

The cDNA-deduced Amino Acid Sequence for Trichohyalin, A Differentiation Marker in the Hair Follicle, Contains a 23 Amino Acid Repeat

Michael J. Fietz, Richard B. Presland, and George E. Rogers

Commonwealth Centre for Gene Technology, Department of Biochemistry, University of Adelaide, Adelaide, South Australia, 5001 Australia

Abstract. Trichohyalin is a highly expressed protein within the inner root sheath of hair follicles and is similar, or identical, to a protein present in the hair medulla. In situ hybridization studies have shown that trichohyalin is a very early differentiation marker in both tissues and that in each case the trichohyalin mRNA is expressed from the same single copy gene. A partial cDNA clone for sheep trichohyalin has been isolated and represents ~40% of the full-length trichohyalin mRNA. The carboxy-terminal 458 amino acids of trichohyalin are encoded, and the first 429 amino acids consist of full- or partial-length tandem repeats

of a 23 amino acid sequence. These repeats are characterized by a high proportion of charged amino acids. Secondary structure analyses predict that the majority of the encoded protein could form α -helical structures that might form filamentous aggregates of intermediate filament dimensions, even though the heptad motif obligatory for the intermediate filament structure itself is absent. The alternative structural role of trichohyalin could be as an intermediate filament-associated protein, as proposed from other evidence (Rothnagel, J. A., and G. E. Rogers. 1986. *J. Cell Biol.* 102: 1419-1429).

THE mammalian hair follicle is a derivative of the epidermis that develops within the dermal layer of the skin. In the follicle bulb, the epidermal cells surround a dermal papilla that is essential for the initial development of the follicle and the growth of the fiber (Oliver and Jahoda, 1989). The positioning of the epidermal cells relative to the papilla determines the follicle layer that they will form, namely the different hair components, the cuticle, cortex, and medulla, and the three layers of the inner root sheath (IRS),¹ Henle, Huxley, and IRS cuticle, which accompany the developing fiber in its outward growth.

The IRS is a cylinder of cells surrounding the developing fiber and is believed to fulfill a structural role within the follicle, supporting and directing the fiber cells (Straile, 1965). The basal IRS cells are positioned near the periphery of the follicle bulb, and the daughter cells mature as they move up the follicle hardening into a rigid sheath that eventually degenerates by an unknown process before reaching the surface of the skin. Electron-dense nonmembrane-bound trichohyalin granules appear at a very early stage within all of the developing IRS cells and contain the protein trichohyalin which, between species, has been found to vary in size from

190 to 220 kD (Rothnagel and Rogers, 1986). As the cells move up the follicle, 8-10-nm-diameter filaments appear in close association with the trichohyalin granules (Rogers, 1964). Finally, the granules disappear, the IRS cells become completely filled with intermediate-like filaments aligned parallel with the direction of hair growth (Rogers, 1964), and these filaments harden into the insoluble contents of the mature IRS cell (Birbeck and Mercer, 1957).

The insoluble proteins within the mature cells of the IRS are readily distinguishable from those of the hardened α -keratin in the hair fiber (Rogers, 1964). The IRS proteins are cross-linked by γ -(ϵ -glutamyl)lysine bonds, formed between glutamine and lysine residues, instead of the disulphide bonds typical of keratin in the hair cortex (Harding and Rogers, 1971). In addition, the IRS proteins also contain the amino acid citrulline, which is posttranslationally formed from arginine by peptidylarginine deiminase (Rogers et al., 1977).

The medulla, when present in hair, is a central core of cells within the fiber that develops from cells covering the dome of the dermal papilla. The maturation process of these cells is similar to that of the IRS cells with one important distinction. Nonmembrane-bound granules, which are immunologically cross-reactive with the IRS trichohyalin (Rothnagel and Rogers, 1986) and thus termed trichohyalin granules, begin to appear within the developing medulla cells soon after they move up from the papilla. The trichohyalin granules finally coalesce to form the interior of the hardened medulla

R. B. Presland's current address is Department of Periodontics, Oral Biology and Medicine/Dermatology, University of Washington, Seattle, WA 98195.

1. *Abbreviations used in this paper:* IF, intermediate filaments; IRS, inner root sheath.

cells (Parakkal and Matoltsy, 1964). Protein-bound citrulline and γ -(ϵ -glutamyl)lysine cross-links, typical of the IRS cells, are also present in the hardened medulla cells (Steinert et al., 1969; Harding and Rogers, 1971, 1972). The essential difference between the IRS and medulla is that the hardened medulla cells are filled with an amorphous protein mass and not the oriented filamentous structures of the IRS.

It is presently unknown what causes the visible differences in the structure of the medulla and the IRS. If trichohyalin is the major precursor in both cell types, it is possible that either the two tissues contain very similar but nevertheless differing forms of trichohyalin or that trichohyalin contains a region capable of folding into an intermediate filament (IF) structure but which is only able to do so within the developing IRS. Alternatively, IF proteins could be expressed within the developing IRS and are cross-linked to the trichohyalin to form the hardened tissue.

In the present paper, we report the isolation and characterization of a cDNA clone encoding a portion of sheep trichohyalin. To determine the probability of trichohyalin forming IFs, the resultant protein sequence was examined for predicted secondary structure and for similarity to known conserved IF sequences. We have also used in situ hybridization analysis to localize trichohyalin expression within the follicle and to examine the relationship of mRNAs coding for medulla and IRS trichohyalin.

Materials and Methods

Peptide Isolation and Sequencing

Sheep and guinea pig trichohyalin were prepared from follicle tissue as described by Rothnagel and Rogers (1986) with a number of modifications. Follicle extracts were concentrated using Centrifo CF25 ultrafiltration cones (Amicon Corp., Danvers, MA) and chromatographed at 0.13 ml/min on a CL-4B gel filtration column (85 \times 1.5 cm; Pharmacia Fine Chemicals, Uppsala, Sweden) that had been equilibrated with 7 M guanidine-HCl, 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM DTT. Trichohyalin-containing fractions were concentrated, and the buffer was changed to 7.5 M urea, 0.5 M guanidine-HCl, 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1% β -mercaptoethanol by serial filtration in Centrifo CF25 cones.

A sample (500 μ g) of pure guinea pig trichohyalin was digested with endoproteinase lysine C (Boehringer Mannheim GmbH, Mannheim, West Germany) for 18 h at 37°C. The resulting peptides were separated on a Superose 12 gel filtration column (30 \times 1 cm; Pharmacia Fine Chemicals) that had been equilibrated with 2 M urea, 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA. A sample containing 130 μ g of the proteolytic digest was chromatographed at a flow rate of 0.2 ml/min, and 0.5-ml fractions were collected. Protein-containing fractions were loaded onto an Aquapore Butyl reverse-phase cartridge (30 \times 2.1 mm; Brownlee Labs, Santa Clara, CA) equilibrated with 0.1% trifluoroacetic acid, and a number of linear gradients of 10–35% acetonitrile were applied over 25–30 min at a flow rate of 0.2 ml/min. The absorbance at 215 nm was recorded, and fractions were collected corresponding to absorbance peaks.

Purified peptides were then sequenced using a gas-phase protein sequencer (470A; Applied Biosystems, Inc., Foster City, CA).

Amino acid analysis of sheep trichohyalin was kindly performed by Dr. R. C. Marshall (Division of Biotechnology, Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia).

Isolation of RNA

Total cellular RNA was isolated from the wool follicles of Merino-Dorset Horn crossed with Border Leicester sheep as described by Powell et al. (1983). High molecular weight genomic DNA was removed from the nucleic acid preparation either by LiCl precipitation (Diaz-Ruiz and Kaper, 1978) or by treatment with RNase-free DNase I (Bresatec, Adelaide, South Australia). Poly A⁺ RNA was then isolated from the total RNA preparation

by oligo(dT)-cellulose chromatography (Boehringer Mannheim GmbH) (Bantle et al., 1976).

Construction and Screening of the Wool Follicle cDNA Library

A cDNA library was constructed from sheep wool follicle poly A⁺ RNA in the expression vector λ gt11 (Huynh et al., 1985). Briefly, double stranded cDNA was prepared essentially as described by Gubler and Hoffman (1983) except that murine Moloney leukemia virus reverse transcriptase (Bethesda Research Laboratories, Gaithersburg, MD) was used to synthesize the first strand. The cDNA was ligated to Eco RI linkers (Bresatec) (Huynh et al., 1985) and size selected on 10–40% sucrose gradients to remove cDNAs <1 kb. The cDNA was ligated into λ gt11, packaged in vitro (Gigapack Plus; Stratagene, La Jolla, CA), and plated on *Escherichia coli* strain Y1090 (Huynh et al., 1985).

The library was screened essentially as detailed elsewhere (Huynh et al., 1985) using a polyclonal antibody that had been raised in rabbits against sheep trichohyalin by the procedure described by Rothnagel and Rogers (1986). Antibody-bound plaques were detected by incubation with a goat anti-rabbit IgG/alkaline phosphatase conjugate (Sigma Chemical Co., St. Louis, MO) followed by colorimetric staining (Forster et al., 1985).

Recombinant phage DNA was prepared by the liquid culture method of Kao et al. (1982).

DNA Subcloning and Sequencing

The two Eco RI fragments from the longest immunopositive clone, λ Str1, were subcloned into pUC19, and a detailed 6-base restriction enzyme map was derived (see Results). DNA fragments, prepared by restriction enzyme digestion or deletion with Bal31 endonuclease (New England Biolabs, Beverly, MA) (Maniatis et al., 1982), were subcloned into appropriate M13 vectors (Messing and Vieira, 1982; Norrander et al., 1983). These clones were sequenced by the dideoxy chain termination method (Messing et al., 1981; Sanger et al., 1980) using the Bresatec dideoxy sequencing kit.

DNA sequence data was compiled and analyzed on a VAX 11-785 computer (Digital Equipment Corp., Marlboro, MA) using the programs ANALYSEQ (Staden, 1984) and DIAGON (Staden, 1982). Secondary structure analyses were obtained using the PREDICT program, which is a modified version of the joint prediction program of Eliopoulos et al. (1982).

Searches of the Genbank and National Biomedical Research Foundation databases were performed using the programs MATCH, MATCH TRANS-LATE, and MATCH FAST (Wilbur and Lipman, 1983; Lipman and Pearson, 1985).

DNA and RNA Hybridizations

Sheep genomic DNA (prepared by Dr. G. R. Cam, Department of Biochemistry, University of Adelaide, Adelaide, Australia) was digested with the appropriate restriction enzymes (Toyobo, Osaka, Japan), electrophoresed on a 1% agarose gel (Sigma Chemical Co.), and transferred to Zeta Probe membrane (Bio-Rad Laboratories, Richmond, CA) by alkaline DNA blotting. After transfer, the membranes were hybridized according to the instructions for Gene Screen (New England Nuclear, Boston, MA) and washed in 0.3 M NaCl, 30 mM tri-sodium citrate, and 0.1% SDS at 65°C.

For Northern blot analysis, total wool follicle RNA was denatured with glyoxal and then fractionated on a 1% agarose gel containing 10 mM sodium phosphate, pH 7.0 (Thomas, 1983). The RNA samples were then blotted onto Zeta Probe membrane, and the filters were hybridized as above and washed in 15 mM NaCl, 1.5 mM tri-sodium citrate, and 0.1% SDS at 20°C.

In Situ Hybridizations to Wool Follicle Sections

High specific activity cRNA probes for in situ hybridization analysis were produced by cloning DNA fragments from λ Str1 into pGEM-2 and transcribing these clones, after appropriate linearization, in the presence of [α -³⁵S]UTP. To obtain the repeat-containing probe, the fragment extending from the Eco RI site in the 5' linker to the second Pst I site (position 119) was cloned into pGEM-2 and transcribed with T7 RNA polymerase (Promega Biotec, Madison, WI). The 3' noncoding probe was produced by cloning the 1.9-kb Eco RI fragment and transcribing the Sac I cut clone (λ Str1 position 1,756) with SP6 RNA polymerase (Bresatec).

The in situ hybridization procedure was based on the method of Cox et al. (1984) with the modifications of Powell, B. C., and G. E. Rogers (manuscript in preparation).

Results

Isolation of Peptide Sequences

To confirm the identity of purified cDNA clones a partial amino acid sequence of trichohyalin was required. Due to low yields and poor proteolysis obtained with sheep trichohyalin, guinea pig trichohyalin, which cross reacts immunologically with sheep trichohyalin, was used. Purified guinea pig trichohyalin was proteolysed with endoproteinase lysine C, and after chromatographic purification three of the resultant peptides were subjected to automated Edman degradation. The three peptide sequences showed considerable cross-homology (Fig. 1). All three peptides contain the sequence glutamine-leucine, which is surrounded by a region of charged amino acids. In addition, two peptides also begin with phenylalanine-arginine. The homology observed suggests that at least a portion of the trichohyalin protein consists of repeats. Note that protease-specific cleavage indicates that each peptide should be preceded by a lysine residue in the total protein sequence (see Fig. 1).

Isolation of Sheep Trichohyalin cDNA Clones and Northern Blot Analysis

To increase the probability of isolating full-length trichohyalin cDNA clones, estimated from the size of sheep trichohyalin (190 kD) to be 5–6 kb in length, cDNA that had been prepared by oligo-dT priming of sheep follicle poly A⁺ RNA was size selected to remove cDNAs <1 kb before cloning into λ gt11. The resultant clones were screened with a polyclonal anti-trichohyalin antibody. This screening yielded three positive cDNA clones, and the longest of these, λ sTr1, was 2.4 kb long, which is less than half the expected size for a full-length trichohyalin cDNA clone.

The nucleotide sequence of λ sTr1, together with the deduced protein sequence, is shown in Fig. 2 A. λ sTr1 is 2,408 bp long, with an open reading frame spanning the first 1,375 bp. This is flanked by a 3' noncoding region of 1,030 bp of which the final 5 bases possibly belong to the poly(A) tail. A putative polyadenylation signal is present at position 2,385 (Fig. 2 A).

Northern blot analysis of sheep follicle RNA demonstrated that λ sTr1 hybridizes to a 6-kb mRNA species (Fig. 3) which, as stated above, is sufficient to encode the estimated 1,600 amino acids of sheep trichohyalin.

```
B      k F R U E P F L A X D R E E Q L R R R X E
D      k F R E E E Q L R L E S E E E
F1     k L Y T R P G Q R E Q L R E E E Q L E R E S R R Q E R D R R F H E E K
```

Figure 1. The amino-terminal sequences of three guinea pig trichohyalin peptides (B, D, and F1) are shown. They were purified from endoproteinase lysine C digests of pure trichohyalin and sequenced by the gas-phase method. Unassigned residues are indicated by the letter X. A lysine residue (small letters) has been placed at the amino end of each peptide because of the site-specific cleavage of endoproteinase lysine C. The peptides have been aligned to emphasize their similarity. Each peptide contains at least one QL combination (boxed) which is surrounded by a highly charged region (underlined). Two peptides also begin with FR (boxed), indicating an association of these residues with lysine, a substrate amino acid for transglutaminase. The sequence similarity suggests the presence of a repeat structure within trichohyalin.

Predicted Amino Acid Sequence of Trichohyalin

The 1,375-bp open reading frame of λ sTr1 encodes a protein of 458 amino acids with a predicted molecular mass of 60 kD. Therefore, the cDNA clone encodes ~30% of the native 190-kD trichohyalin. The deduced protein is hydrophilic, with 59% of the amino acids being charged (Table I). Analysis of the amino acid composition of the deduced protein indicates that glutamic acid/glutamine, arginine, leucine, and lysine are, in decreasing order, the most abundant amino acids, and this feature corresponds with an analysis of total sheep trichohyalin (Table I). The mole percents of glutamic acid/glutamine and arginine are much higher than in the total protein, suggesting that they are enriched within this coded segment. Interestingly, there are no sulphur-containing amino acids in the deduced sequence, which correlates with their low levels in total trichohyalin and also indicates that no disulphide cross-links, which are typical of hair IF proteins, are present within this region of the trichohyalin molecule. Portions of the deduced protein sequence show strong homology with all three guinea pig peptide sequences (Fig. 2 A), confirming that the cDNA clone does indeed encode trichohyalin.

Secondary structure predictions, performed with the program PREDICT, indicate that the majority of the protein could adopt an α -helical structure (Fig. 4 A). Comparison of the trichohyalin and available IF amino acid sequences (Conway and Parry, 1988) showed no significant homology between the two, indicating that this portion of trichohyalin is unrelated to epithelial, epidermal, or hair IF proteins.

A dot matrix plot of the λ sTr1 amino acid sequence that had been compared with itself revealed numerous diagonals spaced in the main by ~25 amino acids (Fig. 5). These diagonals, which indicate internal repeats, cover a region that extends from the beginning of the protein sequence to ~50 amino acids from the carboxy terminus. More detailed analysis of the sequence reveals that there is a 23 amino acid repeat, and a consensus sequence, as determined from the initial 14 repeats, is shown at the top of Fig. 6. The deduced protein sequence was aligned with the consensus sequence and contains 25 full or partial length repeats (Fig. 6). Secondary structure analysis of the consensus sequence predicts that the complete 23 amino acid sequence could form an α -helical rod, although the region from aspartate (at position 1) to phenylalanine (at position 4) could also form random coil (Fig. 4 B).

Detection of Trichohyalin Sequences in Sheep Genomic DNA

Sheep genomic DNA samples were digested with Bam HI, Eco RI, and Hind III, blotted, and hybridized with two probes, namely the 1.9-kb (predominantly coding) and 0.47-kb (3' noncoding) Eco RI fragments of λ sTr1 (see Fig. 2 B). The coding probe detected a single band within all three tracks (Fig. 7), indicating that the sheep genome contains only a single gene encoding the trichohyalin repeat. Additionally, the hybridization to one, rather than two, Hind III fragments indicates that the genomic sequence appears to differ from that of the cDNA clone since the Hind III site within the 1.9-kb Eco RI fragment of λ sTr1 is not detected in the genomic DNA. Upon hybridization with the 0.47-kb probe to the Hind III genomic digest, three bands were de-

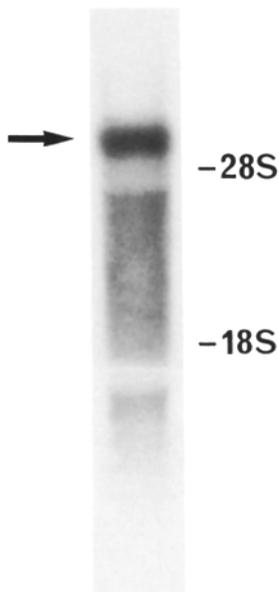


Figure 3. Northern blot analysis of sheep follicle RNA. A sample of total sheep RNA (5 μ g) was denatured with glyoxal and fractionated on a 1% agarose gel. The resultant filter was probed with the 1.9-kb Eco RI fragment of λ Tr1. Full-length trichohyalin mRNA (6 kb) is detected (arrow) together with partially degraded message. Ribosomal RNA marker positions are indicated.

tected (Fig. 7, lane 6), namely the 13-kb band bound by the coding probe (not seen in Fig. 7 but visible after longer exposure) and also two additional fragments that are both >0.47 kb (4 and 1.5 kb). This suggests that an intron is present within the 3' noncoding region of the trichohyalin gene.

Localization of Trichohyalin mRNA in Sheep Wool Follicles

The expression of trichohyalin within the follicle was exam-

Table I. Amino Acid Composition of the Deduced Trichohyalin Protein and Native Wool Follicle Trichohyalin

Amino acid	Deduced trichohyalin sequence	Wool follicle trichohyalin*
	mole percent	mole percent
Asp/Asn [‡]	3.3/0.2	6.4
Thr	0.0	2.9
Ser	1.7	5.4
Glu/Gln [§]	26.0/18.1	28.0
Pro	1.1	3.3
Gly	1.1	5.3
Ala	1.3	4.7
½-Cys	0.0	0.6
Val	1.3	4.0
Met	0.0	0.1
Ile	0.2	2.5
Leu	10.9	10.0
Tyr	0.9	2.1
Phe	3.7	2.4
Lys	5.0	6.7
His	1.1	1.7
Arg	23.8	13.7
Trp	0.2	0.2

* Amino acid analysis of purified wool follicle trichohyalin was performed by Dr. R. C. Marshall (Division of Biotechnology, Commonwealth Scientific Industrial Research Organization, Melbourne, Australia).

[‡] Aspartic acid and asparagine are denoted separately for the deduced sequence but combined for the native trichohyalin analysis.

[§] Glutamic acid and glutamine are denoted separately for the deduced sequence but combined for the native trichohyalin analysis.

ined by in situ hybridizations performed on skin sections from Merino and Tuki-dale sheep. Merino wool follicles, which are equivalent to those from which the trichohyalin cDNA was derived, produce fibers that are nonmedullated, whereas a percentage of the Tuki-dale wool follicles produce medullated fibers. Hybridizations to the Tuki-dale follicles therefore enabled a comparison of the trichohyalin mRNA expression in the IRS with that in the medulla. cRNA probes were made from both the coding repeat and the 3' noncoding region by cloning fragments from λ Tr1 into pGEM-2 and synthesizing transcripts with either SP6 or T7 RNA polymerase. Hybridization of the repeat-containing probe to the Tuki-dale follicle sections produced a strong signal over both the IRS and the medulla cells (Fig. 8, A and B). Within both layers, the signal extended from the basal cells within the follicle bulb (Fig. 8 A) to the cells positioned immediately beneath the zone of hardening (Fig. 8 B). The signal intensity over the medulla cells was similar to that covering the IRS cells. No signal was detected over any layer of the Tuki-dale epidermis (data not shown). Hybridization with the 3' noncoding probe produced a pattern identical to that seen with the repeat-containing probe, although the signal was less intense due to the presence of a nonrepeated target sequence (Fig. 8 C). With the Merino wool follicle sections, both probes bound to the IRS and produced signals of similar positioning and intensity to that seen in the IRS of the Tuki-dale follicles (data not shown).

Discussion

We report here the isolation and characterization of a cDNA clone, λ Tr1, that encodes 2.408 kb of the \sim 6-kb mRNA of sheep wool follicle trichohyalin. This clone contains the partial coding sequence (1,375 bp) and probably the complete 3' noncoding region that is >1 kb in length. The coding region encodes 458 amino acids of trichohyalin, extending up to the carboxy terminus, and would form a partial protein with a predicted mass of 60 kD. The amino acid composition of the encoded protein is similar to that of intact sheep trichohyalin (Table I), and regions within the deduced protein sequence are nearly identical to the sequence of purified peptides isolated from guinea pig trichohyalin (Fig. 2 A).

Proteolytic peptide sequences (Fig. 1) and two-dimensional gel analysis of trichohyalin breakdown products (Rothnagel and Rogers, 1986) predicted that peptide repeats are present within trichohyalin. Analysis of λ Tr1 indicates that trichohyalin contains a 23 amino acid repeat and that 95% of the deduced protein sequence consists of full- or partial-length repeats (Fig. 6). The 23 amino acid consensus sequence is highly charged, containing 19 charged or polar residues. Although 15 of these residues are charged, the repeat has a net charge of only -1 . Arginine, glutamine, and lysine residues, which are the substrate amino acids for the follicle enzymes peptidylarginine deiminase and transglutaminase, are present within the repeat unit, and some or all of these residues may act as substrates for the enzymes. Since only a single lysine residue is present within the repeat, it is possible that the adjacent highly conserved phenylalanine residue and also the surrounding charged environment are necessary for transglutaminase function. In addition, the glutamine at position 17 within the repeat (Fig. 6) is in a charged environment (EEQLRR) which is similar to that of

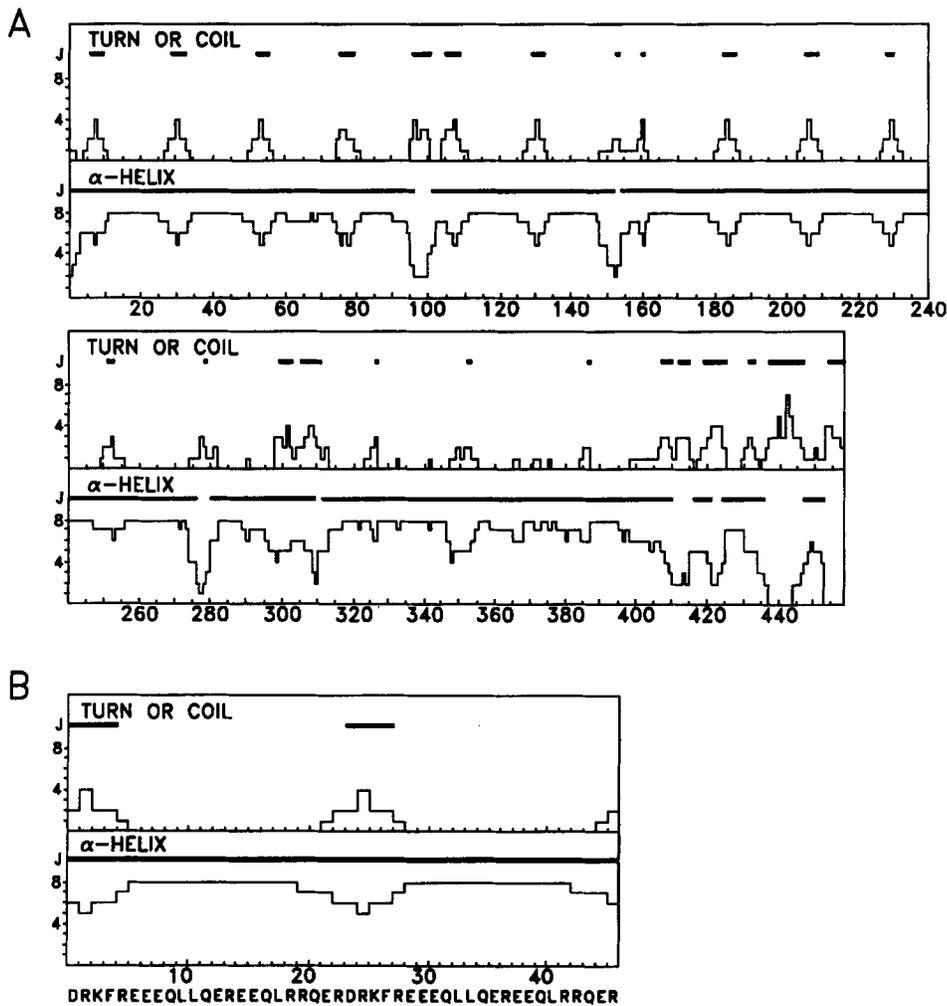


Figure 4. The secondary structure of the deduced amino acid sequence was analyzed by the program PREDICT (see Materials and Methods), and the number of predictions for α -helix and for turn or coil are graphed. Predicted regions of the given secondary structure are indicated by the solid bars at the top of each section (*J*). (*A*) Secondary structure predictions for the total deduced protein sequence. Note that α -helix is predicted for the vast majority of the sequence and that small regions of coil, which predominantly overlap regions of α -helix, are also predicted. (*B*) The predicted structure for two consecutive repeats is enlarged (e.g., from residue 183 to 208). Note that the region from asp(1) to phe(4) (positions 1-4 and 24-27) could be either α -helix or random coil.

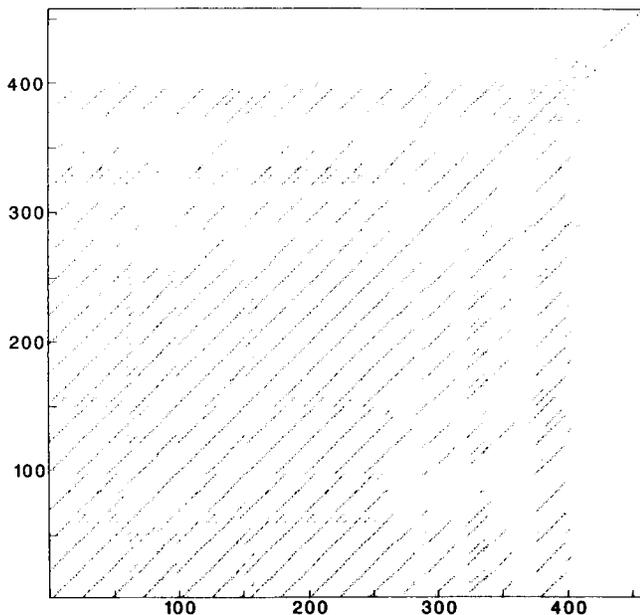


Figure 5. The predicted amino acid sequence of λ sTr1 was analyzed for similar internal sequences using the computer program DIAGON (Staden, 1982) with a window length of 23. The output of this comparison was then plotted. Internal similarities are represented by the lines parallel to the central diagonal. The axes are labeled in residue numbers.

the cross-linked glutamine residue within fibrin (EGQQHH), another transglutaminase substrate (Chen and Doolittle, 1971). Thus, gln (at position 17) may therefore be cross-linked by the follicle transglutaminase. Interestingly, the two proposed sites for transglutaminase activity are nearly identical in sequence to the two common sequences present in the guinea pig peptides (Fig. 1). Enzymatic studies on synthetic oligopeptides or cDNA-derived proteins will be necessary to determine the exact sites of transglutaminase and peptidylarginine deiminase action.

The repeat region of the protein can be divided on the basis of the length of each repeat and the level of amino acid conservation into amino- and carboxy-terminal segments. The 14 repeats within the amino-terminal segment are mainly full length and are highly conserved, with 75-100% of the amino acids within each repeat identical to those of the consensus sequence (Fig. 6). Of the 11 repeats within the carboxy-terminal segment, all but one are of partial length and together they show a lower degree of conservation with the consensus sequence (40-89%), although each does retain an eight amino acid stretch (residues 15-22; see Fig. 6). This overall structure suggests that within the amino-terminal segment of the deduced sequence the full 23 amino acid repeat is functionally required, whereas at the carboxy terminus only a smaller region is necessary. The differences in the respective levels of arginine and glutamic acid/glutamine in the deduced sequence and in the native sheep trichohyalin

Residue Position 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

CONSENSUS D R K F R E E E Q L L Q E R E E Q L R R Q E R

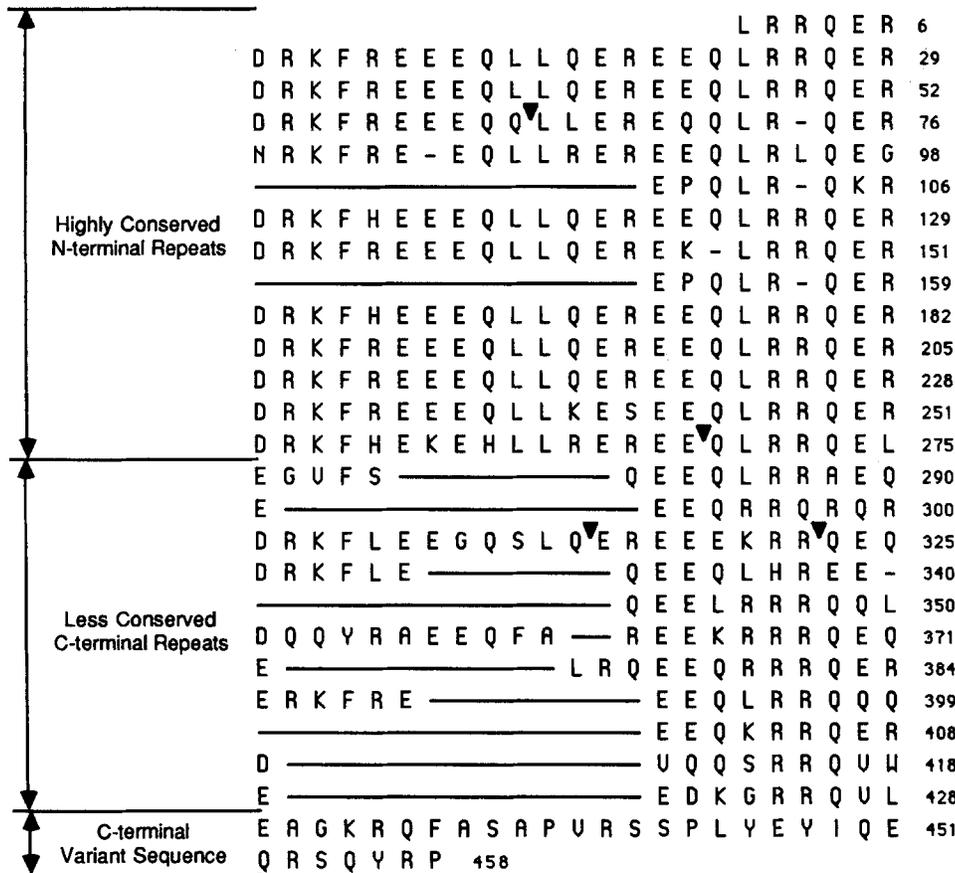


Figure 6. The predicted trichohyalin amino acid sequence is aligned with respect to the 23 amino acid consensus sequence (top) which was derived from the amino-terminal segment of highly conserved repeats. Dashes and arrowheads indicate space insertions or sequence deletions that have been introduced for optimal alignment. Each arrowhead indicates the removal of only one or two amino acids.

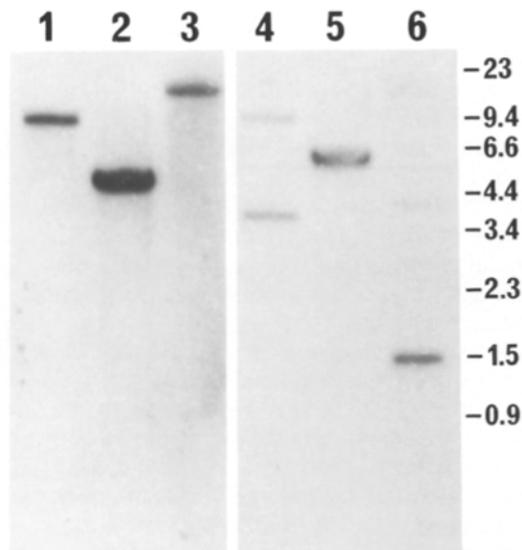


Figure 7. Southern blot analysis of sheep genomic DNA. Total sheep genomic DNA (4 μ g/lane) was digested with Bam HI (lanes 1 and 4), Eco RI (2 and 5), and Hind III (3 and 6). Lanes 1-3 were probed with the 1.9-kb Eco RI fragment of λ sTr1 (coding) and lanes 4-6 were probed with the 0.47-kb Eco RI fragment (3' noncoding). After prolonged exposure of lane 6, an additional band is seen and is the same size as the band present in lane 3 (13 kb). The autoradiograph of lanes 1-3 was exposed for 20 h and that of lanes 4-6 for 6 d. Size markers are shown (in kilobase pairs).

(Table I) suggest that the 23 amino acid repeat is not present throughout the whole of the trichohyalin molecule and that sequences distinct from it may be present.

The repeat structure of trichohyalin is similar in certain respects to that of involucrin, an epidermal transglutaminase substrate. Human involucrin contains a highly conserved central region consisting of 39 tandem repeats of a 10 amino acid sequence and this region is flanked by segments that have little homology with the 10 amino acid repeat (Eckert and Green, 1986). The involucrin repeat unit contains a large proportion of hydrophilic residues (70%). Of these, glutamic acid and glutamine residues constitute 50% of the involucrin repeat and this is very similar to the trichohyalin repeat where they account for 48% of the consensus sequence. Although there is no homology between the two protein repeats, glutamic acid residues are positioned within close proximity to the glutamine residues in both repeat sequences, suggesting that neighboring glutamic acid residues may be required for glutamine to be cross-linked by transglutaminase. For example, the involucrin repeat contains the sequence EGQLKH, in which the glutamic acid positioning and charged environment are similar to those within the trichohyalin and fibrin sequences mentioned above. In addition to the similar amino acid composition, there are also similarities in the nucleotide compositions, with both highly conserved repeat segments having a very low T content, being 9% in trichohyalin and 8% in involucrin (Eckert and Green, 1986).

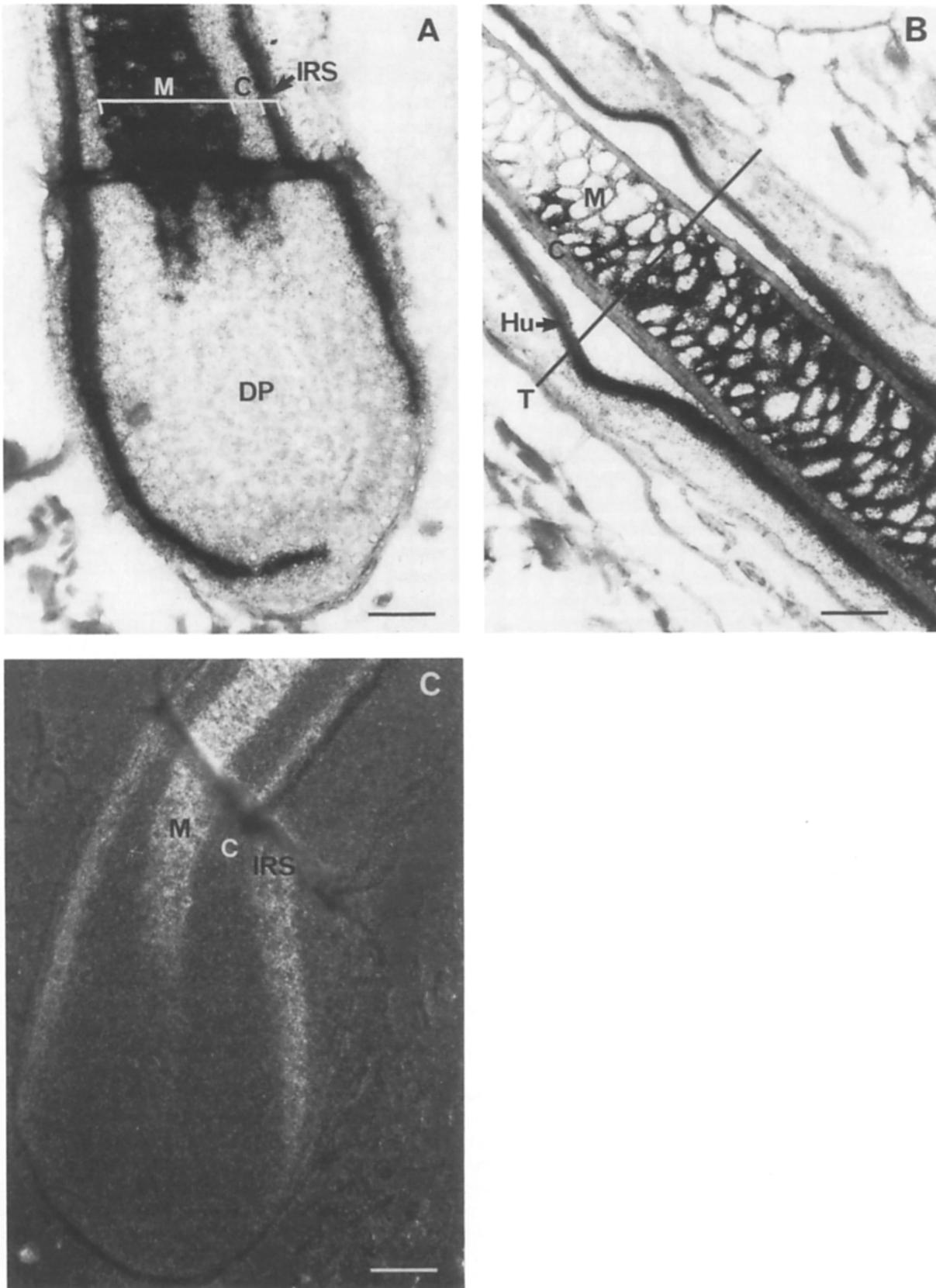


Figure 8. In situ hybridization analysis of Tukidale wool follicles. (A) The bulb region of a medullated Tukidale follicle is shown after hybridization with the repeat-containing cRNA probe. Note that the hybridization to the medulla begins with the cells lining the tip of the dermal papilla (DP). The hybridization to the IRS extends from the edge of the bulb on one side and the base of the bulb on the other. The latter appearance is caused by the obliqueness of the follicle section. There is no hybridization to the fiber cortex. The dark line present immediately above the dermal papilla is artefactual and caused by folding of the tissue during sectioning. (B) Hybridization of the repeat-

The similarity of the proteins within the trichohyalin granules of the IRS and the medulla and the functions of these proteins has long been questioned (Rogers, 1962, 1963, 1983; Rogers and Harding, 1976; Rogers et al., 1977; Rothnagel and Rogers, 1986). Genomic southern blot analysis has shown that the repeat region of λ Str1 hybridizes to only a single fragment when sheep DNA is digested with either Bam HI, Eco RI, or Hind III (Fig. 7), indicating that only a single gene encodes the repeat sequence found in λ Str1. In situ hybridization analysis has shown that both repeat and 3' non-coding probes derived from λ Str1 hybridize to mRNAs in both the medulla and IRS of Tukidale wool follicles (Fig. 8). Therefore, the trichohyalin protein within the medulla and IRS is encoded by the same single copy gene. It remains possible that differential splicing of an intron positioned in the 5' end of the trichohyalin transcript may lead to variant mRNA species in the IRS and the medulla and that the structure and function of the resultant proteins may be somewhat different. Once the complete gene has been purified, examination of the gene structure will establish whether introns are present in the coding region. If they are, then in situ hybridization with fragments corresponding to the 5' end of the mRNA will be required to determine whether differential splicing occurs.

The trichohyalin that is synthesized in both the IRS and the medulla appears to be the same. The difference between these two hardened tissues could be due to either the additional synthesis of IF proteins in the IRS but not the medulla, such that in the IRS trichohyalin acts as an IF-associated protein (Rothnagel and Rogers, 1986), or a different chemical environment occurring in the two tissues so that trichohyalin itself forms filaments only in the IRS. Evidence for the presence of IF proteins and the independent synthesis of IF in IRS cells has come from Heid et al. (1988) who showed that monoclonal antibodies to the hyperproliferative epidermal IF proteins K6 and K16 bind to the IRS in the human hair follicle. Further, Ito et al. (1986a,b) and Lane et al. (1986) have shown that certain IF antibodies also bind to the IRS. However, contrary evidence to the presence of IF proteins has come from two in situ hybridization studies using, respectively, a sheep cDNA probe encoding a K6-like protein (Whitbread, L., personal communication) and a probe derived from the conserved α -helical region of a follicle IF gene (Powell, B., personal communication). In both instances, mRNAs were not detected in the IRS cells.

With the primary structure for ~33% of the trichohyalin protein available from the present study, the secondary structure was examined for its theoretical capacity to form α -helical coiled coils with the dimensions specific for the formation of IFs. We have found that although the majority of the trichohyalin sequence and the complete consensus sequence

could adopt an α -helical formation (Fig. 4) the predicted trichohyalin sequence has no significant homology with the primary structure of the IF α -helical region. This region is characterized by tandem heptads (a-b-c-d-e-f-g) where the presence of hydrophobic residues at a and d allows the formation of α -helical coiled coils (Parry et al., 1977; McLachlan and Karn, 1982). The α -helical trichohyalin repeats contain a different heptad where only the first residue (phe 4, leu 11, leu 18) is hydrophobic (Fig. 6). This suggests that the trichohyalin repeats, although not involved in coiled coil formation, could nevertheless form a linear cluster of α -helical rods producing filaments with an 8–10 nm diameter, characteristic of genuine IFs. Alternatively, the secondary structure predictions indicate that random coil is also possible between asp(1) and phe(4) (Fig. 4B), producing short α -helical units that could, by random aggregation, form the nonfilamentous structure seen in the medulla cells. Since only 30% of the complete sequence is available, the structural predictions are not yet conclusive and it remains possible that trichohyalin could act as either an IF or an IF-associated protein in the IRS. Despite this equivocal situation, we believe that the follicle enzymes transglutaminase and peptidylarginine deiminase play a central role in the divergent development of the IRS and medulla (Rogers et al., 1989). Differences in the timing of enzyme expression and the level of enzyme activity within the two tissues could allow trichohyalin to adopt different conformations in the two hardened structures.

It is important to note that the secondary structure predictions are based on the precursor protein and that conversion of certain arginine residues to citrulline may alter the overall secondary structure. Further studies will be required to determine which arginine residues are converted and to assess the effect that these changes have on the secondary structure. In addition, complete sequence analysis may determine whether regions within the remaining 70% of the trichohyalin protein are capable of forming IFs.

The in situ hybridization studies have revealed that in both the medulla and the IRS follicle layers, trichohyalin mRNA is expressed within the basal cells. It remains within the developing cells until immediately before tissue hardening (Fig. 8) and this could be due to either the continuous expression or a prolonged half-life of the trichohyalin mRNA. Therefore, trichohyalin mRNA is present during almost all levels of cell development emphasizing the large amount of trichohyalin produced and its importance in growth and differentiation within the follicle. The absence of trichohyalin mRNA in the epidermis suggests that trichohyalin is a follicle-specific protein, although further in situ analysis using other keratinised tissues is required to confirm this. It is clear that in the follicle trichohyalin is a very early

containing cRNA probe to a medullated follicle shaft at the level of conversion from granular to hardened cells. The developing cells of the medulla and the Huxley layer of the IRS (*Hu*) reach, in their movement up the follicle, a transition level (marked approximately by the line *T*) at which the trichohyalin granules disappear and are replaced by the hardened-cellular material. The hybridization to the trichohyalin mRNA terminates in both tissues at the same level as the disappearance of the trichohyalin granules. Note that on both sides of the follicle the fiber cuticle and the IRS cuticle have separated during the sectioning procedure. (C) The longitudinal section of a medullated Tukidale follicle was photographed under dark-field illumination to highlight the low signal strength obtained with the 3' noncoding probe (*white grains*). The probe hybridizes to both the IRS and the medulla and spans the same regions as the repeat-containing probe (A and B). The line present above the bulb was caused by folding of the tissue during sectioning. C, cortex; IRS, inner root sheath; M, medulla. Bars, 50 μ m.

differentiation marker, produced much earlier than the IF proteins or the IF-associated proteins present in the cortical cells of the hair fiber (Kopan et al., 1989; MacKinnon, 1989; Powell, B. C., J. L. Arthur, L. A. Crocker, M. J. Fietz, A. Fratini, R. A. Keough, P. J. MacKinnon, A. Nesci, M. T. O'Donnell, and G. E. Rogers, manuscript in preparation) and is apparently expressed immediately after the initiation of differentiation. The factors involved in the initiation of differentiation are at present unknown and their expression may be dependent upon the relative positioning of the basal cells to the dermal papilla. We anticipate that the further investigation of factors controlling the expression of trichohyalin will prove to be of great importance in understanding the initiation and control of differentiation in the hair follicle.

We are very grateful to Dr. B. Powell for help with the in situ hybridizations and for critical analysis of the manuscript. We thank Dr. R. C. Marshall for the amino acid analysis and Mrs. Guo Xiao Hui for assistance with the cDNA sequencing.

This work was supported by a Research Associateship awarded by the Australian Research Council (R. B. Presland), a Commonwealth Postgraduate Research Award (M. J. Fietz), and funds from a Commonwealth Special Centre Grant and the Australian Wool Corporation (G. E. Rogers).

Received for publication 14 August 1989 and in revised form 9 October 1989.

References

Bantle, J. A., I. H. Maxwell, and W. E. Hahn. 1976. Specificity of oligo(dT)-cellulose chromatography in the isolation of polyadenylated RNA. *Anal. Biochem.* 72:413-427.

Birbeck, M. S. C., and E. H. Mercer. 1957. The electron microscopy of the human hair follicle. III. The inner root sheath and trichohyalin. *J. Biophys. Biochem. Cytol.* 3:223-230.

Chen, R., and R. F. Doolittle. 1971. γ - γ cross-linking sites in human and bovine fibrin. *Biochemistry.* 10:4486-4491.

Conway, J. F., and D. A. D. Parry. 1988. Intermediate filament structure. III. Analysis of sequence homologies. *Int. J. Biol. Macromol.* 10:79-98.

Cox, K. H., D. V. DeLeon, L. M. Angerer, and R. C. Angerer. 1984. Detection of mRNAs in sea urchin embryos by *in situ* hybridization using asymmetric RNA probes. *Dev. Biol.* 101:485-502.

Diaz-Ruiz, J. R., and J. M. Kaper. 1978. Isolation of viral double-stranded RNAs using a LiCl fractionation procedure. *Prep. Biochem.* 8:1-17.

Eckert, R. L., and H. Green. 1986. Structure and evolution of the human involucrin gene. *Cell.* 46:583-589.

Eliopoulos, E., A. J. Geddes, M. Brett, D. J. C. Pappin, and J. B. C. Findlay. 1982. A structural model for the chromophore-binding domain of ovine rhodopsin. *Int. J. Biol. Macromol.* 4:263-268.

Forster, A. C., J. L. McInnes, D. C. Skingle, and R. H. Symons. 1985. Non-radioactive hybridization probes prepared by the chemical labelling of DNA and RNA with a novel reagent, photobiotin. *Nucleic Acids Res.* 13:745-761.

Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene (Amst.)* 25:263-269.

Harding, H. W. J., and G. E. Rogers. 1971. ϵ -(γ -glutamyl)lysine cross-linkage in citrulline-containing protein fractions from hair. *Biochemistry.* 10:624-630.

Harding, H. W. J., and G. E. Rogers. 1972. The occurrence of the ϵ -(γ -glutamyl)lysine cross-link in the medulla of hair and quill. *Biochim. Biophys. Acta.* 257:37-39.

Heid, H. W., I. Moll, and W. W. Franke. 1988. Patterns of expression of trichocytic and epithelial cyokeratins in mammalian tissues. I. Human and bovine hair follicles. *Differentiation.* 37:137-157.

Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in λ gt10 and λ gt11. In *DNA Cloning: A Practical Approach*. Vol. 1. D. M. Glover, editor. IRL Press, Oxford. 49-78.

Ito, M., T. Tazawa, K. Ito, N. Shimizu, K. Katsuomi, and Y. Sato. 1986a. Immunological characteristics and histological distribution of human hair fibrous proteins studied with anti-hair keratin monoclonal antibodies HKN-2, HKN-4 and HKN-6. *J. Histochem. Cytochem.* 34:269-275.

Ito, M., T. Tazawa, N. Shimizu, K. Ito, K. Katsuomi, Y. Sato, and K. Hashimoto. 1986b. Cell differentiation in human anagen hair and hair folli-

cles studied with anti-hair keratin monoclonal antibodies. *J. Invest. Dermatol.* 86:563-569.

Kao, F. T., J. A. Hartz, M. L. Law, and J. N. Davidson. 1982. Isolation and chromosomal localization of unique DNA sequences from a human genomic library. *Proc. Natl. Acad. Sci. USA.* 79:865-869.

Kopan, R., and E. Fuchs. 1989. A new look into an old problem: keratins as tools to investigate determination, morphogenesis, and differentiation in skin. *Genes & Dev.* 3:1-15.

Lane, E. B., J. Bartek, P. E. Purkis, and I. M. Leigh. 1985. Keratin antigens in differentiating skin. *Ann. NY Acad. Sci.* 455:241-258.

Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science (Wash. DC).* 227:1435-1441.

MacKinnon, P. J. 1989. Molecular Analysis of the Ultra-High-Sulphur Keratin Proteins. Ph.D. thesis. University of Adelaide, Adelaide, Australia. 94 pp.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.

McLachlan, A. D., and J. Karn. 1982. Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacing in muscle. *Nature (Lond.)* 299:226-233.

Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene (Amst.)* 19:269-276.

Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9:309-321.

Norrander, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene (Amst.)* 26:101-106.

Oliver, R. F., and C. A. B. Jahoda. 1989. The dermal papilla and maintenance of hair growth. In *The Biology of Wool and Hair*. G. E. Rogers, P. J. Reis, K. A. Ward, and R. C. Marshall, editors. Chapman and Hall Ltd., London. 51-67.

Parakkal, P. K., and A. G. Matoltsy. 1964. A study of the differentiation products of the hair follicle cells with the electron microscope. *J. Invest. Dermatol.* 43:23-34.

Parry, D. A. D., W. G. Crewther, R. D. B. Fraser, and T. P. MacRae. 1977. Structure of α -keratin: structural implications of the amino acid sequences of the type I and type II chain segments. *J. Mol. Biol.* 113:449-454.

Powell, B. C., M. J. Sleight, K. A. Ward, and G. E. Rogers. 1983. Mammalian keratin gene families: organisation of genes coding for the B2 high-sulphur proteins of sheep wool. *Nucleic Acids Res.* 11:5327-5346.

Rogers, G. E. 1962. Occurrence of citrulline in proteins. *Nature (Lond.)* 194:1149-1151.

Rogers, G. E. 1963. The localization and significance of arginine and citrulline in proteins of the hair follicle. *J. Histochem. Cytochem.* 11:700-705.

Rogers, G. E. 1964. Structural and biochemical features of the hair follicle. In *The Epidermis*. W. Montagna and W. Lobitz, editors. Academic Press Inc., New York. 179-236.

Rogers, G. E. 1983. The occurrence of citrulline in structural proteins of the hair follicle. In *Biochemistry and Physiology of the Skin*. Vol. 1. L. A. Goldsmith, editor. Oxford University Press, New York. 511-521.

Rogers, G. E., and H. W. J. Harding. 1976. Molecular mechanisms in the formation of hair. In *Biology and Disease of the Hair*. T. Kobori and W. Montagna, editors. University of Tokyo Press, Tokyo. 411-435.

Rogers, G. E., H. W. J. Harding, and I. J. Llewellyn-Smith. 1977. The origin of citrulline-containing proteins in the hair follicle and chemical nature of trichohyalin, an intracellular precursor. *Biochim. Biophys. Acta.* 495:159-175.

Rogers, G. E., E. S. Kuczek, P. J. MacKinnon, R. B. Presland, and M. J. Fietz. 1989. Special biochemical features of the hair follicle. In *The Biology of Wool and Hair*. G. E. Rogers, P. J. Reis, K. A. Ward, and R. C. Marshall, editors. Chapman and Hall Ltd., London. 69-85.

Rothnagel, J. A., and G. E. Rogers. 1986. Trichohyalin, an intermediate filament-associated protein of the hair follicle. *J. Cell Biol.* 102:1419-1429.

Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161-178.

Staden, R. 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* 10:2951-2961.

Staden, R. 1984. Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Res.* 12:521-538.

Steinert, P. M., H. W. J. Harding, and G. E. Rogers. 1969. The characterization of protein-bound citrulline. *Biochim. Biophys. Acta.* 175:1-9.

Straille, W. E. 1965. Root sheath-dermal papilla relationships and the control of hair growth. In *Biology of the Skin and Hair Growth*. A. G. Lyne and B. F. Short, editors. Angus and Robertson Ltd., Sydney. 35-57.

Thomas, P. S. 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. *Methods Enzymol.* 100:255-256.

Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA.* 80:726-730.