Derived Amino Acid Sequence of an Ultra-high Sulfur Keratin Gene A murine genomic library was screened with a synthetic oligodeoxynucleotide (AMC-16) that had been derived from position 854 to 890 of a published DNA sequence of the B2A high sulfur keratin from sheep [12]. A recombinant λ phage that was isolated contained three genes as seen by hybridization to DNA Southern blots (data not shown). After subcloning of Eco R1 restriction DNA fragments, we sequenced the

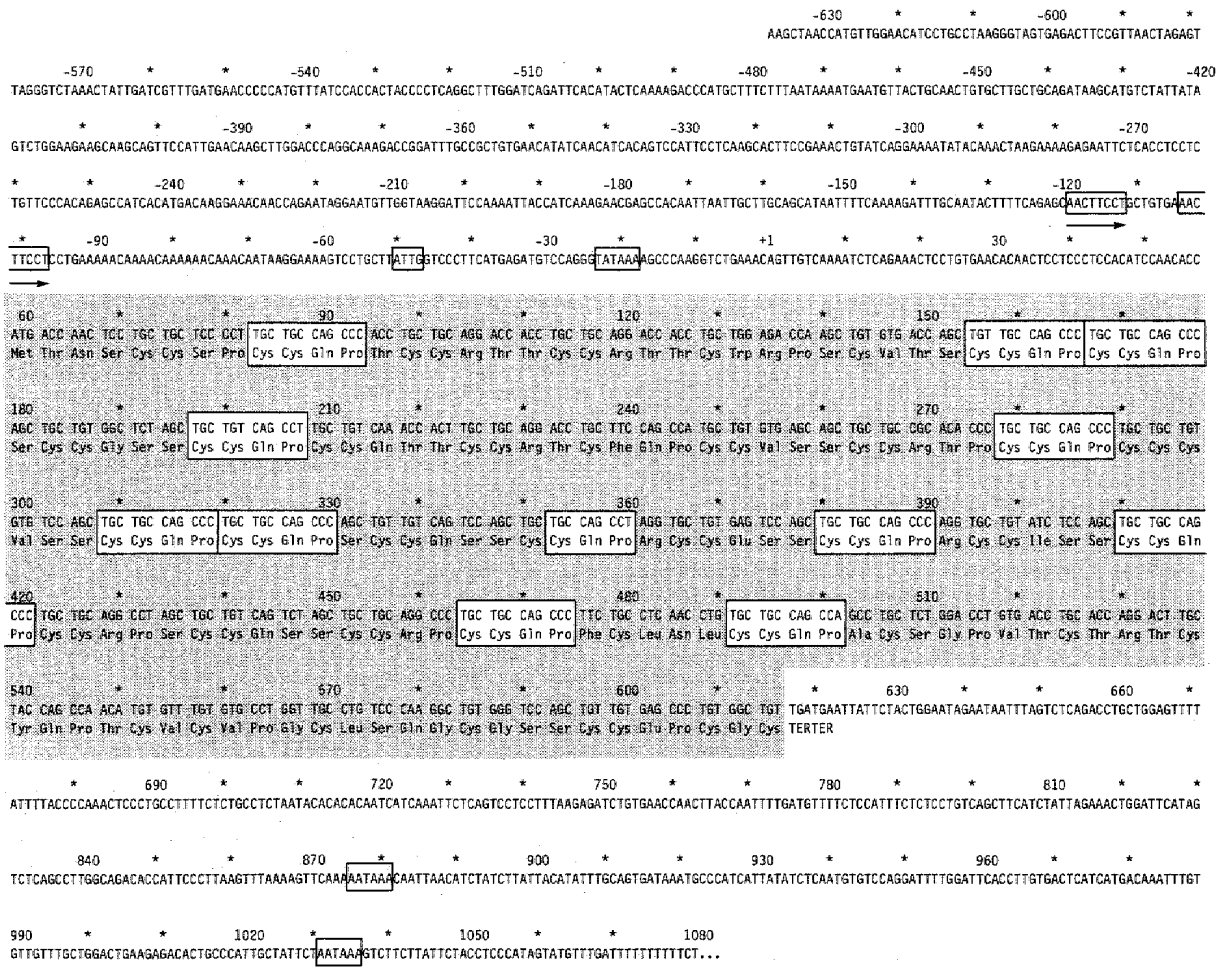

Figure 2. The nucleotide sequence and the derived amino acid sequence of gUHSK-704Eco-pUC. The coding region is *shaded;* the Cys-Cys-Gln-Pro repeat is *boxed.* In the 5′ prospective promoter region the elements (CAAT and TAAT) that are usually indicative of a eukaryotic promoter are *boxed.* Two 8 bp repeats in the prospective promoter region between position −120 to −95 are also *boxed.*

subclone gUHSK-704Eco (sequencing strategy see Fig 1). The DNA sequence and the derived amino acid sequence is shown in Fig 2. The gene has no intervening sequences. 558 nucleotide codes for 186 amino acids with 37% cysteine, 13% serine, 11% proline, and 9% glutamine. A Cys-Cys-Gln-Pro repeat is found 12 times within the coding region. DNA sequence analysis shows at the 5′ end of the coding sequences the elements (TATAA and CAAT box) expected from an eukaryotic promotor region. At the 3′ side of the gene, after the translation termination, we find the expected poly A addition sites. The oligonucleotide AMC-16, used to isolate the ultra-high sulfur keratin genes, has 80% similarity at the nucleotide level at sequences in position 260–295 (see Fig 2):

```
     260       270       280       290
GAGCAGCTGCTGCCGCACACCCTGCTGCCAGCCCTGCTGCTGTGT
    ::::: :::    :    ::::::::::::   ::::
   3'-ACGACGGCGGGTCGGACGACGACGGTCGGGTGGACG-5' AMC-16
```

Accumulation of mRNA for the Ultra-high Sulfur Keratin Gene To measure the amount of ultra-high sulfur keratin mRNA, total RNA isolated from skin was hybridized on dot blots with the synthetic oligonucleotide AMC-12 (see Fig 3*a*). The oligo-nucleotide AMC-12 was derived from the complementary DNA sequence at position 232–273 (Fig 2). Figure 3*a* shows that the intensity of the labeling and thus the amount of ultra-high sulfur keratin mRNA increases 3 d after birth. The amount of mRNA reaches a peak at day 9, stays at a steady level until day 18, and then declines as the hair cycle comes to an end at around day 22. Thus the amount of mRNA for ultra-high sulfur keratin in total skin follows the time course of the hair cycle.

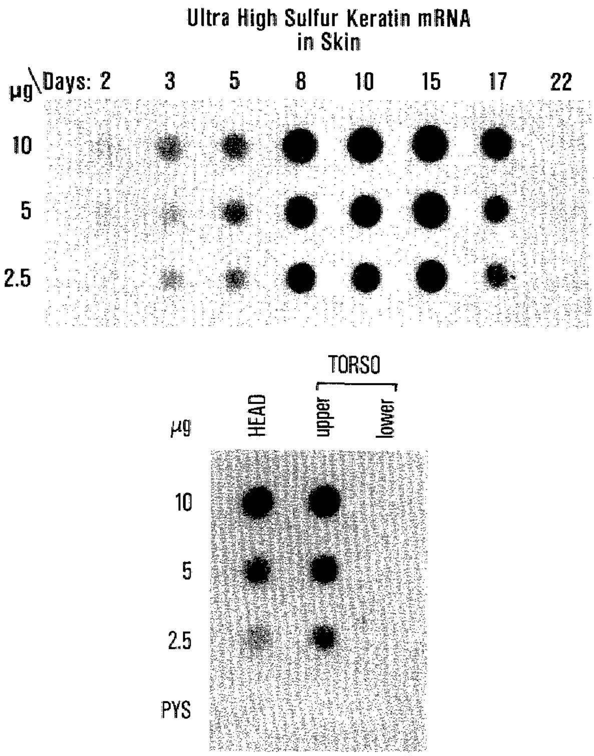

Figure 3. RNA dot blot from skin of mice. Total RNA isolated from the skin of young mice was applied onto NEN gene screen plus and hybridized with the labeled oligonucleotide AMC-12. *a*: Whole skin RNA was applied, and the age is listed above the dots. The amount of RNA loaded is listed. *b*: The skin of a 3-d-old mouse was divided into three parts (Head, upper torso, and lower torso) and applied to dot blots. *PYS* is a lane with control RNA.

Figure 3*b* shows the amount of ultra-high sulfur keratin mRNA in total RNA from skin isolated from different positions of a mouse at day 3. The spatial patterns in mice of other ages was not analyzed. At 3 days, the hair cycle progressed to where the expression of the ultra-high sulfur keratin gene was just starting (Fig 3*a*). Figure 3*b* shows that the scalp has substantial mRNA for ultra-high sulfur keratin, and so does the shoulder region, whereas the skin of the lower back has no hybridization to ultra-high sulfur keratin mRNA.

DISCUSSION

To study hair growth at the molecular level we are isolating molecular probes for hair specific proteins. Because hair is a derivative of the skin, many of the proteins found in hair are also present in skin [3]. To overcome this problem, we have isolated a murine ultra-high sulfur keratin gene and show that it is expressed specifically in the hair during the hair cycle.

Ultra-high Sulfur Keratin The predicted amino acid sequence of the genomic clone gUHSK-704Eco fits well into the proposed domain structure of ultra-high sulfur hair keratin proteins [3]. The foremost feature of our gene is the high cysteine content (37%), the presence of a 4 amino acid repeat unit (Cys-Cys-Gln-Pro), and the complete absence of any introns. A number of genes in higher eukaryotes without introns have been reported; for example, for histones [20], the adrenergic receptor [21], calmodulin [22], and also significantly for high sulfur sheep keratins [2,3,12] and high glycine-tyrosine keratins [23]. The genomic organization of the murine ultra-high sulfur keratin gene resembles the high sulfur genes of sheep; neither has intervening sequences and both have four amino acid repeats that are related (Cys-Cys-Gln-Pro compared to Cys-Cys-Arg-Pro in sheep) [3,12]. Also, the partial cDNA clone for an ultra-high sulfur gene (K4) from sheep [3] shows a high (68%) similarity at the nucleotide level with the mouse gene. These overall similarities account for the fact that we isolated the ultra-high sulfur keratin gene with an oligonucleotide probe (AMC-16) derived from a published high sulfur keratin sequence [12]. Thus the gene contained in gUHSK-704Eco belongs to the group of ultra-high sulfur keratins [3].

Prospective Promoter Elements and Untranslated Region By DNA sequence analysis, we have identified the prospective promoter region of this gene and compared it with other eukaryotic promoter regions [24]. At position −50, we find an inverted CAAT box (5′-ATTG) going in the opposite direction; such CAAT elements have been reported for a variety of promoters [24] and seem to be functionally involved in the activity of the promoter. At the 3′ end of the gene we find the expected poly A addition sites. If these sites actually serve as poly A addition sites was not analyzed.

Hair Specific Expression of Ultra-high Sulfur Keratin Genes gUHSK-704Eco was isolated as one of three genes contained in a 18 Kb DNA genomic insert (data not shown). We have not yet determined if the ultra-high sulfur keratin gene family is more complex. Because the Southern analysis with AMC-12 shows cross hybridization with the three genes present in the initial λ clone we had to determine if clone gUHSK-704Eco was actively expressed or if it was an inactive pseudogene [22]. To show that gUHSK-704Eco is actively expressed, we used a specific oligonucleotide derived from the unique sequence of the 3′ untranslated region (AMC-48, position 626–664) as hybridization probe. We showed that AMC-48 is unique for gUHSK-704Eco by partial sequence analysis of the other two genes (data not shown). When AMC-48 was hybridized to the dot blot filters (Fig 3), the autoradiograms (data not shown) showed exactly the same pattern as presented in Fig 3. Thus, the genomic clone gUHSK-704Eco is actively expressed during the hair cycle. The determination of the level of expression of the three genes has to await the complete sequence analysis of the other members of this gene family. Our results demonstrate that mRNA specific for gUHSK-704Eco accumulates in the skin only at times when hair organogenesis is well

under way and mRNA accumulation stops when active hair growth terminates.

The proof that ultra-high sulfur keratin (gUHSK-704Eco) is expressed specifically in hair stems from the dot blot pattern that is restricted to the skin with actively growing hair follicles. In areas where the hair is not growing, no mRNA for ultra-high sulfur keratin is detected. Thus, the clone gUHSK-704Eco provides a molecular probe to separate events during the hair growth cycle from events in interfollicular skin.

We would like to thank Dr. Seidman (NIH) for the genomic mouse library, Lee Giordano for technical assistance, and Nancy Burr for the preparation of the manuscript.

REFERENCES

1. Chase HB: Growth of the hair. Phys Rev 34:113–126, 1954
2. Rogers GE: Genes for hair and avian keratin. Ann NY Acad Sci 455:403–425, 1985
3. Powell BC, Rogers GE: Hair keratin: Composition, structure and biogenesis. In: Breiter-Hahn J, Matoltsy AG, Richards KS (eds.). Biology of the Integument. Springer-Verlag, Berlin, vol 2, pp 696–721
4. Marshall RC: Characterization of the proteins of human hair and nail by electrophoresis. J Invest Dermatol 80:519–524, 1983
5. Chapman RE, Gemmell RT: Stages in the formation and keratinization of the cortex of the wool fiber. J Ultra Res 36:342–354, 1971
6. Ley KF, Crewther WG: The proteins of wool cuticle. In: 6th Int. Wool Text Res. Conf., S. Afr. Wool Text Res. Inst., Pretoria, 1980, vol II, pp 13–28
7. Silvers WK: The Coat Colors of Mice. Springer Verlag, New York, 1979, pp 45–124
8. Jahoda CAB, Horne DA, Oliver RF: Induction of hair growth by implantation of cultured dermal papilla cells. Nature 311:560–562, 1984
9. Kollar EJ: The induction of hair follicles by embryonic dermal papillae. J Invest Dermatol 55:374–378, 1970
10. Vogeli G, Horn E, Carter J, Kaytes PS: Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen. FEBS lett 206:29–32, 1986
11. Wood L, Theriault N, Vogeli G: cDNA clones completing the nucleotide and derived amino acid sequence of the $\alpha 1$ chain of basement membrane (type IV) collagen from mouse. FEBS lett 227, 5–8, 1988
12. Powell BC, Sligh MJ, Ward KA, Rogers GE: Mammalian keratin gene families: organization of genes coding for the B2 high-sulfur proteins of sheep wool. Nucl Acids Res 11:5327–5346, 1983
13. Vogeli G, Kaytes PS: Amplification, storage, and replication of libraries. Methods Enzymol 1987: 152:407–415
14. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, New York, 1982
15. Marko MA, Chipperfiled R, Birnboim HC: A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder. Anal Biochem 121:382–387, 1982
16. Maxam AM, Gilbert W: Sequencing end-labeled DNA with base-specific chemical cleavages. In: Grossman L, Moldave K (eds.). Methods of Enzymology, Academic Press, New York, 1980, vol 65, pp 499–560
17. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467, 1977
18. Zagursky RJ, Baumeister K, Lomax N, Berman ML: Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. Gene Anal Techn 2:89–94, 1985
19. Chomczynski P, Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159, 1987
20. Sittman DB, Chiu IM, Pan CJ, Cohn RH, Kedes LH, Marzluff WF: Isolation of two clusters of mouse histone genes. Proc Natl Acad Sci USA 78:4078–4082, 1981
21. Kobilka BK, Frielle T, Dohlman HG, Bolanowski MA, Dixon RA, Keller P, Caron MG, Lefkowitz RJ: Delineation of the intronless nature of the genes for the human and hamster $\beta 2$-adrenergic receptor and their putative promoter regions. J Biol Chem 262:7321–7327, 1987
22. Nojima H, Sokabe H: Structure of rat calmodulin processed genes with implications for a mRNA-mediated process of insertion. J Mol Biol 190:391–400, 1986
23. Kucek ES, Rogers GE: Sheep wool (glycine + tyrosine)-rich keratin genes. A family of low sequence homology. Eur J Biochem 166, 79–85, 1987
24. Dorn A, Bollekens J, Staub A, Benoist C, Mathis D: A multiplicity of CCAAT Box-binding proteins. Cell 50:863–872, 1987