

Analysis of the Sheep Trichohyalin Gene: Potential Structural and Calcium-binding Roles of Trichohyalin in the Hair Follicle

Michael J. Fietz, Clive J. McLaughlan, Mary T. Campbell,* and George E. Rogers

Department of Biochemistry, University of Adelaide, Adelaide, S.A., 5001, Australia; and *School of Business and Technology, University of Western Sydney Macarthur, Campbelltown, N.S.W., 2560, Australia

Abstract. Trichohyalin is a structural protein that is produced and retained in the cells of the inner root sheath and medulla of the hair follicle. The gene for sheep trichohyalin has been purified and the complete amino acid sequence of trichohyalin determined in an attempt to increase the understanding of the structure and function of this protein in the filamentous network of the hardened inner root sheath cells. Sheep trichohyalin has a molecular weight of 201,172 and is characterized by the presence of a high proportion of glutamate, arginine, glutamine, and leucine residues, together totaling more than 75% of the amino acids.

Over 65% of trichohyalin consists of two sets of tandem peptide repeats which are based on two different consensus sequences. Trichohyalin is predicted to form an elongated α -helical rod structure but does not contain the sequences required for the formation of intermediate filaments. The amino terminus of trichohyalin contains two EF hand calcium-binding domains indicating that trichohyalin plays more than a structural role within the hair follicle. In situ hybridization studies have shown that trichohyalin is expressed in the epithelia of the tongue, hoof, and rumen as well as in the inner root sheath and medulla of the hair follicle.

THE mammalian hair follicle is a specialized epidermal appendage consisting of an array of up to six different cellular layers. Upon division, the follicle cells move up the shaft of the follicle and give rise to either the cortex, cuticle, medulla, or the inner root sheath (IRS)¹ which accompanies the developing fiber in its outward growth. The IRS is a cylinder of cells surrounding the developing fiber and is believed to play a structural role within the follicle, supporting and directing the fiber cells (Straile, 1965). The developing IRS cells are characterized by the appearance of electron-dense nonmembrane-bound trichohyalin granules which have been shown to contain trichohyalin, a protein with a mass of \sim 200 kD (Rothnagel and Rogers, 1986). As the cells move up the follicle, filaments with a diameter of 8–10 nm appear in close association with the trichohyalin granules (Rogers, 1964). Finally, the granules disappear, the IRS cells become 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 cells (Birbeck and Mercer, 1957).

The insoluble protein network of the IRS cells are characterized by two posttranslational modifications. The IRS proteins are cross-linked by ϵ -(γ -glutamyl)lysine isopeptide bonds (Harding and Rogers, 1971) which are formed by the enzyme transglutaminase (Chung and Folk, 1972; Harding and Rogers, 1972). In addition, arginine residues within the IRS proteins are converted to citrulline by peptidylarginine deiminase (Rogers et al., 1977). The activities of both enzymes are calcium-dependent.

The developing medulla cells also contain trichohyalin granules although the mature cells do not contain filaments but are filled with an amorphous protein mass. Both of the posttranslational modifications also occur within the medulla (Steinert et al., 1969; Harding and Rogers, 1971; Harding and Rogers, 1972).

The role of trichohyalin in the IRS and medulla has long been enigmatic. Immunological studies suggest that trichohyalin acts as an intermediate filament-associated protein (IFAP) in the IRS cells (Rothnagel and Rogers, 1986; O'Guin et al., 1992) with K1 and K10-like intermediate filament (IF) proteins forming the IF (Stark et al., 1990). It is therefore postulated that the lack of IF protein synthesis in the medulla would allow cross-linked trichohyalin to form the major component of the protein mass of the mature medulla.

A partial trichohyalin cDNA clone has been previously isolated and \sim 30% of the protein sequence of sheep tricho-

Dr. M. J. Fietz's present address is ICRF Developmental Biology Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom.

Please address all correspondence to Professor G. E. Rogers, Department of Biochemistry, University of Adelaide, GPO Box 498, Adelaide, South Australia 5001, Australia.

1. *Abbreviations used in this paper:* IFAP, intermediate filament-associated protein; IF, intermediate filament; IRS, inner root sheath; UWGCG, University of Wisconsin Genetics Computer Group.

hyalin determined (Fietz et al., 1990). This portion of the protein does not contain any of the central core sequence typical of IF proteins but consists of an array of tandem 23-amino acid repeats which are predicted to form α -helix. In the present paper, we report the isolation and characterization of the single copy sheep trichohyalin gene. The complete protein sequence is analyzed and the functional implications discussed. Further, the expression of trichohyalin within other sheep keratinized epithelia is examined by *in situ* hybridization.

Materials and Methods

Screening and Subcloning of the Sheep Trichohyalin Gene

A sheep genomic cosmid library, which was constructed in the vector pWE15 by Clontech Laboratories Inc. (Palo Alto, CA) using DNA provided by our laboratory, was screened with a 1.9-kb sheep trichohyalin cDNA *EcoRI* fragment (Fietz et al., 1990) using the method of Hanahan and Meselson (1980). Southern blot analysis of the purified clone was performed using the method described in Fietz et al. (1990) except that the transfer was performed using an LKB Vacugene Blotting System (Bromma, Sweden). The desired fragments were initially subcloned into either the *BamHI* or *EcoRI* site of pGEM-7Zf(+) (Promega Corp., Madison, WI) for further analysis.

Genomic Sequence Analysis

Fragments spanning the trichohyalin gene were inserted into either pGEM-3Zf(+) (Promega Corp.), pGEM-7Zf(+) or pBluescript II SK (Stratagene Inc., La Jolla, CA). Deletions were performed using *Bal31* endonuclease (New England Biolabs, Beverly, MA) (Maniatis et al., 1982), subcloned into appropriate M13 vectors (Yanish-Perron et al., 1985), and sequenced by the dideoxy chain termination method (Sanger et al., 1980; Messing et al., 1981) using the Sequenase Version 2.0 DNA Sequencing Kit (Un. States Biochem. Corp., Cleveland, OH). Sequences at the beginning of the fragments were obtained using double-strand DNA sequencing methods (Chen and Seeburg, 1985).

DNA sequence data and the derived protein sequence data were analyzed using programs from both the Staden Package (Staden, 1982, 1984) and the Sequence Analysis Software Package of the University of Wisconsin Genetics Computer Group (UWGCG) (Devereux et al., 1984). Database searches were performed using the UWGCG package while secondary structure analyses were obtained using the PREDICT program (Eliopoulos et al., 1982) and UWGCG programs.

Sequence Analysis of Trichohyalin mRNA

To obtain the sequence of the 5' end of the trichohyalin mRNA it was attempted to clone the 5' terminus using the "PCR-RACE" technique of Frohman et al. (1988). PCRs were used to amplify this region from cDNA synthesized from total wool follicle RNA, prepared by the method of Chomczynski and Sacchi (1987), using an oligonucleotide primer (Tr3) located ~370 bp from the 5' end of the mRNA. The full-length product could not be cloned but sequence data was obtained from M13mp18 clones containing partial inserts and from double-strand sequence analysis (Chen and Seeburg, 1985) of the amplified fragment primed with the oligonucleotides Tr3 and Tr5 (see Fig. 2).

The size of Exon 1 was determined by RNase protection analysis using the method of Krieg and Melton (1987). A 186-bp *HinfI* fragment (1035-1220) was cloned into the *Sma* I site of pBluescript II SK, transcribed with T7 RNA Polymerase (Bresatec Ltd., Thebarton, S. Australia), and used to protect wool follicle total RNA. The cap site was located by primer extension analysis (McKnight et al., 1981) of wool follicle total RNA using Tr5 to prime DNA synthesis by AMV Reverse Transcriptase.

In situ Hybridization of Sheep Epithelia

In situ hybridization analysis was performed using a cRNA probe derived from the COOH-terminal repetitive region of trichohyalin as described in Fietz et al. (1990). The probe was hybridized to sections from sheep tongue,

hoof, skin, esophagus, and rumen using a procedure based on the method of Cox et al. (1984) with the modifications of Powell and Rogers (1990).

Results

Purification and Analysis of Sheep Trichohyalin Genomic Clones

We have previously reported the purification and analysis of a 2.4-kb sheep trichohyalin cDNA clone (λ sTr1) encoding the COOH-terminal 30% of trichohyalin (Fietz et al., 1990). To further the examination of trichohyalin, a sheep genomic cosmid library was screened with a 1.9-kb *EcoRI* fragment derived from λ sTr1 and a single clone isolated (sTr10). The fragments encoding the COOH terminus of trichohyalin and the orientation of the gene were determined by Southern analysis using probes derived from λ sTr1. The region expected to contain the trichohyalin gene was then mapped and found to abut the cosmid vector (Fig. 1). Fragments spanning the gene were subcloned and the nucleotide sequence covering the complete region (9,344 bp) was determined (Fig. 2).

Determination of the Structure of the Trichohyalin Gene

Examination of the obtained sTr10 sequence showed that the coding region present in the cDNA clone λ sTr1 occurs at the 3' end of an open reading frame of 4,521 bp. The open reading frame does not contain a methionine residue and has only a single probable intron splice site 12 bp from the beginning of the reading frame. To confirm the presence of the splice site and to determine the upstream exon sequences, the cloning of the 5' end of the trichohyalin cDNA was attempted using the PCR protocol of Frohman et al. (1988). Unfortunately the full-length amplified product would not clone using a number of different vector/host combinations. Therefore regions of the upstream exon sequences were determined by cloning portions of the amplified product and by directly sequencing the product using internal primers. From this analysis, it was determined that the trichohyalin gene contains two upstream exons, the first containing only upstream non-coding sequence (see Fig. 1). The size of Exon 1 was determined by RNase protection of follicle RNA (Fig. 3) while the location of the cap site was confirmed by primer extension analysis (data not shown).

The trichohyalin gene contains three exons: Exon 1 contains 52 bp of 5' noncoding sequence; Exon 2 contains 166 bp (28 bp 5' noncoding, 138 bp coding); and Exon 3 contains 5,529 bp (4,512 bp coding, 1,017 bp 3' noncoding), which are separated by introns of 1,808 bp and 586 bp, respectively. It encodes a protein of 1,549 amino acids with a molecular weight of 201,172 which roughly corresponds with the size of 190,000 estimated from PAGE (Rothnagel and Rogers, 1986).

Analysis of the Predicted Amino Acid Sequence

The amino acid composition of the deduced trichohyalin protein sequence is shown in Table I and is compared with that obtained from a previous amino acid analysis of purified trichohyalin (Fietz et al., 1990). Trichohyalin is a hydrophilic protein with over 57% of the amino acids having ionizable side chains. It is further characterized by the four amino

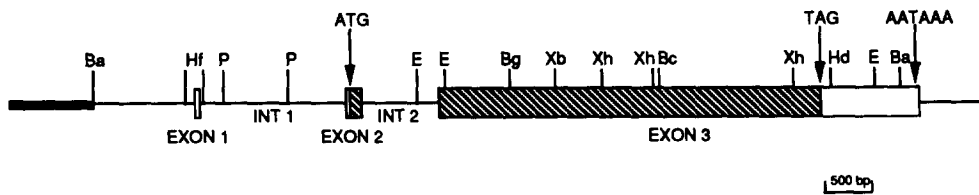


Figure 1. Structure of the sheep trichohyalin gene. A partial map of the region of sTr10 containing the trichohyalin gene is shown. The predicted start (ATG) and stop (TAG) codons are indicated as is the position of the putative polyadenylation

signal (AATAAA). Boxed regions indicate the exons and the hatched area represents the predicted coding region. The portion of the cosmid vector which abuts the 5' flanking region of the gene is represented by the black bar. The restriction sites shown were used in the sequencing and analysis of the gene. Note that not all of the sites for each restriction enzyme are necessarily indicated. Ba, BamHI; Bc, BclII; Bg, BglII; E, EcoRI; Hd, HindIII; Hf, HinfI; P, PstI; Xb, XbaI; Xh, XhoI.

acids, glutamic acid, glutamine, arginine, and leucine, constituting over 75% of the protein and has high levels of the substrate amino acids for peptidylarginine deiminase (arginine) and transglutaminase (glutamine and lysine). Although the relative levels of the different amino acids are similar in the deduced sequence and amino acid analysis, the absolute amounts of numerous amino acids differ significantly. The levels of glutamic acid/glutamine and arginine are markedly higher in the deduced amino acid sequence while the levels of threonine, proline, glycine, alanine, and isoleucine, as determined by amino acid analysis of the purified protein, are at least three times greater than in the deduced amino acid sequence. At present we have no logical explanation for these significant differences between the two amino acid compositions. Examination of the remainder of the amino acid composition shows that trichohyalin has a very low sulphur content containing only a single methionine and five cysteine residues.

Computer analysis of the overall charge ratio, using the computer program ISOELECTRIC (Devereux et al., 1984), has indicated that sheep trichohyalin has a pI of ~5.6 due to the excess of acidic residues within the protein. This corresponds with the estimation of a pI of 5.3 obtained from earlier isoelectric focusing analysis (Fietz, M., unpublished observations).

Closer examination of the trichohyalin protein sequence has revealed that it can be subdivided into three separate domains based on amino acid composition. The majority of the protein (residues 96–1,517) has the high proportion of hydrophilic residues indicated by the overall amino acid analysis (Table I) while small domains at the NH₂ terminus (residues 1–95) and COOH terminus (1,518–1,549) have a much higher proportion of hydrophobic residues. The NH₂-terminal domain has homology with a family of calcium-binding proteins (discussed below) but the COOH-terminal domain does not bear any close relationship with any other known protein sequence.

The secondary structure of sheep trichohyalin was analyzed using the program PREDICT (Eliopoulos et al., 1982) and PEPTIDESTRUCTURE (Devereux et al., 1984), both of which predicted that almost all of the large hydrophilic domain of trichohyalin has the capacity to form α -helix (data not shown). In addition, the high ratio of charged to apolar residues within the hydrophilic domain (3.3 compared with 0.9–1.4 for rod shaped molecules including IF proteins and 0.3–0.6 for globular proteins [Cohen and Parry, 1986]) suggests that trichohyalin will have a rod structure highly stabilized by ionic interactions. Together these data indicate that trichohyalin may form a single-stranded α -helical molecule.

The trichohyalin amino acid sequence does not contain the central region of heptad repeats required for the formation of the double stranded coiled-coil rod domain, characteristic of IF proteins. In agreement with the deductions from analysis of rabbit trichohyalin (Rogers et al., 1991), trichohyalin cannot form intermediate filaments.

The amino acid sequence deduced from the sheep trichohyalin cDNA clone is characterized by the presence of a tandem array of a 23-amino acid repeat. To further characterize the repetitive nature of sheep trichohyalin, a dot matrix plot of the complete trichohyalin sequence compared with itself was produced (Fig. 4). In addition to the repetitive region seen in λ sTr1 (seen to span residues 920–1,380 of the complete protein) there is a separate, apparently unrelated, repetitive region which appears to extend between residues 400 and 800 (see Fig. 4). To further examine this, the complete amino acid sequence of trichohyalin is arranged with respect to the two repetitive regions in Fig. 5. The central repetitive region contains 16 tandem repeats which are based on a 28-amino acid consensus sequence (Fig. 6) although five of the repeats also contain an insert of seven extra amino acids (see Fig. 5). The COOH-terminal repetitive region contains 28 repeats based on the 23-amino acid consensus sequence present in λ sTr1 (Fig. 6) (Fietz et al., 1990). Although the two consensus sequences have similar compositions, the sequences are considerably different with the major similarity being the identity of two five amino acid stretches in each repeat (EEEQL and RRQER, see Fig. 6). Both consensus sequences are predicted to form α -helical structures (data not shown) and also contain the substrate amino acids for peptidylarginine deiminase and transglutaminase. Note that the terminal repeats of each repetitive region are poorly conserved with respect to the consensus sequence.

The Calcium-binding Domain

The NH₂ terminus of trichohyalin is characterized by the presence of a much higher proportion of hydrophobic amino acids than the bulk of the protein. A database search with this sequence, using the UWGCG program FASTA, yielded significant similarities with a number of calcium-binding proteins which are characterized by the presence of the EF-hand calcium-binding motif typified by calmodulin (Vaman et al., 1977).

Examination of the trichohyalin sequence has shown that it contains two EF-hands, the first is encoded by Exon 2 and the second is encoded at the 5' end of Exon 3. Both the protein sequence and gene structure of the region encoding the

ggatcctacaaaatctcatcaaaattcagaaaacctatcttgcctctgttggaacaaggagagagaaatgccaaaaccactgaatcatttaagtctccatgaaccttttaca -1011
Bam HI
accaaaagtctcatctggctgaaatgtaagggtgagaatcaaccgggataaaaatcagggtcttacaataaaatagctgatgtgcaaaaatcaccagaggtgagaaaaccctga -891
cagacataatgggtcagaatgactgatctcctaagtgtctcccgagaccagactcaaggttttaggtgtcataatcatcagtaactggggtgcaacaagatgattatactgcaat -771
aatgcattacaccagctcttttccaccggggcatcagatgctaatgagaaacagaaatgagcaaatctttagatagaatcatcctaataaaatcaccagattcaagttctc -651
ttatgacaatttgatgaaaacatatttttaagaccctcaaaaatttttaaggacatgatctccaatctgtaggaaaatcataagccaaaagaatagatgcaaatggtttccc -531
accagcatttttagagaaaataaaactagttcttcaaaaactctgttgacctagaatggcctcaggcagctccatggactgtcatatcccggaaaagaatcatctgggtctcatta -411
gtttatgcaaaaattttgtttaacctcagcctagaatttatatcttttataaagttaagtgtttacaggtcactactttgtaagaaattccaattaactaagattctgaacatgaatt -291
tctttatttaatttaaacagttaagaaaatttagagattctgccaatattctatcctttcatgttaggtgagccagactctgtcaagaggaatggtctgtgaacctcgtgccaaaga -171
gtcactgtgagcaaacagcaagataaacaagtcatctcgtggagacagaggtggagctgttttagagcaaacctccacagagtggtgctctcctttcaagtgtcctt -51
cggggcacaattaaaacctctctaaaggccagctcccagctgtcagctACACCTGCCAGCAGTGCAGCGAGCATTGACCTCTTCTCTGGTGAAGTGGgaagtcacatctcggtg 70
gatcatgtctcctcgtatgatttctcatttccagctgtgtagatgttggtggggcagtggttaaggagggtctatgtgacagatggcatcagatcttacagatacactgctgctgct 190
gctgctgtaagtcaactctcgttctcaactctgtgtgacccatagacacagaggtcaccaggtcagagcctcagcctcaccagcctaccagcgtcctcgtccatgggattttccaggcaag 310
ttctctaagtcatgaaagtgaaaagtgaaagtgaagtgcctcagtcagctgactcttagcgacccatggagctcagcctaccagcgtcctcgtccatgggattttccaggcaag 430
tactggagtggtgctgttgcctctccaacagataatacctaattcctactaactccatgtcatctctatagagcacttaactctataagtttggtcggacgtcggagtgaggat 550
ggtcataatgactgatgtggctcatctccataaaaagtgttgccagcctcagttctcttttatttattgcttctacccaataggactatgttcatgtggttatggccactgagga 670
ttgaagtgagaagcaactaaagaagcagctttccaggttcagctgactcaccaaaatctggcagccagtcagaaaagcactaacagaaactcaaatgattgaatcattgtatacat 790
aagaataaggctaaacacgaactttgttttctgaaatgatataataaattacaataccacotctgactaggaagtagtacttaagcgtggtgacttacgaatttactcttag 910
atttcaaggactatatacagttagaaataaccacatgccctttttattgtccagttaccacagatgttttgcctacaattaggttacttaccatttttagacagcactgca 1030
tggcacaggcaatttacaggcttccctggtggctcagatggtaaaaatgctgctcagctcagagagaccagatgtgacctgggtgggagatccccgggagaggaatggcaac 1150
ccactccagctacttgcctgaaaatccatggaactgaggagcctgtaggtctacagctcagttcaggtggtgcgaaagagtcagacatgactgagcaactcactttatccatttaagatcc 1270
ttcacttttgttatccactatcctcttctgctcaattttagagaaaaattcttagtagtagaacctcaaatcaaatctcaaatctcactgtctgattgaaaacactaggga 1390
ataggaagaaagagaaggatgatgataatacaaaactggtagaaaaacaaaactaacacagtagcctgagatgcatccctgactgaaatttttcagatctcttcccctttata 1510
agcactacatgaaagtgaactcagttagttgtgctcgaagcctgtgactccatggactgtgactgctgctccttttccatgtaattctccagcaagagtagtggatgggtgc 1630
catttcttctccaggatcttcccccccaggatagaaaccaggtctcctgtgtgacggcagactcttaccacactgagcatttcaagaataaagtcttgaaggtcccagttg 1750
aaaactcttggtaaagtcaaattacttaggactcgcctgaaactgtagaaaagcctcccagaatcagaaacagagtttgaactctgaaatccaatttattgtctaggtTTACTGGA 1870
ACTTGAAGGAGAAAATGTCTCCACTTGTGAGAGCATCTTAAATACACTAAAATTTCAATCAATATGGCTCAGTATGTGATGGGACAACTAAGCAAGAAAAGACCTGAAG 1990
35 N L L E R E F G E I L R
AATCTCCTTGGAGGGAAATTCGAGAAATCTTCGGgaagaataacaagaaaaggaaatctctcactgatttagagctatgaaatcttctcagggagaccagagcaaggaaatg 2110
tgactttatattcatttattgcatcagctctcaccatttcccactctcagctcctgctcagctgaaatgtaataatttttaacaggtcctctctagtaaaatgg 2230
cattctcgaatggtttgttctgtttcacaacatgggataaaaatacaatcagcactgctgatttccagacagaatggctgctgcaaatcttcataaatttccactgcttctctct 2350
tgaggatgaattcatttatttaccagctataaactgtttttttagtcagcattaaataagatgcaaaaatttccacgaatgtgtgttaaaattaggtgtgctcacttttaca 2470
tgctacaagactccatcaaaactggttctcctaatgttcttattgctttttagtaccacatattataaaatatacagtggtggtggaagtttactacctgtaaatagcattgta 2590
47 K P H D P E T V D L V L E L Q D R D R D G L I D F H E F L A I V F
aaagtttatttctttttcagAAACCACGACCCGAAACTGTAGACCTAGTCTGGAACCTCAGGATCGGACCGTACGGGGTAAATCGATTTCCATGAATCTCGCAATCGTTTT 2710
80 K V A A C C Y Y A L G Q A S G L N E K E A K C E R K A G N P L Q D R R R E L D Q R R
CAAAGTGGCTAGACTTGTATTGCTCTGGCCAGGCTCGGGCTAAATGAGAAGAGCCAAAGTGTGAGAGAAAGGGAACCCGTTACAGATCGCAGGAGGGAAGCAAGAGG 2830
120 F E P Q D R Q L E E R R L K R Q E L E L A E E E E L R E K Q V R R E Q R L Q R
ATTCGAGCCCCGAGCAGCACTAGAAGAACGCGAGGCTGAAGAGGAGGAACTGGAGGAGCTCGCTGAGGAAGGAGGCTGAGGGAGGAGCAGTAAAGAGCTGAGCAGAGGCTGCAAG 2950
160 R E Q E E Y G G E E E L Q Q R P K G R E L E E L L N R E Q R F E R Q E Q R E R Q
GCGAACAAGAAAGATATGGTGGGGAGGAAGAGCTGCAGCAAAAGGCCAAGGCTCGGGAGCTGGAGGAGCTTGAACCAGAGCAAAAGTTCGAGAGGCGAGCAACGGGAGCGACA 3070
200 R L Q V E Q Q R Q R G E L R E R Q E E V Q L Q K R E T Q E L Q R E R L E E E Q
GCGCTCCAGGTGGAACGCAACGCGCAGGAGCTGCGGAGCGCCAGGAAGAAGTGCAGCTCCAAAAGCGGGAGACTCAGGAGCTCCAGAGGAGCGCTGGAGGAGAGCA 3190
240 Q L Q K Q K R G L E E R L L E Q E R R E Q E L R R K E Q E R R E Q Q L R Q E Q E
GCAACTGCAAAAAGCAAGGCGGGCTCGAGGAGCGGCTCTGGAGCAGGAGGCGGGAGCGAGCTGCGGGCAGGAGCAAGAAAGCGGAGCAGCAGCTAAGCAAGAGCAGGA 3310
280 E A T Q E E I S E R G E S R T S R C Q W Q L E S E A D A R Q R K V Y S R P H R Q
GGAGGGAGCGCAGGAGAGATCAGTGCAGGGGGCAGGAGCGCACCTCGAGGTGTCACTGGCAGCTAGAAAAGCGGCGGAGCGCCGCTCAGCGCAAGTCTACTCCAGGCCCCAGGCA 3430
320 E Q Q S R R Q E E Q E L L E R Q Q G E Q Q I S E E V Q S L Q E D Q G G R Q R L K Q E Q
AGAGCAGCAGAGCCCGCCGAGGAGGAGCTGCTGAGAGCAGCAGGAGCAGCAGATCAGCGAGGAGGTGACAGGCTTCAAGAAAGACCGGGGAGGCAACGACTCAAGCAAGAGCA 3550
360 R Y D Q N W R W Q L E E E S Q R R R Y T L Y A K P A Q R E Q V R E E E Q L R L K
ACGCTACGACCAAACTGAGGAGGCAACTAGAGGAGGAAAGCAAGAGCAGCGCTACACTGTACCGCAAGAGCAGTCAACGGGAGCAGGTGAGGAGGAGAGCAGCTCCGCTTAA 3670
400 E E K L Q R E K R R Q E R E R Q Y R E A V E L Q R E E E R L Q R E E E Q L Q R E E
AGAGGAAAGCTGACGGGAGGAAGAGGCGCCAGGAGAAAGGCAAGTATCGGGAAGTGGAGCTGCAGAGAGGAGGAGGAGTACAGCGGGAGGAGCAGCTGCAGAGAGGAGGA 3790
440 R E K R R R Q E R E K Q Y L E K V E L W E E E Q L Q R E E R E R E K R R Q E R E K Q
ACGCGAAGAGGAGGCGCCAGGAGCCTGAGAAGCAATACCTAGAGAAAGTGGAGCTGTGGGAGGAGGAGCAGTTGCAGAGAGGAAACCGGAGAGGCGCCAGGAGCGCGAAGAGCA 3910
480 Y L E K V E L R E E E Q L Q R Q E R E K R R Q E R E R Q Y L E K V E L Q E E E Q
ATATAGAGAAAGTGGAGCTGCGGGAGGAGGAGCAGCTGCAGAGCAGGAAACCGGAGAGGCGCCAGGAGCGGAGAGGCAATATCTAGAGAAAGTGGAGCTGCAGGAGGAGGAGCA 4030
Xba I

Figure 2. The nucleotide sequence and predicted amino acid sequence of the sheep trichohyalin gene. The 9,344-bp sequence extending from the cosmid arms of sTr10 to 86 bp downstream of the polyadenylation signal of the trichohyalin gene was determined. The exon sequences and 5' flanking sequences are unambiguous. Exon sequences are shown in upper case while intron and flanking sequences are

```

520  L Q R E E R E K R R R Q E R E R Q Y L E K V E L Q E E E Q L Q R Q E R E K R R R Q E
    GCTGCAGAGAGAGAAACCGAGAAAGAGGCCGCCAGGAGCGCGAGAGGCATATCTAGAGAAAGTGGAGCTGCAGGAGAAAGAGCAGCTGCAGAGACAGGAAACCGAGAAAGAGGCCAGAA 4150
560  R E K Q Y L E K V E L Q E E E Q L Q R Q E R Q K R R Q E R E K Q Y L E K V E L Q
    GCGCGAGAGCAATATCTAGAGAAAGTGGAGCTGCAGAGAGAGGAGCGAGCTGCAGAGACAGGAAACCGCAGGAGGCGCCAGGAGCGCGAGAAACCAATATCTAGAGAAAGTGGAGCTGCA 4270
600  E E E Q L Q R Q R E R E K R R R Q E R E R Q Y L E K V E L Q E E E E Q V Q R Q E R E K
    GGAGGAAAGACAGCTGCAGAGACAGGAAACCGGAGGAGGCGCCAGGAGCGCGAGAGCCAAATCTAGAGAAAGTGGAGCTGCAGAGGAGGACGAGGTGCAGAGACAGGAAACCGCGAGAA 4390
640  R R Q R E R E R Q Y L E K V E L Q R Q E E R L Q E E E Q L L R E R E R E K R R R Q E R
    GAGGCCAGGAGCGCGAGAGGCAAGTACTTCGAGAAAGGAGCTGCAGGCGAGGAAAGCAGCTGCAGGAAAGAGGAGCAGCTGCTGAGGAGGAAACCGAGAAAGAGGCCAGGAGCGCGA 4510
680  R Q Y L E K V E L Q E E E Q L Q R E R E R E K R R R Q E R E R Q Y L E K E E L Q R Q
    GAGGCAGTATCTAGAGAAAGTGGAGCTGCAGGAAAGAGGAGCACTTCGAGAGAGAGGAAACGAGAAAGAGGCCAGGAGCGTGCAGAGGAGCAGTATCTCAGAGAGGAGGAGCTGCAGCGCA 4630
720  E E R L Q R E K E Q L Q R E D R E K R R Q V R E R K Y L E E E E Q Q E E D R L Q
    GGAAGAGCAGCTGCAGAGAGAGGAGCAAGTGCAGAGAGGAGCCGCGAGAGAGGCGCCAGGTGCAGGAGAGAGTATCTCAGGAGGAGCTGCAGAGGAGGAAAGCCGAGTTGCA 4750
760  R E K Q L L R E D R E K R Q Y L E K V E L Q R E E E Q L Q R E K R R R Q E R E R Q
    GAGAGAGAAAGCAGCTGCAGAGAGGAGTGCAGAGAGAGCAGTATCTAGAGAAAGTGGAGCTGCAGCGGAGAGGAGCAGCTGCAGAGGAGAAAGGAGCCCAAGAGCGGAGAGAGCA 4870
800  Y R E E E L L R E E E R L H R K E Q Q L Q R E E C E K R R R R Q E L E R Q L E E E
    ATATCCGGAAGAGGAGCTGCTGCGGGAGGAAGAGCAGTGCAGAGAAAGGAGCAGCAGCTGCAGAGAGAGGAAATGTGAGAGAGGAGGAGCGCCAGGAGCGCGAGAGCAGCTAGAGGAGAA 4990
840  E L Q R L D R K R Q F R D D D Q H Q N E V R N S R V Y S K H R E N K E K S R Q L
    GGAGCTGCAGCGCTTGCAGAGAGAGGCAATTCGCGGATGATGATCAGCCAAACCGAATTCGTAATAGCAGGTTTACTCCAAACCCGAGAGAAAGAGAAAGAGCGCGCACT 5110
880  D D S W V R E S Q F Q Q D L R P L Q D E Q E E K R E R E Q E W R S R Q K R D S Q
    GGATGATTCTGGGTCGGGAGGAGTCAATTCAGCAGGATCTGCAGCAGTGCAGGATGAACAAGAGAGAAAGAGAAACCGCAGAGAGTGGAGGAGCCCGCAAAAGCGTGACAGCA 5230
920  F P A E Q L L E R E A Q Q K E T E R R D R K F R E E E Q L L K G Q R E E K I R Y L
    ATTTCAGCTGAACAACTGCTGAGAGGAGCAGCAAGGAGGAAACCGAGGAGCAGCAAGGAAATCCAGAGGAGGAGGAGCAGGACTGCTGAAAGGCGAAAGAGGAGGAGGAGAAATACGCTATCT 5350
960  E E D R K F R E E E Q Q L R L E R E Q Q L R Q E R D R K F R E E L S R Q E R D
    GGAAGAGACAGAAAGTTCGCGGAGAGGAAAGAGCAGTGCAGCGAGCAACAGCTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5470
1000 R K F R E E E Q L L Q E R E E Q L R R R Q E R D R K F R E E E Q L L Q E R E E Q L
    CAGAAATTCGGTGAAGAGAGCAGCTCTCAGGAGAAAGGAAAGACAGCTGCCGCCCAAGAGCGCGACAGAAAGTTCGGTGAGGAGGAACAGCTCCTGCAGGAAAGGAAAGAGCAGCT 5590
1040 R R Q E R D R K F R E E E Q L L Q E R E E Q L R R R Q E R D R K F R E E E Q Q Q L R
    GCGCGCGAAAGCGCAGCAAGAGTTCGGTGAAGGAGGAGCAGCTCCTGCGAGAAAGGAAAGCAAGCTGCGCGCGCCAGGAAACCGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5710
1080 L L E R E Q Q L R Q E R N R K F R E E Q L L R E R E E Q L R L Q E G E P Q L R Q
    CCTCTGGAACGCGCAGCAACAGCTAGCCAAAGAGCAAGAAATAGAAATTCGCGGAGAAAGCAAGCTCCTGCGAGAAAGGAGAAAGCAAGCTGCGCCTCCAGGAGGCGAGCGCCAGCTTCGCCA 5830
1120 K R D R K F H E E E Q L L Q E R E E Q L R R Q E R D R K F R E E A Q I L K E R E
    GAAGCGCATAGAAAGTTCATGAGGAGGAAACAGCTCCTGAAAGAAAGAGAAAGCAGCTGCAGCGGAGGAAAGCGCGAGCAAGAGTTCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5950
1160 E Q L R R Q E R D R K F R E E E Q L L Q E R E E L R R R Q E R E P Q L R Q E R D
    AGAACAGCTCCGCGCGAGGAGCGCAGCAAGTTCGGTGAAGGAGGAAAGCAGCTCCTGCAAGAAAGAGAGAGTGCAGGCGCGAGGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 6070
1200 K F R E E E Q L L Q E R E K L R R R Q E R E P Q L R Q E R D R K F H E E E Q L L Q
    AAAGTTCGGTGAGGAGGAAAGCAGCTCCTGCAAGAAAGAGAAAGCAGCTGCCGCCAGGAGGCGCGGAGCACAAGTTCGCGAGGAAAGCGCAGAGAAAGTTCATGAGGAGGAAAGCAGCTCCTGCA 6190
1240 E R E E Q L R R Q E R D R K F R E E E Q L L Q E R E E Q L R R R Q E R D R K F R E
    GGAAAGGAAAGACAGCTGCGCGGAGGAGGAAAGGAGTTCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6310
1280 E E Q L L Q R R Q E R E E Q L R R Q E R D R K F R E E E Q L L Q E R E E Q L R R Q E R
    GGAGGAAAGCAGCTCCTGCGAGGAAAGGAGGAAAGCAGCTGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6430
1320 D R K F R E E E Q L L K E S E E Q L R R R Q E R D R K F H E K E H L L R E R E E Q
    CGACAGAAAGTTCGGTGAGGAGGAAACAGCTCCTGAAAGAAAGCAGAAAGCAGCTGCCGCCAGGAGGCGCGAGCAGGAAAGTTCATGAGAAAGAAACAGCTCCTTCGAGAAAGGAGGAGAA 6550
1360 Q L R R Q E L E G V F S Q E E Q L R R A E Q E E E Q R R Q R Q R D R K F L E E E
    GCAGCGCCCGCTCAGGAACTTGAGGGGCTTCTCCCAAGGAGAAAGCTGAGGGCGGCGAGCAAGAGAAACAGCAGCTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6670
1400 Q S L Q R E R E E E K R R V Q E Q D R K F L E Q E E Q L H R E E Q E E L R R R Q
    GCAAGCTCCAGCGCGAGGAGGAAAGAAAGCGCGGCTCCAGGAGGAGGAGGAAAGTTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6790
1440 Q L D Q Q Y R A E E Q F A R E E K R R R R Q E Q E L R Q E E Q R R R R Q E R E R K F
    GCAGCTAGACCAGCAGTACCGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6910
1480 R E E E Q L R R Q Q E E Q K R R E R D V Q Q S R R Q V W E E D K G R R Q V L
    CCGGAAAGAAAGCTCCGCGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7030
1520 E A G K R Q F A S A P V R S S P L Y E Y I Q E Q R S Q Y R P *
    GGAGGCTGGCAAGCGGAGTTCGCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 7150
    CCGAAGCTTCGAGCAAAAGGAAATGAGAACTACTGAGTACCAATGACTCTGGTGTGGGAAACTCTGGTGTAGACTAAGCTCTTTTACAAAATCTTTAATCTATATCTTTTCATGT 7270
    GCTTTGFACTCTGCCTTTTATCTTCTTAAATAGTCTTGTAGATGTCTTGTCTTGTGTGCGAGATTTGGTGCAATTTTTTAAAACATAAAAAGCCATTTAATTTGTTTAAGGAAT 7390
    TTTGTTGGGAAACAGTTCAATCATTGCTTCCAGAAAGTAAACAAAATATTGTGTCCATTTGAGATTCAAAAGAAATGGTCCAGTCTTTTATTGTTGATCCATCTTATAGAGAGCTCAGA 7510
    TATTTTTATGTTCAAGTGTATTTCTTCTGCGCATAAATATGTTAATTTATCTCCAAAAGAGCTCCCACTTTTTGTTGCATAATTAGCAGAGTTCGCAAGGAGGAGCGAAATT 7630
    TTCAAAAGAACCCCTGAAAGTGCAGGAAAGTGGTCTTCCAGAAAAAGTCTGAAATTCAGCCCAATAATCGGAGAAATATCTTTAAGACTGAAATAGATTTCTTTTCCATTTAATT 7750
    CCATATATACTTATGAAGCAAAAAGCAGGCGTCTCTCAAACAACCCAGTTCGAGTCTAAAAGTCTTTAACAACCTGCTGCTGAGTCTGTGACTTCAAGTCTCTTCCCAAGTCTTGGATGAG 7870
    CTTGAGACTTACTGCTTTGTTAATTTAGCTGAGAGGAGCTGCTCAGGATCTGTAACACTTTAAAACACTGCTGCTGTAGCTGTGAGTCTGAGACTTCAAGTCTCTTCCCAAGTCTTGGATGAG 7990
    GGGAAAGTTTGTGGAGAGGAGGCAAGTCCCTGAGGAGCTTGAAGAGCGCTGCAGCCCAAGTCTTGAAGCCTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 8110
    TAAAAAATAATATAATAATTTTTGGCAATTTTTCCTCTTTTTactgtgcatcaaacataactcaagatttaagtgtgtggaactgacataactttcc 8214

```

in lower case. The binding sites for the primers Tr3 and Tr5, used for analysis of the 5' end of the trichohyalin gene, are underlined, as are the polyadenylation signal (position 8123) and GT- and T-rich regions (position 8143) required for processing of the mRNA species. The probable TATA and CAAT boxes are indicated while possible AT-1, AP-2, and NF- κ B binding sites (detected by the computer program SIGSCAN [Devereux et al., 1984]) are boxed. Note that immediately upstream of the CAAT box the gene contains a perfect 8-bp inverted repeat with the sequence between the two repeats containing two 6 bp direct repeats (indicated by *arrows*). The complete predicted amino acid sequence of trichohyalin is shown. Restriction sites denoted in Fig. 1 are underlined. These sequence data are available from EMBL/GenBank/DBJ under accession number Z18361.

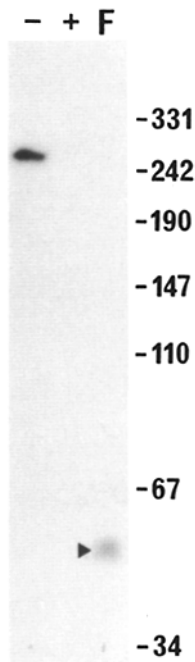


Figure 3. Determination of the size of Exon 1. The 186-bp *Hinf* I fragment surrounding the predicted location of Exon 1 was cloned into pBluescript II SK+. The resultant plasmid was linearized and transcribed with T7 RNA Polymerase producing radiolabeled anti-sense cRNA. The resultant cRNA probe was hybridized to total wool follicle RNA (20 μ g) and digested with RNase A and T1 (Krieg and Melton, 1987). The protected fragment (▶) was resolved by denaturing PAGE: (-) track, no RNase digestion; (+) track, control with yeast RNA rather than follicle RNA; F, RNase protection of wool follicle RNA.

EF-hands are very similar to those of the S100 family of small calcium-binding proteins (Fig. 7). As is the case for these proteins, trichohyalin consists of two EF-hands, the first of which is two amino acids longer than the standard EF-hand of 28 amino acids (Szebenyi et al., 1981). The trichohyalin gene, like those encoding the S100 proteins, contains two introns, one in the 5' noncoding region and the other positioned at a conserved site between the two hands

Table I. Amino Acid Composition of Sheep Trichohyalin Determined by Amino Acid Analysis and from the Sheep Gene Sequence

Amino acid	Wool follicle trichohyalin [§]	Sheep gene sequence
	<i>mole percent</i>	
Asp/Asn*	6.4	3.2/0.7
Thr	2.9	0.6
Ser	5.4	1.9
Glu/Gln [‡]	28.0	26.1/17.2
Pro	3.3	1.0
Gly	5.3	1.2
Ala	4.7	1.5
Cys	0.6	0.3
Val	4.0	2.0
Met	0.1	0.1
Ile	2.5	0.7
Leu	10.0	11.0
Tyr	2.1	1.8
Phe	2.4	2.3
Lys	6.7	6.1
His	1.7	0.8
Trp	0.2	0.5
Arg	13.7	21.2

* Aspartic acid and asparagine are denoted separately for the gene sequence composition but are combined for the wool follicle analysis.

[‡] Glutamic acid and glutamine are denoted separately for the gene sequence composition but are combined for the wool follicle analysis.

[§] Taken from Fietz et al. (1990).

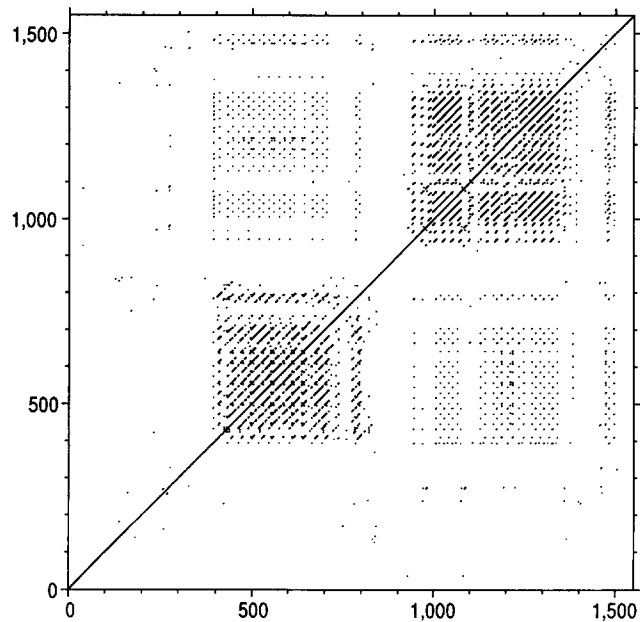


Figure 4. Dot matrix comparison of the sheep trichohyalin amino acid sequence. The sheep trichohyalin amino acid sequence was compared with itself in order to search for internal repetitive sequences using the computer program COMPARE (word size of 5). The results were plotted by the program DOTPLOT. The analysis has detected two repetitive regions, indicated by the arrays of lines parallel to the central diagonal, spanning from residues 400 to 800 and from residues 930 to 1,400. The two blocks of dots away from the diagonal indicate points of homology between the two repetitive regions. Axes are labeled in residue numbers.

(Lagasse and Clerc, 1988). Although trichohyalin is more closely related to the S100 proteins, the glycine within the first hand occurs at the conserved site for the remainder of the EF-hand proteins, for example, calmodulin, parvalbumin, and troponin C, rather than at the alternative position found in the S100 proteins (see Fig. 7).

The Noncoding and Flanking Regions of the Trichohyalin Gene

The promoter of the trichohyalin gene contains a number of transcriptional control sequences including a typical TATA box and a modified CAAT box (Fig. 2). In addition the sequences immediately upstream of the CAAT box and at the 5' end of the first intron contain several putative control sequences (see Fig. 2).

The transcribed message contains a single polyadenylation signal (position 8123, Fig. 2), corresponding to the site located in λ sTr1, and has the GT-rich and T-rich regions (nucleotides 8143–8156, Fig. 2) required for the efficient and accurate formation of a correct 3' mRNA terminus (Gil and Proudfoot, 1987). The sequence immediately preceding the initiating ATG is markedly different from the Kozak consensus sequence (Kozak, 1987) and consists of six consecutive adenosine residues.

Comparison with cDNA Sequence

Comparison of the deduced trichohyalin protein sequence with that derived from the cDNA clone has revealed that the sequence from sTr10 contains two large inserts introducing

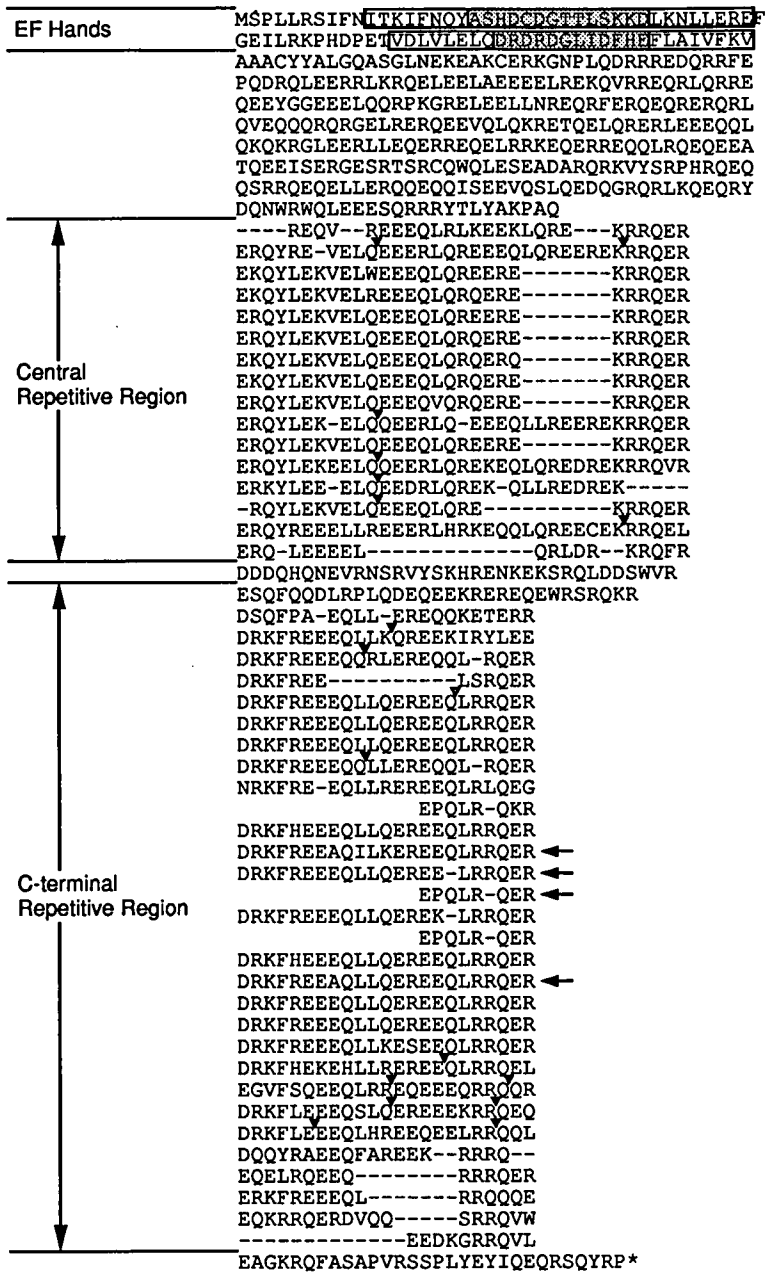


Figure 5. Amino acid sequence of sheep trichohyalin. The amino acid sequence of sheep trichohyalin is arranged with respect to the consensus sequences of both the central and COOH-terminal repetitive regions (see Fig. 6). Both repetitive regions are marked. To optimize alignment of the sequence with the consensus sequences, spaces have been introduced (dashes) and stretches of one or two amino acids removed (arrowheads). The arrows to the right of the COOH-terminal repetitive region denote repeats which are not present in the previously isolated trichohyalin cDNA clone (Fietz et al., 1990) which encodes the region extending from the open arrowhead through to the 3' end of the mRNA. Also shown are the helix (open box) and turn (hatched box) regions of the proposed EF hands.

four extra repeats or partial repeats into the COOH-terminal repetitive region of the deduced cDNA sequence (see Fig. 5). This would give the protein an extra mass of ~10 kD. Apart from these insertions there is only one alteration in the amino acid sequence (replacement of a glutamate with a glycine at residue 1399).

The genomic 3' noncoding region is almost identical to

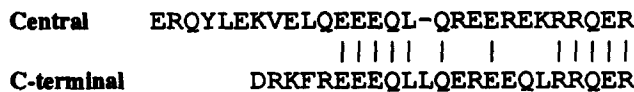


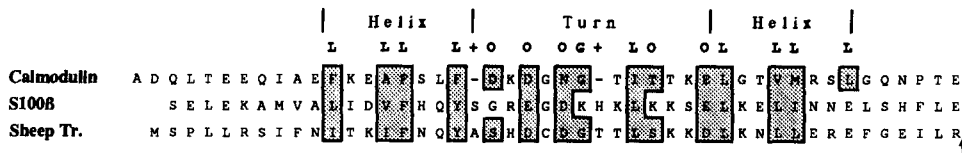
Figure 6. Comparison of the consensus sequences for the central and COOH-terminal repetitive regions. The consensus sequences for the central repetitive region and the COOH-terminal repetitive region of sheep trichohyalin are aligned. Note that two stretches of five amino acids are present in both consensus sequences.

that of the cDNA sequence and contains only five single base changes spread throughout the noncoding region and an 8-bp deletion. Comparison of the 5' noncoding sequences available from the PCR amplification of the 5' end of the mRNA has shown that there are at least three single base changes in the noncoding sequences in Exon 1. All of these differences between the cDNA and genomic sequences are attributed to the fact that the DNAs originated from different strains of sheep.

In Situ Hybridization to Sheep Keratinized Epithelia

A variety of sheep epithelia were probed with a cRNA probe derived from the COOH-terminal repetitive region of sheep trichohyalin to examine the extent of trichohyalin expression. Previous work had shown that the same probe hybridized to the medulla and IRS of the wool follicle (Fietz et al., 1990).

FIRST EF HAND



SECOND EF HAND



Ygen-containing; G, glycine. Note that trichohyalin and S100 β contain a variant first hand which contains two single amino acid inserts (+). The arrow at the end of the first hand indicates the conserved location of the intron occurring in the genes encoding trichohyalin and a number of other S100-like proteins (Lagasse and Clerc, 1988).

Initial work confirmed the expression of trichohyalin in the papillae of sheep tongue (Fig. 8) and the developing sheep hoof (Fig. 9) as originally determined by immunohistochemical analysis (O'Guin and Manabe, 1991). Further analysis

determined that sheep rumen epithelium also contains small amounts of trichohyalin mRNA (Fig. 10) and trichohyalin granules similar to those present in the follicle IRS and medulla. No signal was detected in esophageal epithelium or the epidermis (data not shown).

Discussion

The cells of the IRS of the hair follicle form a novel system, the structure, and differentiation of which is incompletely understood. The developing cells are filled with a large number of nonmembrane-bound granules which contain the protein trichohyalin. A rapid transformation, involving the disappearance of the trichohyalin granules, produces an organized array of intermediate-like filaments aligned parallel with the direction of hair growth. The mature cells contain the amino acid citrulline (Steinert et al., 1969) and the proteins are cross-linked by ϵ (γ -glutamyl)lysine isopeptide bonds (Harding and Rogers, 1971). Radioactive tracing of the fate of trichohyalin has shown that it remains in the mature cells (Rogers et al., 1977). It is currently believed that trichohyalin acts as an IFAP within the mature IRS cells, intermingling with separately synthesized IFs, however, the evidence for this is still limited. Numerous studies have detected IF proteins within the IRS using a variety of different antibodies (Lane et al., 1985; Ito et al., 1986a,b; Heid et al., 1988; Stark et al., 1990) yet almost all of these detect similar levels in both the IRS and medulla thus giving no clear indication why the two tissues are markedly different in structure. Separate immunoelectron microscopic studies using polyclonal and monoclonal antibodies to trichohyalin (Rothnagel and Rogers, 1986; O'Guin et al., 1992) have suggested that trichohyalin acts as an IFAP in the mature IRS.

Within the present study we have determined the complete amino acid sequence of sheep trichohyalin in an attempt to further analyze the role of trichohyalin. The protein is 1,549 amino acids long and has a molecular weight of 201,172. It does not contain the central region of heptad repeats, required for the formation of a coiled-coil molecule, which is present in all IF proteins indicating that trichohyalin cannot form an IF. Nevertheless, the prediction of α -helix over almost the complete length of trichohyalin together with the very high ratio of charged to apolar residues (3.3) suggests

Figure 7. Comparison of the EF hands of trichohyalin with those of calmodulin and S100 β . The sequences of the first and second EF hands of sheep trichohyalin were aligned with the corresponding regions of bovine calmodulin (Vanaman et al., 1977) and human S100 β (Jensen et al., 1985) which is a member of the family of S100-like proteins. The helix and turn regions of each hand are denoted and the amino acids corresponding to the conserved residues of the EF hands are boxed: L, hydrophobic; O, ox-

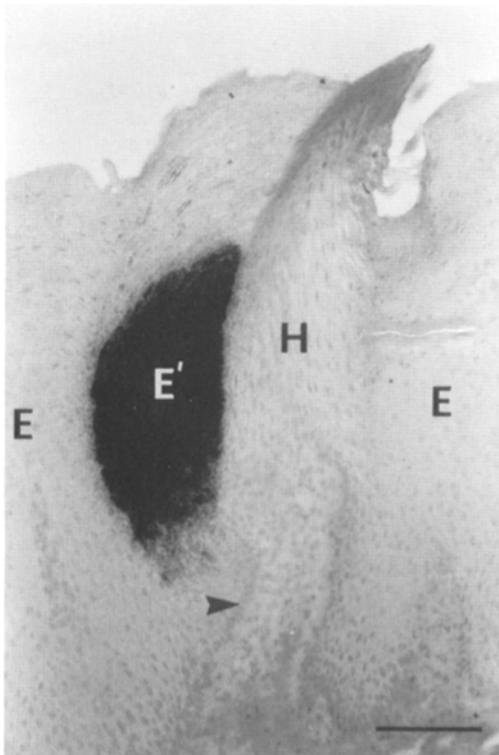


Figure 8. In situ localization of trichohyalin expression in the sheep tongue. Sheep tongue sections were hybridized with ³⁵S-labeled anti-sense and sense (produced random signals, data not shown) RNA probes derived from the COOH-terminal repetitive region of the sheep trichohyalin gene (Fietz et al., 1990). A bright-field micrograph of a filiform papilla is shown and labeled on the basis of the previously detected expression of the esophageal (E) and hair (H) IF proteins within the respective regions (Dhouailly et al., 1989). The region containing the trichohyalin mRNA appears to correspond with the E' region which has been shown to contain nonesophageal proteins in addition to the esophageal IF proteins. The arrow indicates the basal cells for the E' region. Bar, 100 μ m.

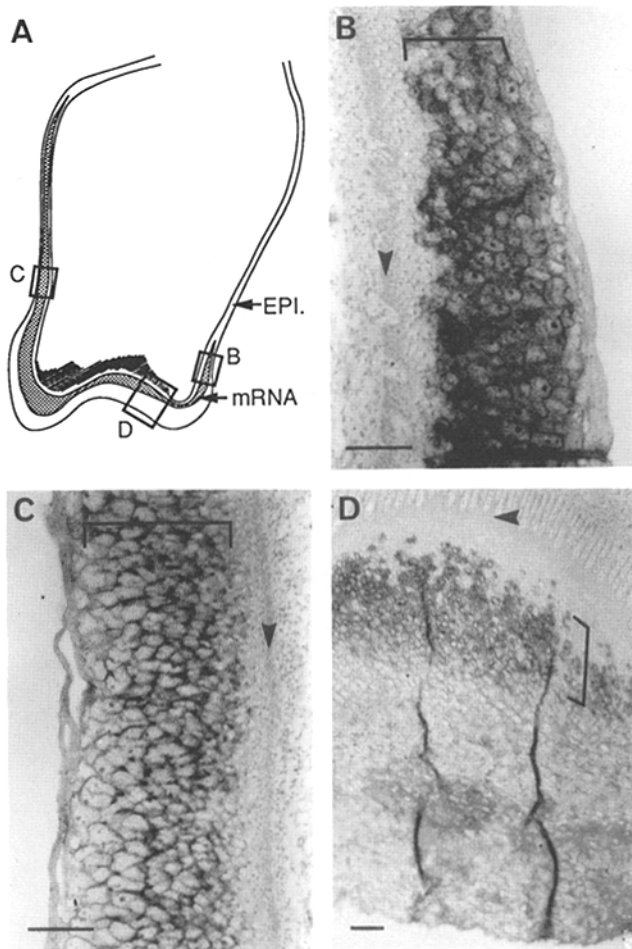


Figure 9. In situ localization of trichohyalin expression in fetal sheep hoof. Sheep fetal hoof sections were hybridized with ^{35}S -labeled anti-sense and sense (produced random signals, data not shown) RNA probes derived from the COOH-terminal repetitive region of the sheep trichohyalin gene (Fietz et al., 1990). (A) A schematic diagram indicating the hybridization of the trichohyalin anti-sense cRNA probe to a fetal hoof section. The extent of hybridization is depicted by the hatched region (mRNA) which is contained within the hoof epithelium (EPI). The locations of the micrographs in B, C, and D are shown. (B) Bright-field micrograph of the lower posterior portion of hoof epithelium. The region expressing trichohyalin mRNA is bracketed and the basal epithelial cells are arrowed. (C) Bright-field micrograph of the ventral hoof epithelium. The region expressing trichohyalin mRNA is bracketed and the basal epithelial cells are arrowed. (D) Bright-field micrograph of the anterior hoof epithelium. The region expressing trichohyalin mRNA is bracketed and the basal epithelial cells are arrowed. Bars, 100 μm .

that trichohyalin may be able to form an elongated rod shaped molecule in agreement with the electron micrographic studies of Hamilton et al. (1992). Although this structure may well act as an IFAP, it remains possible that an organized array of these molecules could be involved in the formation of the intermediate-like filaments of the IRS. If this is the case, there is no present indication why such filaments would not be able to form in the follicle medulla. The studies of Hamilton et al. (1992) have also suggested that trichohyalin is stable as a dimer. The presence of five cysteine residues may allow this to occur with sheep trichoha-

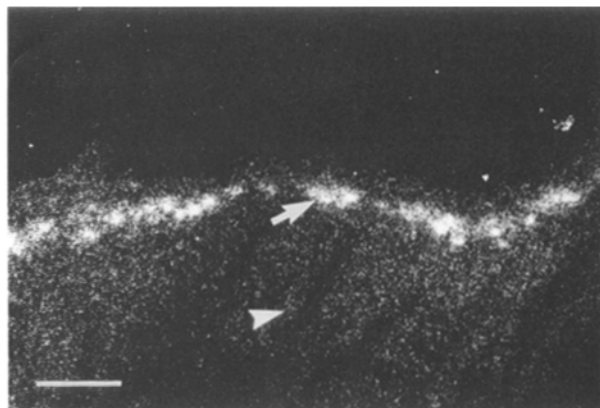


Figure 10. In situ localization of trichohyalin expression in sheep rumen. Sheep rumen sections were hybridized with ^{35}S -labeled anti-sense and sense (produced random signals, data not shown) RNA probes derived from the COOH-terminal repetitive region of the sheep trichohyalin gene (Fietz et al., 1990). A dark-field micrograph of a portion of the rumen epithelium is depicted. Trichohyalin expression occurs in many, though not all, of the cells in the granular layer of the rumen epithelium (arrowed). The arrowhead indicates the position of the basal epithelial cells. Bar, 100 μm .

lin thus containing two internal disulphide bridges leaving a single cysteine to form an intermolecular bridge producing a trichohyalin dimer.

All previous examinations of the role of trichohyalin have assumed that it plays purely a structural role within the IRS and medulla. The discovery of two EF-hand calcium-binding domains at the NH_2 terminus of trichohyalin strongly suggests that, in addition, trichohyalin plays a role in the control of calcium levels within the cells of the differentiating IRS and medulla. Furthermore, as both peptidylarginine deiminase and transglutaminase, two enzymes in the hair follicle known to act on trichohyalin, are calcium-dependent enzymes, it is possible that trichohyalin either stores calcium for each of these enzymes at the site of their substrate molecule or releases calcium for the activation of the enzymes. The role of calcium within the IRS and medulla is further emphasized by the expression of calcyclin, another calcium-binding protein, in both of these follicle layers (Wood et al., 1991). Calcyclin is also a member of the S100 family of EF-hand calcium-binding proteins. Together trichohyalin and calcyclin may control the availability of calcium in the IRS and medulla.

The amino acid sequence of sheep trichohyalin is characterized by the presence of two separate repetitive regions which together constitute $\sim 70\%$ of the total protein (Fig. 5). The COOH-terminal repetitive region extends only marginally more NH_2 -terminal than the sequence deduced from the cDNA clone λSTr1 (Fietz et al., 1990) (see Fig. 5) and is based on a 23-amino acid consensus sequence (Fig. 6). The central repetitive region is based on a 28-amino acid consensus sequence which has a very similar amino acid composition to the COOH-terminal consensus sequence although the individual sequences are markedly different. Both consensus sequences contain the substrate amino acids for both peptidylarginine deiminase (arginyl residues) and transglutaminase (glutamyl and lysyl residues). Although sheep trichohyalin contains two repetitive regions, it is probable that both perform a similar function since the central

repetitive region is not present in rabbit trichohyalin which has a much larger COOH-terminal repetitive region (Rogers et al., 1991; Fietz, M. J., manuscript in preparation). It is therefore probable that the function of the repetitive regions is to provide an ordered array of substrate amino acids for peptidylarginine deiminase and transglutaminase in an α -helical environment. In addition, the total number of repeats within trichohyalin may be controlled to maintain the overall length of trichohyalin which in those mammalian species examined, ranges from 190 to 220 kD (Rothnagel and Rogers, 1986).

The studies of the predicted structure of the deduced trichohyalin sequence do not take into account the changes in charge introduced into trichohyalin by the deimination of arginyl residues. These modifications in the charge profile will alter the structure of trichohyalin and its arrangement within the protein network of the mature IRS and medulla cells. To predict the structure of the modified trichohyalin molecule it will be necessary to determine the position of the substrate residues for peptidylarginine deiminase. This could in part be obtained by treating synthesized peptides containing the two consensus sequences with peptidylarginine deiminase and determining the modified residues by sequence analysis. Further examination of the role of trichohyalin could be obtained by the treatment of purified trichohyalin with peptidylarginine deiminase in the presence and absence of IF proteins. The mixtures could then be examined for the production of aggregates of IF. However, the most direct approach to the structure and function of trichohyalin would be the ablation of the mouse gene using embryonic stem cell technology.

The purification of the trichohyalin gene has allowed a long-standing question regarding the trichohyalin protein to be answered. That is, what causes the protein doublet seen upon the initial purification of sheep trichohyalin (Rothnagel and Rogers, 1986) and pig trichohyalin (Hamilton et al., 1992)? Comparison of the protein sequences deduced from the genomic and cDNA clones has shown that the sequence deduced from the genomic clone has four full or partial repeats inserted within the COOH-terminal repetitive region of the cDNA sequence. This would produce a difference in mass of 10 kD. Thus it appears that the overall number of repeats may be altered by incorrect crossovers during meiosis producing alleles encoding proteins of differing molecular weights. Purification of trichohyalin from a heterozygous animal would therefore produce a doublet, explaining the results originally seen in the purifications of Rothnagel and Rogers (1986) and of Hamilton et al. (1992).

Complete analysis of the trichohyalin gene structure has further demonstrated similarities between trichohyalin and a number of epidermal structural proteins, namely, involucrin, profilaggrin, and loricrin. All four proteins are substrates for transglutaminase (Rogers et al., 1977; Rice and Green, 1979; Simon and Green, 1984; Richards et al., 1988; Mehrel et al., 1990), contain quasi-repetitive structures (Eckert and Green, 1986; Rothnagel et al., 1987; Fietz et al., 1990; Haydock and Dale, 1990; Mehrel et al., 1990; Rogers et al., 1991) and are encoded by genes located at the same position on chromosome 1 (McKinley-Grant et al., 1989; Simon et al., 1989; Fietz et al., 1992). Furthermore, all four genes have a similar structure with each characterized by the presence of a single intron in the 5' noncoding region (Eckert

and Green, 1986). However, in contrast to involucrin and loricrin, profilaggrin more closely resembles trichohyalin in that there is an additional intron within the coding region of its gene and it has two EF hands at the NH₂ terminus (Presland et al., 1992). This feature may have been brought about by the fusion of the gene of an S100-type calcium-binding protein to an epithelial structural gene related to involucrin and loricrin thereby producing the trichohyalin and profilaggrin genes. These strong similarities suggest that all four proteins may be members of a super family of epithelial structural proteins (Fietz et al., 1992).

The in situ hybridization results are generally in agreement with earlier immunological work performed on various human epithelia (Hamilton et al., 1991; O'Guin and Manabe, 1991; Hamilton et al., 1992). Trichohyalin mRNA is present in a subset of filiform papillae keratinocytes (Fig. 8) in a region labeled the "E" region on the basis of co-expression of IF proteins normally present in the esophagus (Dhouailly et al., 1989). These cells also contain granules with similar staining characteristics to the trichohyalin granules of the IRS and medulla (Fietz, 1991), and have also been shown to express filaggrin (O'Guin and Manabe, 1991). Trichohyalin mRNA is also present in the developing cells of the fetal hoof (Fig. 9) which again corresponds with the presence of trichohyalin-like granules (Fietz, 1991) and the immunological detection of trichohyalin in the human nail bed (O'Guin and Manabe, 1991). Trichohyalin has previously been detected within foreskin epidermis (O'Guin and Manabe, 1991) and other nonhair-bearing epidermis (Hamilton et al., 1991), but was not detected in the wool-bearing epidermis of the sheep. It is possible that nonwool-bearing sheep epidermis may contain cells expressing trichohyalin. Trichohyalin expression was also detected within a number of cells within sheep rumen epithelium (Fig. 10). The expression appears to be in the granular layer of the rumen epithelium which may correlate with the detection of trichohyalin in the granular layer of nonhair-bearing skin (Hamilton et al., 1991; O'Guin and Manabe, 1991). The role of trichohyalin in each of these tissues is presently uncertain although each express sets of IF proteins suggesting that trichohyalin may in each case act as an IFAP. Alternatively, mature nail epithelial cells contain a cellular envelope which differs in amino acid content from the epidermal cellular envelope (Baden and Fewkes, 1983) and is believed to contain ϵ -(γ -glutamyl)lysine cross-links (Baden and Fewkes, 1983; Shono and Toda, 1983) suggesting that trichohyalin may be involved in the formation of the cellular envelope of the hoof epithelial cells.

This work was supported by a grant from the Australian Research Council.

Received for publication 14 December 1992 and in revised form 1 February 1993.

References

- Baden, H. P., and J. Fewkes. 1983. The nail. *In* *Biochemistry and Physiology of the Skin*. Vol. 1. L. A. Goldsmith, editor. Oxford University Press, New York. 553-566.
- 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, E. Y., and P. H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA (NY)*. 4:165-170.
- Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Bio-*

- chem. 162:156-159.
- Chung, S. I., and J. E. Folk. 1972. Transglutaminase from hair follicle of guinea pig. *Proc. Natl. Acad. Sci. USA.* 69:303-307.
- Cohen, C., and D. A. D. Parry. 1986. α -Helical coiled coils—a widespread motif in proteins. *Trends Biochem. Sci.* 11:245-248.
- 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.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387-395.
- Dhouailly, D., C. Xu, M. Manabe, A. Schermer, and T.-T. Sun. 1989. Expression of hair-related keratins in a soft epithelium: subpopulations of human and mouse dorsal tongue keratinocytes express keratin markers for hair-, skin- and esophageal-types of differentiation. *Exp. Cell Res.* 181:141-158.
- 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.
- Fietz, M. J. 1991. Purification and analysis of the trichohyalin gene: an examination of the role of trichohyalin in the inner root sheath. Ph.D. thesis. University of Adelaide, Adelaide, South Australia. 103 pp.
- Fietz, M. J., R. B. Presland, and G. E. Rogers. 1990. The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker in the hair follicle, contains a 23 amino acid repeat. *J. Cell Biol.* 110:427-436.
- Fietz, M. J., G. E. Rogers, H. J. Eyre, E. Baker, D. F. Callen, and G. R. Sutherland. 1992. Mapping of the trichohyalin gene: co-localization with the profilaggrin, involucrin, and loricrin genes. *J. Invest. Dermatol.* 99:542-544.
- Frohman, M. A., M. K. Dush, and G. R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA.* 85:8998-9002.
- Gil, A., and N. J. Proudfoot. 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β -globin mRNA 3' end formation. *Cell.* 49:399-406.
- Hamilton, E. H., R. E. Payne Jr., and E. J. O'Keefe. 1991. Trichohyalin: presence in the granular layer and stratum corneum of normal human epidermis. *J. Invest. Dermatol.* 96:666-672.
- Hamilton, E. H., R. Sealock, N. R. Wallace, and E. J. O'Keefe. 1992. Trichohyalin: purification from porcine tongue epithelium and characterization of the native protein. *J. Invest. Dermatol.* 98:881-889.
- Hanahan, D., and M. Meselson. 1980. Plasmid screening at high colony density. *Gene (Amst.)*. 10:63-67.
- 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. Formation of the ϵ -(γ -glutamyl)lysine cross-link in hair proteins. Investigation of transamidases in hair follicles. *Biochemistry.* 11:2858-2863.
- 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.
- Haydock, P. V., and B. A. Dale. 1990. Filaggrin, an intermediate filament-associated protein: structural and functional implications from the sequence of a cDNA from rat. *DNA Cell Biol.* 9:251-261.
- Heid, H. W., I. Moll, and W. W. Franke. 1988. Patterns of expression of trichocytic and epithelial cytokeratins in mammalian tissues. I. Human and bovine hair follicles. *Differentiation.* 37:137-157.
- 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 antigen hair and hair follicles studied with anti-hair keratin monoclonal antibodies. *J. Invest. Dermatol.* 86:563-569.
- Jensen, R., D. R. Marshak, C. Anderson, T. J. Lukas, and D. M. Watterson. 1985. Characterization of human brain S100 protein fraction: amino acid sequence of S100 β . *J. Neurochem.* 45:700-705.
- Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15:8125-8148.
- Krieg, P. A., and D. A. Melton. 1987. *In vitro* RNA synthesis with SP6 RNA polymerase. *Methods Enzymol.* 155:397-415.
- Lagasse, E., and R. G. Clerc. 1988. Cloning and expression of two human genes encoding calcium-binding proteins that are regulated during myeloid differentiation. *Mol. Cell Biol.* 8:2402-2410.
- 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.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
- McKinley-Grant, L. J., W. W. Idler, I. A. Bernstein, D. A. D. Parry, L. Cannizzaro, C. M. Croce, K. Huebner, S. R. Lessin, and P. M. Steinert. 1989. Characterization of a cDNA clone encoding human filaggrin and localization of the gene to chromosome region 1q21. *Proc. Natl. Acad. Sci. USA.* 86:4848-4852.
- McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel. 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. *Cell.* 25:385-398.
- Mehrel, T., D. Hohl, J. A. Rothnagel, M. A. Longley, D. Bundman, C. Cheng, U. Lichti, M. E. Bisher, A. C. Steven, P. M. Steinert, S. H. Yuspa, and D. R. Roop. 1990. Identification of a major keratinocyte cell envelope protein, loricrin. *Cell.* 61:1103-1112.
- Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9:309-321.
- O'Guin, W. M., and M. Manabe. 1991. The role of trichohyalin in hair follicle differentiation and its expression in nonfollicular epithelia. *Ann. NY Acad. Sci.* 642:51-63.
- O'Guin, W. M., T.-T. Sun, and M. Manabe. 1992. Interaction of trichohyalin with intermediate filaments—three immunologically defined stages of trichohyalin maturation. *J. Invest. Dermatol.* 98:24-32.
- Powell, B. C., and G. E. Rogers. 1990. Cyclic hair-loss and regrowth in transgenic mice overexpressing an intermediate filament gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:1485-1493.
- Presland, R. B., P. V. Haydock, P. Fleckman, W. Nirunskisiri, and B. A. Dale. 1992. Characterisation of the human epidermal profilaggrin gene. Genomic organization and identification of an S-100-like calcium binding domain at the amino terminus. *J. Biol. Chem.* 267:23772-23781.
- Rice, R. H., and H. Green. 1979. Presence in human epidermal cells of a soluble protein precursor of the cross-linked envelope: activation of the cross-linking by calcium ions. *Cell.* 18:681-694.
- Richards, S., I. R. Scott, C. R. Harding, J. E. Liddell, G. M. Powell, and C. G. Curtis. 1988. Evidence for filaggrin as a component of the cell envelope of the newborn rat. *Biochem. J.* 253:153-160.
- 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., H. W. J. Harding, and I. J. Llewellyn-Smith. 1977. The origin of citrulline-containing proteins in the hair follicle and the chemical nature of trichohyalin, an intracellular precursor. *Biochim. Biophys. Acta.* 495:159-175.
- Rogers, G. E., M. J. Fietz, and A. Fratini. 1991. Trichohyalin and matrix proteins. *Ann. NY Acad. Sci.* 642:64-81.
- Rothnagel, J. A., and G. E. Rogers. 1986. Trichohyalin, an intermediate filament-associated protein of the hair follicle. *J. Cell Biol.* 102:1419-1429.
- Rothnagel, J. A., T. Mehrel, W. W. Idler, D. R. Roop, and P. M. Steinert. 1987. The gene for mouse epidermal filaggrin precursor: its partial characterization, expression, and sequence of a repeating filaggrin unit. *J. Biol. Chem.* 262:15643-15648.
- 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.
- Shono, S., and K. Toda. 1983. The structure proteins of the human nail. In *Normal and Abnormal Epidermal Differentiation.* M. Seiji and I. A. Bernstein, editors. University of Tokyo Press, Tokyo, Japan. 317-326.
- Simon, M., and H. Green. 1984. Participation of membrane-associated proteins in the formation of the cross-linked envelope of the keratinocyte. *Cell.* 36:827-834.
- Simon, M., M. Phillips, H. Green, H. Stroh, K. Glatt, G. Bruns, and S. A. Latt. 1989. Absence of a single repeat from the coding region of the human involucrin gene leading to RFLP. *Am. J. Hum. Genet.* 45:910-916.
- 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.
- Stark, H.-J., D. Breiterkreutz, A. Limat, C. M. Ryle, D. Roop, I. Leigh, and N. Fusenig. 1990. Keratins 1 and 10 or homologues as regular constituents of inner root sheath and cuticle cells in the human hair follicle. *Eur. J. Cell Biol.* 52:359-372.
- 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, Australia. 35-57.
- Szebenyi, D. M. E., S. K. Obendorf, and K. Moffat. 1981. Structure of vitamin D-dependent calcium-binding protein from bovine intestine. *Nature (Lond.)*. 294:327-332.
- Vanaman, T. C., F. Sharief, and D. M. Watterson. 1977. Structural homology between brain modulator protein and muscle TnC. In *Calcium-Binding Proteins and Calcium Function.* R. H. Wasserman, R. Corradine, E. Carafoli, R. H. Kretsinger, D. MacLennan, and F. Siegel, editors. North-Holland, New York. 107-116.
- Wood, L., D. Carter, M. Mills, N. Hatzenbuehler, and G. Vogeli. 1991. Expression of calyculin, a calcium-binding protein, in the keratogenous region of growing hair follicles. *J. Invest. Dermatol.* 96:383-387.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains—nucleotide sequences of the M13mp18 and pUC 19 vectors. *Gene (Amst.)*. 33:103-119.