



Molecular Analysis of the Ultra-High-Sulphur Keratin Proteins

Philip James MacKinnon B.Sc. (Hons.)

A thesis submitted for the degree of Doctor of Philosophy in the University of Adelaide

Department of Biochemistry

January, 1989

Contents

Summary

Declaration

Acknowledgements

Abbreviations

<u>Chapter 1 Introduction</u>	1
1.1 <u>General Introduction</u>	1
1.2 <u>Formation and Morphology of the Follicle</u>	1
1.3 <u>Fine Struncture of the Fibre</u>	2
.1 Medulla	3
.2 Cortex	3
.3 Cuticle	3
1.4 <u>Proteins of the Cortex and Cuticle</u>	4
.1 Low Sulphur Proteins	4
.2 High-Glycine/Tyrosine Proteins	5
.3 High-Sulphur Proteins	6
.4 Ultra-High-Sulphur Protein Sub-Family	7
.5 Cuticle Proteins	8
1.5 <u>Follicular Location of Keratin Synthesis</u>	8
1.6 <u>Molecular Biological Studies of the Follicle</u>	9
1.7 <u>Variation in Wool and Hair Growth and Composition</u>	11
.1 Nutritional Regulation of Wool Growth	11
.2 Genetic Regulation	15
.3 Physiological Regulation	17
1.8 <u>Aims of this Thesis</u>	17

Chapter 2 Materials and Methods

2.1 Materials

.1 Tissue	19
.2 Bacterial strains	19
.3 Bacteriophage Strains	19
.4 Plasmid Strains	19
.5 Enzymes	20
.6 Radiochemicals	20
.7 General Chemicals	20
.8 Media and Buffers	21
A Growth media	21
B General Buffers	21
.9 Miscellaneous	22

2.2 Methods

.1 Protein Methods	22
A Extraction and Alkylation of Wool High-Sulphur Proteins with 4-Vinyl Pyridine	22
B Reduction and S-carboxymethylation of Total Wool Protein and Isolation of Total High-Sulphur Protein	23
C Protein Estimation	23
D Cleavage with Endoproteinase Glu-C	24
E ¹⁴ C Labelling of Wool and Hair Keratin	24
F Two-dimensional PAGE	24
G SDS PAGE	25
H Gel Staining	25
I Fluorography	25
J Autoradiography	25
.2 DNA Methods	
A General Methods	26
B Preparation of labelled DNA	26

C	Screening Cosmid Library	27
D	Screening Lambda Library	27
E	DNA preparation	27
F	Cleavage by Restriction Endonucleases	28
G	Polyacrylamide Electrophoresis	28
H	Agarose electrophoresis	28
I	Isolation of DNA from Polyacrylamide Gels	29
J	Isolation of DNA from Agarose Gels	29
K	Subcloning into Plasmids and Phage	29
L	Bal31 deletions	30
M	Preparation of ssDNA for Sequencing	31
N	DNA sequencing	31
O	Southern Transfer	31
.3	<u>RNA Methods</u>	
A	RNA Preparation from Wool Follicles	32
B	Northern Transfer	33
C	RNA Protection	33
D	<i>In vitro</i> Transcription	33
E	<i>In vitro</i> Translation	34
F	Tissue <i>In Situ</i> Hybridization	34
.4	<u>Computer Programmes</u>	
A	General	35
B	Predict	35
.5	<u>Containment Facilities</u>	36

Chapter 3 Wool UHS Protein Purification

3.1	<u>Introduction</u>	37
3.2	<u>Results</u>	
.1	High-Sulphur Protein Derived using 4-Vinylpyridine	38
.2	Ion-Exchange-Chromatography	38

.3 Peptide isolation	40
.4 Peptide Sequence Analysis	40
3.3 <u>Discussion</u>	40
3.4 <u>Summary</u>	42

Chapter 4 Characterization of a Human UHS Gene

4.1 <u>Introduction</u>	43
4.2 <u>Results</u>	
.1 Cosmid Library Screening	43
.2 Sequence of the 1.0 kb EcoRI Fragment	44
.3 Gene Product Analysis	44
.4 Southern Transfer Analysis	45
.5 Northern Transfer Analysis	46
.6 Chromosomal <i>In Situ</i> Hybridization	47
4.3 <u>Discussion</u>	
.1 5' flanking and noncoding regions	47
.2 3' Noncoding Region	48
.3 Lack of Introns	49
.4 Conceptual Protein Analysis	49
.5 Amino Acid Repeats in the Conceptual Protein	50
.6 <i>In Vitro</i> Translation	52
.7 Southern Transfer Experiments	54
.8 Northern Transfer Experiments	54
4.4 <u>Summary</u>	55

Chapter 5 Characterization of a Sheep UHS Gene

5.1 <u>Introduction</u>	56
5.2 <u>Results</u>	
.1 Lambda Library Screening	56
.2 Sequence of the Sheep UHS Gene	57

.3 <i>In vitro</i> Expression of the Sheep UHS Gene	57
.4 Genomic Southern Experiments	58
.5 Northern Transfer Experiments	59
5.3 <u>Discussion</u>	
.1 5' Flanking/Noncoding Region	59
.2 3' Untranslated/Flanking Region	61
.3 Conceptual Protein Analysis	61
.4 <i>In Vitro</i> Translation Experiments	63
.5 Genomic Southern Transfer Experiments	64
.6 Northern Transfer Experiments	65
5.4 <u>Summary</u>	66

Chapter 6 Comparison of UHS Gene Sequences

6.1 <u>Introduction</u>	67
6.2 <u>5' Noncoding/Flanking Regions</u>	67
6.3 <u>3'Flanking/Noncoding Regions</u>	68
6.4 <u>Coding Regions</u>	68
.1 Conceptual Protein Size and Amino Acid Content	68
.2 Codon Usage	70
6.5 <u>Comparison of Coding Region Sequences</u>	70
.1 Comparison of Cysteine-Rich Repeats	70
.2 Glycine-Rich and Serine-Rich Repeats	72
.3 An Alignment of the Human and Sheep UHS gene coding regions	72
.4 K4 Coding Region Structure	73
.5 Mouse UHS gene	74
.6 Summary of Coding Region Sequence Comparisons	74
6.6 <u>Repeats and Protein Function</u>	75
6.7 <u>Protein Structure</u>	76
6.8 <u>Summary</u>	78

Chapter 7 UHS Gene Expression

<u>7.1 Introduction</u>	80
<u>7.2 Results</u>	
.1 Sheep Infusion	80
.2 Northern Transfer Analysis	82
.3 RNA Protection Experiments	83
.4 Follicle Tissue <i>in situ</i> Hybridization	83
<u>7.3 Discussion</u>	
.1 Sheep Infusion Experiment	85
.2 Northern Transfer and RNA Protection Analysis	86
.3 Tissue <i>in situ</i> Hybridization	87
<u>7.4 Summary</u>	89

Chapter 8 Final Discussion 91

Bibliography

Appendix A; Human UHS Gene Sequencing Strategy

Appendix B; Sheep UHS Gene Sequencing Strategy

Thesis Summary

This thesis undertakes an analysis at both the protein and gene level of a sub-group of the high-sulphur proteins present in wool and hair. The high-sulphur proteins are cysteine-rich, with the major proportion of these proteins in wool containing an average of 20 mole% of this amino acid. The sub-class of this family described in this thesis, known as the ultra-high-sulphur (UHS) proteins, contain in excess of 30 mole% of cysteine.

In sheep it has been demonstrated that when the supply of cysteine is increased to the follicle, there is a dramatic increase in the rate of wool growth, and commensurate with this is an increase in the relative abundance of the UHS proteins. The relative proportions of the other major protein components of the wool fibre, including the high-sulphur proteins of lower cysteine content, do not alter significantly under these conditions.

Whilst there is a considerable body of knowledge in the literature concerning the other wool keratin proteins and the genes which code for them, the only information available for the UHS proteins is derived from partially purified fractions, and no sequence data has been published.

Initially, the thesis describes an attempt to purify a UHS protein component and to obtain partial sequence. This was successful to the extent that some peptides were isolated from UHS protein and partially sequenced.

Next, the thesis describes the isolation, sequencing, and characterization of two UHS protein genes, one from the human genome and one from the sheep genome. The conceptual proteins encoded by these genes were both found to contain repeated amino acid sequences. Comparison of the coding regions of the two genes to each other revealed that most of the repeated sequences of one gene are related at both protein and nucleic acid levels to those of the other gene. The repeated sequences present in the UHS protein gene coding regions were found to be related to repeated sequences previously observed in the B2 and BIIIa high-sulphur proteins.

The products of both genes were obtained by *in vitro* translation and their behaviour when fractionated by two-dimensional polyacrylamide gel electrophoresis was correlated to that of authentic UHS protein.

Subclones from both genes were used as probes for Southern transfer analysis of sheep and human genomic DNA and it was found that there are a number of UHS protein genes in both genomes.

The same probes were used for northern transfer analysis of wool follicle RNA. These experiments indicated that the sheep gene is expressed in the wool follicle, and that the human gene cross-hybridizes at high stringency to mRNA species present in the wool follicle.

Experiments are described in which both the sheep and human UHS protein genes were used as probes for *in situ* hybridization to wool and hair follicle tissue. Probes from these genes were found to hybridize to mRNA present in the nascent hair and wool fibres, with strongest hybridization nearest to the fibre cuticle.

Finally, the thesis describes an experiment in which both UHS protein genes were used for quantitative analysis of UHS protein mRNA isolated from sheep before and after cysteine infusion *per abomasum*. These experiments indicated that there is no change in the UHS protein mRNA when sheep are infused with sufficient cysteine to elevate the expression of these proteins. This finding led to the consideration of several possibilities by which UHS protein synthesis might be controlled in the wool follicle.

DECLARATION

This thesis contains no material which has been accepted for the award of a degree or diploma in any University and in my belief, contains no material which has been published by another person, except where due reference is given.

Philip James MacKinnon

ACKNOWLEDGEMENTS

I wish to thank Professor W.H. Elliott for the opportunity to undertake these studies in this department. I also wish to thank Professor G.E. Rogers for his patient supervision of my work. I am indebted to Dr. Barry Powell for all his advice, support and encouragement during the course of this work, and also for his critical reading of this thesis. I extend special thanks to Dr. Richard D'Andrea for his critical reading of this thesis.

I would also like to express my appreciation to the following people:

Lesley Crocker for her excellent preparation of the photographic diagrams for this thesis and her technical assistance during the course of this work.

Dr. A.V. Sivraprisad for his advice on the computing conducted during this study.

Dr. Phil Hynd and co-workers of the Waite Institute for their aid with the animal work.

Michael Calder for his technical assistance during the sheep infusions.

To the members of Professor G.E. Rogers research group, past and present, who all contributed to the stimulating and interesting atmosphere of this group.

Finally I would like to thank my wife, Ruth, for all her patience and support during my Ph.D..

During the course of this work I was supported by a Postgraduate Award from the Australian Wool Corporation.

Abbreviations

bp:	base pair
BSA:	bovine serum albumin
cDNA:	DNA complementary to mRNA
Ci:	curie
cRNA:	RNA complementary to DNA (synthesized <i>in vitro</i>)
DNA:	deoxyribonucleic acid
dNTP:	deoxynucleoside triphosphate
DTT:	dithiothreitol
HGT:	high-glycine/tyrosine
HS:	high-sulphur
IF:	intermediate-filament
IFAP:	intermediate-filament-associated protein
kb:	kilobase pair
PAGE:	polyacrylamide gel electrophoresis
pfu:	plaque forming unit
poly(A):	polyadenylic acid
RNA:	ribonucleic acid
mRNA:	messenger RNA
rRNA:	ribosomal RNA
tRNA:	transfer RNA
rpm:	revolutions per minute
S:	Svedberg coefficient of sedimentation
ssDNA:	single-stranded DNA
tris:	2-amino-2(hydroymethyl)-1,3-propandiol
UHS:	ultra-high-sulphur
UV:	ultraviolet

Chapter 1

Introduction

1.1 General Introduction

Over the past thirty years or more a vast amount of knowledge has been gleaned upon the structure and function of the hair follicle (Orwin, 1979; Gillespie, 1983; Powell and Rogers, 1986, for reviews). However, the process by which the undifferentiated cells of the follicle bulb form the mature wool or hair fibre and its associated structures is still poorly understood. This process involves a series of coordinated activities including controlled cellular division, movement, and shape changes. At the molecular level hair follicle keratinization involves the synthesis and arrangement of a plethora of different proteins (Orwin, 1979).

The histology (Mercer, 1961; Rogers, 1964; Orwin, 1979; Fraser *et al.*, 1980) and protein chemistry (Crewther, 1976; Gillespie, 1983) of the follicle and fibre have been studied extensively and a large body of literature exists, especially on the wool follicle. More recently the isolation and characterization of the genes coding for various keratin components have been initiated, and have resulted in the isolation of representative genes from most of the predominant classes of wool protein (Powell and Rogers, 1986; Powell *et al.*, 1988; Rogers *et al.*, 1988).

To fully understand the process of keratinization, it will be necessary to expand our knowledge, especially at a molecular level, to include: (a) more protein and gene sequence data, particularly from the protein families for which no sequence data has yet been obtained; (b) the organization of the keratin genes in the genome; (c) the location of keratin gene transcription and translation and how these two processes are regulated; and (d) the process by which the nascent proteins interact to form the fully keratinized structure.

This thesis describes research into both the protein chemical and molecular biological aspects of wool fibre development and a review of literature relevant to the aims of this thesis follows.

1.2 Formation and Morphology of the Follicle

Hair and wool follicles are epidermal invaginations into the dermis which form embryogenically as downgrowths of epidermal cells from the basal layer (Mercer, 1961).

The events leading to the formation and fine structure of hair and wool follicles have been reviewed extensively elsewhere (Mercer, 1961; Rogers, 1964; Orwin, 1979; Fraser *et al.*, 1980; Holbrook *et al.*, 1988). The structures of the follicle are shown in figure 1.1.

The follicle has been divided into a number of zones to facilitate observation of the stages of differentiation (Orwin, 1976, 1979; Orwin and Thomson, 1972). The stages, as described in these references, are summarized below and depicted in figure 1.1.

Zone A:

Undifferentiated stem cells undergo mitosis and move upwards. This zone covers the area from the base of the follicle to the top of the papilla. There is no well-defined germinative layer and mitoses may be observed throughout this region (Epstein and Maibach, 1969).

Zone B:

In this region cells start to differentiate to form the various components of the fibre and associated structures. Differentiation is observed as changes in cell shape along with the appearance of large numbers of ribosomes, polysomes and mitochondria. This zone finishes when Henle's layer of the inner root sheath hardens.

Zones C and D:

These two zones comprise the keratogenous region in which the cells are fully differentiated. The cortical cells undergo major protein synthesis and asymmetry is first observed in these cells with the appearance of the ortho- and paracortex.

Zone E:

Keratinization is completed here. The organelles and nuclei are degraded and finally the cells dehydrate. Beyond this region the cells are fully keratinized, being packed with protein, and are no longer metabolically active.

1.3 Fine Structure of the Fibre

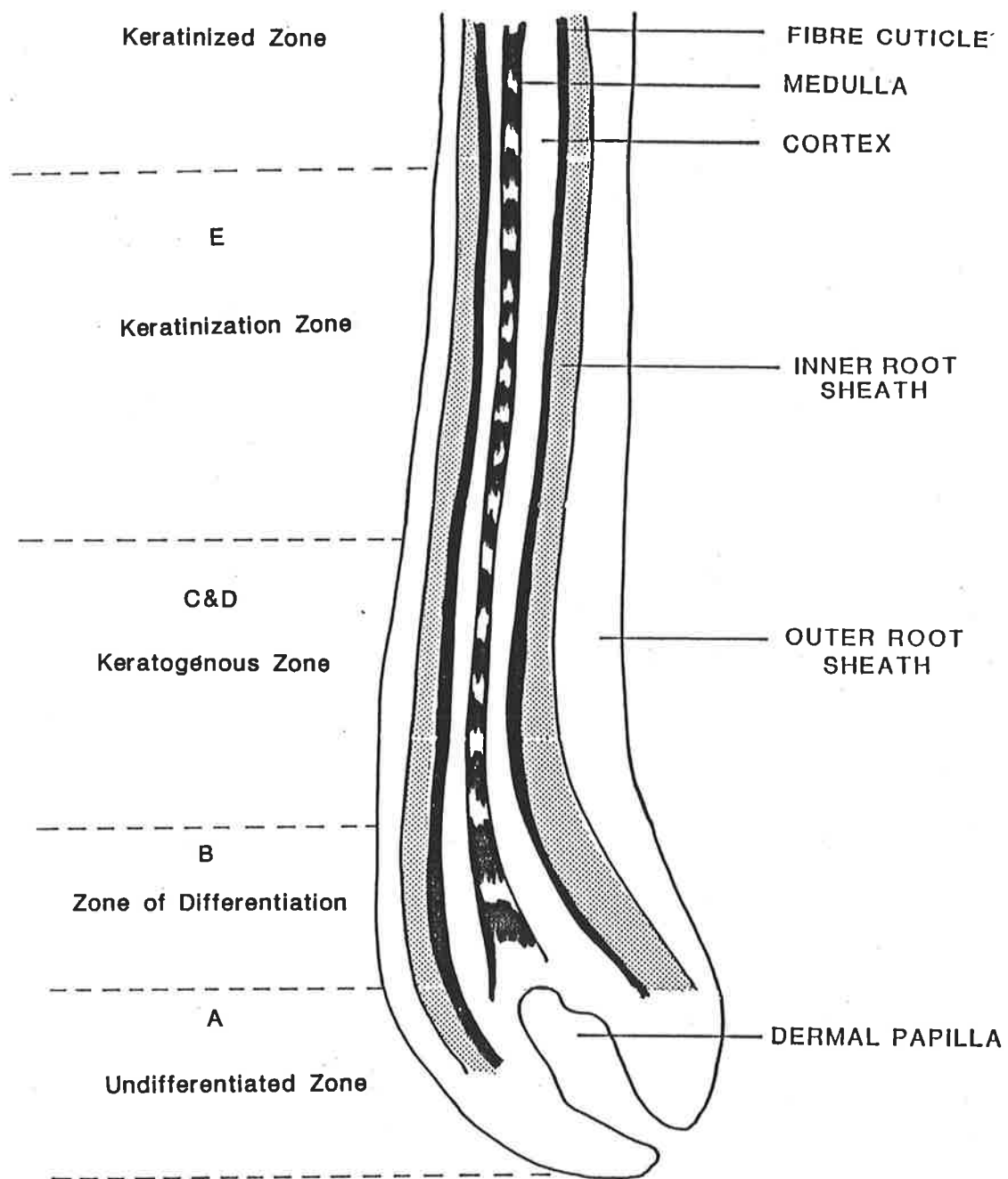
The fine structure and function of each of the structures of the wool and hair fibre have been described at length elsewhere and for the purposes of this thesis may be summarized as follows.

Figure 1.1

The Wool Follicle

A schematic diagram of a medullated wool follicle showing the major histological components, and the developmental zones.

Modified from Orwin (1976)



1.3.1 Medulla

The medulla is not always present and where present may be uni- or multicellular in width and continuous or discontinuous along the length of the fibre (Orwin, 1979). Where present the medulla is believed to improve thermal insulation and also reduces the mass of thick fibres (Fraser *et al.*, 1980).

The mature cells of the medulla are highly vacuolated (Orwin, 1979). The cells contain amorphous granules of trichohyalin, a protein which is also present in the inner root sheath (Rogers, 1964). This protein is rich in arginine which during differentiation is converted to citrulline (Rogers, 1963).

1.3.2 Cortex

The cortex accounts for most of the volume of wool and hair, reaching almost 90% in merino wool, and is responsible for the mechanical properties of the fibre such as strength, elasticity and crimp (Fraser *et al.*, 1980; Orwin, 1979).

The cortex is comprised of microfibrillar material oriented in the direction of fibre growth, which is embedded in a proteinacious matrix (Fraser, 1980).

During differentiation two types of cortical cell become obvious, the orthocortex where the microfibrillar aggregates known as macrofibrils remain discrete and are surrounded by matrix proteins, and the paracortex in which the macrofibrils fuse (Fraser *et al.*, 1980; Chapman and Gemmell, 1971 and 1973). The paracortex is believed to have a larger amount of the sulphur-rich matrix than the orthocortex (Chapman and Gemmell, 1973). In merino wool the cell types are grouped to form a bilateral configuration which is associated with crimp formation (Kaplin and Whiteley, 1978). Coarser fibres frequently lack this definite arrangement and the features of each cortex type are often less obvious (Fraser *et al.*, 1980).

1.3.3 The Cuticle

The cuticle is a layer of cells which forms a protective shell around the fibre. It is a single cell thickness in merino wool but may be much thicker in other species such as man where it is 8 to 10 cells thick (Swift, 1976). The cuticle contains aggregates of keratin

protein which form initially furthest from the central axis of the fibre during differentiation and these migrate to the periphery of each cell and eventually give rise to the exocuticle (Fraser *et al.*, 1980). The remainder of the cell lies towards the cortex and is termed the endocuticle (Orwin, 1979).

1.4 Proteins of the Cortex and Cuticle

Twenty-five years of intensive study of the components of the wool fibre have resulted in the isolation and sequencing of representative components of most of the main classes of cortical protein. Little is known of the proteins of the other histological components of wool and hair.

Most studies have involved isolation of wool protein components by reduction of the numerous disulphide linkages which bind the fibre together, followed by blockage of the thiol groups by carboxymethylation using iodoacetate (Crewther, 1976). This method gives a minimum of side reactions and the products are soluble over a wide pH range (Gillespie, 1983). The proteins of the medulla (Rogers, 1964), and to a lesser extent the cuticle (Ley and Crewther, 1980), are not extracted by this method, and proteolysis is required for the isolation of the former. The proteins studied by reduction and S-carboxymethylation must therefore derive predominantly from the cortex.

On the basis of amino acid content, particularly cysteine, and molecular weight, the wool keratin proteins have been categorized into three main classes: the low-sulphur proteins, the high-sulphur proteins and the high-glycine/tyrosine proteins. When fractionated using two-dimensional PAGE, the three major classes constituting the cortex and cuticle protein are readily visualized (fig 1.2). Each of the classes of wool proteins are divided into subfamilies of many closely related components, and it has been estimated that there are at least 100 proteins, of which less than a third have been sequenced (Powell and Rogers, 1986).

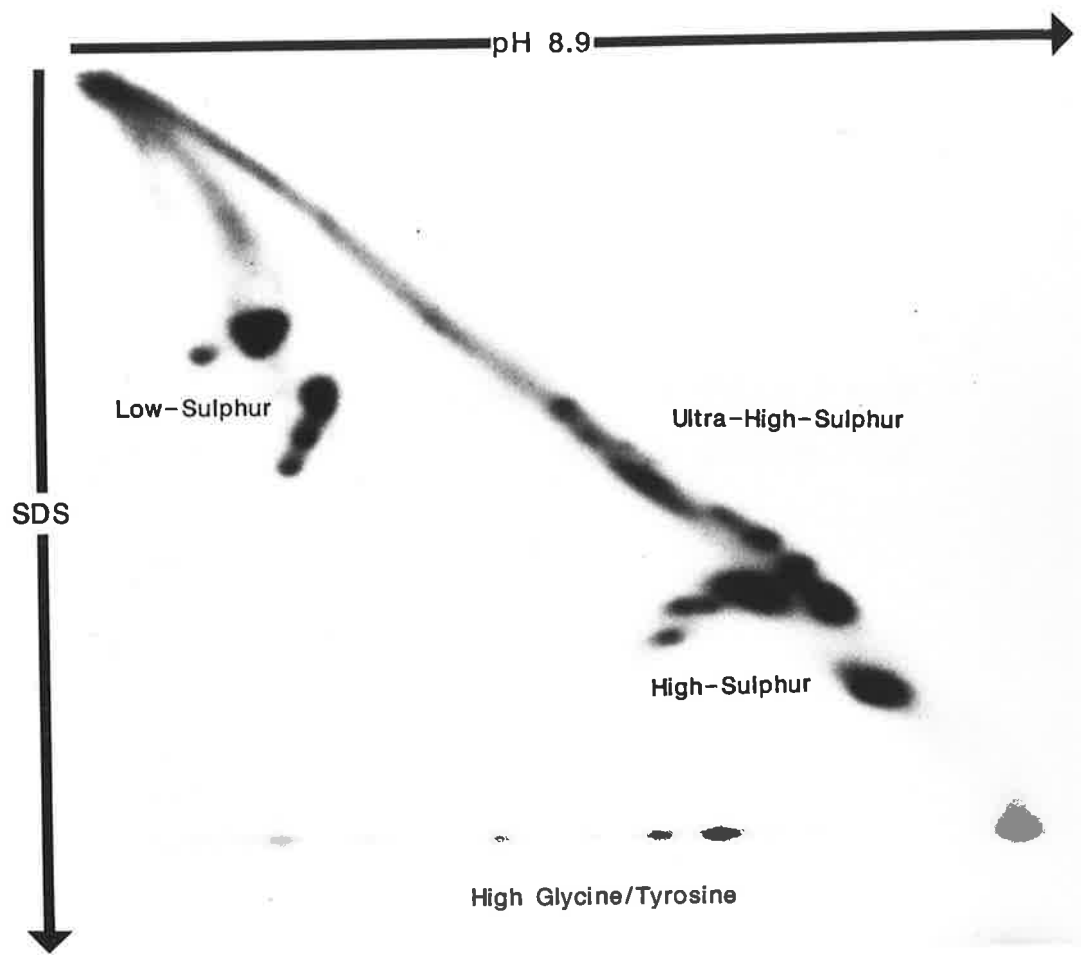
1.4.1 The Low-Sulphur Proteins

It has been shown conclusively that the low-sulphur proteins form the microfibrils (Jones, 1975, 1976). There are two sub-families, termed Type I (formerly component 8)

Figure 1.2

Two-Dimensional PAGE of Total Wool Protein

Two-dimensional polyacrylamide gel pattern of extracted and S-carboxymethylated wool keratin protein. The proteins were labelled with ^{14}C -iodoacetic acid (section 2.2.1 E) and fractionated in the first dimension at pH 9.5, and in the second dimension in the presence of SDS (section 2.2.1 F). This fractionation technique allows the visualization of all the major classes of wool keratin protein including the UHS proteins which do not resolve by conventional single dimensional electrophoresis.



and Type II (formerly components 5 and 7), each consisting of a number of proteins. The general characteristics of the two families are summarized in Table 1.1.

The low-sulphur proteins are a subclass of cytokeratins which in turn belong to the group of structural proteins known as intermediate filaments (IF) which are general cytoplasmic proteins. Like other intermediate filaments the low-sulphur proteins consist of both helical and nonhelical domains (Crewther *et al.*, 1983).

Physicochemical and cross-linking studies have demonstrated that dimers are formed consisting of one type I and one type II protein in a two-stranded coiled-coil arrangement and that two dimers pair to give rise to a tetramer (Ahmadi and Speakman, 1978; Woods and Gruen, 1981). Exactly how the dimers and tetramers are arranged in the microfibril is yet to be determined conclusively.

1.4.2 The High-Glycine/Tyrosine Proteins

The high-glycine/tyrosine (HGT) proteins are small (<10,000 MW) constituents of the intermicrofibrillar matrix, and are rich in the aromatic amino acids and glycine (Gillespie, 1983). They vary in abundance from almost zero in human hair and lincoln wool up to 30-40% in echidna quill (Gillespie and Frenkel, 1974). They are divided into two sub-families (type I and II) on the basis of amino acid content and solubility. Type I HGT proteins contain higher levels of phenylalanine and lower levels of cysteine than the type II subfamily (Marshall *et al.*, 1980).

The HGT protein family is apparently highly heterogeneous with at least 30 components being identified by various methods (Gillespie, 1983). It is thought that single amino acid substitutions are responsible for much of the heterogeneity, although it has been suggested that some may arise from either conformational changes related to the aromatic amino acid content (Powell and Rogers, 1986), or from nonspecific partial deamination (Marshall *et al.*, 1980).

The peptide sequence for three type I components has been determined, but as yet no type II sequence has been ascertained. Two of the sequences (C2 and C3) derive from the same type I subgroup and have identical sequences (Marshall *et al.*, 1980); the third sequence (F) (Dopheide, 1973) appears to be quite different.

Table 1.1

Summary of the Nomenclature and Characteristics of the Major Wool Keratin Protein Classes.

The data in this table are from Crewther (1976), Swart *et al.* (1976), and Gillespie, (1983).

- 1 The prefix SCMK indicates that the proteins are S-carboxymethylated derivatives of the protein. The numbers in parentheses represent estimates for the number of individual chains in each family.
- 2 The proteins of this group have not been characterized.
- 3 There are some protein-chemical data which suggest that the type I and type II high-glycine/tyrosine proteins should be considered as subclasses, each containing several families of proteins. However, the numbers of different polypeptide chains may be a conformational anomaly.

The table was taken from Powell and Rogers (1986).

Protein class and location in fibre	Characteristic amino acid content	Protein families within the class	Family notation ¹		Molecular weight
Low-Sulphur (microfibrils, i.e. intermediate filaments)	Cysteine (1-3 mol%)	3	SCMKA-5	(1)	56,000
			SCMKA-7	(3)	58,000
			SCMKA-8	(4)	38-43,000
High-Sulphur (matrix)	Cysteine (20-30 mol%)	4	SCMKB-1/IA	(?) ²	23-26,000
			SCMKB-2/TB	(7)	19,000
			SCMKB-III A	(11)	16,000
			SCMKB-III B	(4)	11,000
Ultra-High-Sulphur ² (matrix and cuticle)	Cysteine (>30 mol%)	?	-		>16,000
High-Gly/Tyr (matrix)	Glycine (20-40 mol%) Tyrosine (12-21 mol%)	2	Type I	(10) ³	6-9,000
			Type II	(5) ³	6-9,000

The relative amount of HGT protein in the hair or wool of an individual has been shown to vary under certain circumstances such as the administration of depilatory compounds (Frenkel *et al.*, 1975). This will be discussed in more detail in section 1.6.3.

1.4.3 The High-Sulphur Proteins

The high-sulphur (HS) proteins are cysteine-rich components of the matrix. They vary in abundance from 7% in horns to 50% in some hairs (Gillespie, 1983).

This family exhibits a remarkable degree of heterogeneity both in size and amino acid content, especially cysteine (Gillespie, 1983). In wool there are four subfamilies, known as SCMKBI, SCMKB2, SCMKBIIIA and SCMKBIIIB (Crewther, 1976), and the characteristics of these are summarized in table 1.1. The cysteine content of wool high-sulphur protein varies from 17% in the BIIIB group to 23% in the B2 fraction (Crewther, 1976). The cysteine content of some of the high-sulphur components of mouse hair is in excess of 30 mole%, and these are thought to be analogous to the ultra-high-sulphur proteins present in wool (see section 1.4.4)(Marshall and Gillespie, 1976a).

With the exception of SCMKBI and the ultra-high-sulphur proteins (see section 1.4.4), peptide sequences have been determined for all the high-sulphur protein subfamilies of wool (Crewther, 1976). Perhaps the most interesting observation is the presence of either a decapeptide (Elleman *et al.*, 1973) or a pentapeptide (Swart, 1973) repeat in the sequence of all but the SCMKBIIIB high-sulphur subfamilies. The pentapeptide repeat takes the form Cys-Cys-(Arg/Gln)-Pro-(Ser/Thr/Val). This repeat occupies a third to a quarter of the peptide chains of the BIIIA and B2 components which have been sequenced (Powell and Rogers, 1986). Elleman *et al.* (1973) suggest that the high-sulphur proteins evolved through repeated duplication of a primordial gene coding for a shorter sulphur-rich protein.

The amino acid repeat observed in the BIIIA and B2 components is likely to be structural in function with theoretical evidence indicating that the high-sulphur proteins have a highly convoluted structure stabilized by disulphide linkages between the cysteine residues of the repeat (Parry *et al.*, 1979). Powell and Rogers (1986) suggest that the

cysteine residues in the repeat may form disulphide cross-links to the cysteine-rich domains of the microfibrillar low-sulphur proteins.

The decapeptide repeat may also have a function in protease inhibition since this sequence resembles known inhibitors, and also because of the observation that cortex cells isolated using formic acid inhibit protease action (Crewther, 1976).

1.4.4 The Ultra-High-Sulphur Protein Sub-family

The ultra-high-sulphur (UHS) proteins are an ill-defined subfamily of the high-sulphur group. Their distinguishing characteristics are their extreme cysteine content which is in excess of 30 mole% (Gillespie and Broad, 1972), and the dramatic increase in their abundance observed in wool when sheep are infused *per abomasum* with cysteine or methionine (Gillespie and Reis, 1966) (section 1.6.1). The UHS proteins localize as an unresolved diagonal smear when fractionated using two-dimensional PAGE (Gillespie *et al.*, 1982)(figure 1.2). Other than this little is known of these proteins. Studies using protein chemical techniques have yielded little information because it has been difficult to obtain adequate amounts of these proteins from wool, and because there are small differences in the properties of the UHS proteins and the normal high-sulphur proteins (Marshall and Gillespie, 1976a). Lanthionine, a nonreducible cross-link formed between two cysteine residues, is known to form under the conditions used for reductive isolation of wool proteins (Thompson and O'Donnell, 1962) and it is possible that this may contribute to the difficulty experienced in purifying components of the UHS protein subfamily.

The histological location of the UHS proteins is uncertain but it is likely they are predominantly within the matrix of cortical cells and also in the cuticle (see section 1.4.5). Rogers (1964), using electron microscopy, observed that sulphur-enriched wool containing elevated levels of the ultra-high-sulphur proteins had greatly increased amounts of matrix protein that filled large regions of the cortex in which there was no microfibrillar material.

There is evidence for the presence of these proteins in the hair of a diverse range of species including four species from the order Artiodactyla (Gillespie and Broad, 1972), mice (Gillespie and Inglis, 1965; Marshall and Gillespie, 1976 a, b), and man (for example see Gillespie, 1983).

1.4.5 The Cuticle Proteins

A body of evidence exists which suggests that the cuticle has a higher content of sulphur than the rest of the fibre (Fraser *et al.*, 1980). Fractionation of the cuticle from human hair into its components indicates that the A layer and exocuticle contain high levels of cysteine (Swift, 1976). Two-dimensional PAGE has revealed that protein extracted from the cuticle migrated on a diagonal in a similar position to the UHS proteins (Ley and Crewther, 1980). More recently Marshall and Ley (1986), using two-dimensional PAGE, have demonstrated that the high-sulphur proteins of the cuticle are different to the high-sulphur proteins of the cortex.

1.5 Follicular Location of Keratin Synthesis

The determination of the location of keratin protein synthesis in the follicle is important because it will help in understanding the way in which the various proteins interact to form the fibre structure. It may also help in the elucidation of the regulatory mechanisms of keratin gene expression.

The early radioisotope work of Downes *et al.* (1963) and the electron microscopy of Mercer (1961) and Rogers (1964) indicated that there is a two stage synthesis of the cortical proteins, with the low-sulphur proteins forming in the proximal regions of the follicle (zone A) followed by the high-sulphur proteins in the distal regions (zones C and D). Subsequent work using radiolabelled methionine to specifically locate low-sulphur protein synthesis (Wilson *et al.*, 1971), and immunological techniques (Lynch *et al.*, 1986), lend weight to this hypothesis.

However, Fraser (1969, a, b) examined keratin protein at three different levels in the follicle and concluded that the two groups of proteins are synthesized concurrently. He observed that low-sulphur protein synthesis occurs linearly throughout the prekeratin regions of the follicle, whilst the high-sulphur proteins are synthesized exponentially with the greatest synthesis occurring in the more distal regions (zone D).

By observation of differences between the ortho- and paracortexes, Chapman and Gemmell (1971, 1973) and Chapman (1976) have refined the two stage keratin synthesis

hypothesis. They suggest that synthesis occurs sequentially in the paracortex, with high-sulphur protein production lagging behind that of the low-sulphur protein, but increasing sharply in the higher regions of the follicle (zone C) where synthesis occurs concurrently. In the orthocortex synthesis of both groups of proteins occurs concurrently.

Little is known of the position of synthesis of the HGT proteins or the ultra-high-sulphur proteins. Future studies of the synthesis and distribution of the various keratin components will be facilitated by the use of mRNA *in situ* hybridization using gene probes, perhaps in conjunction with the immunological techniques used by Lynch *et al.* (1986).

1.6 Molecular Biological Studies of the Follicle

Keratin synthesis was initially studied using homologous follicle cell-free translation systems from a range of species including guinea pig, mouse and sheep (Steinert and Rogers, 1971; Wilkinson, 1971; Steinert and Rogers, 1973; Ward and Kasmarik, 1980; Bertolino *et al.*, 1982). These studies demonstrated that keratin synthesis was ribosome dependant and the protein products synthesized *in vitro* had similar electrophoretic and immunological behaviour to authentic keratin proteins. Thus, fibre keratin synthesis must occur by conventional mRNA translation, and no extensive modification of the peptide chains occurs *in vivo* after synthesis.

With the exception of the ultra-high-sulphur proteins nucleic acid sequence data is available for all the major classes of wool fibre keratin. These sequences were initially obtained from follicle cDNA libraries (Ward and Sleight, unpublished results; Kuczek and Rogers, 1985) and subsequently the corresponding genomic sequences have been obtained. Complete gene sequences are known for one intermediate filament type I component (Ward *et al.*, 1988), two type II components (Powell *et al.*, in preparation), proteins from three of the four high-sulphur families (B2, BIIIA and BIIIB) (Powell *et al.*, 1983; Frenkel, 1988; Powell, unpublished results) and for HGT type I components C2 and F (Kuczek and Rogers, 1987).

The low-sulphur type I and II genes are arranged in separate clusters which are apparently unlinked (Powell *et al.*, 1986). Sequence data reveal that both types contain introns.

The genes encoding the high-sulphur and HGT protein families are unusual in that they do not have introns. It is possible that this is a general feature of matrix protein genes. Interestingly, a lack of introns may also be a feature of the intermediate-associated protein genes of the inner root sheath (R. B. Presland and M.J. Fietz, unpublished results) and also the epidermis (Dale *et al.*, 1988).

The genes which encode the SCMKB2A and SCMKB2D high-sulphur proteins are separated by only 1.9 kb (Powell *et al.*, 1983), hence it appears that at least some of the high-sulphur protein genes are clustered. Sequence comparison between the three SCMKB2 genes that have been sequenced (B2A, B2C and B2D) reveals that there is extensive sequence conservation within the coding region, which includes a low number of conservative base substitutions (Powell *et al.*, 1983). Powell *et al.* (1983) suggest that this reflects strong codon selective pressure and/or frequent gene conversion during the evolution of these genes. The 3' untranslated regions of the three genes are divergent with little sequence homology and differing lengths, indicating that duplication of these genes must have occurred at a distant evolutionary time (Powell *et al.*, 1983, Powell and Rogers, 1986).

With the exception of the BIIIA gene, extensive sequence homology is observed in the 5' untranslated regions of the high-sulphur genes (Powell *et al.*, 1983). This is particularly strong in the 18 residues contiguous with the methionine start codon (figure 1.3) and extends to both the HGT genes that have been sequenced (Kuczek and Rogers, 1987). This sequence is not found in other sheep genes including the wool intermediate-filament genes (Powell and Rogers, 1986; Powell *et al.*, in preparation) and it may be specific to wool matrix genes. Powell *et al.* (1983) postulate that this region has a role in the regulation of translation of the wool matrix genes.

The 5' flanking sequence of the HGT-F gene contains a region which may form Z-DNA and also a sequence which is similar to the simian virus core enhancer element

Figure 1.3

Comparison of the 5' Untranslated Regions of Six Sheep HS and Two HGT Genes

Comparison of the 18 bp contiguous with the methionine initiation codon in the 5' non-coding regions of the wool intermediate-filament-associated protein (IFAP) genes. The B2A gene is the model sequence against which the remaining genes are compared. Deletions and insertions were introduced to maximize homology. A common nucleotide is indicated by a star (*), a deletion by a dash (-), and insertions as a superscript. (Diagram was adapted from Kuczek and Rogers, 1987: data from Powell *et al.*, 1983; Kuczek and Rogers, 1985, 1987; Frenkel *et al.*, 1988; Ward *et al.*, 1988; Powell *et al.*, in preparation)

	-10	-1
<u>B2A:</u>	AACCCAACTCCTGACACCATG	
<u>B2C:</u>	*G**T***C*****ATG	
<u>B2D:</u>	*G*****ATG	
<u>BIIIA:</u>	TCGG*CC*A*ACAC****ATG	
<u>BIIIB"2":</u>	*****G*****ATG	
<u>BIIIB"4":</u>	***A**G ^T ***CA*****ATG	
<u>HGT-F:</u>	CG*****-**A*****ATG	
<u>HGT-C2:</u>	TT***-*****C***AATG	

(Kuczek and Rogers, 1987). These elements may function in the regulation of transcription of this gene.

Both the HGT-F and HGT-C2 genes appear to be unique, hence much of the apparent heterogeneity of the HGT proteins observed from protein-chemical analysis must be artefactual (Kuczek and Rogers, 1987).

Although no definite sequence data is available on the ultra-high-sulphur proteins, a cDNA clone has been isolated which contains an incomplete coding region of 75 codons, of which 37% code for cysteine (K. Ward, unpublished results)(figure 1.4). This is above the minimum observed for ultra-high-sulphur protein fractions (Gillespie and Broad, 1972), hence this clone may contain UHS protein coding sequences. The coding region of this clone contains four contiguous repeats of 10 codons which are strikingly similar, both at the peptide and nucleotide sequence levels (figure 1.4), to the repeat observed in the high-sulphur proteins (see section 1.4.3). The remainder of the coding sequence, located towards the 5' end, doesn't appear to be repetitive but is still rich in cysteine codons.

1.7 Variation in Wool and Hair Growth and Composition

Wool and hair are tissues which vary greatly in both growth rate and composition (Marshall and Gillespie, 1988). In general, compositional variation, both quantitative and qualitative, focuses on both matrix protein families: the high-sulphur proteins, particularly the UHS proteins; and the HGT proteins. The intermediate filament proteins appear to remain relatively constant both in type and amount (Gillespie and Reis, 1966; Frenkel *et al.*, 1974, 1975; Gillespie and Marshall, 1980; Marshall, 1981b; Gillespie *et al.*, 1982; Marshall and Gillespie, 1988). A number of factors which influence synthesis and constitution have been identified and are discussed below.

1.7.1 Nutritional Regulation of Wool Growth

Reis and Schinckel (1963, 1964) demonstrated that infusion *per abomasum*, of sheep consuming a diet of roughage, with cysteine or methionine resulted in a dramatic increase in the rate of wool growth. Similar observations have since been made after administration of these amino acids parentally or post-ruminally (Reis, 1979 for review).

Figure 1.4

Sequence of the Sheep Presumptive UHS Encoding cDNA Clone, K4

The partial nucleotide sequence of the sheep putative UHS cDNA clone, K4 (Dr. K. Ward, unpublished results). The conceptual protein sequence (containing 37 mole% cysteine) in the one-letter code is shown above the nucleotide sequence. The decapeptide repeat observed in this protein has been outlined.

Below is an alignment of the consensus amino acid and nucleotide sequences of the decapeptide repeats of the K4 and B2 family (Powell *et al.*, 1983). A star (*) indicates a nucleotide match between the two nucleotide sequences.

This figure was adapted from Powell and Rogers (1986).

C C V T V C C C V P A C S C S S C G K G G C S W C G C S Q S S C C R P C C S Q S
CTGCTGTGTGACCGTGTGCTGTGTGTGCCAGCCTGCTCCTGCTCCAGCTGTGGCAAAGGGGGCTGCAGCTGGTGTGGTGTCTCCAGTCCAGCTGCTGCAGACCCTGCTGCTCCAGT 120

S C C R P C C S Q S S C C R P C C S Q S S C C V P A C C Q R R D L R C *
CCAGCTGCTGCAGACCCTGCTGCTCCAGTCCAGCTGCTGCAGACCCTGCTGCTCCAGTCCAGCTGCTGCGTCCCTGCTTGTCTCAGCGCAGAGATCTGAGATGCTGAGCTCAGACCT 240

CAGGCGGTTCCAGCCAGTCTCAGCTCCCTGACTGTGAGCAGTGACACGTCCAGACCCCCAGGGCACACTCCATTCTCTCCAAGGTAGTGAAGTCTGAATCTTCGGGACCAGAGATCCTGT 360

TCCACACTTCTGATGGAGAACCCCTTCCCCTTGTACTGCCTGAGCCATCTGAGGATTGACTGCCACTTTCAATCCAAGAGAAGTTCTTCTTGGCAGCTGAACTCCACTGCAGACGC 480

CTTAGGGGCTAAACCACAGAAACCACTCAACCTCTGATGGGCACAGAGCTGTTCCAACCATTTCTTGGCTTCTTGCTGCCTAGCCTTGTTCTCACATAACCTCAGGGCTGGTGTGACC 600

TCTCGGGCTGGGCCCCAAATAATCCTCCCCTCTTCTGTGACTCCACCTCCTTCTGATCTCGGTCTAGCAAATTCTACTCTGGGCTTGTGAAGCCGCTAAATAAATAGACTGTCTTCAAC 720

K4

Q S S C C R P C C S
CAGTCCAGCTGCTGCAGACCCTGCTGCTCC
*** ***** ** **** *
CAGACCAGCTGCTGCCAGCCAACCTGCATC
Q T S C C Q P T C I

B2

Smaller changes are observed when sheep are shifted from low to high nutritional levels (Reis, 1965; Gillespie and Reis, 1966). Oral administration of sulphur-containing amino acids to sheep generally has no effect on wool growth since the bacterial flora in the rumen catabolize these amino acids thereby preventing their absorption and passage to the follicle (Reis, 1979).

It was found that concomitant with the increased growth rate there is an increase in the sulphur content of the wool (Reis and Schinckel, 1963, 1964), and subsequently this has been demonstrated to result from increased synthesis of the ultra-high-sulphur proteins (Gillespie *et al.*, 1964). The remaining components, including the high-sulphur proteins, appear to remain constant (Gillespie *et al.*, 1964; Gillespie and Reis, 1966; Gillespie *et al.*, 1969).

As the follicle proteins are rich in cysteine, it is believed that the sulphur-containing amino acids are limiting for wool growth. The increased supply of these amino acids allows synthesis of the ultra-high-sulphur proteins and increased rates of wool growth (Gillespie and Reis, 1966; Reis, 1979; Gillespie, 1983; Reis, 1988).

The role of methionine in wool growth is unclear. Parentally administered tracers of [^{35}S]methionine are efficiently incorporated into wool protein (Downes *et al.*, 1964, 1970). Methionine is not a major constituent of wool protein, but it may be converted to cysteine by the transulphuration pathway (Reis, 1979, 1988). However this is unlikely to be the sole role of methionine, as differences in the response of the wool follicle to methionine and cysteine have been observed. It was found that when sheep are infused *per abomasum* with mixtures of the essential amino acids, omission of methionine failed to stimulate wool growth and that replacement with cysteine had no effect (Reis, 1982). Reis (1988) suggests that the role for methionine in the control of wool growth, which cysteine could not fulfill, is either to act as a methyl donor or as a precursor for polyamines via S-adenosylmethionine.

Reis (1988) presented experimental evidence which indicates that treatments which result in inhibition of polyamine synthesis also reversibly reduce wool growth. Also, he found that the methionine analogue, ethionine, which is a substrate of the enzyme methionine adenosyl-transferase, caused a cessation of wool growth, which could not be

alleviated by infusion with cysteine (Reis and Tunks, 1982). Both these observations, Reis (1988) suggests, implicate methionine as being necessary for polyamine synthesis if wool growth is to occur.

Hynd (1986, 1988) has shown through the use of colchicine as a block of mitosis that one of the effects of an increased supply of the sulphur-containing amino acids is to increase the rate of cell division in the follicle bulb. Hynd (1988) also found that methionine, cysteine and a mixture of the two were equally effective in stimulating mitosis, indicating that if methionine has a special role as Reis suggests, it must be met at a low level, or it must require a balance of the other essential amino acids, such as those used by Reis, to be apparent.

When large doses of methionine or cysteine (8-10 g/day) were administered to sheep the rate of wool growth was not stimulated as much as smaller doses (2 g/day), and in the case of methionine may even have depressed wool growth (Reis, 1967). Reis (1967) suggests that this effect may have been due to amino acid imbalance or toxicity. Interestingly, it was found that regardless of whether wool growth was increased or not, infusion with cysteine or methionine resulted in elevated synthesis of the ultra-high-sulphur proteins (Reis, 1967). This further indicates that methionine has a multiple effect on fibre development.

Since monogastric animals such as man lack a ruminal microflora the supply of the sulphur-containing amino acids to the hair follicle isn't limiting as it is in ruminants (Gillespie, 1983). Human hair contains high levels of the ultra-high-sulphur proteins under normal dietary conditions (Gillespie, 1983). Evidence exists which indicates that the reverse of sulphur enrichment may occur in malnourished individuals. Children suffering from kwashiorkor produce hair which is deficient in ultra-high-sulphur protein (Gillespie, 1983). The composition of the hair of such individuals returned to normal when they were fed an adequate diet (Gillespie, 1968, 1983). It would appear that a deficiency of the sulphur-containing amino acids, resulting from malnutrition, causes a relative reduction in the synthesis of the ultra-high-sulphur proteins of human hair, a situation which mimics the normal condition of the wool fibre (Gillespie, 1983).

Marshall and Gillespie (1988) suggest two mechanisms for sulphur enrichment of wool:

1. Synthesis of the ultra-high-sulphur proteins is dependent on the supply of cysteine to that region of the follicle in which they are synthesized. At low levels of cysteine intake or during high rates of protein synthesis only small amounts of cysteine reach the upper regions of the follicle where it is believed the ultra-high-sulphur proteins are synthesized. Under conditions where cysteine supply to the follicle is greatly increased the concentration of cysteine in the ultra-high-sulphur protein synthesis region rises, allowing enhanced expression of these proteins.

2. Cysteine regulates the differentiation of the stem cells, and at higher levels causes more cells to form the paracortex. The paracortex is believed to contain higher levels of high-sulphur protein than the orthocortex (see section 1.3.2).

Hynd (1988) has shown that there is a decrease in the proportion of orthocortical cells when sheep are shifted from low to high nutritional levels, suggesting that the latter mechanism may in part explain sulphur enrichment. However, the observation of Rogers (1964; personal communication) that there is an apparent large increase in the amount of paracortical matrix in sulphur-enriched wool indicates that there are intracellular changes, as well as changes in the proportions of cortical cell types. The first mechanism should result in larger cortical cells as a result of the presence of additional UHS protein. This is consistent with the observations of Rogers.

Hynd (1988) has suggested a more elaborate mechanism, in which improved nutrient supply increases the cortical cell volume until a genetically determined maximum is reached. Any subsequent improvement in nutrient supply results in a switch in cell differentiation from ortho- to paracortex. This mechanism can explain both the observations of Rogers and those of Hynd.

Depression of wool growth is observed when sheep are infused with the unbalanced protein zein which is devoid of lysine or tryptophane (Reis, 1979). The adverse effect of this protein could be alleviated by addition of lysine to the infusion. Infusion with mixtures of amino acids lacking lysine had a similar effect to zein (Reis, 1988), which is indicative of the important role of this amino acid for wool growth. Reis

(1988) suggests that like methionine, lysine may not be required purely as a substrate for keratin synthesis, but rather it may be needed to sustain bulb cell division.

Hynd (1988) investigated the effect of zein on follicle bulb mitotic rate, and found that there was a depression during zein infusion. Hynd also observed that there may have been an increase in the proportion of cells entering the fibre as opposed to the inner root sheath, and he suggests that this may be a result of the higher lysine requirement of the latter.

Frenkel *et al.* (1974) found that one of the effects of zein infusion was a reduction in the levels of the HGT proteins and similar observations were made with other unbalanced proteins such as wheat and maize gluten and also amino acid mixtures lacking lysine or phenylalanine. Frenkel *et al.* (1974) found that addition of lysine, tryptophane or tyrosine to zein failed to reverse the inhibition of the HGT proteins, indicating that the regulation of synthesis of the HGT proteins is not dependent on supply of tyrosine to the follicles. This is in contrast to the relationship that exists between cysteine and the ultra-high-sulphur proteins.

Many of the treatments that result in elevated levels of the ultra-high-sulphur proteins also result in decreased levels of the HGT proteins. This relationship may be explained in part by the ortho/para switch which occurs in response to nutritional variation (Marshall and Gillespie, 1988), but it cannot be fully explained by this mechanism since the levels of the HGT proteins are usually unaffected by sulphur enrichment (Gillespie, 1983). An alternative possibility is that reduction in wool growth rate associated with inhibition of the HGT proteins liberates more cysteine for UHS protein synthesis (Frenkel *et al.*, 1975).

1.7.2 Genetic Regulation

Some of the variation, both quantitative and qualitative, in wool and hair components, can be attributed to genetic causes (Marshall and Gillespie, 1988). This variation may have no observable effect on the quality of the fibre produced, but there are several well documented mutations which have deleterious consequences on hair and wool formation (Gillespie and Marshall, 1988).

Significant variation between individuals in the amount of the two matrix protein families is observed in human hair and wool. In man sufficient variation in the high-sulphur proteins has been observed to allow identification of individuals for forensic purposes (Lee *et al.*, 1978). In sheep, variation is observed both between and within breeds (Gillespie and Marshall, 1980).

Genetic influences on the wool fibre extend to variation in both the molecular size and composition of the proteins of the wool and hair fibres and is reported to be most extensive in the high-sulphur family (Gillespie, 1983; Gillespie and Marshall, 1980; Marshall and Gillespie, 1988). In man an intermediate filament variant has been identified which is reported to have a frequency of 8% in the population assayed (Marshall and Gillespie, 1988). This variant may be normally expressed in the nails of all individuals. Thus it is believed that expression in hair may be the result of a mutation in a regulatory gene (Marshall, 1983; Marshall and Gillespie, 1988).

Deleterious mutations affecting man, sheep and mice have been documented, which result in major disruption of the matrix proteins, particularly the ultra-high-sulphur proteins and the HGT proteins (Tenenhouse and Gold, 1976; Gillespie, 1983; Gillespie and Marshall, 1988). These are thought to result from disruption of the regulatory process of keratin gene expression rather than a direct alteration in the keratin genes themselves. In sheep the felting lustre mutant (Gillespie and Marshall, 1988) results in elevated levels of the ultra-high-sulphur proteins and almost complete loss of the HGT proteins. The wool of these sheep consists mainly of paracortical cells (Rogers, 1959; Gillespie and Marshall, 1988).

The semi-dominant mutation in mice, termed naked through the loss of much of the fur of affected individuals, also greatly reduces the levels of the HGT proteins (Tenenhouse and Gold, 1976; Gillespie and Marshall, 1988). Two-dimensional PAGE reveals major disruption of components of all the cortical families in mice homozygous and heterozygous for this trait (Gillespie and Marshall, 1988).

Finally, in humans, a presumptive mutation termed trichothiodystrophy (TTD) has been described which results in the loss of the ultra-high-sulphur proteins and the

appearance of anomolous high-sulphur protein (Gillespie, 1983; Gillespie and Marshall, 1988).

1.7.3 Physiological Regulation

Most information on the physiological influences on wool growth and structure have come from the use of various depilatory compounds, including cyclophosphamide, mimosine, ethionine and epidermal growth factor. After cessation of wool and hair growth using these compounds, the regrowth fibre has significantly altered protein constitution, including decreased levels of the HGT proteins and increased levels of the ultra-high-sulphur proteins (Frenkel *et al.*, 1975; Gillespie *et al.*, 1980, 1982; Marshall, 1981b; Reis and Gillespie, 1985). This is believed to be not simply the result of residual amounts of these compounds remaining in the body of the animal, but rather to reflect the authentic constitution of newly growing fibres, since similar findings have been made with follicles regenerating after plucking, and newly operating follicles (Gillespie and Marshall, 1980).

Hormones exert control over wool growth, with hypophysectomy resulting in cessation of wool growth (Wallace, 1979). Administration of thyroxine to hypophysectomized sheep resulted in recovery of wool growth to pre-operative rates, indicating that the major function of the pituitary gland in wool growth is to stimulate the thyroid (Wallace, 1979). Hynd (1988) found that thyroidectomy of sheep resulted in depressed follicle mitotic rate.

1.8 Aims of this Thesis

The aim of this thesis was to undertake an analysis of the ultra-high-sulphur protein components of wool. As described in section 1.4.3 it is the only major fibre component for which no protein or gene sequence has been obtained. This, combined with their regulation by the supply of cysteine to the follicle, made the ultra-high-sulphur proteins of particular interest. A two pronged approach was undertaken to yield information on these proteins. Firstly, it was proposed to purify a UHS protein component and to obtain from it at least a partial amino acid sequence. Since initial attempts to purify the ultra-high-sulphur proteins were made twenty years ago (Lindley *et al.*, 1968; Marshall and Gillespie,

1976a and b), there have been considerable advances made in protein fractionation methods, so that a renewed attempt to purify these components was more practical. Obtaining a partial sequence should facilitate the second aspect of the project, which was to isolate the gene(s) coding for these proteins.

An alternative approach for isolating the ultra-high-sulphur protein genes was through the use of a putative UHS cDNA clone (section 1.6) as a probe.

There are several questions of interest about the UHS proteins which could be addressed once a gene sequence was identified: (a) the primary structure of one or more of the components of this family; (b) the number of components in the genome and their relative position; (c) the site of expression of the genes in the follicle; and (d) the mechanism of regulation by the sulphur-containing amino acids.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Tissue

RNA, DNA, and wool samples used in the experiments described in this thesis were isolated from domestic sheep *Ovis aries*. For most experiments, Merino-Dorset Horn X Border Leicester breed sheep were used. Experiments involving experimentation directly on sheep were approved by the Animal Ethics Committee of the University of Adelaide. Unless stated otherwise, sheep were maintained indoors in individual pens, being fed 800 g of lucerne chaff per day and water *ad libitum*.

Merino Wool samples used for protein purification were a gift of Dr. J.M. Gillespie (CSIRO Division of Protein Chemistry, Parkville, Melbourne).

Human hair and scalp samples were obtained from a caucasian male.

2.1.2 Bacterial strains

The following *E coli* K12 strains were used in the experiments described in this thesis:

MC1061: ara D139, Δ (ara, leu)7697, Δ lacX74, gal U⁻, gal K⁻, hsr⁻, hsm⁺, str A (Casadaban and Cohen, 1980).

JM101: lac, pro, sup E, thi, F' traD36, pro AB, lac Iq, ZAM15 (Messing, 1979).

LE392: F⁻, hsd R514(rk⁻, mk⁻), supE44, supF58, lacY' or (lacIZY)6 galK2, galT22, metBI trpR55, λ ⁻.

2.1.3 Bacteriophage Strains

M13mp18 and M13mp19 (Norranders *et al.*, 1983) were used for subcloning and sequencing of restriction and BAL31 deletion fragments.

2.1.4 Plasmid Strains

pTZ19R: obtained from Pharmacia, and was used as a general cloning vector.

pGEM-2: is a derivative of pSP64 (Melton *et al.*, 1984) and was obtained from Promega.

This was used for SP6 and T7 *in vitro* transcription.

2.1.5 Enzymes

All restriction endonucleases were purchased from Toyobo Chemical Co..

E. coli DNA polymerase I (Klenow fragment) and T4 DNA ligase were purchased from Bresatec Pty. Ltd..

Lysozyme, Ribonuclease T1 (RNase T1) and Ribonuclease A (RNase A) were purchased from Sigma Chemical Co..

Endo-Glu-C (V8 protease) was purchased from Boehringer-Mannheim.

T7 RNA polymerase and BAL31 exonuclease were obtained from Promega.

2.1.6 Radiochemicals

[α -³²P]dATP, [α -³²P]dCTP, [α -³²P]dGTP, and [α -³²P]UTP (specific activity, 3,000 Ci/mmol) were purchased from Bresatec Pty. Ltd..

L-[4,5-³H]Serine (specific activity, 28 Ci/mmol), iodo-[2-¹⁴C] acetic acid (specific activity, 54 mCi/mmol) were purchased from The Radiochemical Centre, Amersham.

5' [α -³⁵S]UTP (specific activity, 1,350 Ci/mmol) was purchased from New England Nuclear.

2.1.7 General Chemicals

The following chemicals were purchased from the Sigma Chemical Co.:

acrylamide and bis-acrylamide, ampicillin, BCIG (5-bromo-4-chloro-3-indoyl-3-D- β -galactoside), cysteine-HCl, EDTA (ethylenediaminetetraacetic acid), guanidine hydrochloride (grade 1), IPTG (isopropylthiogalactoside), 2-mercaptoethanol, salmon sperm DNA, SDS (sodium dodecyl sulphate), 1-ethyl-2-(3-(1-ethylnaphtho[1,2-*d*]thiazolium-2-ylidene)-2-methylpropenyl)naphthol[1,2-*d*]thiazolium bromide (Stainsall), tetracyclin.

HPLC grade acetonitrile and isopropanol were obtained from Waters Associates Pty. Ltd..

4-Vinylpyridine and thioglycolic acid were gifts of Dr. A Inglis (CSIRO Division of Protein Chemistry, Parkville, Melbourne) and were redistilled before use.

Agarose (type 1) was purchased from BRL.

Polyethylene glycol (6,000), and formamide were purchased from BDH Laboratories.

Dextran sulphate and ficoll were purchased from Pharmacia.

Ultra-pure grade urea was obtained from Merck.

Both liquid and powder Vertex denture material was obtained from Dentimex Zeist, Holland.

General chemicals not listed above were obtained from one of the following suppliers: Ajax Chemicals Pty. Ltd., BDH Chemicals Pty. Ltd., May and Baker Pty. Ltd., Merck, Pharmacia, or Sigma Chemical Co.. Chemicals were of the highest purity available.

2.1.8 Media and Buffers

A Growth media

The following growth media were made as described in Miller (1972).

L-Broth: was used for growth of *E. coli* MC1061 containing plasmids and cosmids.

Minimal Media: was used for JM101 growth.

2xYT: this broth was used for the growth of JM101, when infected with M13.

Agar plates and agar/agarose overlay were made using L-broth or minimal media as described by Maniatis *et al.* (1982). Antibiotics were added where appropriate as described by Maniatis *et al.* (1982).

B General Buffers

The following were buffers used regularly during the study and all were made as 20x stocks. The final concentrations of these buffers are listed below:

SSC: 150mM NaCl, 15mM Na citrate, 1mM EDTA, pH 7.0.

SSPE: 150mM NaCl, 20mM sodium dihydrogen phosphate, 1mM EDTA pH 7.4.

TAE: 40mM Tris-acetate pH 8.2, 20mM Na acetate, 1mM EDTA.

TBE: 130mM Tris, 50mM Boric acid, 2.5mM EDTA, pH 8.3.

The following buffers were made as 1x stocks:

TE: 10mM Tris-HCl pH 8.0, 0.1mM EDTA.

PSB: 100mM NaCl, 10mM MgCl₂, 0.05% gelatine, 10mM Tris-Cl pH 7.4.

Bidistilled H₂O was used for making solutions, and solutions were sterilized by autoclaving. Where autoclaving was not possible solutions were made using sterile H₂O and filter sterilized using a Millex-GS 0.22 µM filter (Waters Associates).

2.1.9 Miscellaneous

Zetaprobe membrane used for Southern and northern transfer experiments was obtained from Bio-Rad Laboratories.

Dialysis tubing was obtained from BDH Chemicals Ltd.

Sartorius nitrocellulose filter paper was obtained from Schleicher and Schuell.

X-ray film: Fuji and Kodak (Aust) Pty Ltd..

High molecular weight genomic DNA: Sheep DNA (made from merino ram semen) was a gift of Dr. G.R. Cam, and human DNA (made from placenta) was a gift of Dr. M. Bawden (both of the Dept. of Biochemistry, The University of Adelaide). Both were made by the method of Maniatis *et al.* (1982).

Amino acid depleted rabbit reticulocyte lysate used for *in vitro* translation experiments was obtained from Amersham International plc.

2.2 Methods

2.2.1 Protein Methods

A Extraction and Alkylation of Wool High-Sulphur Proteins with 4-Vinyl Pyridine

Reduced wool high-sulphur protein was preferentially extracted by immersion of wool in 0.8M thioglycolic acid for 18 hours at 0°C as described by Harrap and Gillespie (1963).

The solubilized high-sulphur protein was separated from the residue by vacuum filtration through Whatman 541 paper, and precipitated by addition of 1/30 volume of

glacial acetic acid. The precipitate was recovered by centrifugation (18K rpm, 15 minutes, 4°C, Beckman JA20 rotor). The pellet was washed twice with ethanol and lyophilized.

Alkylation of the reduced total wool high-sulphur protein with 4-vinylpyridine was conducted essentially as described by Friedman *et al.* (1970). After dialysis the protein was either lyophilized or used immediately for HPLC analysis.

B Reduction and S-carboxymethylation of Total Wool Protein and Isolation of Total High-Sulphur Protein

Total wool keratin was reduced and S-carboxymethylated by a modification of the method of Harrap and Gillespie (1963). 100mg of wool was added to 10mL of a solution containing 0.1M 2-mercaptoethanol, 8M urea, 0.5M ethanolamine-HCl pH 10.5, and incubated at 37°C for 1 hour. The wool was then homogenised in this solution using an ultraturrax, and incubated for a further 2 hours. 2.0mL of a solution containing 30% (w/v) iodoacetic acid, 0.3M Tris-Cl pH 8.0, final pH 8.2 adjusted using KOH, was added to the reduced wool, and allowed to react at room temperature for 20 minutes. 0.3mL 14.3M 2-mercaptoethanol was added and the reaction incubated for a further 2 hours. Another 2mL of the iodoacetic acid solution was added. The reaction was immediately centrifuged (18K rpm, 25°C, 30 minutes, Beckman JA20 rotor). 0.4mL 14.3M 2-mercaptoethanol was added to the supernatant and the solution was dialysed against monodistilled H₂O.

The intermediate filament protein and high-glycine/tyrosine proteins were precipitated using zinc acetate as described by Gillespie (1964). The remaining high-sulphur proteins were dialysed against monodistilled H₂O, and freeze-dried.

C Protein Estimation

Where sufficient lyophilized protein was available the amount of protein was determined by weighing.

Where there was insufficient material or the protein was in solution, the method of Bradford (1976) as described by Rylatt and Parish (1982) was used.

D Cleavage with Endoproteinase Glu-C

Protein was cleaved using endoproteinase Glu-C essentially as described by Drapeau (1976). An enzyme to substrate ratio of 1:100 (w/w) was used and the reaction conducted in 100mM Tris-Cl pH 7.8. Under these conditions the enzyme will cleave on the carboxyl terminal side of either aspartate or glutamate.

E ¹⁴C Labelling of Wool and Hair Keratin

Extraction, ¹⁴C-labelling, and S-carboxymethylation of hair and wool samples for analytical purposes was conducted by a modification of the method of Marshall (1981a), reported in Gillespie *et al.* (1982). 2-4mg wool/hair was solubilized in 300μL 8M urea, 0.05M Tris pH 8.0 and 0.05M DTT for 18 hours at room temperature. To a 10μL aliquot 1.0μCi of iodo[2-¹⁴C]acetic acid was added and after 5 minutes at room temperature 5μL of 1M iodoacetic acid, 2.3M Tris pH 8.5 was added. After 5 minutes at room temperature 1μL of 2-mercaptoethanol was added. An aliquot was then taken for electrophoresis and the remainder stored at -20°C.

F Two-dimensional PAGE

Two-dimensional PAGE was conducted by a modification of the method of Marshall (1981a). Alkaline (pH 8.9) electrophoresis was conducted in the first dimension. To the *in vitro* translation products 4μl of 10mM Tris-phosphate pH 7.0, 4μl 80% glycerol and 1μl 1% bromophenol-blue (BPB) were added prior to loading onto the first-dimension gel. To other samples an equal volume of a buffer containing 5mM Tris-phosphate pH 7.0, 0.1% bromophenol blue, and 50% glycerol was added. A slab gel (140mm x 1mm containing 100mm 7.5% separating gel and 30mm 4% stacking gel) was used for the first dimension. Electrophoresis was conducted at 10mA. Strips containing protein samples were cut out after the BPB had migrated 80mm through the separating gel, and placed in a solution containing 0.125M Tris-Cl pH 7.0, 0.1% SDS for 15 minutes.

SDS PAGE in the second dimension was conducted as described by Laemmli (1970), with the following modifications. Electrophoresis was conducted in a Hoefer slab gel apparatus which could accommodate up to four gel slabs. The excised strips from the

first dimension were placed on top of a 140mm x 140mm x 1.5mm slab gel and held in place with 1% agarose. A 10% separating gel was used. Prestained molecular weight markers (BRL low molecular weight range) were loaded onto second dimension gels in a well formed in the stacking gel beside the first-dimension strips. Electrophoresis was conducted at 30mA until the tracking dye had migrated 80mm through the separating gel. Fluorography was then conducted (section 2.2.1 I).

Samples fractionated at the same time on different gel slabs could be directly compared by alignment of the molecular weight markers.

G SDS PAGE

One dimensional PAGE was conducted using the second dimension described above except that wells were formed in the stacking gel. An equal volume of a buffer containing 0.25M Tris pH 7.0, 0.2% SDS, 10mM DTT, 50% glycerol was added to the samples prior to electrophoresis.

H Gel Staining

Protein fractionated in polyacrylamide gels were either silver stained by the method of Merril *et al.* (1984) to produce a positive image of the protein, or stained using "stainsall" by the method of Cutting (1984).

I Fluorography

After electrophoresis the gels were soaked with agitation in 25% ethanol, 5% acetic acid for 30 minutes at room temperature. The gels were transferred to a 1M sodium salicylate solution for 30 minutes with agitation at room temperature. The gels were removed from this solution and dried onto Whatman 3MM paper, using a Hoeffer gel drying apparatus. Dried gels were set up for autoradiography as described below.

J Autoradiography

Gels, nitrocellulose filters and Zetaprobe were autoradiographed using Fuji X-ray film at either room temperature or -80°C for empirically determined exposure times. With

the exception of DNA sequencing gels, a Cronex lightning Plus intensifying screen (Dupont) was used. Where quantitative estimates of band intensity were required, the X-ray film was flashed by exposing to an Ilford Wratten series 1A safety light at a distance of 300mm for 60 seconds.

2.2.2 DNA Methods

A General Methods

Ethanol precipitation was conducted by the method of Zeugin and Hartley (1985). Phenol extraction was conducted as described by Maniatis *et al.* (1982).

End-Filling 5' overhangs was conducted as described by Davis *et al.* (1986) using *E. coli* DNA polymerase I. 3' overhangs were removed by exposure to *E. coli* DNA polymerase I under the same conditions as used for endfilling.

B Preparation of labelled DNA

Nick Translation

Nick translation radiolabelling of DNA was conducted by the method of Rigby *et al.* (1977), using a kit obtained from Bresatec according to the manufacturer's instructions. After radiolabelling the DNA was purified using "Gene-clean" (2.2.2 J).

Oligo-labelling

Oligo-labelling of DNA restriction fragments was conducted by the method of Feinberg and Volgelstein (1983), using a kit obtained from Bresa, according to the manufacturer's instructions. After labelling the DNA was purified using "Gene-clean" (2.2.2 J).

End-labelling

Radiolabelling of molecular weight markers was conducted by end-filling 3' overhangs resulting from restriction endonuclease activity. A solution containing the DNA to be labelled was made 20 μ M in each of the deoxynucleotides other than that used for labelling, 50mM Tris-Cl pH 7.6. In a 10 μ l reaction 50 μ Ci [α -³²P]dNTP (S.A. about 3,000

Ci/mmol) was used. 1 unit of DNA polymerase I (Klenow fragment) was added and the reaction was allowed to proceed for 10 minutes at 37°C.

C Screening Cosmid Library

The human genomic cosmid library was a gift of Dr. K. H. Choo and was constructed as described by Choo et al. (1986). The library was screened at high density (50,000 per 150 mm filter) for recombinant clones at high density by the method of Hanahan and Meselson (1980). Positive clones were rescreened at low density until all colonies on a filter hybridized to the probe (2 cycles were usually sufficient).

D Screening Lambda Library

The sheep genomic lambda library was a gift of Dr. R Crawford from the Howard Florey Institute of Experimental Physiology and Medicine, Melbourne. The library was screened at high density (30,000 pfu per 150mm filter) by plaque hybridization as described by Benton and Davies (1977). Positive plaques were rescreened until all plaques on a plate hybridized to the probe.

E DNA preparation

Recombinant plasmid DNA was prepared by a modification of the procedure of Birnboim and Doly, (1979) as described by Maniatis *et al.* (1982), using CsCl density gradient equilibrium centrifugation (Radloff *et al.*, 1967) for final purification of the plasmid DNA.

Analytical preparation of DNA from plasmid and cosmid agar plate colonies was conducted essentially as described by Ish-Horowicz and Burke (1981).

DNA was prepared from lambda recombinant phage which had been eluted in 10mL PSB from confluent 150mm plates, using CsCl step gradient centrifugation as described by Davis *et al.* (1986).

F Cleavage by Restriction Endonucleases

Restriction endonuclease cleavage of DNA was conducted under the conditions described for the particular enzyme in the New England Biolabs catalogue. Usually 1-2 enzyme units were used per μg of DNA, and 60 minutes reaction time was sufficient. Genomic DNA often required 5-10 fold excess of enzyme and 18 hour reaction times.

G Polyacrylamide Electrophoresis

Polyacrylamide electrophoresis of nucleic acids was carried out essentially as described by Maniatis *et al.* (1982). TBE running buffer was generally used.

Non-denaturing electrophoresis was conducted in vertical gel apparatus containing a slab 140mm x 140mm x 1.0mm in size, and samples were loaded in 10% glycerol, 1mM EDTA, 10mM Tris-Cl pH 7.4, 0.2% bromophenol blue, 0.2% xylene cyanol. Electrophoresis was conducted with a current of between 20 and 30mA, until the tracker dye had migrated the required distance.

Denaturing electrophoresis was conducted by the addition of 7M urea to the acrylamide solutions. DNA and RNA samples were loaded in a solution containing 50% formamide, 10mM EDTA pH 8.0, 0.1% bromophenol blue, 0.1% xylene cyanol. Samples were heated to 100°C for 2 minutes prior to loading.

After electrophoresis, unlabelled samples were detected by staining with 0.1% (w/v) ethidium bromide for approximately 10 minutes and then viewed under short wave UV light. Photography was conducted using a polaroid camera and type 665 (ASA 80) or 667 (ASA 3,000) Polaroid film.

Radioactive samples were detected by autoradiography. DNA sequencing gels were fixed prior to autoradiography by washing in 20% ethanol, 12% acetic acid and dried at 100°C.

H Agarose electrophoresis

Was conducted in horizontal apparatus essentially as described by Maniatis *et al.* (1982). For analytical purposes gels 75mm (width) x 55mm (length) in size were used.

For large scale preparative electrophoresis, or where higher resolution was required, Pharmacia type GNA-100 or GNA-200 apparatus were used.

TAE buffer was used throughout, samples were loaded in the same buffer as used for nondenaturing polyacrylamide electrophoresis. A current of 50-100mA was applied to the gels. Fragments were visualized by staining with ethidium bromide as described above.

I Isolation of DNA from Polyacrylamide Gels

The region of the gel containing the sample to be isolated was removed from the remainder of the gel and placed in TE, 0.1% SDS and the DNA allowed to elute at 37°C for 3-18 hours. The DNA was recovered by ethanol precipitation.

J Isolation of DNA from Agarose Gels

Restriction fragments were extracted from agarose by a method described by Vogelstein and Gillespie (1979), using a "Gene-clean" kit obtained from BIO 101, according to the manufacturer's instructions.

K Subcloning into Plasmids and Phage

Vector Preparation

Plasmid and M13 cloning vectors were prepared by linearizing with the appropriate restriction endonuclease(s), followed by removal of the 5' terminal phosphate groups as described by Maniatis *et al.* (1982). The vectors were purified by agarose electrophoresis prior to use.

Ligation

The DNA fragment and appropriate vector were combined in 3:1 molar ratio (estimated by intensity of staining with ethidium bromide), in a solution containing 50mM Tris-Cl pH 7.5, 1mM ATP, 10mM MgCl₂, 10mM DTT. 0.5-1.0 units of T4 DNA ligase were added, and ligation allowed to proceed overnight at 14°C. 10-20ng of vector DNA in

a 10-20 μ l reaction usually allowed the isolation of sufficient recombinant clones for practical purposes.

Plasmid Transformation

E. coli strain MC1061 cells were made competent and transformed by a modification of the method of Mandel and Higa (1970) as described by Dagert and Ehrlich (1979)

Generally the transformation of half of one ligation reaction produced sufficient colonies containing recombinant plasmids for practical purposes.

M13 Transfection

Preparation of competent cells and transfection of recombinant M13 phage was conducted essentially as described by Messing (1983). One half of a ligation reaction was used in a single transfection.

L Bal31 deletions

The method for preparation of progressively deleted DNA fragments was adapted from a procedure reported in Davis *et al.* (1986). The clone of interest was linearized with an appropriate restriction endonuclease, and pilot reactions conducted in a solution containing 0.6M NaCl, 12mM CaCl₂ and MgCl₂, 50mM Tris-Cl pH 8.0, 1mM EDTA at 37°C. Aliquots were drawn from the reaction at fixed intervals and mixed with buffer saturated phenol. When the reaction was complete the aliquots were individually phenol extracted and ethanol precipitated. The progressively deleted inserts were excised by cleavage with an appropriate enzyme, and their size was assessed by agarose electrophoresis.

Once the desired reaction conditions had been found, a full-scale reaction was conducted in a similar manner, with the ratio of reaction volume to DNA concentration to enzyme concentration being identical to that of the successful pilot reaction. After the deletion reaction had been conducted, all the time-aliquots were combined in a single tube, phenol extracted, and ethanol precipitated. The deleted ends were endfilled (2.2.2 A) prior

to excision of the progressively deleted inserts from the vector. The fragments were size-fractionated by agarose electrophoresis, and the appropriately sized fragments were isolated and subcloned into a suitable M13 vector.

M Preparation of ssDNA for Sequencing

Single-stranded M13 template DNA for dideoxy sequencing was prepared essentially as described by Winter and Fields (1980).

N DNA sequencing

Sequencing of single-stranded M13 DNA templates was conducted by the dideoxynucleotide chain termination method of Sanger *et al.* (1977 and 1980). A kit obtained from Bresatec Pty. Ltd. was used for most sequencing reactions according to the manufacturer's instructions, with the exception that polymerisation reactions were conducted in titre plates.

Some sequencing reactions were conducted using a modified form of T7 DNA polymerase (Tabor and Richardson, 1987), using a "Sequenase" kit obtained from the United States Biochemical Corporation. Reactions were conducted according to the manufacturer's directions.

[α -³²P]dATP was used with both kits for labelling of extension products.

Denaturing PAGE of sequencing reactions was conducted using a 6% 400mm x 200mm x 0.25mm acrylamide gel. Electrophoresis was conducted at 1,200V/30mA maximum, after the addition of 1/10 volume 3M Na acetate to the lower tank. After electrophoresis the gels were fixed (2.2.2 G) and autoradiographed in the absence of an intensifying screen at room temperature (2.2.1 J).

O Southern Transfer

Transfer of restriction endonuclease cleaved plasmid and genomic DNA to Zetaprobe was conducted by a modification of the method of Southern (1977), reported by Reed and Mann (1985). DNA was transferred after agarose electrophoresis to Zetaprobe using 0.4M NaOH as the transfer buffer.

Hybridization was conducted in the presence of dextran sulphate and formamide as described by Reed and Mann (1985).

After hybridization, the membranes were washed at low stringency in 2 changes of 2x SSPE, 0.1% SDS for 10 minutes with agitation at room temperature. All filters were rewashed at 65°C in the same solution. Autoradiography was conducted at this stage.

Some filters were rewashed at high stringency in 0.2x SSPE, 0.5% SDS, 65°C, with agitation for a minimum of 60 minutes. Autoradiography was then reconducted.

Membranes which were to be reused were washed in two changes of boiling 0.1x SSPE, 1.0% SDS. The initial wash was agitated until the solution had cooled to room temperature, and then repeated. Cleaned filters were autoradiographed before being reused.

2.2.3 RNA Methods

A RNA Preparation from Wool Follicles

Sheep were sheared using Oster clippers with a size 40 comb. Xylocaine local anaesthetic was injected subcutaneously between the spinal cord and the region where follicles were to be plucked, in 0.25mL amounts (10mL total). A 1:1 mixture of vertex powder and liquid was applied to the side of the sheep, in areas of approximately 100mm X 50mm. This was overlaid with a strip of glassfibre tape, and once set, the strips were removed and immediately placed in liquid nitrogen. Follicle patches were either stored at -80°C or used immediately.

Follicular material was collected by scraping into liquid nitrogen. Extraction of RNA from this tissue was conducted by a modification of the method of Brooker *et al.* (1980). The following buffer volumes refer to 1g of starting material. To the follicle tissue 10mL of a buffer containing 6M guanidine hydrochloride, 0.2M Na acetate pH 5.5, and 0.1M 2-mercaptoethanol, and homogenized using an ultraturrax. The material was ethanol precipitated, and resuspended in 10mL of 7M urea, 0.1M Tris-Cl, 5mM EDTA and 0.1% SDS. This solution was phenol extracted until the interface was clear, and then ethanol precipitated. High molecular weight DNA was removed by the method of Diaz-Ruis and

Kaper (1978). The RNA was ethanol precipitated twice and resuspended in an appropriate volume of sterile TE.

B Northern Transfer

Glyoxylated RNA was fractionated through 1.5% agarose as described by Thomas (1983), and transferred to Zetaprobe by a modification of the method of Southern (1977). RNA was transferred without pretreatment of the gel using 10x SSC as the transfer buffer. After transfer, the Zetaprobe membrane was washed in 10mM NaOH for 5 minutes, and neutralized in 2x SSPE. Hybridizations and washings were conducted under the same conditions as described for Southern transfer (2.2.2 O).

C RNA Protection

RNA protection experiments using sheep follicle RNA were conducted by the method of Krieg and Melton (1987). Total yeast RNA was used as both carrier for ethanol precipitations and in negative control hybridizations.

D *In vitro* Transcription

Transcription of inserts in pGEM-2 using either SP6 or T7 RNA polymerases was conducted by the method of Krieg and Melton (1987), using a kit obtained from Bresatec Pty. Ltd.. Both unlabelled RNA and RNA labelled to high specific activity using [α - ^{32}P]UTP or [α - ^{35}S]UTP were produced in this manner.

Where applicable the *in vitro* transcripts were capped at the 5' end by the addition of G(5')ppp(5')G to the reaction mix as described by Konaska *et al.* (1984).

Transcripts were phenol extracted and ethanol precipitated prior to further use. For RNA protection experiments radiolabelled transcripts were purified by denaturing PAGE (2.2.2 G).

E *In vitro* Translation

Translation of the *in vitro* transcription products was conducted by the method of Pelham and Jackson (1976) using amino acid depleted rabbit reticulocyte lysate obtained from Amersham Int. plc. 15 μ L reactions containing 10 μ L rabbit reticulocyte lysate, 5 μ Ci L-[2,3-³H]serine (37 Ci/mmol, Amersham Int. plc), 135mM potassium acetate, 100 μ M 19 amino acids minus serine (Amersham Int. plc), and one tenth of the RNA from a single 20 μ L *in vitro* transcription reaction, were incubated for 90 mins. at 30°C. The reactions were either S-carboxymethylated or made 0.05M in DTT and then either used immediately for PAGE analysis or stored at -20°C.

In vitro translation products were S-carboxymethylated in the following manner. The 15 μ L translation reactions were made 0.05M DDT and 0.05M Tris (pH 8.0), and incubated for 30 mins at 37°C. 4 μ L of a solution containing 1M iodoacetic acid and 2.3M Tris (pH 8.5) was added. After 10 minutes at room temperature the excess iodoacetic acid was allowed to react with 2 μ L of 2-mercaptoethanol.

F Tissue *In Situ* Hybridization

In situ hybridizations were based on the method of Cox *et al.* (1984) with modifications by D. Roop (personal communication) and B. Powell (personal communication).

The paraffin embedded sections were deparaffinized by soaking in Histo-Clear (2x2 minutes), and rehydrated by immersing consecutively in 95% (1 minute), 80% and 70% ethanol (2 minutes each). The sections were acetylated by immersing in 0.1M triethanolamine-HCl pH 8.0, 0.25% (v/v) acetic anhydride (10 minutes). The slides were washed twice in 2x SSPE (2x 1 minute) and allowed to air dry prior to hybridization.

Negative control sections were exposed to a solution containing 100 μ g/mL RNase A and 1 unit/mL RNase T1 in 2x SSPE at 37°C for 2 hours. RNase treatment was conducted under a coverslip in a humid chamber containing 50% formamide, 0.25M NaPO₄ pH 6.5.

Hybridization was conducted in 50% formamide, 10% PEG, 2xSSPE, 1% SDS, 1mg/ml yeast RNA, 0.5mg/ml salmon sperm DNA, 0.5% Blotto and 50mM 2-

mercaptoethanol. A total of 55µl of hybridization solution was used per 22x22mm area. Hybridizations were conducted under coverslips in a humid chamber overnight at 50°C.

Hybridized sections were washed twice in 50% formamide, 2xSSPE at 52°C (1x 5 minutes; 1x 20 minutes), followed by two washes at room temperature in 2xSSPE (2x5 minutes). The sections were then RNased treated as described above for 30 minutes. The slides were then washed twice in 2xSSPE at 52°C for 5 minutes followed by 0.1xSSPE at 52°C for 30 minutes. The sections were dehydrated by soaking consecutively in 70%, 80% and 95% ethanol. The slides were air dried prior to autoradiography.

Autoradiography of sections was conducted as described by Rogers (1979) using Ilford K5 and Ilford L4 emulsion. Kodak D19 developer and Ilford Hypam Rapid Fixer was used to develop autoradiographed tissue sections.

Tissue sections were stained by the SACPIC staining procedure of Auber (1950).

Photography was conducted using a Zeiss photomicrographic apparatus using Kodak "Kodacolor" ASA 100 or ASA 400 film.

2.2.4 Computer Programmes

A General

All sequence analysis programmes were run on a VAX 11/750 computer using the VMS operating system. General nucleic acid and protein sequencing programmes from the Standen package were used. Some programmes obtained from NIH were also used.

B Predict

Predict is a suite of 10 secondary structure prediction programmes which were assembled at the Dept. of Biophysics, University of Leeds. The component programmes are as follows:

Garnier (Garnier *et al.*, 1978).

Lim (Lim *et al.*, 1974).

Fasman (Chou and Fasman, 1974).

Nagano (Nagano, 1973).

Burgess (Burgess *et al.*, 1974).

Dufton (Dufton and Hider, 1977).

Kabat (Kabat, 1973).

McLach.

Joint (Eliopoulos *et al.*, 1982).

Seplot.

The output of the first 8 programmes was processed by JOINT which produced a joint prediction of secondary structure. This was used by SEPLIT to draw a joint prediction histogram.

2.2.5 Containment Facilities

All work involving recombinant DNA was carried out under C1 containment conditions for work involving viable organisms and C0 containment conditions for work not involving viable organisms, as defined and approved by the Australian Academy of Science Committee on Recombinant DNA and by the University Council of the University of Adelaide.

Chapter 3

Wool UHS Protein Purification

3.1 Introduction

This chapter describes an attempt to isolate a pure fraction of wool UHS protein and to obtain sufficient amino acid sequence to allow the construction of an oligonucleotide probe specific for a UHS protein gene or genes.

Previous fractionation attempts by other workers (Lindley *et al.*, 1968; Dr R.C. Marshall, personal communication) have not been successful in purifying any components from this family and consequently no protein sequence is available. Fractionation of the UHS proteins from the high-sulphur proteins of lower cysteine content proved difficult because of the similarity in the properties of both types of proteins (Marshall and Gillespie, 1976a). Also, analytical identification of the purity of UHS protein fractions was difficult, as the only method of detection of these proteins was as a partially resolved peak when fractionated by moving boundary electrophoresis (Gillespie and Reis, 1966). Other electrophoretic methods failed to resolve these proteins.

By an extensive extraction procedure which involved: (a) anion exchange chromatography; (b) fractional precipitation using potassium carbonate; and (c) gel filtration, an enriched fraction of UHS protein was obtained (Lindley *et al.*, 1968). No subsequent work describing the isolation of a single component from this fraction has ever been published.

Because of the difficulty encountered with wool UHS protein, attempts were made to purify UHS protein components from mouse hair (Marshall and Gillespie, 1976 a, b), which contains higher levels of these proteins than wool (Gillespie and Inglis, 1965). These studies revealed that mouse hair UHS proteins are highly heterogeneous in both size and amino acid content. However, as with wool UHS protein, no mouse hair UHS protein component was ever purified and sequenced.

The aim of the work presented in this chapter was to apply the more recent techniques of protein chemistry to the purification of a fraction of wool UHS protein.

3.2 Results

3.2.1 High-Sulphur Protein Derived using 4-Vinylpyridine.

Initial attempts to purify the UHS proteins were conducted using wool derived with 4-vinylpyridine. This reagent will react with the sulphydryl group of cysteine, and has an advantage over alkylation with other reagents such as iodoacetate, in that it should allow the relative cysteine content of a protein to be determined spectrophotometrically by comparing its absorption at 215nm (peptide bonds) to that at 260nm (pyridine ring) (Freidman and Krull, 1969).

Prior to reaction with 4-vinylpyridine, a preliminary purification of total high-sulphur protein was conducted by preferential extraction at pH10.3 with 0.8M potassium thioglycollate at 0°C as described by Gillespie (1962) and Harrap and Gillespie (1963).

Upon reaction with 4-vinylpyridine (for method see chapter 2, section 2.2.1 A), it was found that the total fraction of high-sulphur protein was partially insoluble over a wide pH range. When fractionated by reversed-phase chromatography, poor and inconsistent fractionation patterns were achieved.

In view of these difficulties, use of 4-vinylpyridine was discontinued and carboxymethylation was employed for all subsequent work.

3.2.2 Ion-Exchange-Chromatography

Total high-sulphur protein (chapter 2, section 2.2.1 B) was fractionated by anion exchange chromatography using a 5mL Pharmacia "Mono Q" column, which contains a quarternary ammonium exchange resin. Chromatography was conducted at pH10.5 with an increasing gradient of Cl⁻ counter ions (figure 3.1). To locate the UHS proteins, the elution profile of high-sulphur protein isolated from sulphur-enriched wool was compared to that of wool clipped prior to infusion. Figure 3.1 reveals that the sulphur-enriched sample contained elevated amounts of material in the final major peak. It was concluded that the UHS proteins lay within this region, and material was collected from this region as indicated in figure 3.2. The results of amino acid analysis on four fractions collected from this region of the elution profile are presented in table 3.1. These indicated that material which eluted at lower salt concentration had a lower cysteine content, similar to that of the

Figure 3.1

Anion Exchange Chromatography of Total High-Sulphur Protein

The elution profile of total high-sulphur protein isolated from sulphur-enriched wool (4.06% w/w elemental sulphur) when fractionated by anion exchange through a 5mL Pharmacia "Mono Q" column. The dashed line indicates where the profile of total high-sulphur protein from control wool (3.49% sulphur) differs from the sulphur-enriched fraction.

Buffer conditions:

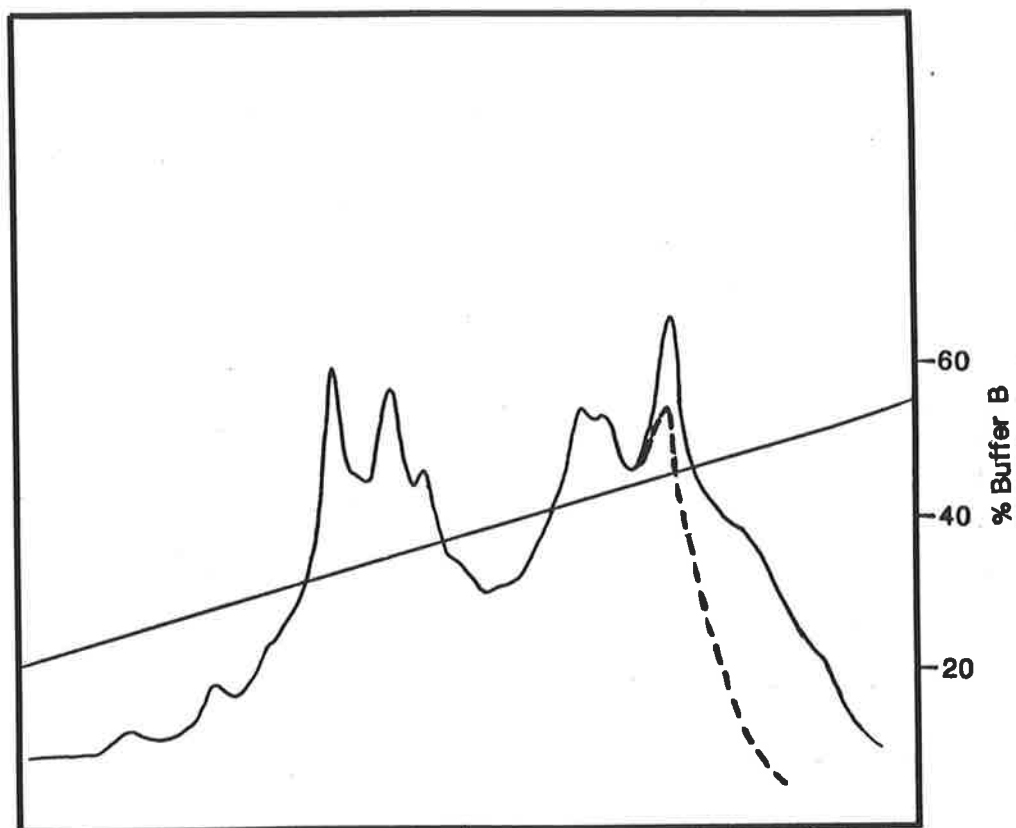
Buffer A: 40mM Sodium Borate pH10.5, 8M urea.

Buffer B: Buffer A + 1M NaCl.

Gradient:

0 to 55% buffer B over an elution volume of 20mL and a flow rate of 1 mL/min.

Relative Absorbance 215nm

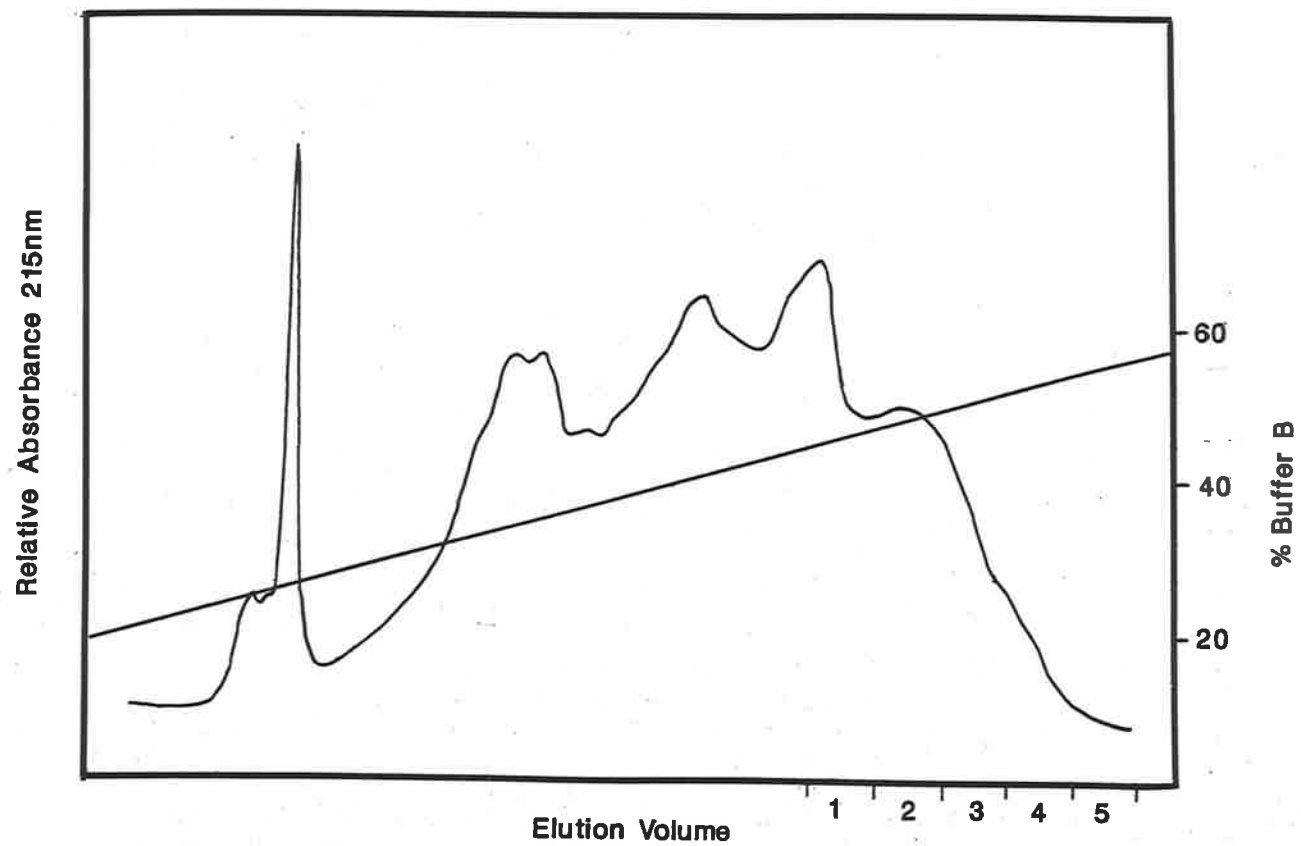


Elution Volume

Figure 3.2

Preparative Anion Exchange Chromatography

A preparative fractionation of 5 mg of high-sulphur protein isolated from sulphur-enriched wool. The same column, buffers and buffer gradient were used as described in figure 3.1. 1mL fractions of the material which is elevated in sulphur-enriched wool were collected as indicated. These were each pooled with fractions taken from the same position from a number of subsequent preparative fractionations. These were dialysed and lyophilized. The results of amino acid analysis of fractions 1, 3, 4 and 5 are presented in table 3.1.



high-sulphur protein which is not induced in response to cysteine infusion (Gillespie, 1983). The value for the cysteine content of material later in the profile was closer to that previously observed for the UHS proteins (Gillespie and Broad, 1972). The fractions from the dietary-enriched peak which contained higher levels of cysteine were pooled as indicated in figure 3.2. The pooled fraction was dialysed and lyophilized before subsequent fractionation procedures were conducted.

An attempt was made to fractionate the UHS protein which had been partially purified by anion exchange chromatography using reversed-phase chromatography (figure 3.3). However, all of the material in this fraction eluted as a single peak, indicating that reversed-phase chromatography would not improve the resolution achieved using anion exchange chromatography.

The pooled dietary-enriched fraction isolated by anion exchange chromatography was refractionated using the same column and buffer conditions, but a shallower counter-ion gradient was used (figure 3.4). The fractions collected from the second cycle of anion exchange chromatography were analyzed by SDS PAGE (figure 3.5). This indicated that all fractions appeared to contain the same protein species but minor differences existed in the relative proportions of the species. In particular, there appeared to be an increase in the proportion of the high molecular weight smear in fractions eluting at higher counter-ion concentrations.

Material from the second cycle of anion exchange chromatography was pooled as indicated in figure 3.4. Amino acid analysis was conducted upon the material within this region of the elution profile and it was found to contain approximately 30 mole% cysteine (table 3.1). Overall, the levels of the other amino acids were similar to that observed previously for the UHS proteins (Gillespie and Broad, 1972) (table 3.1).

The pooled protein from the second cycle of anion exchange chromatography was fractionated by two-dimensional PAGE (figure 3.6), and this revealed that the fraction contained only material which migrated as a diagonal smear emanating from the origin. It has been observed previously that the UHS proteins migrate in this region when fractionated by two-dimensional PAGE (Gillespie *et al.*, 1982).

Figure 3.3

Reversed-Phase Chromatography of a Crude Fraction of UHS Protein

The reversed phase chromatographic profile of a combined fraction of the protein which increases in response to sulphur-enrichment which was isolated by anion exchange chromatography (fractions 1-5, figure 3.2).

Column:

Brownlee C18 reversed-phase guard column.

Buffer conditions:

Buffer A: 0.1% TFA.

Buffer B: 0.1% TFA in acetonitrile.

Gradient:

20%-50% buffer B over 15mL and a flow rate of 1 mL/min.

Graduations on the X axis mark 2mL intervals in the elution volume.

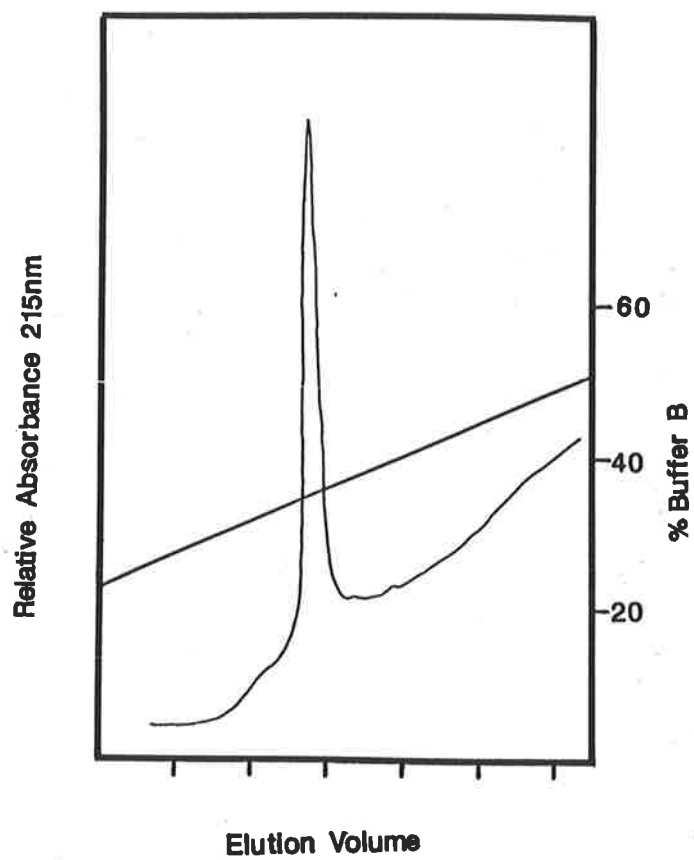


Figure 3.4

Anion Exchange Chromatography of Crude Fraction of UHS Protein

Refractionation using a Mono Q column of the combined fractions 2-5 isolated from total high-sulphur protein using the same column, as delineated in figure 3.2.

Column:

5mL Pharmacia "Mono Q" anion exchange column.

Buffers:

Buffer A: 40mM Sodium Borate pH 10.5, 8M urea.

Buffer B: Buffer A + 1M NaCl.

Gradient:

35% B to 55% B over 40mL, with a flow rate of 1 mL/min.

Initially 1mL fractions (1 to 5) were collected and subsequently 2mL fractions (6 to 9) were collected. The SDS PAGE profiles of fractions 1 to 8 are presented in figure 3.5.

For the purposes of peptide isolation early fractions (1 and 2) in this and subsequent fractionations were discarded and the remainder were pooled.

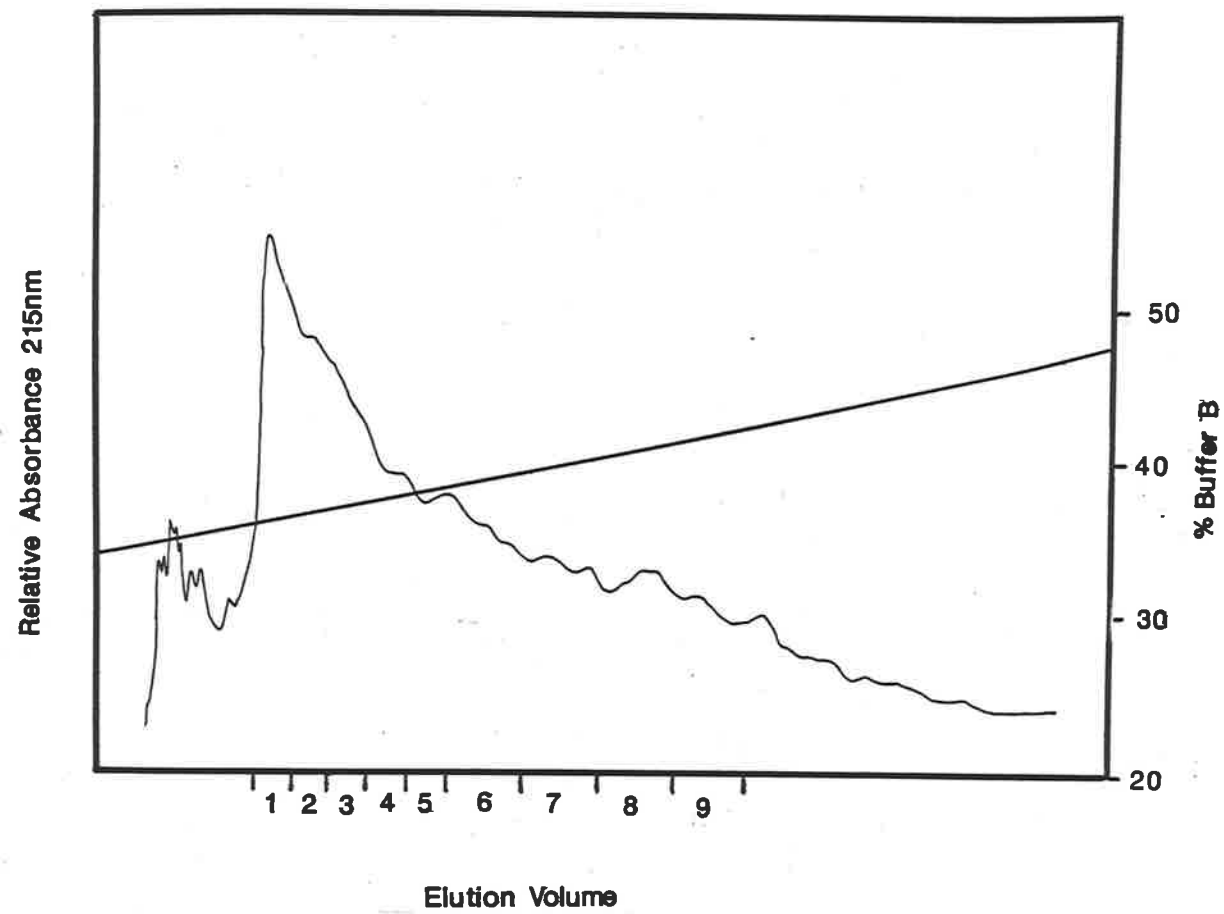


Figure 3.5

SDS PAGE of Refractionated UHS Protein

The profile of fractions 1 to 8 of the refractionated dietary enriched high-sulphur protein when electrophoresed through 10% acrylamide in the presence of SDS (2.2.1 G). The first track entitled "Total" contains protein from the combined fractions 2 to 5 of the initial anion exchange fractionation (figure 3.2) which was the starting material for refractionation delineated in figure 3.4.

The protein was visualized by silver staining (2.2.1 H).

Early → Late

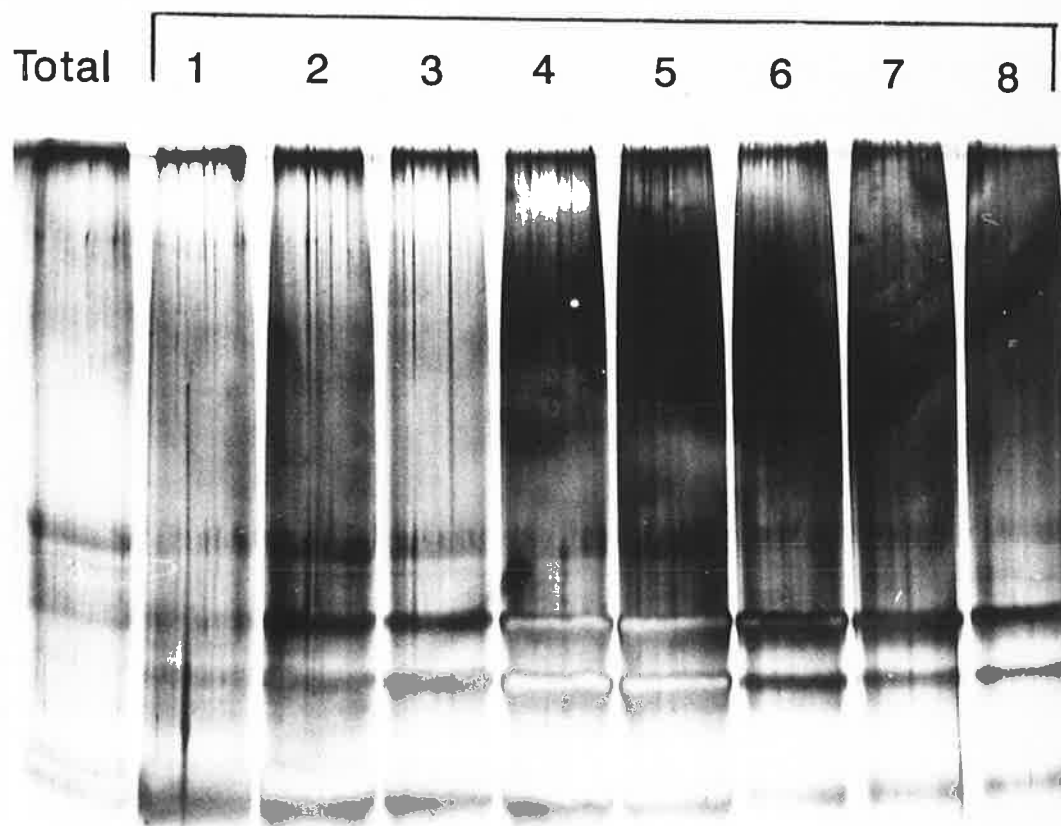


Table 3.1

Amino Acid Analysis

The results of amino acid analysis on the following fractions: the 1st, 4th and 5th fractions from the initial anion exchange fractionation (figure 3.2); and a combined fraction of protein after a second cycle of anion exchange fractionation (see figure 3.4). Also presented is the published amino acid analysis of a purified fraction of UHS protein (from Gillespie and Broad, 1972). The proportions of the amino acids are expressed as a % of total N. Amino acid analyses were conducted by the CSIRO Division of Protein Chemistry, Parkville.

Amino Acid Analysis

Amino Acid	Published UHSP	1st cycle				Combined 2nd Cycle
		1	3	4	5	
Ala	2.0	5.4	2.4	2.4	0.8	2.1
Arg	6.9	5.5	5.9	6.0	5.8	4.2
Asp	0.6	1.0	0.9	0.9	0.8	0.7
Cys	29.9	19.0	25.2	27.2	27.3	30.1
Glu	7.9	9.1	9.3	9.4	9.3	9.4
Gly	4.2	6.8	6.2	4.7	8.7	4.7
His	1.3	1.0	1.3	1.4	1.1	1.0
Ile	1.7	1.6	1.6	1.6	1.6	1.8
Lys	0.9	1.0	1.1	1.1	1.1	0.3
Leu	1.4	1.7	1.4	1.4	1.3	1.4
Met	0	0	0	0	0	0
Phe	0.5	0.3	0.4	0.4	0.3	0.5
Pro	12.8	15.5	11.9	12.1	11.8	11.0
Ser	12.7	11.5	10.3	10.6	10.2	11.5
Thr	11.1	9.4	10.0	10.8	10.8	9.9
Tyr	1.8	1.7	1.7	1.7	1.7	1.4
Val	4.3	2.9	3.5	3.6	3.4	3.2

Figure 3.6

Two-Dimensional PAGE

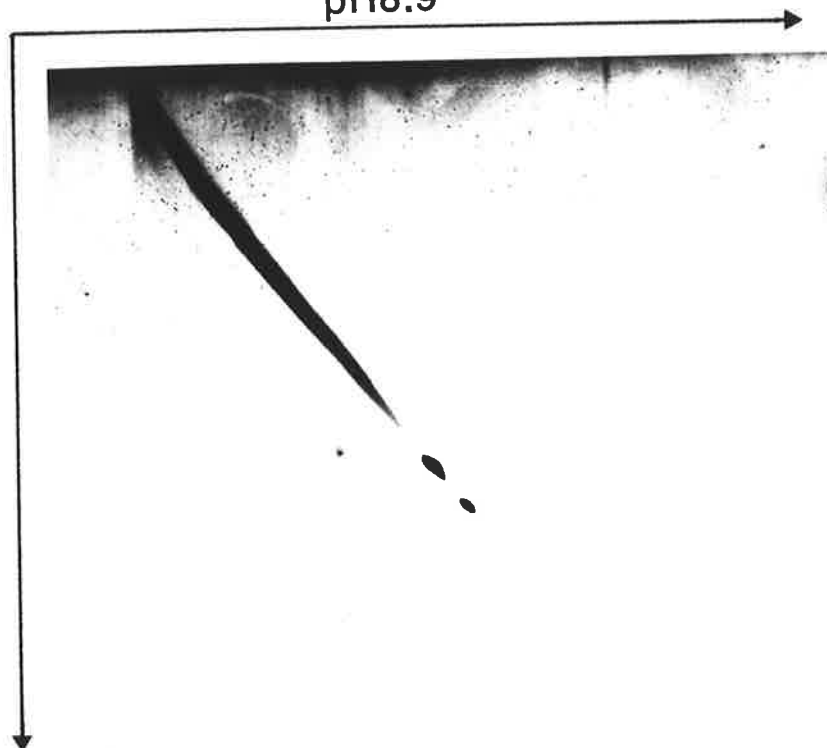
Electrophoretic pattern of the purified UHS protein fraction described in the text when fractionated by two-dimensional PAGE (2.2.2 F). All the protein in this fraction migrated upon a 45° diagonal upon which the UHS proteins are believed to migrate (see text). After electrophoresis the protein was visualized by silver staining (2.2.1 H).

Below is a two-dimensional PAGE fractionation of total wool, S-carboxymethylated using [¹⁴C]iodoacetic acid (2.2.1 E). The line indicates the approximate position of the components present in the purified fraction of UHS protein above. The exact location of the purified fraction upon the diagonal is uncertain as electrophoresis of the total wool and UHS protein fractions were conducted at different times.

Purified UHS Fraction

pH8.9

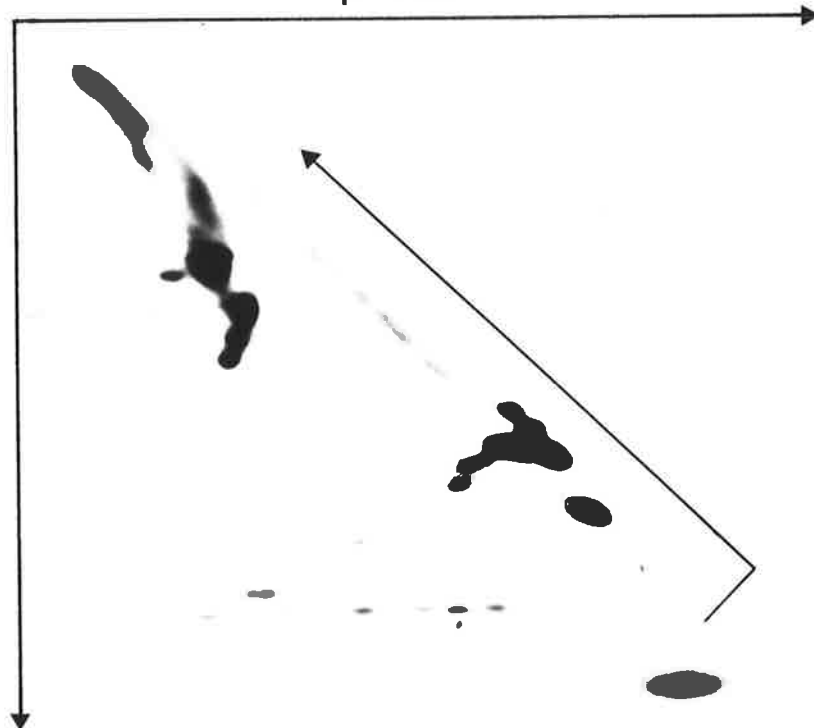
SDS



Total Wool

pH8.9

SDS



3.2.3 Peptide isolation

The UHS protein fraction purified as described in section 3.2.2 was used as substrate for endo-Glu-C, which is a protease that cleaves on the C-terminal side of glutamic and aspartic acid residues (chapter 2, 2.2.1 D). This protease was used since aspartate and glutamate residues are not present in the amino acid repeat observed in the high-sulphur family (see chapter 1, section 1.4.3), thereby allowing the isolation of longer peptides if a similar repeat is present in the UHS proteins. Peptides so produced were fractionated by reversed-phase chromatography (RPC). All the major peptide species obtained from the UHS protein fraction eluted from the column at a similar position in the buffer gradient (figure 3.7). Peptide material collected from this fractionation was analyzed by SDS PAGE (figure 3.8), and by this technique it was clear that all fractions appeared to contain more than one peptide species. However, some fractions appeared to contain a peptide species which was of higher concentration than the remaining peptides.

3.2.4 Peptide Sequence Analysis

Sequence analysis was attempted on the fractions indicated (figure 3.8). This resulted in the identification of the sequence of the dominant peptides present in each fraction and these are presented in figure 3.9. Comparison of the sequences of each of the fractions revealed that they were very similar. Fraction 15 produced the longest sequence, which extended for approximately 30 residues. Inspection of its sequence revealed that it contained 30 mole% cysteine, consistent with the amino acid content of the UHS proteins.

3.3 Discussion

Initial attempts to fractionate total high-sulphur protein derived with 4-vinylpyridine were not successful. The rationale for using this reagent was that it would have allowed easier identification of the UHS protein components during chromatography without the need to resort to the more laborious techniques of two-dimensional PAGE or amino acid analysis. The poor fractionation patterns and insolubility of the wool protein derived with this agent prevented the use of this method to isolate the UHS proteins.

Figure 3.7

Reversed Phase Chromatography of UHS Protein Peptides

The profile of the endo-Glu-C cleavage products of a purified fraction of UHS protein when fractionated by reversed-phase chromatography.

Column:

Brownlee C18 "Aquapore" column.

Buffers:

Buffer A: 10mM sodium acetate pH 4.0.

Buffer B: acetonitrile.

Gradient:

0% to 60% buffer B over 60mL with a flow rate of 1 mL/min.

0.5mL fractions were collected as indicated. Each graduation on the X axis contains two fractions. Figure 3.8 contains the electrophoretic patterns of selected fractions.

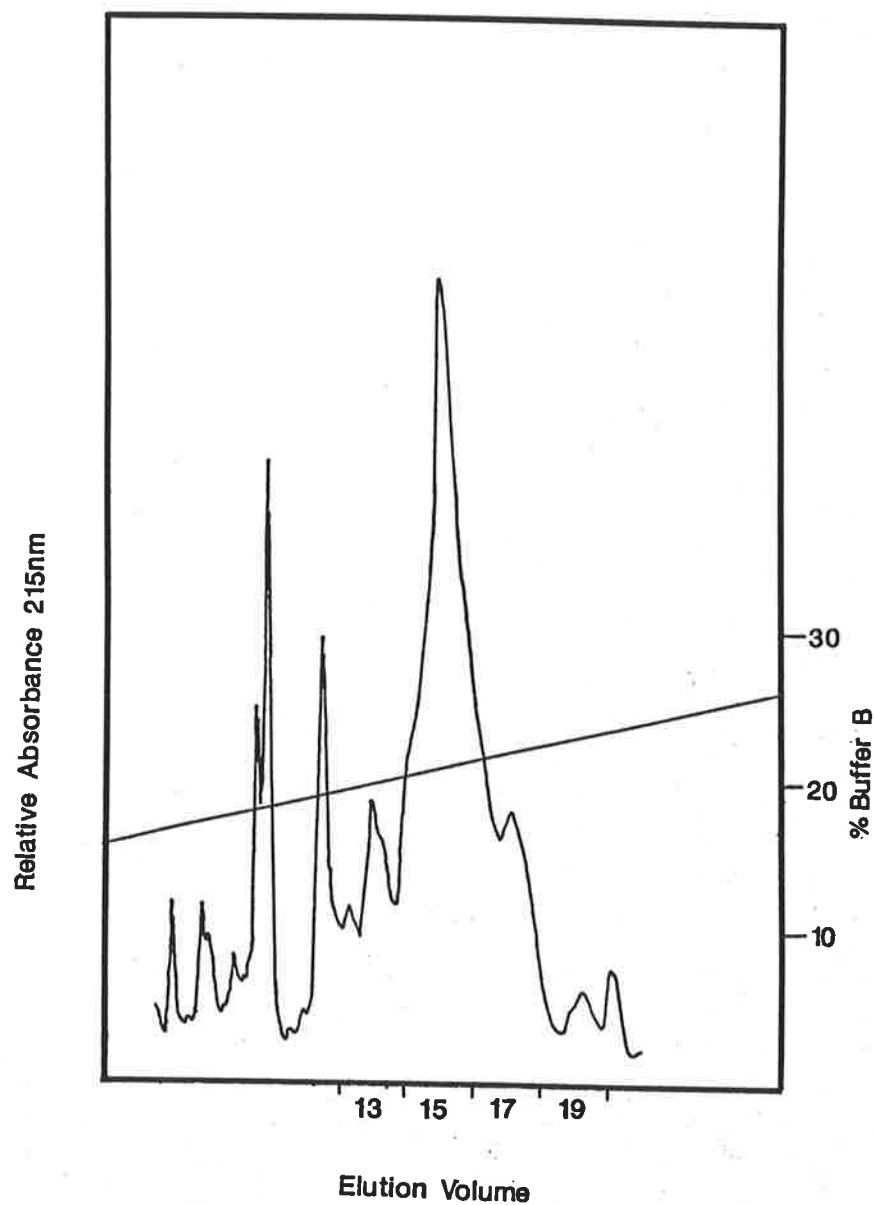


Figure 3.8

SDS PAGE of UHS Protein Peptides

The electrophoretic patterns of fractions 15 to 18 from the reversed phase fractionation of UHS protein endo-Glu-C cleavage products. The protein was separated on 15% acrylamide in the presence of SDS (2.2.1 G) and the protein visualized using "stains-all" (2.2.1 H). Track T contains the total purified UHS protein cleaved with endo-Glu-C. The track U contains a fraction of total high-sulphur protein which was not cleaved with endo-Glu-C. Tracks T and U were fractionated on different gels to the reversed phase fractions.

Sequencing was conducted on material from fractions 15, 16 and 17, and the results are presented in figure 3.9.

T

15

16

17

18

U



Figure 3.9

UHS Peptide Sequence

- A The results of sequence analysis of peptide fractions 15, 16 and 17. Where the identity of a residue is not certain the alternative residue is presented below. A consensus sequence derived by aligning the three sequences is presented. The distances between cysteine residues are indicated in the consensus sequence.
- B The pentapeptide repeat observed in the B2 and BIIIA proteins (Swart, 1973). Similar sequences to this repeat are observed in the UHS protein sequences.
- C An alignment of part of the sequence of peptide fraction 15 with part of the decapeptide repeat observed in the sheep presumptive UHS cDNA clone K4 (chapter 1, 1.6).

A: Peptide Sequence

Fraction 15

G T C C R P T C S K P C C Q T T C C Q P S S G G S S C C Q P
 K S V C A Y

Fraction 16

H T C C R P T C S K S C C Q
 T

Fraction 17

H T C C A P T C
 T

Consensus Sequence

 1 5 1 4 1 5 1 10 1
H T C C R P T C S K S C C Q T T C C Q P S S G G S S C C Q P
 C A Y

B: B2 and BIIIA High-Sulphur Protein consensus Repeat

C C R P X
 Q

where X= S, T, or V

C: K4 Repeat Comparison

K4: S S C C R P

Peptide 15: S S C C Q P

Difficulty was encountered in staining the S-carboxymethylated high-sulphur and UHS proteins after fractionation by polyacrylamide gel electrophoresis using Coomassie blue. This was most likely to be due to the high density of negatively charged carboxyl groups present on the protein after carboxymethylation. This problem was largely overcome by use of silver staining techniques, or by use of "stainsall", which carries a positive charge (Green *et al.*, 1973; Cutting, 1984).

Purification of the UHS protein family was most successful by employing a two-cycle anion exchange chromatographic fractionation. Protein purified by this method satisfied the three criteria by which the wool UHS proteins are defined (Gillespie, 1983): (a) they were present in elevated amounts in sulphur-enriched wool; (b) by two-dimensional PAGE the protein migrated as a diagonal smear; and (c) the protein fractions contained about 30 mole% cysteine.

Gel electrophoresis also revealed that none of the UHS protein fractions obtained by anion exchange chromatography contained a single component. Attempts to further purify individual UHS protein components using reversed-phase chromatography were not successful.

The similarity in the behaviour of the UHS protein components when fractionated by anion exchange and reversed-phase chromatography indicated that to purify an individual component would have required an extensive purification procedure, and indeed may have been unsuccessful, particularly when the efforts of previous workers are taken into account. Since only sufficient sequence was required to construct an oligonucleotide probe, it was decided to attempt the isolation of peptides from the UHS protein without proceeding with a more extensive purification.

Fractionation of peptides produced from the total UHS protein fraction by reversed-phase chromatography, indicated that like the parent proteins, the peptides would be difficult to separate. However, some peptide fractions obtained by this method appeared sufficiently pure to attempt sequence analysis, and some sequence was obtained from these fractions.

Inspection of the most extensive peptide sequence, obtained from fraction 15, revealed that it contains 30 mole% cysteine, and that with a single exception these residues

occur in pairs, spaced at a distance of either five or ten residues. This is consistent with there being either a penta- or decapeptide repeated sequence in the UHS protein family, similar to that observed in the B2 and BIIIA high-sulphur protein family (Elleman *et al.*, 1973; Swart, 1973; Parry *et al.*, 1979). In each case the sequence surrounding the cysteine pair closely resembles the consensus sequence determined for the repeated sequence of the B2 and BIIIA proteins (figure 3.9). Of particular interest is the observation that the region surrounding one cysteine-pair resembles the repeated amino acid sequence which is present in the conceptual protein coded by the sheep presumptive UHS cDNA clone, K4 (figure 3.9), and adds confidence to the belief that it is a UHS protein encoding clone. The next chapter of this thesis describes the successful use of this clone to isolate a UHS protein gene from the human genome. As this approach was successful, oligonucleotide probes constructed from the partial UHS protein sequence were not employed for the isolation of UHS protein cDNA or genomic clones.

3.4 Summary

The purification of a fraction of wool UHS protein has been described. This fraction satisfied the three criteria by which the UHS proteins have been defined: (a) the amount of the fraction present in the wool fibre was dependent on the supply of cysteine to the follicle; (b) the material migrated on a diagonal by two-dimensional PAGE; and (c) the protein contained about 30 mole% cysteine.

Partial sequence was obtained from peptide material produced from the purified UHS protein fraction. Analysis of this sequence revealed the presence of repeats similar to those observed in the B2 and BIIIA high-sulphur proteins, and also in the sheep presumptive UHS cDNA clone, K4.

Chapter 4

Characterization of a Human UHS Gene

4.1 Introduction

This chapter describes the isolation and characterization of a human UHS gene. This gene was isolated using the sheep putative UHS cDNA clone, K4, which has been described in chapter 1 (section 1.6). Prior to the isolation of the human gene a number of sheep genomic libraries constructed in both lambda and cosmid vectors had been screened using K4 (chapter 1, 1.6) as a probe. No positive clones were isolated from these screenings, so it was decided to attempt to isolate UHS clones from a human library.

Apart from obtaining valuable information on the UHS proteins and their genes, isolation of a human gene had two additional advantages: (a) obtaining a complete coding region for a UHS gene might facilitate the isolation of a sheep clone; and (b) it would allow cross-species comparisons to be made when a sheep gene was eventually isolated and characterized. The latter is of particular relevance in determining what features of the protein are important for its function, and may help in elucidating the mechanism by which UHS gene expression is regulated.

Given only a limited amount of peptide sequence, identification of a sequence as a UHS gene required the correlation of the product of the predicted UHS gene with that of authentic UHS protein isolated from wool and hair. Experiments to determine this for the putative human UHS gene are described in this chapter.

Finally, genomic Southern and northern transfer experiments are described in which the human gene was used as a probe.

4.2 Results

4.2.1 Cosmid Library Screening

A total of 6.0×10^5 recombinant cosmids, corresponding to five genome equivalents, were screened using a 310 bp *Pst*I fragment isolated from K4, described in chapter 1 (section 1.6), as a probe. This fragment included 65 bp from the repetitive region of the putative coding region; the remainder was from the 3' noncoding region (figure 4.1). In excess of 30 positive colonies were detected in the first round screening. After rescreening twice at low density, 16 positive colonies were purified.

Figure 4.1

K4 Sequence

The nucleotide sequence of the insert of the sheep putative UHS cDNA clone, K4 (Dr. K. Ward, unpublished observations). The 311 bp *Pst*I fragment which was used as a probe to screen the human cosmid library is underlined. It contained 65 bp of coding sequence including one complete decapeptide repeat (see chapter 1, 1.6) and the remainder of the fragment contained sequence from the 3' noncoding region.

C C V T V C C C V P A C S C S S C G K G G C S W C G C S Q S S C C R P C C S Q S
CTGCTGTGTGACCGTGTGCTGCTGTGTGCCAGCCTGCTCCTGCTCCAGCTGTGGCAAAGGGGGCTGCAGCTGGTGTGGCTGCTCCAGTCCAGCTGCTGCAGACCCTGCTGCTCCAGT 120

S C C R P C C S Q S S C C R P C C S Q S S C C V P A C C Q R R D L R C *
CCAGCTGCTGCAGACCCTGCTGCTCCAGTCCAGCTGCTGCAGACCCTGCTGCTCCAGTCCAGCTGCTGCGTCCCTGCTTGCTGTCAGCGCAGAGATCTGAGATGCTGAGCTCAGACCT 240

CAGGCGGTTCCAGCCAGTCTCAGCTCCCTGACTGTGAGCAGTGACACGTCCAGACCCCCAGGGCACACTCCATTCTCTCCAAGGTAGTGAAGTCTGAATCTTCGGGACCAGAGATCCTGT 360

TCCCACTTCTGATGGAGAACCCCTTCCCTTGTACTGCCTGAGCCATCTGAGGATTGACTGCCACTTTCAATCCAAGAGAAGTTCTTCTTGGCAGCTGAACTCCACTGCAGACGC 480

CTTAGGGGCTAAACACAGAAACCACTCAACCTCTGATGGGCACAGAGCTGTTCCAACCATTTCTTGGCTTCTTGCTGCCTAGCCTTGGTTCTCACATAACCTCAGGGCTGGTGTGACC 600

TCTCGGGCTGGGCCCCAAATAATCCTCCCCTCTTCTGTGACTCCACCTCCTTCTGATCTCGGTCTAGCAAATTCTACTCTGGGCTTGTGAAGCCGCTAAATAAATAGACTGTCTTCAAC 720

Southern transfer analysis was conducted on DNA prepared from one of these clones restricted with various enzymes, and the K4 310 bp *Pst*I fragment was used as a probe (figure 4.1). The salient feature of this blot is that with all the restriction enzymes used at least two restriction fragments from the cosmid hybridized strongly to K4 (figure 4.2). This indicates that there is probably more than one UHS gene in the cosmid insert. The 1.0 kb *Eco*RI fragment was isolated and subcloned into a pTZ18 *Eco*RI vector.

4.2.2 Sequence of the 1.0 kb *Eco*RI Fragment

The strategy used to determine the sequence of the 1.0 kb *Eco*RI fragment is summarized in figure 4.3. The sequencing of this fragment was conducted by subcloning restriction fragments and BAL31 deletion fragments into M13 vectors. Full details of the sequencing strategy appear in Appendix A. The complete sequence of the 1.0 kb *Eco*RI fragment appears in figure 4.4.

Conceptual translation of the nucleotide sequence of this fragment in all six frames revealed an ATG codon that was followed by an open reading frame of 168 codons, of which 36% encoded cysteine (figure 4.4). This observation along with the presence of; (a) a TATA and possibly a CCAAT box, and (b) the Kozak sequence surrounding the initial ATG codon, strongly suggest that a UHS gene is present in the 1.0 kb *Eco*RI fragment (see discussion).

4.2.3 Gene Product Analysis

The electrophoretic mobility by one- and two-dimensional PAGE was determined for the putative human UHS gene translation product.

The protein product of the gene was synthesized by *in vitro* transcription of a DNA template followed by *in vitro* translation. A clone for *in vitro* transcription was constructed by subcloning into a pGEM-2 *Sma*I vector, a fragment in which the 5' flanking sequences had been deleted using BAL31 exonuclease, to a position 74 bp upstream from the putative ATG initiation codon. This fragment contained no ATG codons prior to the start codon of the cysteine-rich open reading frame.

Figure 4.2

Southern Blot Analysis of the Selected Cosmid

Cosmid DNA was prepared (2.2.2 E) and digested (2.2.2 F) using the restriction enzymes indicated. The digestion products were electrophoresed through 1% agarose (2.2.2 H). A *Hind*III digestion of lambda DNA and an *Eco*RI digestion of SPP-1 DNA (both obtained from Bresatec) were co-electrophoresed as marker DNA. After staining and photography, the DNA fragments were transferred to Zetaprobe (2.2.2 O) and hybridization conducted (2.2.2 O) using the 311 bp fragment from the sheep presumptive UHS cDNA clone (section 4.2.1) as a probe. No difference was observed in the hybridization pattern after washing at low stringency (2.0 x SSPE, 0.1% SDS, 65°C) and at high stringency (0.2 x SSPE, 0.1% SDS, 65°C). An autoradiograph after washing at high stringency is presented.

Tracks photographed prior to transfer are paired with the corresponding hybridization pattern. The sizes of fragments to which the probe hybridized are given in kilobases (kb).

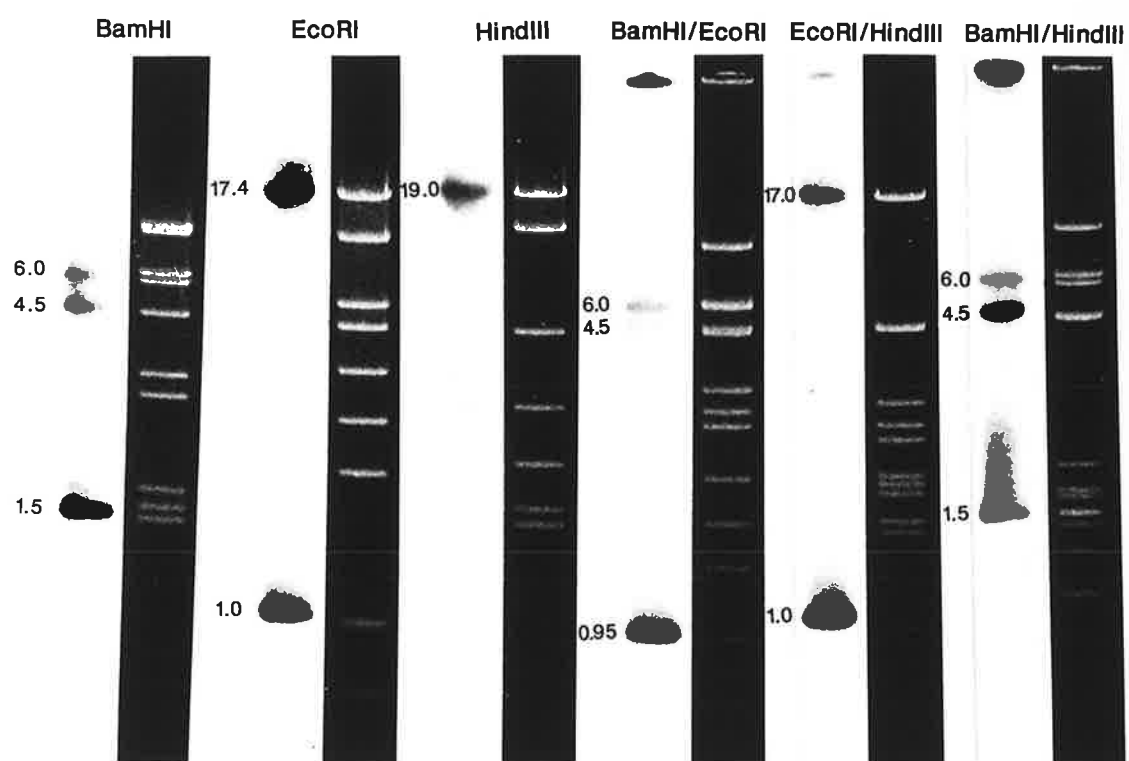
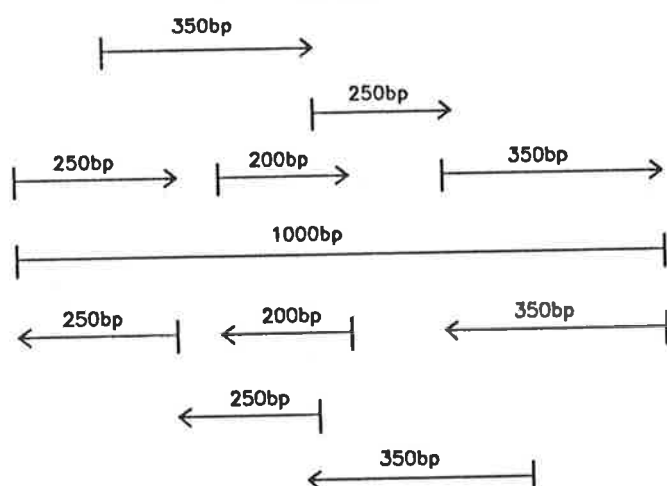


Figure 4.3

Sequencing Strategy

A summary of the strategy used to sequence the 1.0 kb *Eco*RI fragment. The M13mp18/19 subclones from which the sequence was obtained are indicated in the left hand margin. The details of the subcloning strategy are presented in appendix A. The 1.0 kb *Eco*RI fragment was sequenced completely in both orientations.

SEQUENCING STRATEGY



SUBCLONES USED

- (5) Bal31
- (3) Mbol/HindIII
- (2) PvuII
- (2) PvuII
- (3) Mbol/EcoRI
- (4) Bal31

Figure 4.4

Sequence of the 1.0 kb *EcoRI* fragment

The DNA sequence of the 1.0 kb *EcoRI* fragment. The derived protein sequence of the UHS gene appears above the DNA sequence. The conceptual protein is 168 residues in length and contains 36 moles% cysteine. The possible CCAAT homologies in the 5' flanking region (4.3.1) are underlined. The TATA homology is boxed. The Kozak homology surrounding the presumptive methionine initiation codon (4.3.1) is underlined. Possible CAP sites are as indicated.

The three repeats found in the coding region of this gene (4.3.5) have been outlined as follows:

- Red: Cysteine-rich decapeptide repeats.
- Green: Serine-rich nonapeptide repeats.
- Blue: Glycine-rich repeats.

The arrow 74 bp upstream from the methionine initiation codon indicates the first base of the fragment that was cloned into pGEM-2 for *in vitro* transcription experiments (4.2.3).

AATTCCACAAGGAAATCATCTCAGGAGGAAGGGCTCATACTTGGATCCAGAAAATATCAACATAGCCAAAGAAAAACAATGACATACCTCCAGGAGCTGTGTAACAGCAACCGGAAAGAG 120

CAP Site

AAACAATGGTGTGTTCTATGTGGGATATAAAGAGCCGGGGCTCAGGGGGCTCCACACCTGCACCTCCTTCTCACCTGCTCCTCTACCTGCTCCACCTCAATCCACGAAACCATGGGC 240

C C G C S G G C G S S C G G C D S S C G S C G S G C R G C G P S C C A P V Y C C
TGCTGTGGCTGCTCCGGAGGCTGTGGCTCCAGCTGTGGAGGCTGTGACTCCAGCTGTGGGAGCTGTGGCTCTGGCTGCAGGGGCTGTGGCCCCAGCTGCTGTGCACCCGCTACTGCTGC 360

K P V C C C V P A C S C S S C G K R G C G S C G G S K G G C G S C G C S Q C S C
AAGCCCGTGTGCTGTGTGTTCCAGCCTGTTCTGCTCTAGCTGTGGCAAGCGGGGCTGTGGCTCCTGTGGGGGCTCCAGGGAGGCTGTGGTTCTTGTGGCTGCTCCAGTGCAGTTGC 480

C K P C C C S S G C G S S C C Q C S C C K P Y C S Q C S C C K P C C S S S G R G
TGCAAGCCCTGCTGTTGCTCTTCAGGCTGTGGGTATCCTGCTGCCAGTGCAGCTGTGCAAGCCCTACTGCTCCAGTGCAGCTGTGTAAGCCCTGTTGCTCCTCCTCGGGTCGTGGG 600

S S C C Q S S C C K P C C S S S G C G S S C C Q S S C C K P C C S Q S R C C V P
TCATCCTGCTGCCAATCCAGCTGTGCAAGCCCTGCTGCTCACTCAGGCTGTGGGTATCCTGCTGCCAGTCCAGCTGTGCAAGCCCTGCTGCTCCAGTCCAGATGCTGTGTCCT 720

V C Y Q C K I *
GTGTGCTACCAGTGCAAGATCTGAGGCTCTAGTGGGAAACCTCAGGTAGCTCCCGAAGATCTGTGCTTTCCAACAAGTGACTACCTTGAAGCACATCCCCTTCTGAATCTGAAAAGAGC 840

CTTGGCTCAGGGCGTCTTTTCCAGCCCTGAGGAAATGGAATGAACCACTCCCTGCCCATTCCCTATAAGAATATCCAAGACCCAGGCAATTTTGCCCTCTTTCCACATGCCCCCA 960

TATGCTGAGCCAACTGCACTGGGGGCTGCCTAGTCCAAGCAAGAGCCTGGGAATT 1015

The orientation of this insert was determined by cleavage with *Pst*I which cuts in the pGEM-2 polylinker and once in the insert (figure 4.5). One subclone (called pGEM-HUHS) was selected in which the fragment was oriented with the 5' end of the open reading frame adjacent to the SP6 promoter of the vector. DNA was prepared (section 2.2.2 E) from this clone and was linearized by cleavage with *Hind*III to allow transcription with SP6 RNA polymerase (figure 4.5).

RNA transcripts were prepared as described in section 2.2.3 D. *In vitro* translation of these transcripts was conducted using rabbit reticulocyte lysate as described in section 2.2.3 E. The translation products were labelled using [³H]serine, since this amino acid is also abundant (23 mole%) in the predicted protein, and unlike cysteine has a relatively inert side group.

The translation products in the sulphydryl form and after S-carboxymethylation were fractionated using SDS PAGE (figure 4.6). The carboxymethylated form was fractionated by two-dimensional PAGE (figure 4.7). As a comparison, total wool and hair keratin were fractionated by two-dimensional PAGE after S-carboxymethylation using [¹⁴C]iodoacetic acid (figure 4.7).

SDS PAGE reveals that the sulphydryl and SCM forms of the translation product of the human UHS gene have differing electrophoretic mobilities. The sulphydryl form has an apparent MW of approximately 20×10^3 , and the SCM form 30×10^3 . Possible reasons for this will be discussed below. Two-dimensional PAGE of the SCM form of the human UHS gene product reveals that it comigrates on the diagonal upon which it has been observed that authentic UHS protein migrates (Gillespie *et al.*, 1982).

4.2.4 Southern Transfer Analysis

Genomic Southern blot analysis was conducted using the human UHS gene as a probe in order to determine: (a) the number of related genes in the human and sheep genomes; and (b) the relationship of the UHS genes to the high-sulphur B2 subfamily.

Sheep and human DNA were digested using both *Eco*RI and *Hind*III prior to fractionation on 1% agarose and transfer to "Zetaprobe" (2.2.2 O). The complete 1.0 kb *Eco*RI fragment, excised from the pTZ18 vector, was used as a probe. The filter was

Figure 4.5

UHS gene *In Vitro* Transcription

A map of the pGEM-2 clone (pGEM-HUHS) described in section 4.2.3, which was used for *in vitro* transcription. The thick line represents vector DNA and continues in both directions to form a closed circle. S indicates the vector *Sma*I site into which the fragment was inserted. The orientation of the fragment was determined using the internal asymmetric *Pst*I site. The initiation codon of the presumptive coding region was adjacent to the SP6 promoter.

A template for transcription was prepared from this clone by cleavage at the *Hind*III site in the pGEM-2 polylinker. Run-off transcripts were synthesized in the direction indicated by the arrow, using SP6 polymerase (2.2.3 D), and these were used in *in vitro* translation experiments (4.2.3).

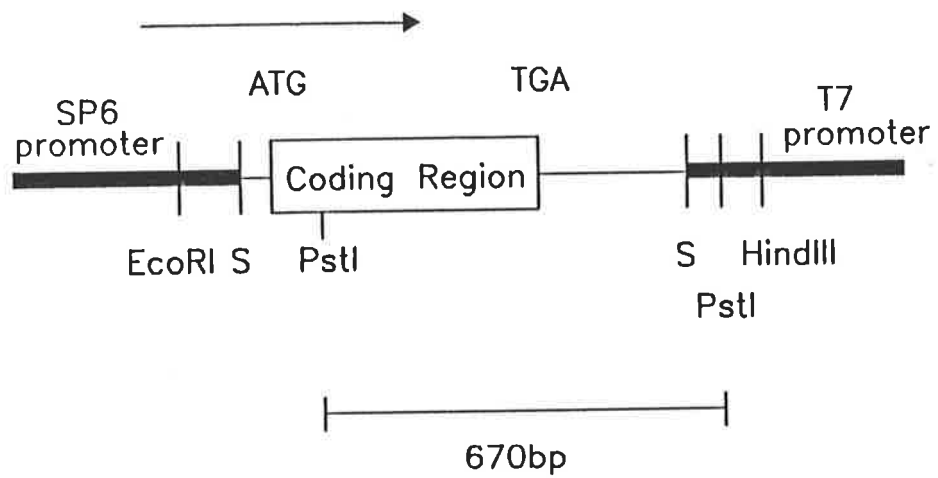


Figure 4.6

SDS Polyacrylamide Gel Electrophoresis of Human UHS Gene *In Vitro* Translation

Product

SDS PAGE was conducted on the sulphhydryl (-SH) and SCM (derivatized) forms of the human UHS gene *in vitro* translation product. The size of MW markers co-fractionated with the *in vitro* translation products are indicated on the right-hand side. The sulphhydryl form had an apparent MW of 20,000 and the SCM form 30,000.

Cys1 was a yeast gene *in vitro* transcription and translation product prepared by the same method as the UHS gene product (Gift of Dr. R. D'Andrea), and was included as a positive control.

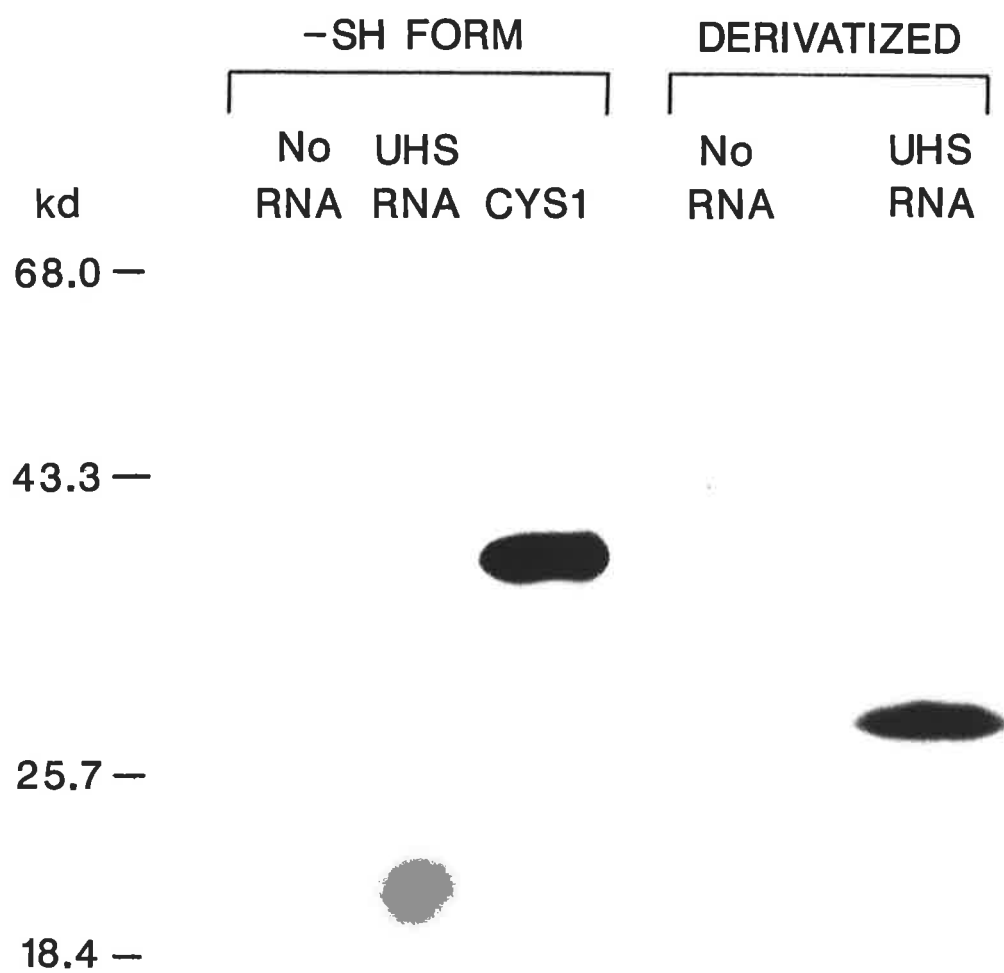


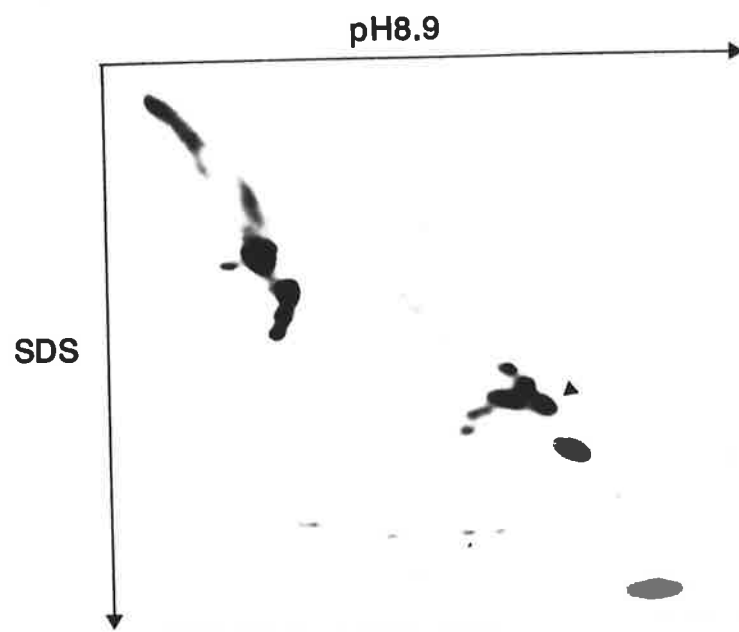
Figure 4.7

Two-Dimensional Polyacrylamide Gel Electrophoresis

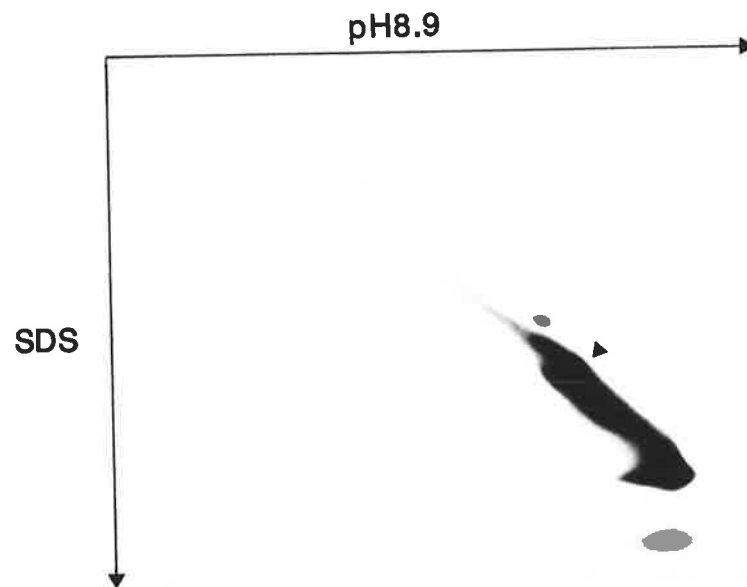
- A** Total wool protein was reduced and S-carboxymethylated using [^{14}C]iodoacetic acid (2.2.1 E) and fractionated by two-dimensional PAGE (2.2.1 F). The proteins were visualized by fluorography (2.2.1 I). The arrow indicates the position of the human UHS gene *in vitro* transcription/translation product.
- B** Total human hair protein isolated, alkylated and fractionated as described above. The arrow indicates the position of the human UHS gene *in vitro* transcription/translation product.
- C** The human gene *in vitro* transcription product which was produced as described in the text (4.2.3) was fractionated by two-dimensional PAGE (2.2.1 F) and was visualized by fluorography (2.2.1 I).

Electrophoretic fractionation of all three samples was conducted in the same apparatus under identical electrophoretic conditions so that the patterns of each were directly comparable to the others.

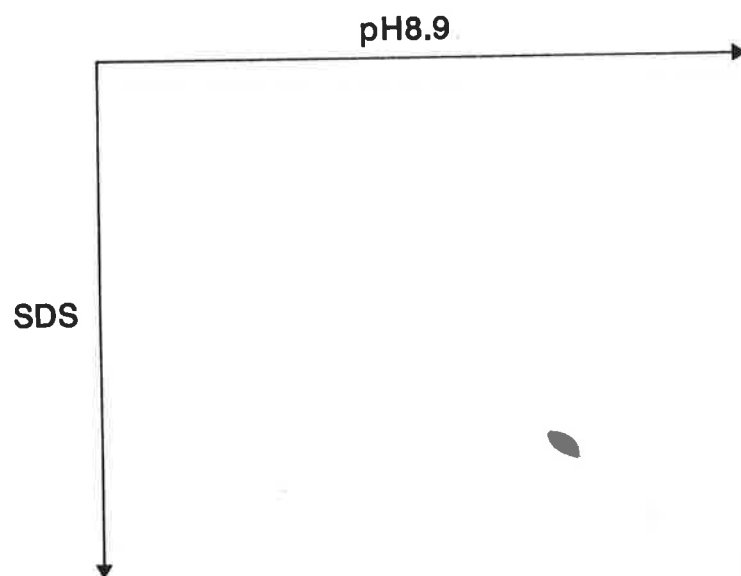
A



B



C



autoradiographed after washing at low stringency, and again after washing at high stringency (figure 4.8).

The same filter was probed using a DNA fragment from the coding region of the high-sulphur B2A gene. This fragment included sequences coding for the amino acid repeat observed in this family (Elleman, 1973; Powell and Rogers, 1983). The filter was autoradiographed after washing at low stringency, and again after washing at high stringency (figure 4.8).

Southern transfer indicated that there were a number of *Eco*RI and *Hind*III fragments in both the human and sheep genome that hybridized to the human UHS gene probe, indicating the presence of a number of related sequences in both genomes. The high-sulphur B2A probe did not appear to cross-hybridize to any of the fragments to which the UHS gene hybridized. This indicates that there is insufficient homology between the two genes to allow significant cross-hybridization, and also that the UHS and B2 genes do not reside on the same *Eco*RI and *Hind*III restriction fragments.

The pattern of hybridization for both probes was identical after washing at low and high stringency.

4.2.5 Northern Transfer Analysis

To determine whether or not a gene closely related to the human UHS gene was expressed in the wool follicle, northern transfer experiments were conducted using total wool follicle RNA.

The RNA blots were probed using the complete 1 kb *Eco*RI fragment containing the human UHS gene. As comparisons, equivalent RNA blots were probed with: (a) the K4 *Pst*I fragment used to screen the human cosmid library; and (b) the high-sulphur fragment containing the repeat as described in 4.2.4. All blots were autoradiographed firstly after washing at low stringency and again after washing at high stringency. The results of northern transfer analysis are presented in figure 4.9.

Northern transfer indicated that the human UHS gene cross-hybridized to wool follicle mRNA of approximately 1.5 kb, and that the sheep presumptive cDNA clone K4 hybridized to mRNA of the same size. The B2A probe hybridized to mRNA that was

Figure 4.8

Southern Analysis of Human and Sheep Genomic DNA

Human and sheep genomic DNA was restricted with *Eco*RI (E) and *Hind*III (H), fractionated by electrophoresis through 1% agarose (2.2.2 H), and transferred to Zetaprobe (2.2.2 O). Transferred DNA was initially hybridized to a fragment containing the human UHS gene coding region (4.2.4), and autoradiography was conducted after the filter was washed at low stringency (2.0 x SSPE, 0.1% SDS; 65°C) and is presented here. Rewashing at high stringency (0.2 x SSPE, 0.1% SDS; 65°C) revealed no changes in the hybridization pattern.

After removal of the human UHS gene probe (2.2.2 O) the same filter was hybridized to a high-sulphur (HS) B2A coding region probe (4.2.4) and washed under the same conditions as described above.

The estimated sizes of the positive fragments are given in kilobases (kb).

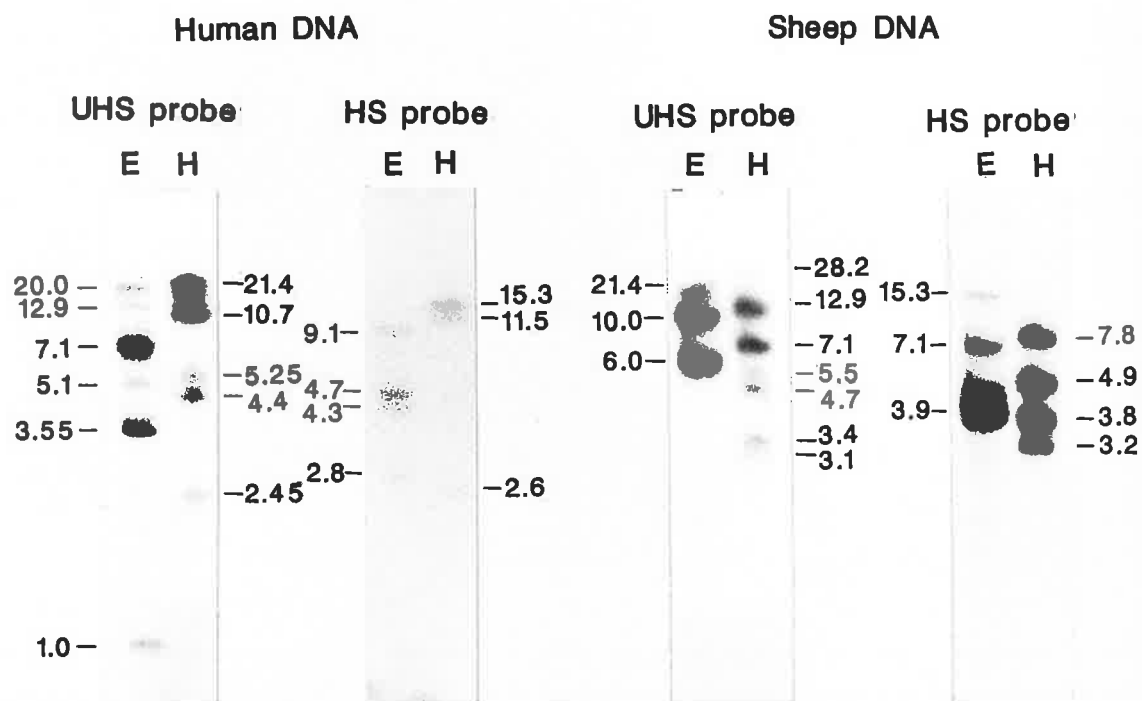


Figure 4.9

Northern Transfer of Total Wool Follicle RNA

Total wool follicle RNA was prepared (2.2.3 A) and fractionated through 1.5% agarose after glyoxylation (2.2.3 B). RNA was transferred to Zetaprobe (2.2.3 B) and equivalent tracks were hybridized as follows:

- A** A coding region probe from the B2A gene (Gift of Dr B.C. Powell).
- B** The 310 bp *Pst*I fragment from the sheep putative UHS cDNA clone K4 which was used to isolate the cosmid clone.
- C** the 1.0 kb *Eco*RI fragment containing the human UHS gene coding region.

Autoradiographs after washing at high stringency (0.2 x SSPE, 0.1% SDS; 65°C) are presented and were indentical to autoradiographs after washing at low stringency (2.0 x SSPE, 0.1% SDS; 65°C).

The sizes of the bands were determined relative to the positions of the 18S and 28S rRNA species.

A B C

1.2



1.5



approximately 1.1 kb in size. The pattern of hybridization for both probes was identical after washing at both low and high stringency.

4.2.6 Chromosomal *In Situ* Hybridization

In situ hybridization was conducted by Mrs. E.G. Baker and Dr. G.R. Sutherland (Adelaide Childrens Hospital) with the human karyotype. The complete *Eco*RI fragment containing the human UHS gene cloned into pTZ18 was used as a probe. Hybridization was found to occur at two positions on chromosome 11: 11p15 and 11q13-q14 (figure 4.10).

4.3 Discussion

The isolation of a human cosmid using the sheep cDNA clone K4 revealed a fragment which contained a methionine codon followed by an open reading frame of 168 codons of which 36% coded for cysteine (figure 4.4). A number of lines of evidence implicate the cysteine-rich open reading frame as being the complete coding region of a functional UHS gene. The evidence to support this belief is discussed below.

4.3.1 5' Flanking and Noncoding Regions

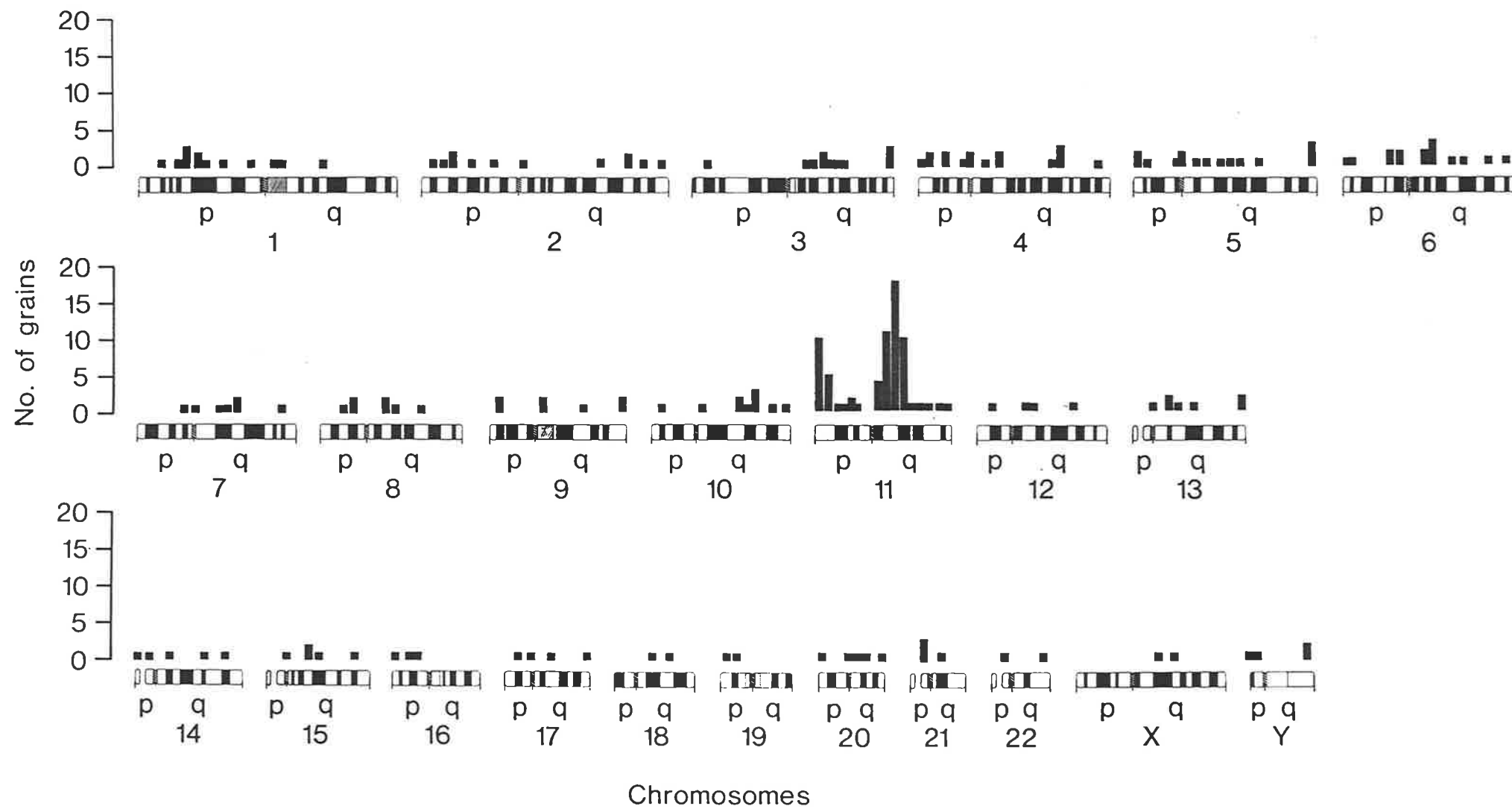
When a large number of eukaryotic genes from a number of different species are compared similarities are observed in the region surrounding the site of translation initiation (Kozak, 1981 and 1984; Cavener, 1987). Kozak (1984) compared 211 messenger RNAs from higher eukaryotic cells and arrived at the CC(A/G)CCAUG(G) consensus sequence, where the underlined residues correspond to the initiation codon and A and G are alternatives at the -3 position. Cavener (1987) has derived the CANCAUG consensus sequence specifically for vertebrate messenger RNA molecules. Kozak (1986), by observing the effect upon translation of mutations in this region of the preproinsulin gene, identified ACCATGG as being the optimal sequence for translation, with the presence of a purine at the -3 position having a dominant effect.

Figure 4.10

Chromosomal *In Situ* Hybridization

In Situ hybridization was conducted by Mrs. E.G. Baker and Dr. G.R. Sutherland (Adelaide Childrens Hospital), using the 1.0 kb *Eco*RI fragment including the UHS gene coding region.

The positions of silver grains over all chromosomes were determined for 26 metaphases. Lines above the chromosomes indicate the number of silver grains that were observed in that position. DNA sequences that hybridized to the UHS gene coding region were localized to 11p15 and 11q13-q14.



The human UHS gene contains the GAACCATGG sequence surrounding the putative initiation codon. This conforms well with the consensus sequences of both Kozak and Cavener, particularly with the presence of an A in the -3 position.

A possible TATA box is located about 80 bp upstream from the initiation codon (for review see Breathnach and Chambon, 1981). The TATAAA homology occurs in almost every gene which has been analyzed (Nevins, 1983) and *in vitro* mutation experiments have revealed that it is required for accurate initiation of transcription at a specific start point (Corden *et al.*, 1980; Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Breathnach and Chambon, 1981; Ghosh *et al.*, 1981; Shenk, 1981; McKnight and Kingsbury, 1982).

Another homology which has been observed in the 5' flanking regions of some genes 70 to 80 bp upstream from the site of transcription initiation is the 5' CCAAT 3' sequence (Benoist *et al.*, 1980; Efstratiadis *et al.*, 1980). Three possible CCAAT homologies occur 20, 40 and 70 bp upstream from the TATA box (figure 4.4).

Apart from the TATA and CCAAT boxes, the 5' flanking region appears to lack any other well recognized sequence motifs which are known to be involved in the regulation of transcription of some genes by RNA polymerase II, such as the SPI binding site (Dyner and Tjian, 1985; Briggs *et al.*, 1986), or the SV40 enhancer core sequence (for review see Khoury and Gruss, 1983). It is possible that in common with many other eukaryotic genes, this gene lacks these sequences. Alternatively, it is possible that regulatory elements may lie beyond the region of this gene which has been sequenced. This is especially true for enhancer core elements, since enhancing elements have been shown to exert an influence on transcription at considerable distance (>10 kb) (Khoury and Gruss, 1983).

4.3.2 3' Noncoding Region

An AATAAA sequence is observed 10 to 30 bp upstream from the polyadenylation site of most genes (Proudfoot and Brownlee, 1976) and is required for polyadenylation of the nascent transcripts (for reviews see Birnstiel *et al.*, 1985 or Platt, 1986). The AATAAA sequence is absent from the sequenced 3' noncoding region of the human UHS

gene. The most likely explanation for the lack of this sequence is that it lies beyond the *EcoRI* site which marks the boundary of sequence analysis of this gene, although it is possible that the gene lacks this signal.

4.3.3 Lack of Introns

It is very likely that the 168 bp open reading frame is the complete coding region of this gene, and hence the gene doesn't contain any introns. Without exception, the wool matrix genes that have been sequenced do not contain introns (Powell and Rogers, 1983; Kuczek and Rogers, 1987; Frenkel *et al.*, 1988). This is in marked contrast to the genes that code for the wool intermediate filament proteins, which do contain introns (Powell *et al.*, in preparation; Ward *et al.*, 1988).

The most direct evidence that this gene doesn't contain introns comes from the observation that the open reading frame codes for a protein product that when produced by *in vitro* transcription comigrates with the bulk of the high-sulphur protein isolated from wool and hair, and has a molecular weight within the range that has been observed for the UHS proteins (see section 4.3.6). Also, if introns do occur in this gene and the sequence obtained merely contains the first exon of a longer coding region, then the (C/A)AG:GT(A/G)AGT consensus donor splice site sequence (Mount, 1982) should occur near the 3' end of the cysteine-rich open reading frame. Examination of the sequence in this region failed to reveal the consensus donor sequence.

4.3.4 Conceptual Protein Analysis

The conceptual protein product of this gene contains 35.5 mole% cysteine (table 4.1). This level is in excess of estimates made on fractions of UHS isolated from wool (Gillespie and Broad, 1972). There could be three reasons for this: (a) it is possible that the UHS protein fraction was heterogeneous and contained high-sulphur protein of lower cysteine content; (b) cysteine may have been degraded during extraction and/or during amino acid analysis of the protein fraction; and (c) there could be specific differences, with human UHS protein having a higher cysteine content than sheep UHS protein.

Table 4.1

Amino Acid Content of Human Gene

Presented are the proportions of the various amino acids, expressed as a percentage of the total number of residues in: (a) the complete human UHS conceptual protein (Total); (b) the first 77 amino acids from the N-terminus of the conceptual protein (N); and (c) the remaining 91 residues towards the C-terminus (C). The first column contains the results of amino acid analysis of a fraction of wool UHS protein (Gillespie and Broad, 1972).

Amino Acid	Protein	Human Gene		
		Total	N	C
Ala	2.0	1.2	2.6	0
Arg	6.9	2.4	2.6	2.2
Asn	-	0	0	0
Asp	0.6	0.6	1.3	0
Cys	29.9	35.5	32.0	39.0
Gln	-	4.1	0	7.5
Glu	7.9	0	0	0
Gly	4.2	16.0	27.6	6.5
His	1.3	0	0	0
Ile	1.7	0.6	0	1.1
Leu	1.4	0	0	0
Lys	0.9	5.0	4.0	6.5
Met	0	0	0	0
Phe	0.5	0	0	0
Pro	12.8	6.0	5.3	6.5
Ser	12.7	23.1	18.4	27.0
Thr	11.1	0	0	0
Trp	-	0	0	0
Tyr	1.8	1.8	1.3	2.2
Val	4.3	3.0	4.0	2.2

Table 4.1 reveals that the content of the remaining amino acids is quite distinct from that observed for the UHS fractions of wool. There are elevated levels of serine, glycine and lysine and decreased levels of proline and arginine in the human UHS gene. Asparagine and threonine, which constitute 3.3 and 11.1 mole% respectively of the wool UHS protein fraction, are absent from the conceptual human protein (table 4.1). These differences may be accounted for by specific variation between human and sheep UHS protein, however a more likely reason is the first given for the discrepancy for cysteine content, that is, the protein fractions analyzed were heterogeneous and represented the average amino acid content of many different components.

4.3.5 Amino Acid Repeats in the Conceptual Protein

The amino acid content of the conceptual protein product of this gene is polarized, with the first 80 residues from the N-terminal region containing higher levels of glycine and lower levels of cysteine and serine than the remaining 90 residues from the C-terminal region (table 4.1). Inspection of the C-terminal region of the conceptual protein reveals that the reason for the polarity is that it is composed entirely of two distinct reiterated sequences, both with a high cysteine and serine content (figure 4.4). One of these (called serine-rich), present in three positions in the C-terminal region, is nine residues long, and with the exception of a single arginine solely contains serine, cysteine and glycine (figure 4.11). The other (called cysteine-rich) is ten residues in length and contains high levels of cysteine, but doesn't contain glycine (figure 4.11). The two repeats are intermingled, with the serine-rich nonapeptide repeats being separated from each other by either one or two cysteine-rich repeats.

Examination of the nucleotide sequence of each of the reiterated sequences reveals a high degree of homology. When the three occurrences of the serine-rich repeated nonapeptide are compared in a pairwise fashion the nucleotide homology ranges from 90 to 97%, with an average of 91%, (figure 4.11). Of the four positions at which base changes occur when the three repeats are compared, three occur in the third position of the corresponding codon and do not result in an amino acid substitution. The fourth base change occurs in the first position of the codon and results in the substitution of arginine

Figure 4.11

Amino Acid Repeats in Human Conceptual UHS Protein

- A** An alignment of the three serine-rich nonapeptide repeats present in the human UHS gene conceptual protein (see text, 4.3.5). They are presented in the order in which they occur in the protein commencing from the N-terminus (see figure 4.4).

Below are the corresponding nucleotide sequences coding for the repeats and presented in the same order. Positions at which base substitutions occur are indicated by stars (*), and where the base substitutions result in amino acid substitutions by a vertical line (|).

- B** An alignment of the six cysteine-rich decapeptide repeats present in the sheep human conceptual protein presented in the order that they occur in the protein commencing from the N-terminus (see figure 4.4).

Below is an alignment of the nucleotide sequences coding for the cysteine-rich repeats. Positions at which base substitutions occur are indicated as described above.

A

Serine-Rich Nonapeptide Repeat

S	S	G	C	G	S	S	C	C
S	S	G	R	G	S	S	C	C
S	S	G	C	G	S	S	C	C

Nonapeptide Repeat Nucleotide Sequence

TCT	TCA	GGC	TGT	GGG	TCA	TCC	TGC	TGC
TCC	TCG	GGT	CGT	GGG	TCA	TCC	TGC	TGC
TCC	TCA	GGC	TGT	GGG	TCA	TCC	TGC	TGC
	*	*	*					

B

Cysteine-Rich Decapeptide Repeat

Q	C	S	C	C	K	P	C	C	C
Q	C	S	C	C	K	P	Y	C	S
Q	C	S	C	C	K	P	C	C	S
Q	S	S	C	C	K	P	C	C	S
Q	S	S	C	C	K	P	C	C	S
Q	S	R	C	C	V	P	V	C	Y

Decapeptide Repeat Nucleotide Sequence

CAG	TGC	AGT	TGC	TGC	AAG	CCC	TGC	TGT	TGC
CAG	TGC	AGC	TGC	TGC	AAG	CCC	TAC	TGC	TCC
CAG	TGC	AGC	TGC	TGT	AAG	CCC	TGT	TGC	TCC
CAA	TCC	AGC	TGC	TGC	AAG	CCC	TGC	TGC	TCA
CAG	TCC	AGC	TGC	TGC	AAG	CCC	TGC	TGC	TCC
CAG	TCC	AGA	TGC	TGT	GTC	CCT	GTG	TGC	TAC
	*		*		*		*	*	*

for cysteine in the second repeat (figure 4.11). This is the only amino acid substitution which occurs within the three iterations of this sequence.

The nucleotide homology of the cysteine-rich decapeptide repeat ranges from 60 to 93% with an average homology of 79% (figure 4.11). The last cysteine-rich repeat contiguous with the C-terminus is more divergent than the remaining repeats, and if excluded from comparisons the average nucleotide homology between the five remaining iterations rises to 88%. The C-terminal iteration possibly belongs to a different family of cysteine-rich repeats which are observed in the sheep UHS gene (see chapter 6). Base changes occur at nine positions when the five more conserved iterations of this sequence are compared. Six of the base changes occur in the third codon position, and do not result in a change in the peptide sequence. The remaining three base changes occur in the second position of the corresponding codons and result in the three amino acid substitutions observed in the five repeats. Two of the substitutions are between serine and cysteine. Serine, whilst having a physically similar side group to cysteine, cannot form disulphide linkages. The remaining substitution is the aromatic amino acid tyrosine for cysteine.

The cysteine-rich decapeptide repeat is homologous at both the amino acid and nucleotide level to the repeat observed in the high-sulphur proteins (Elleman 1973; Powell and Rogers, 1983) and to similar repeats observed in the cDNA clone K4, a sheep UHS gene (to be described in chapter 5) and a mouse putative UHS gene. The relationships between these various genes will be discussed in chapter 6.

The 80 residues contiguous with the N-terminus of the conceptual protein do not contain the striking iteration that is present in the C-terminal region, and the primary structure of this region appears to be irregular. However some repetition is observed, with a peptide sequence of about ten residues, which contains about 40% glycine, being repeated three times in this region (figure 4.12). Comparison of the nucleotide sequences coding for these sequences reveals about 80% homology. These sequences are homologous to a highly conserved glycine-rich decapeptide repeat present in the sheep UHS gene coding region to be described in the next chapter (5.3.3).

Figure 4.12

Glycine-Rich Repeats

An alignment of the glycine-rich repeats observed in the N-terminal region of the human UHS gene conceptual protein (4.3.5). A deletion in each of the last two iterations has been introduced to maximize homology. The repeats are presented in the order they occur commencing from the N-terminus (see figure 4.4). The corresponding nucleotide sequences are presented below.

Glycine-Rich Sequences

S	G	G	C	G	S	S	C	G	G	C
K	R	G	C	G	S		C	G	G	S
K	G	G	C	G	S		C	G	C	S

Nucleotide Sequence

TCC	GGA	GGC	TGT	GGC	TCC	AGC	TGT	GGA	GGC	TGT
AAG	CGG	GGC	TGT	GGC	TCC		TGT	GGG	GGC	TCC
AAG	GGA	GGC	TGT	GGT	TCT		TGT	GGC	TGC	TCC

Unlike the C-terminal region where 70% of cysteine residues occur in pairs, the distribution of this amino acid throughout the N-terminal peptide sequence appears to be irregular.

The implications of the iterated sequence of the conceptual protein product of this gene upon its function as a structural protein of the hair fibre will be discussed along with similar iterations observed in the protein product of a sheep UHS gene in chapter 6.

4.3.6 *In Vitro* Translation

When the *in vitro* translation product of the human UHS gene is fractionated by SDS PAGE (figure 4.6), there is a wide discrepancy in the rate of migration between the sulphydryl and SCM forms of the protein. The large negative charge conferred on the protein by the carboxymethyl moiety may reduce the affinity of the SCM form of the protein for SDS. This would reduce the migration rate of the SCM form through the gel and result in its larger apparent molecular weight. Similar observations have been made with the high-sulphur proteins of lower cysteine content, where molecular weight estimations using SDS PAGE have been in error by up to 150% (Marshall and Gillespie, 1988).

The sulphydryl form resolves poorly, with extensive smearing of labelled material being evident in the track containing this form of the protein. This is not observed when the sulphydryl groups are protected by carboxymethylation, indicating that it is oxidation of cysteine during migration through the gel which causes the poor resolution.

When fractionated by two-dimensional PAGE the *in vitro* translation product of this gene comigrates with a high-sulphur protein component of both human hair and wool (figure 4.7). The precise positioning of the various high-sulphur components in the pattern observed after two-dimensional PAGE is equivocal, however it is believed that most of the UHS fraction migrates as an unresolved diagonal smear (Gillespie, 1983; Gillespie *et al.*, 1982; also see chapter 3). The *in vitro* translation product resides in the lower regions of this diagonal. The minimum size of the UHS proteins is not certain but Buchanan (1977) found that high-sulphur protein with a cysteine content which approximates that of the UHS subfamily occurs in hair with a molecular weight of less than 20,000. Also Marshall

and Gillespie (1976 a, b) found that mouse hair contained UHS protein components with an apparent molecular weight of 16,500. Thus the components of the UHS family have very diverse molecular weights extending from at least a minimum of 16,000 for this protein to in excess of 37,000 (Lindley *et al.*, 1968).

All of the UHS protein may be derived from individual genes, or it is possible that the higher molecular weight material results from cross-linking of shorter components during keratinization or during extraction of the protein from the wool fibre.

In favour of the latter hypothesis is the observation that under the conditions used to extract protein from the wool fibre irreversible cross-links may form between cysteine residues (Thompson and O'Donnell, 1962). The UHS proteins, with their high cysteine content, would be particularly susceptible to the formation of such cross-links. It is also possible that some of the cystine residues are resistant to reduction, resulting in the presence of intermolecular disulphide linkages in the extracted protein. Ley and Crewther (1980) suggest that the presence of hydrophobic amino acids may result in the resistance of some disulphide linkages to reduction. Hydrophobic residues do occur in close proximity to some of the cysteine residues of both this protein and the sheep UHS protein to be discussed in the next chapter.

An alternative possibility is that γ -glutamyl- ϵ -lysine cross-links may form between these proteins as part of the natural process of keratinization. Such cross-links are known to form between keratin proteins of the medulla and inner-root-sheath (Rogers and Harding, 1976) and also the fibre cuticle (Ley and Crewther, 1980). However, whilst the conceptual protein contains both glutamine and lysine in the 10 residue cysteine-rich repeat, lysine is present at very low levels in the total fractions of wool UHS protein (Gillespie and Broad, 1972). Glutamine levels in the UHS protein fractions is unknown as this amino acid is degraded during protein hydrolysis.

If any of the possible modifications suggested above do in fact occur, it is unlikely that the *in vitro* translation product of a particular UHS gene will comigrate with its authentic product isolated from the wool or hair fibre. It is also possible that if intermolecular crosslinking of any UHS protein component is sufficiently extensive, then that protein would not be extracted by standard reduction and S-carboxymethylation.

4.3.7 Southern Transfer Experiments

The results of the genomic Southern transfer experiments indicate that in both the human and sheep genome there are a number of genes closely related to the human UHS gene described in this chapter (figure 4.8). Hybridization to some fragments is more intense indicating the possibility that more than one UHS gene may be present on those fragments. Thus, it is possible that there are six or more genes in both genomes.

The high-sulphur B2 genes have a different pattern of hybridization to the genomic DNA (figure 4.8), indicating that there is not significant cross-hybridization between the genes of the two families and demonstrating that the two kinds of genes do not reside on the same *EcoRI* or *HindIII* restriction fragments.

Chromosomal *in situ* hybridization reveals genes related to the UHS gene described in this chapter are present in two locations, at 11p15 and 11q13-q14 (figure 4.10). Southern blot analysis suggested that at least one more gene was present in the cosmid from which the human UHS gene was isolated. Both of these observations indicate that in the human genome, the UHS genes are closely linked in two clusters which are separated by considerable distance. Close linkage of component genes is a feature that the UHS genes have in common with the other hair and wool components including the intermediate-filament genes (Powell *et al.*, 1986) and the B2 high-sulphur protein genes (Powell *et al.*, 1983).

4.3.8 Northern Transfer Experiments

The human UHS gene cross-hybridizes to sheep follicle RNA species of approximately 1.5 kb and the same size as that to which the sheep UHS cDNA clone, K4 hybridizes (figure 4.9). This hybridization is stable at high stringency (0.2 x SSC; 65°C) and thus a sheep gene closely related to the human gene is expressed in the wool follicle.

A B2 high-sulphur coding region probe hybridizes predominantly to an RNA species 1.1 kb which is smaller than the UHS mRNA.

3.4 Summary

This chapter has described the isolation and characterization of a UHS gene from the human genome. This gene contains higher levels of cysteine than was observed for the protein fraction isolated from wool, and it also contrasts strongly in the content of most of the remaining amino acids. This probably reflects on the heterogeneity of the protein fraction, which is likely to have contained a mixture of different UHS components.

The C-terminal half of the conceptual protein is composed almost entirely of two repeats: a cysteine-rich decapeptide, and a serine-rich nonapeptide. The N-terminal region is less repetitive but contains three iterations of a glycine-rich sequence 10-11 residues in length.

In vitro translation of the gene reveals that the protein product comigrates with authentic wool and hair high-sulphur protein upon the diagonal known to contain UHS protein.

Genomic Southern and *in situ* hybridization reveal that the UHS proteins are encoded by a multigene family in both the human and sheep genome. In the human genome the UHS genes appear to be clustered in two positions on chromosome 11.

Finally, northern transfer analysis revealed that a sheep gene(s) with a coding region closely related to the human gene is expressed in the wool follicle.

Chapter 5

Characterization of a Sheep UHS Gene

5.1 Introduction

The sheep UHS proteins and the genes encoding these proteins are of particular interest because of the regulation of their expression in this species by the sulphur-containing amino acids. Investigation of the mechanism of regulation by cysteine and methionine would be greatly facilitated by the isolation and characterization of a UHS gene from sheep.

This chapter describes the characterization of a sheep UHS gene from a lambda clone which was isolated using the human UHS gene described in the previous chapter as a probe.

Like the human UHS gene, the *in vitro* translation product of this gene was analyzed using two-dimensional PAGE to compare its behaviour to that of authentic wool keratin proteins, and also to the *in vitro* translation product of the high-sulphur BIIIA gene.

Genomic Southern experiments are described in which 5' and 3' noncoding region probes and also a coding region probe from the sheep UHS gene were used.

Finally, to determine whether or not this gene is expressed in the wool follicle, northern transfer experiments were conducted.

5.2 Results

5.2.1 Lambda Library Screening

2.4×10^5 recombinant phage, corresponding to 1.6 genome equivalents, were screened using the 1.0 kb *EcoRI* fragment containing the human gene (refer to chapter 4) as a probe. One positive phage was obtained from the first round screening. This phage was purified by three subsequent rounds of screening at low density. DNA was prepared from the phage and Southern transfer analysis was conducted, using DNA restricted with various enzymes and the human UHS gene as a probe (figure 5.1). A partial map of this clone was constructed and revealed that it probably contained two UHS genes (figure 5.2). The 2.8 kb *BamHI* fragment which hybridized to the probe was subcloned into a M13mp18 *BamHI* vector. The smaller 2.3 kb *EcoRI/HindIII* fragment which was encompassed by the 2.8 kb *BamHI* fragment was endfilled and subcloned into a M13mp19 *SmaI* vector.

Figure 5.1

Southern Transfer Analysis of Selected Lambda Clone

DNA was prepared (2.2.2 E) from the clone isolated from the lamda sheep genomic library (5.2.1) and cleaved (2.2.2 F) with the restriction enzymes indicated. The digestion products were fractionated by electrophoresis through 1% agarose (2.2.2 H). A *Hind*III digestion of lambda and an *Eco*RI digestion of SPP-1 (both obtained from Bresatec) were co-fractionated as molecular weight markers. After staining and photography (2.2.2 H), the DNA was transferred to Zetaprobe (2.2.2 O) and hybridized to the 1.0 kb *Eco*RI fragment containing the human UHS gene (5.2.1). The filter was washed at high stringency (0.2 x SSPE, 0.1% SDS; 65°C) prior to autoradiography.

Tracks photographed prior to transfer are paired with the corresponding hybridization patterns. The sizes of the bands are given in kilobases (kb).

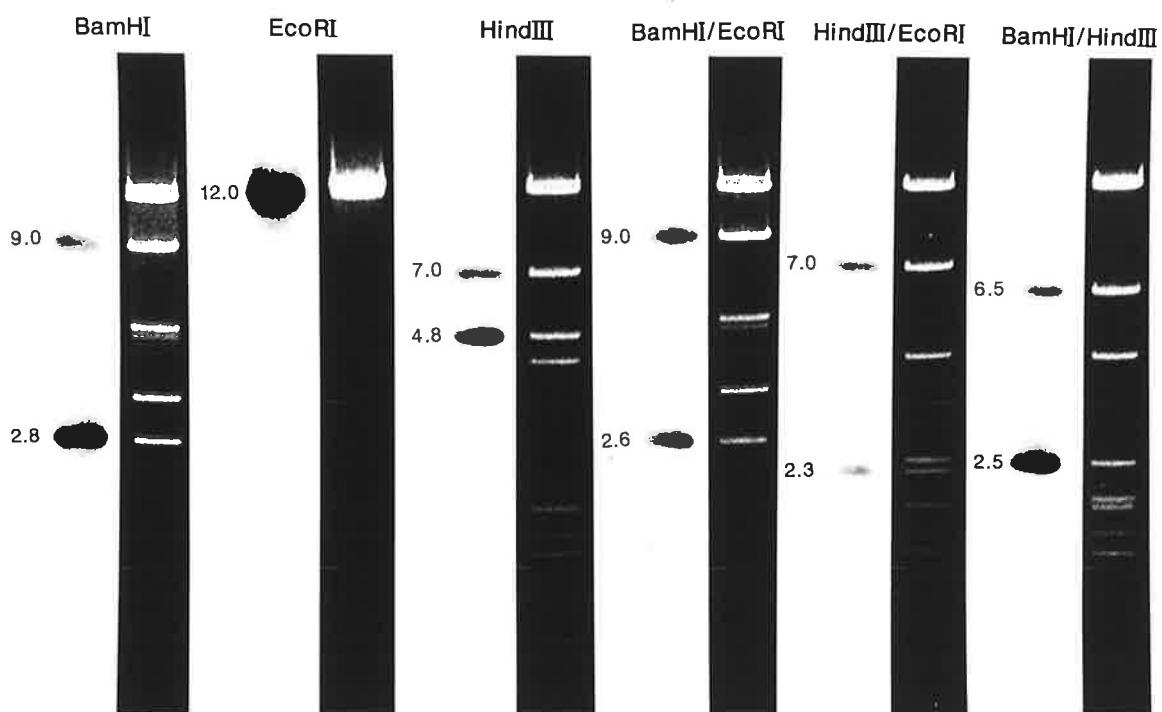


Figure 5.2

Map of Selected Lambda Clone

A partial map of the selected lambda clone was constructed from the information derived from the Southern transfer analysis conducted on this clone (figure 5.1). Regions of the clone which hybridize to the probe are indicated by a thick line. The positions of the 2.8 kb *Bam*HI and 2.3 kb *Eco*RI/*Hind*III fragments which hybridized with the probe and were selected for sequence analysis are as indicated.

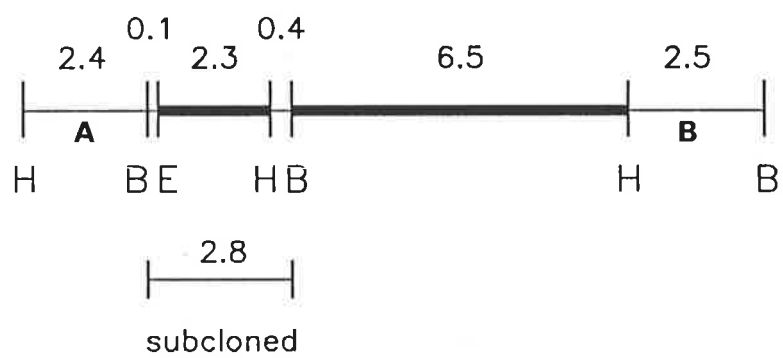
The fragment labelled A contains unknown *Bam*HI sites and the fragment labelled B contains unknown *Hind*III sites.

Sizes of fragments are given in kilobases.

B: *Bam*HI

E: *Eco*RI

H: *Hind*III



5.2.2 Sequence of the Sheep UHS Gene

The sequence of the complete presumptive transcription unit of the sheep UHS gene and also some of the 5' and 3' flanking sequence was determined by subcloning fragments from both the 2.8 kb *Bam*HI fragment and also the 2.3 kb fragment into M13mp18/19 vectors. Both progressively deleted BAL31 deletion fragments and restriction fragments were used. Figure 5.3 contains a summary of the sequencing strategy. Details of the sequencing strategy for the sheep UHS gene appear in appendix B. The sequence of the sheep UHS gene is presented in figure 5.4.

5.2.3 *In vitro* Expression of the Sheep UHS Gene

The UHS gene was expressed *in vitro* in a similar manner to the human UHS gene described in chapter 4.

A BAL31 deletion fragment which was constructed and subcloned into a M13mp18 *Eco*RI/*Sma*I vector as described in appendix B was resected by digestion with *Eco*RI/*Hind*III and subcloned into a pGEM-2 *Sma*I vector. This fragment had been deleted to a position 55 bp upstream from the putative translation initiation codon, and didn't contain any other ATG codons prior to this codon.

DNA was prepared from one clone (pGEM-SUHS) in which the 5' end of the coding region was adjacent to the SP6 RNA polymerase promoter (figure 5.5). SP6 was used to synthesize mRNA coding for the sheep UHS gene from this clone after it had been linearized using *Hind*III (2.2.3 D). Capped mRNA transcripts so produced were translated *in vitro* using rabbit reticulocyte lysate (2.2.3 E) and were labelled with [³H]serine. The translation product was fractionated by two-dimensional PAGE (2.2.1 F), after S-carboxymethylation (2.2.3 E). As a comparison total wool keratin which had been carboxymethylated with [¹⁴C]iodoacetate (2.2.1 E) was fractionated in a similar manner. Also as a comparison, a sheep wool high-sulphur BIIIA gene was expressed *in vitro* and analyzed by two-dimensional PAGE (B.C. Powell and L. Crocker, unpublished observations). The results of two-dimensional PAGE analysis of the various samples is presented in figure 5.6 and demonstrate that the sheep UHS gene *in vitro* translation product migrated on the UHS protein diagonal (see discussion). The sheep high-sulphur

Figure 5.3

Sequencing Strategy

This figure summarizes the position of the sequence obtained from the clones constructed from the 2.8 kb *Bam*HI and 2.3 kb *Eco*RI/*Hind*III fragments as delineated in appendix B. Arrows indicate the direction of sequencing. The clones used to generate the sequence are indicated in the right margin. The presumptive transcription unit of this gene was sequenced in both orientations (see figure 5.4). The 5' noncoding/flanking regions were sequenced in both orientations for 300 bp upstream from the methionine initiation codon. The 3' noncoding/flanking regions were sequenced in both orientations up to the *Bam*HI site which is approximately 100bp downstream from the presumptive polyadenylation signal.

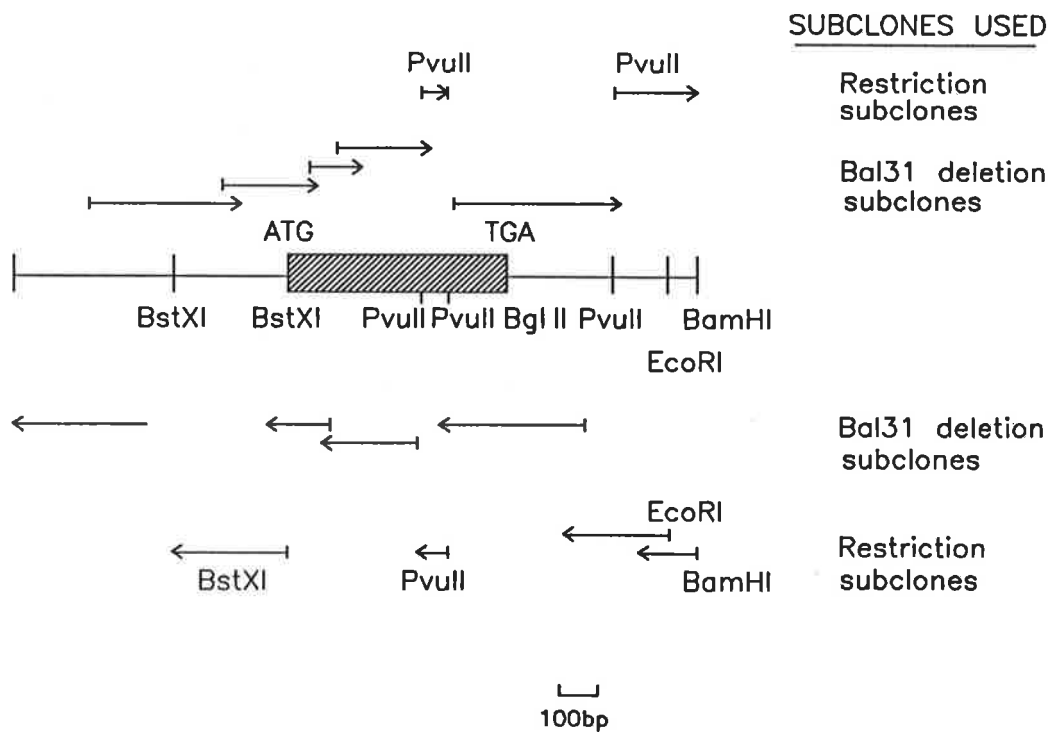


Figure 5.4

Sequence of the Sheep UHS Gene

The DNA sequence obtained from the 2.8 kb *Bam*HI and 2.3 kb *Eco*RI/*Hind*III fragments. Regions of interest in this sequence have been outlined as follows:

Flanking Regions The CCAAT, TATA, and AATAAA homologies and the T/G cluster are outlined in green. The presumptive CAP and poly(A) addition sites are as indicated. Inverted repeats are indicated by arrows under the sequence.

Coding Region (see 5.3.3). The predicted protein sequence appears above the coding region base sequence. The cysteine-rich 21 residue and glycine-rich decapeptide repeats that occur in the conceptual protein have been outlined in brown and blue respectively.

AGGTTGCTAACTACGTCATGAGTAGTGATGCATCACCATCTCTCTGTACCGCTTCTCTCTTGGCCTAGTTTTTCCCAGCATCAGTCCATTTACCAATGAGTCAGATGAGGGGATT 120
 GGGGCTAGATCCTTCTTATCTGGATATTAAAGGAAGTTCGAGAGGCTCGAGTAAAAAGAGGCAAAATCGCCGACATACTCGTTCAACGCCTAGGAGCCAAGCAGAAACACCTGCAGGTG 240
 ACAGGCTCACAGGCAACCTGCTTCTTGGTTTTTTTGTTTTTTTTTTTAACTGTTTCCAGTAAAGACCAAGAATGATCTTTAACTCTACTCATGTAAGAGTGACACCATCGTGGAGAC 360
 GCCAGCACAGCGAAGGCAACAGCGCCGCTGGGACTCCGCCATCTGTAGGTCATCTGAGGTCAGCAGAGAGGCCAGGGGACCAGTGCCCTGGAGGCCCGGGGAAGCGCCAAAGCCTGAG 480
 GTCCTAAGTACAGGGGCCAAGAAAAGGGGAAGAGGACAAAGAGGGGACAAAGCAAGCGGATCTTCTCTGAGCCAGGAGGAAGGCAGCAGCGAAAGGGAAAAACAGGGCTGGAGAAATCA 600
 AAACCACTCAGGAGGAAGTGAGCATGTGCCATCACCACACCTGACCCGGCACAGGTATATAGAGGCTCCCGCCCGGAGGCCCACTGAGCCGCTTCTCTCTCTCCACCTGCTCCT 720
 CAP Site
 CTGACCTACTCCACCCTCAACCCACCAGAACCATGGGCTGCTCTGGCTGCTCCGGAGGCTGCGGCTCCAGCTGTGGGGGCTGGGCTCCAGGTGTGGGGGCTGAGCTCCAGCTGCTGTG 840
 M G C S G C S G G C G S S C G G C G S R C G G C S S S C C V
 P V C C C K P V C C C V P A C S C S S C G K G G C G S S C G G S K G G C G S C G 960
 TGCCCGTCTGCTGCTGCAAGCCCGTGTGCTGTGTGTGCCAGCCTGCTCCTGCTCCAGCTGTGGCAAGGGGGCTGTGGCTCCAGCTGTGGGGGCTCAAGGGGGGCTGCGGCTCCTGTG 1080
 G S K G G C G S C G G C G S S C C K P V C C C V P A C S C S S C G K G G C G S C
 GGGGGTCAAGGGGGGCTGCGGCTCATGTGGAGGCTGGCTCCAGCTGCTGCAAGCCCGTGTGCTGTGTGTGCCAGCCTGCTCCTGCTCCAGCTGTGGCAAGGGGGGCTGCGGCTCCT 1200
 G G S K G G C G S C G G S K G G C G S C G G C G S G C G S S C C V P V C C C V P
 GCGGGGGTCAAGGGGGGCTGCGGCTCCTGCGGGGGTCAAGGGGGGCTGCGGCTCGTGTGGGGGCTGGCTCTGGCTGGCTCCAGCTGCTGTGTGCCCGTGTGCTGTGTGTG 1200
 A C S C S S C G K G G C G S C G C S Q S S C C V P V C C Q R K I *
 CCGCCTGCTCCTGCTCCAGCTGTGGAAGGGGGGCTGCGGCTCCTGTGGCTGCTCCAGTCCAGCTGCTGTGTCCCGTGTGCTGCCAGCGCAAGATCTGAGGCTCTGCCCAGACCTTG 1320
 GGTCCCTCATGTTTGGAGTCTTATCGTCTCATGTTCTCCGGCATCAGCTCCTTTCCACACACCCTCCCTGATTTACCCACAGCCTCCCTCCTCCTGCTCAGGTACACGTCTCCCTCCA 1440
 GGTTCAAGCACTGGGGCTGTGCTCACCCTGCTCTCCAGGGAGATGTGTAATTCTGGTACAGGAGACACATCTCCTTAGAATCAGGTGGCCCTCCAGAGTCACGTGTGGGTACCTTCTG 1560
 CTCTTCTGGGCCAGCTGGGACCCTGCAAGCCAGGGCACGTTCTGAGTGATGGCGAAGGACCCTCCTTCCCTGTGCCCTCTGCTCCCTACGAGAGCTTCTTCTTCTTCTGCTGCT 1680
 CCAGCTCCCTCTGTTTATAAAATAACGACTGAATCCCTCTAATAACATGTCCCATCTCTGACGCCGACTGTGACTGCCTTCTTCACTCTGTTTCTTGGGTGGTGACAATTGGTGC 1800
 Poly(A) Site
 TTCTTTGTAAACGTGCTCTGGGATCC 1824

Figure 5.5

A Southern and Northern Probes

Positions of the fragments used for Southern and northern analysis and also cloned into pGEM-2 as described in the text. Fragment A was used as the 5' noncoding region probe, fragment B as the coding region probe and fragment C as the 3' noncoding/flanking probe. The enzymes used to isolate the fragments are indicated. Fragment B was a BAL31 deletion fragment and is described in 5.2.3.

B pGEM-SUHS

The pGEM clone used for *in vitro* transcription of the sheep UHS gene. Fragment B (above) was cloned and transcribed as described in 5.2.3. The orientation of the fragment was determined using a *Bgl*III/*Eco*RI digestion as indicated.

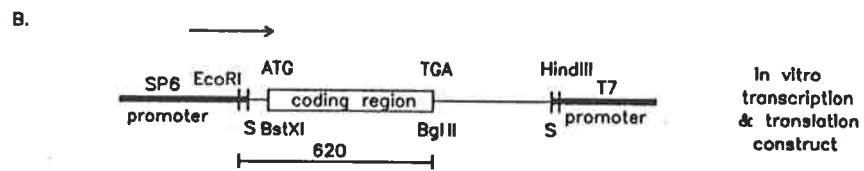
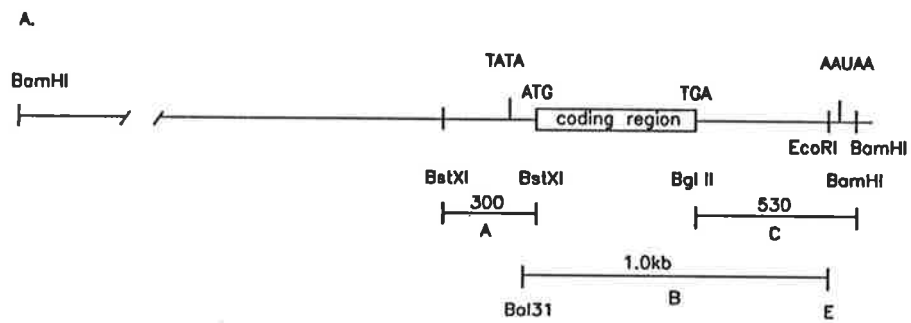


Figure 5.6

Two-Dimensional PAGE Analysis of the Sheep UHS gene *In Vitro* Translation

Product

A Total Wool Protein

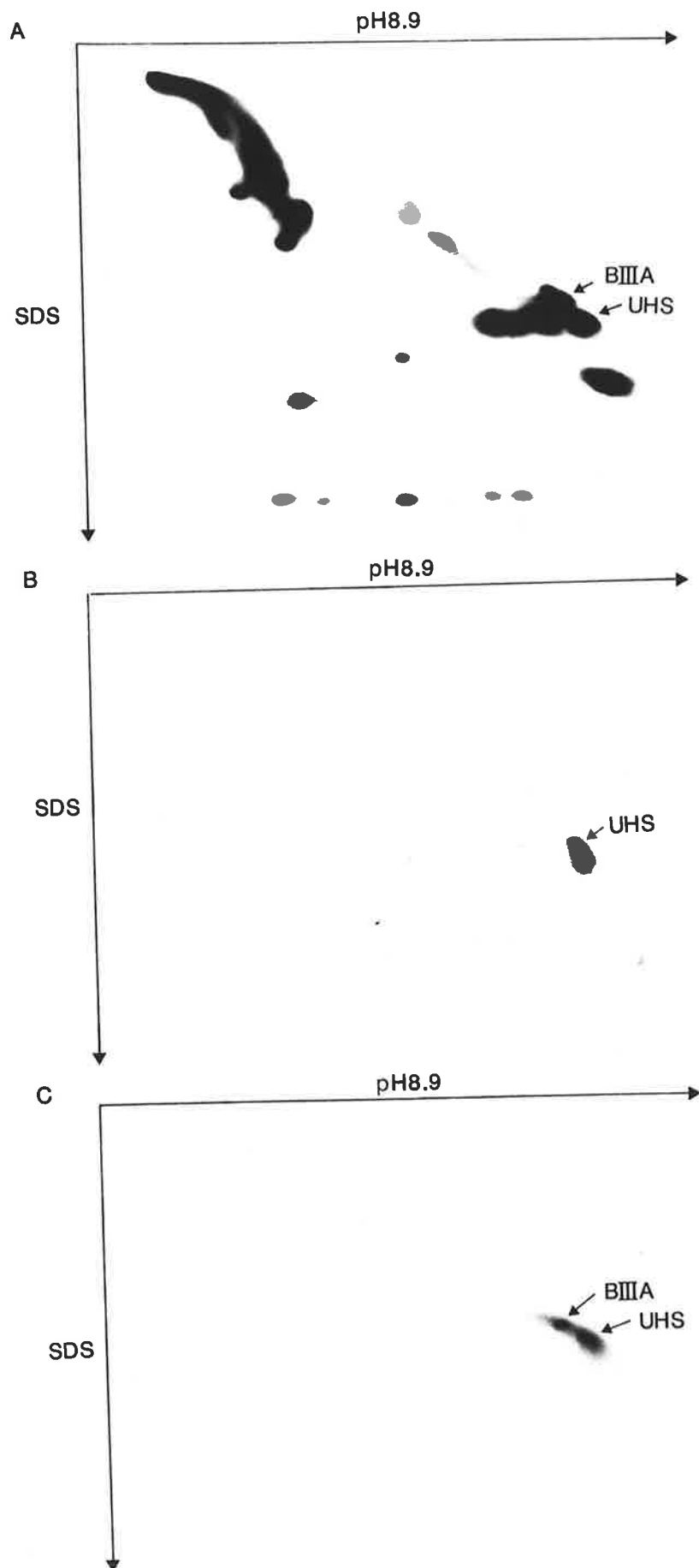
Wool was extracted and S-carboxymethylated with [^{14}C]iodoacetate (2.2.1 E), and fractionated by two-dimensional PAGE (2.2.1 F). The arrows indicate the positions of (A) the BIIIA gene *in vitro* translation product and (B) the sheep UHS gene *in vitro* translation product.

B Sheep UHS Gene In Vitro Translation Product

In vitro translation product of the sheep UHS gene (5.2.3) labelled with [^3H]serine was fractionated as described above after S-carboxymethylation (2.2.3 E).

C BIIIA Gene and Sheep UHS Gene In Vitro Translation Products

The *in vitro* translation products of the sheep BIIIA gene (L. Crocker and B.C. Powell, unpublished results) and the sheep UHS gene were fractionated together. The positions of each were determined by comparison to gels in which the gene *in vitro* translation product of each gene was fractionated individually. The upper spot (A) corresponds to the BIIIA product and the lower (B) to the UHS gene product.



BIIIA gene had a higher apparent molecular weight than the sheep UHS gene. This contradicts the protein and gene sequence data which shows that the BIIIA gene has a lower molecular weight, shorter peptide chain length, and fewer cysteine residues than the sheep UHS protein gene (see discussion 5.3.4).

5.2.4 Genomic Southern Experiments

Sheep and human genomic DNA was cleaved with *EcoRI* and *HindIII* and fractionated by electrophoresis through a 1% agarose gel (2.2.2 H). This DNA was transferred to Zetaprobe as described in chapter 2 (section 2.2.2 O). The filter was probed with: (a) a BAL31 deletion fragment which contained the complete coding region of the UHS gene; (b) a *BstXI* fragment from the 5' flanking and untranslated region; and (c) a *BglII/BamHI* fragment from the 3' untranslated region. The positions of these fragments within the 2.8 kb *BamHI* fragment are delineated in figure 5.5. After hybridization the filters were washed under low and high stringency conditions as described in chapter 2 (section 2.2.2 O). The filters were autoradiographed after both washing conditions. The results of hybridization of the genomic blot with the three probes are presented in figure 5.7, and indicate that the sheep UHS gene coding region probe is identical to the human UHS gene coding region probe.

The sheep gene 5' noncoding/flanking region probe hybridized to all the sheep DNA fragments to which the coding region probe hybridized. The intensity of hybridization was similar to that observed with the coding region probe with there being a slight increase in intensity in those fragments to which the 3' noncoding probe hybridized. The 5' noncoding probe also hybridized weakly to human DNA, but was of insufficient intensity to photograph.

The 3' noncoding region probe hybridized to two sheep DNA *EcoRI* and *HindIII* fragments (figure 5.7). Washing at high stringency didn't appear to alter the hybridization patterns for any of the probes from those observed after washing at low stringency.

Figure 5.7

Southern Analysis of Human and Sheep DNA using the Sheep UHS Gene

The same filter containing *Eco*RI and *Hind*III restrictions of human and sheep DNA as described in the previous chapter (4.2.4) was hybridized to the following probes from the sheep UHS gene:

- A: The 5' noncoding probe (fragment A figure 5.5).
- B: The coding region probe (fragment B figure 5.5).
- C: The 3' noncoding probe (fragment C figure 5.5).

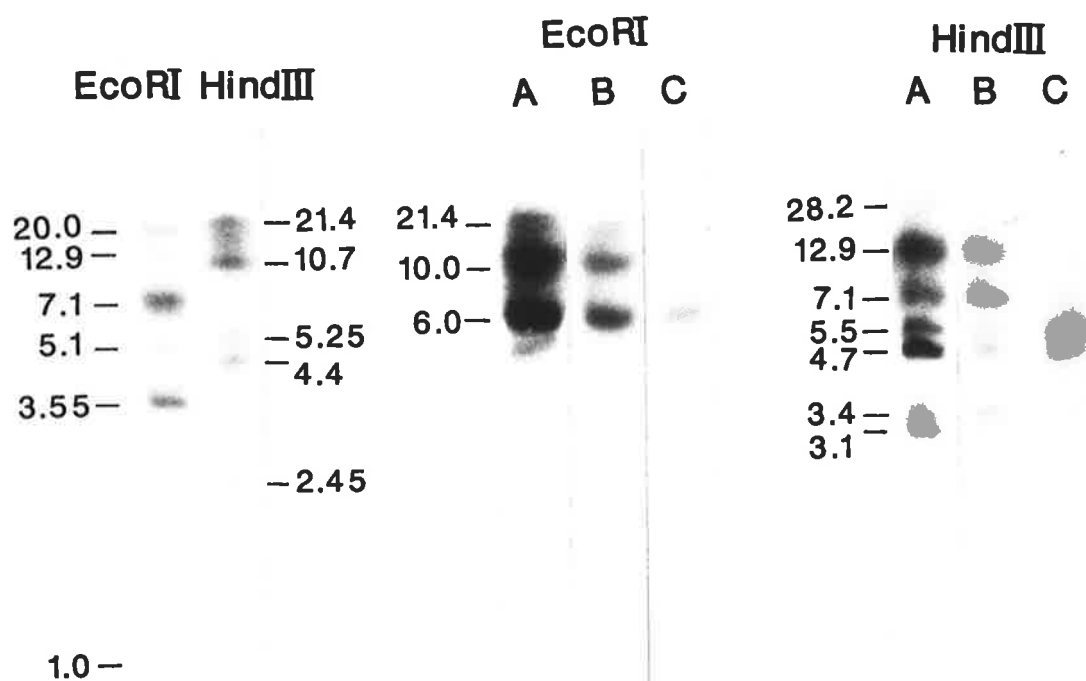
Human genomic DNA was hybridized to the coding region probe (C) only.

Autoradiographs after hybridization with each of the probes and washing at low stringency (2.0 x SSPE, 0.1% SDS; 65°C) are presented. Autoradiography after washing at high stringency (0.2 x SSPE, 0.1% SDS; 65°C) didn't reveal any significant alterations in the hybridization patterns for any of the probes.

The sizes of positive fragments are in kilobases (kb).

Human DNA

Sheep DNA



5.2.5 Northern Transfer Experiments

Northern blots were prepared using total wool follicle RNA as described in chapter 2 (2.2.3 B). The blots were probed using the same three fragments used in Southern analysis (5.2.4). After hybridization, the filters were washed at both low and high stringency (2.2.3 B). The blots were autoradiographed after both washing conditions. The results of hybridization with the three probes are presented in figure 5.8, and indicated that the coding region probe hybridized to mRNA species the same size as that to which K4 and the human UHS gene hybridized (1.5 kb in length). The 3' and 5' noncoding region probes also hybridized to mRNA of this size. In addition the 5' noncoding/flanking probe hybridized to an RNA species the same size as the 28S ribosomal RNA (see discussion; 5.3.6).

5.3 Discussion

The experimental findings presented in this chapter reveal the presence of a UHS gene in the sheep genome which has an uninterrupted coding region of 181 codons, of which 31% code for cysteine, and that this gene is expressed in the wool follicle. The features of the gene, the nature of its translation product, its tissue of expression, and its relationship to similar genes are discussed below.

5.3.1 5' Flanking/Noncoding Region

Inspection of the region surrounding the presumptive translation initiation codon reveals the sequence GAACCATGG. This sequence conforms well to the consensus sequences of Kozak (1981; 1984; 1986) and Cavener (1987) which have been discussed in chapter 4. This sequence is identical to that found in the same position of the human UHS gene discussed previously. The homology between these two genes extends beyond this region in both directions, and will be discussed in chapter 6.

Approximately 90 bp upstream from the initiation codon, the sequence TATATA is observed. This is similar in sequence to the consensus TATA box (Breathnach and Chambon, 1981), and thus, it is likely that this sequence is the functional TATA box for this gene. The CAP site of most eukaryotic genes occurs 25-30 bp downstream from the

Figure 5.8

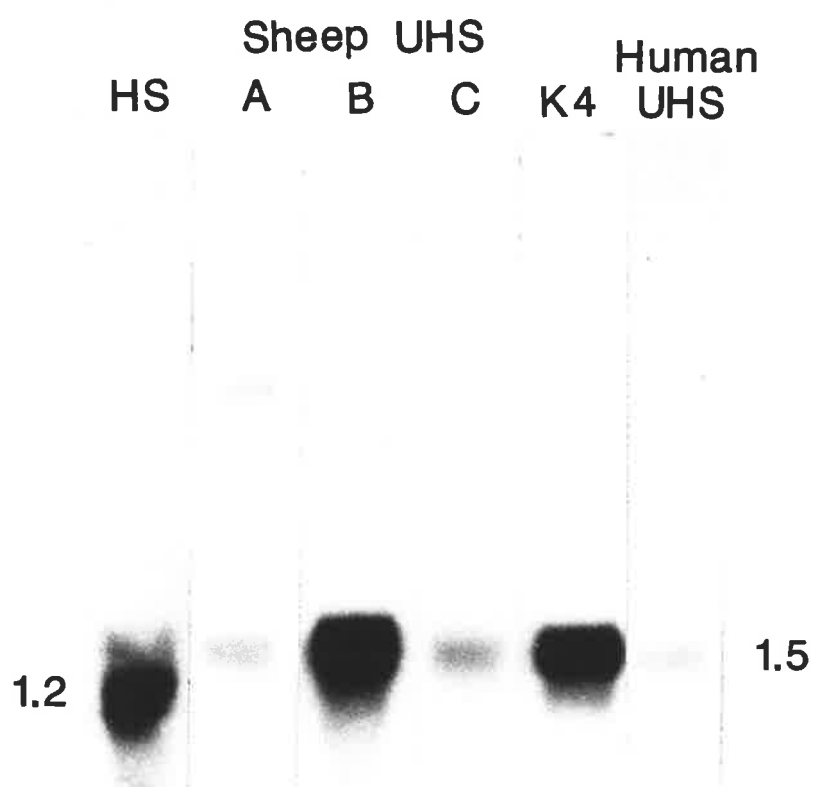
Northern Transfer Analysis of Wool Follicle RNA

The same northern transfer filters that were used in the previous chapter (4.2.5) were used for analysis with the sheep UHS gene. The following probes were used:

- HS:** the B2A coding region probe (4.2.5).
 - A:** sheep UHS gene 5' noncoding probe (figure 5.5).
 - B:** sheep UHS gene coding region probe (figure 5.5).
 - C:** sheep UHS gene 3' noncoding probe (figure 5.5).
 - K4:** 310 bp coding/3' noncoding probe from sheep UHS cDNA clone (4.2.1).
 - Human:** the 1.0 kb *Eco*RI fragment containing the human UHS gene (4.2.1).
- UHS**

The filters were washed at low stringency (2.0 x SSPE, 0.1% SDS; 65°C) prior to autoradiography. An autoradiograph after washing at high stringency (0.2 x SSPE, 0.1% SDS; 65°C) is presented and was essentially identical to the pattern after autoradiography at low stringency.

Approximate sizes of the hybridizing fragments were determined relative to the 18S and 28S ribosomal RNA species. The larger band observed with probe A comigrated with the 28S rRNA.



TATA box (Breathnach and Chambon, 1981). Preliminary RNA protection experiments (not presented) suggest that the CAP site of the UHS gene lies between those distances from the presumptive TATA box.

Approximately 30 bp upstream from the presumptive TATA box, the sequence CCAGT is observed, which may be sufficiently similar to the CCAAT homology present in this region of some eukaryotic genes (Benoist *et al.*, 1980; Efstratiadis *et al.*, 1980), for this sequence to be functionally active as a CCAAT box.

Inspection of the promoter region for the Sp1 consensus decanucleotide binding sequence (Briggs *et al.*, 1986) failed to reveal its presence, indicating that this protein doesn't play a role in the transcription of this gene, or that the sheep Sp1 binding sequence diverges from that of other species.

Four pairs of inverted repeats are observed in the 5' flanking region (figure 5.4). One pair centred about 100 bp upstream from the CAP site are 10 nucleotides in length and are separated by 3 nucleotides. The remaining three pairs all contain substitutions and vary between 9 and 12 bp in length (figure 5.4). It is possible that these motifs have a function in the regulation of transcription through the formation of stem-loop structures. Observations have been made which indicate that factors which enhance stem-loop formation also enhance transcription. Negatively supercoiled DNA, in which the double helix is effectively underwound, has been found to promote the formation of stem-loop structures (Lilley, 1980). Lilley found that an equilibrium existed between hairpin and nonhairpin structures of 9 to 13 bp inverted repeats separated by 2 to 6 bp of nonrepetitious sequence, in supercoiled DNA. Within eukaryotes it has been established *in vitro* (Akrigg and Cook, 1980) and *in vivo* (Weintraub *et al.*, 1986), that negative supercoiling enhances the rate of transcription by RNA polymerase II. The binding of regulatory proteins to inverted repeats similar to those observed in the sheep UHS gene has been demonstrated for the single stranded form of $\phi\chi 174$ (Shlomai and Kornberg, 1980), and also to the iron responsive element of rat ferritin mRNA (Leibold and Munro, 1988). Thus it is possible that supercoiling of the promoter region of the sheep UHS gene allows formation of the stem-loops which in turn may allow regulatory proteins to bind.

5.3.2 3' Untranslated/Flanking Region

A presumptive polyadenylation signal (Proudfoot and Brownlee, 1976) occurs 420 bp downstream from the termination codon of the gene. Polyadenylation occurs 14-30 bp downstream from the polyadenylation signal of eukaryotic genes (Proudfoot and Brownlee, 1976; Breathnach and Chambon, 1981; Birnstiel *et al.*, 1985). Preliminary RNA protection experiments (not presented) suggest that polyadenylation occurs within those distances beyond the presumptive polyadenylation signal.

A G/T rich cluster occurs approximately 50 bp downstream from the presumptive polyadenylation signal (figure 5.4), which is similar to the G/T homology observed in this region of many eukaryotic genes, and may have a function in the correct cleavage and polyadenylation of the nascent transcript (Birnstiel *et al.*, 1985). In some genes this region conforms to the consensus sequence YGTGTTY, where Y is a pyrimidine (McLauchlan *et al.*, 1985). Within the G/T rich region observed in the sheep UHS gene the sequence CCTGTTTC occurs, which conforms in 7 of the 8 base positions to the consensus sequence.

A pair of contiguous inverted repeats occur in the 3' noncoding region centred at 150bp downstream from the termination codon (figure 5.4). A second pair of inverted repeats occur centred about 195bp from the termination codon which are separated from each other by 17bp (figure 5.4). Both pairs of inverted repeats contain base substitutions.

5.3.3 Conceptual Protein Analysis

The gene encodes a protein 181 residues in length, with an empirically determined molecular weight of 15,967. The amino acid content of the conceptual protein is summarized in table 5.1. The sheep gene conceptual protein lacks 10 of the 22 amino acids which are encoded in eukaryotic genomes. Over three-quarters of the protein is composed of cysteine, glycine and serine residues, making it similar in this respect to the human UHS gene translation product. However, inspection of table 5.1 reveals that whilst the combined levels of these three amino acids are similar in both proteins, the sheep protein contains lower levels of cysteine and serine, but higher levels of glycine.

Table 5.1

Amino Acid Content of the Sheep UHS Gene

Presented are the proportions (mole%) of the amino acids in the sheep UHS gene conceptual protein. As comparisons, the amino acid content of the human UHS gene conceptual protein (chapter 4) and the amino acid analysis of a fraction of wool UHS protein (Gillespie and Broad, 1972) are also presented.

Amino Acid	Protein	Human Gene	Sheep Gene
Ala	2.0	1.2	1.7
Arg	6.9	2.4	1.1
Asn	-	0	0
Asp	0.6	0.6	0
Cys	29.9	35.5	31.0
Gln	-	4.1	1.1
Glu	7.9	0	0
Gly	4.2	16.0	28.0
His	1.3	0	0
Ile	1.7	0.6	0.6
Leu	1.4	0	0
Lys	0.9	5.0	5.5
Met	0	0	0
Phe	0.5	0	0
Pro	12.8	6.0	4.4
Ser	12.7	23.1	20.0
Thr	11.1	0	0
Trp	-	0	0
Tyr	1.8	1.8	0
Val	4.3	3.0	6.1

Like the human gene protein, the proportions of the constituent amino acids diverge strikingly from that determined for a fraction of wool UHS protein (table 5.1). In fact, only one similarity is observed in that both have similar levels of cysteine. Possible explanations for these discrepancies will be discussed in the next chapter.

Inspection of the sequence of the protein encoded by the sheep UHS gene, reveals that it is almost completely composed of two distinct reiterated sequences (figure 5.4). One of these is composed of 10 or 11 residues, of which half are glycine (figure 5.10). The other is usually 21 residues long and contains high levels of cysteine (40%), serine (24%) and valine (14%) (figure 5.9). The two repeats are interspersed with each other in a complex fashion (figure 5.4). The N-terminal region of the protein prior to the first cysteine-rich repeat is more divergent from the remainder of the protein, but does contain one glycine-rich repeat.

At the peptide level homology between the cysteine-rich repeats (figure 5.9) is strong, however there are variations in their lengths. The cysteine-rich repeated sequence nearest the N-terminus contains an internal duplication of six residues (figure 5.9). Also, the iteration nearest to the C-terminus contains only the first ten residues of the complete repeat. Substitutions occur at three positions when the four iterations of this repeat are compared (figure 5.9). At two positions lysine, which may carry a positive charge, is substituted for valine, which is hydrophobic. This amino acid substitution will result in a substantial change in the chemical characteristics of the protein, and will be discussed in the next chapter. Three different amino acids are observed in the first position of the repeat. In the first iteration a serine residue is present. The middle two iterations have glycine residues in this position and the iteration nearest the C-terminus has glutamine.

When pairwise comparisons of all combinations of the base sequence of the cysteine-rich repeat were conducted, an average homology of approximately 90% was found (figure 5.9). Base substitutions occur in the third position of four codons, and do not result in amino acid substitutions. The remaining base substitutions result in the amino acid substitutions which are observed at three positions (figure 5.9).

Strong homology is observed between the glycine-rich repeats (figure 5.10), however variation in the length of this repeat is also observed, with the two glycine-rich

Figure 5.9

Cysteine-Rich Amino Acid Repeat

The amino acid sequences of the cysteine-rich repeat (5.3.3) are presented in the order in which they occur from the N-terminus of the conceptual protein. The first repeat contains an internal duplication which has been removed and aligned with the other repeats as indicated.

Below the alignment of the amino acid sequences are the corresponding nucleotide sequences that code for the cysteine-rich repeats. The nucleotide sequences are presented in the same order as the amino acid sequences. Positions at which conservative base substitutions occur are indicated by stars (*) and positions where base substitutions result in amino acid substitutions are indicated by lines (|).

Sheep Gene Cysteine-Rich Repeated Sequence

Peptide Sequence

S	S	S	C	C	V	P	V	C	C	C	K	↑	P	A	C	S	C	S	S	C	G
						<u>P</u>	<u>V</u>	<u>C</u>	<u>C</u>	<u>C</u>	<u>V</u>	↓									
G	S	S	C	C	K	P	V	C	C	C	V		P	A	C	S	C	S	S	C	G
G	S	S	C	C	V	P	V	C	C	C	V		P	A	C	S	C	S	S	C	G
Q	S	S	C	C	V	P	V	C	C												

Nucleotide Sequence

(same order as above)

<u>AGC</u>	TCC	AGC	TGC	TGT	<u>GTG</u>	CCC	<u>GTG</u>	TGC	TGC	TGC	<u>AAG</u>	↑	<u>CCA</u>	GCC	TGC	TCC	TGC	TCC	AGC	TGT	<u>GGC</u>
						<u>CCC</u>	<u>GTG</u>	<u>TGC</u>	<u>TGC</u>	<u>TGT</u>	<u>GTG</u>	↓									
<u>GGC</u>	TCC	AGC	TGC	TGC	<u>AAG</u>	CCC	<u>GTG</u>	TGC	TGC	TGT	<u>GTG</u>		<u>CCA</u>	GCC	TGC	TCC	TGC	TCC	AGC	TGT	<u>GGC</u>
<u>GGC</u>	TCC	AGC	TGC	TGT	<u>GTG</u>	CCC	<u>GTG</u>	TGC	TGC	TGT	<u>GTG</u>		<u>CCC</u>	GCC	TGC	TCC	TGC	TCC	AGC	TGT	<u>GGA</u>
<u>CAG</u>	TCC	AGC	TGC	TGT	<u>GTC</u>	CCC	<u>GTG</u>	TGC	TGC												
					*	*		*			*		*								*

Figure 5.10

Glycine-Rich Repeat

An alignment of the amino acid sequence of the glycine-rich repeat. The first 2 repeats, nearest to the N-terminus, contain an extra serine residue. The remaining repeats have been aligned to maximise homology.

Below are the corresponding nucleotide sequences coding for the glycine-rich repeats, presented in the same order as the amino acid sequence. Positions at which conservative base substitutions occur are indicated by stars (*) and positions where base substitutions result in amino acid substitutions are indicated by lines (|).

Sheep UHS Protein Glycine-Rich Repeat

Peptide sequence

S	G	G	C	G	S	S	C	G	G	C
K	G	G	C	G	S	S	C	G	G	S
K	G	G	C	G	S		C	G	G	S
K	G	G	C	G	S		C	G	G	C
K	G	G	C	G	S		C	G	G	S
K	G	G	C	G	S		C	G	G	S
K	G	G	C	G	S		C	G	G	C
K	G	G	C	G	S		C	G	C	S

Nucleotide sequence

(same order as above)

<u>TCC</u>	GG A	GGC	TG C	GGC	TCC	AGC	TGT	GGG	GGC	<u>TGC</u>
AAA	GG G	GGC	TGT	GGC	TCC	AGC	TGT	GGG	GGC	TCC
AAG	GG G	GGC	TG C	GGC	TCC		TGC	GGG	GGG	TCT
AAG	GG G	GGC	TG C	GGC	TCA		TGT	GG A	GGC	<u>TGC</u>
AAG	GG G	GGC	TG C	GGC	TCC		TGC	GGG	GGG	TCT
AAA	GG G	GGC	TG C	GGC	TCC		TGC	GGG	GGG	TCT
AAG	GG G	GGC	TG C	GGC	TCC		TGT	GGG	GGC	<u>TGT</u>
AAA	GG G	GGC	TG C	GGC	TCC		TGT	GGC	<u>TGC</u>	TCC
*	*		*		*		*	*		* *

repeats nearest the N-terminus each containing an extra serine residue (figure 5.10). Amino acid substitutions occur at three positions (figure 5.10).

Pairwise comparison of the nucleotide sequences of the glycine-rich repeats reveals a high degree of homology which ranges from 73% to 100%, with an average of 84% (figure 5.10). Base substitutions occur at twelve positions of which eight are in the third position of the corresponding codon. The remaining four occur in either of the first two positions of the corresponding codons and coincide with the four positions at which amino acid substitutions occur.

The first ten residues of the cysteine-rich repeat and the corresponding nucleotide sequence are homologous to the cysteine-rich decapeptide repeat observed in the human UHS gene. This region at both the protein and nucleic acid levels, is also homologous to repeats observed in the BIIIA and B2 high-sulphur proteins, the presumptive coding region of the sheep UHS cDNA clone K4, and the putative mouse UHS gene coding region. In addition, both the cysteine-rich and the glycine-rich repeats observed in the sheep gene are present in the N-terminal region of the human UHS gene, and also in the coding region of K4. Comparisons between the various genes will be discussed in chapter 6.

5.3.4 In Vitro Translation Experiments

Like the human UHS gene translation product (4.2.3), the sheep UHS gene translation product lies upon the diagonal on which it is believed the UHS proteins lie when fractionated by two-dimensional electrophoresis (figure 5.6).

The two-dimensional gel in which the *in vitro* translation products of the sheep UHS gene and the high-sulphur BIIIA gene were fractionated together (figure 5.6) demonstrated that the translation product of the BIIIA gene, which is lower in cysteine content (21 mole%), chain length (131 residues), and molecular weight (13,406), had a larger apparent molecular weight than the sheep UHS gene translation product.

It was also observed that the BIIIA translation product is positioned on the diagonal after two-dimensional PAGE, indicating that this region is not the domain solely of the UHS proteins.

These observations demonstrate that there is not a simple relationship between the various physical attributes of a protein and its mobility by polyacrylamide gel electrophoresis. This must be particularly true of proteins with unusual amino acid composition such as the UHS proteins.

To clarify the confusing situation which exists at present as to the exact position of the various high-sulphur protein components it will be necessary to analyse the *in vitro* translation products of as many as possible of the high-sulphur protein genes that have been isolated and characterized. Also, the mobilities of the *in vitro* translation products need to be correlated to those of the same components isolated from wool, as both biological modification, and artefactual modification resulting from extraction conditions, may alter the mobility of the latter.

5.3.5 Genomic Southern Transfer Experiments

The pattern of hybridization of a coding region probe from the sheep UHS gene to sheep and human genomic DNA (figure 5.7) is identical to that of the human UHS gene coding region probe (4.2.4), and is indicative of the close relationship between the two genes. As at least three positive fragments were observed in the *Hind*III track and six in the *Eco*RI track, it confirms the belief that there are a number of similar UHS genes in the sheep genome. Increasing the stringency of the washing conditions did not appear to significantly alter the hybridization pattern.

The strength of the relationship between the various presumptive sheep UHS genes was revealed when the same filter was hybridized to a restriction fragment containing only the 5' flanking and the untranslated sequences from the sheep UHS gene. This probe hybridized to all the sheep genomic restriction fragments to which the coding region probe hybridized (figure 5.7). Washing the filter at high stringency didn't appear to alter the hybridization pattern.

Thus, it would appear that the homology between the various UHS genes at least in the sheep genome must extend beyond the coding region into the 5' flanking regions. It is possible that the sequence conservation in the 5' regions of the UHS genes relates to the regulation of their expression.

Genomic Southern analysis using a fragment isolated from the 3' noncoding/flanking regions of the sheep UHS gene revealed that this probe hybridized to two sheep *EcoRI* and *HindIII* fragments (figure 5.7). As this probe did not straddle a *HindIII* site it indicates that regions homologous to the UHS gene 3' noncoding/flanking region must occur elsewhere in the genome. One possibility is that the sheep UHS gene has been recently duplicated resulting in a second UHS gene with a similar 3' noncoding region.

5.3.6 Northern Transfer Experiments

A coding region probe from the sheep UHS gene hybridizes to wool follicle RNA of the same size as that to which the human UHS gene, and the sheep UHS cDNA clone K4 also hybridize (figure 5.8). The estimated size of this species is 1.5 kb. This is 400 bp longer than the 1.1 kb determined from sequence analysis. This in part may be accounted for by the poly(A) tail which may be up to 300 nucleotides in length when first added (Brawerman, 1981). It may also in part reflect the level of resolution of the northern transfer procedure.

Both the 5' and 3' noncoding probes hybridized to the same species, thus the sheep UHS gene described in this chapter is functional and is expressed in the wool follicle.

Of interest is the observation that the 5' noncoding/flanking probe hybridizes to another species of RNA, which corresponds in size to the 28S ribosomal RNA molecule. Hybridization was stable to washing under high stringency conditions. As no sheep ribosomal sequence was available the sequence of the *BstXI* fragment used as a probe was compared to a mouse 28S gene (Hassouna *et al.*, 1984) and also a human 28S gene (Gonzalez *et al.*, 1985). However no significant homology was revealed by these comparisons. It is possible that if there is strong homology between the regions 5' of the sheep UHS gene coding region and the sheep 28S ribosomal RNA, it must occur in one of the variable regions, which are observed in eukaryotic 28S RNA (Hassouna *et al.*, 1984)

The significance, if any, of the possible homology between the 28S rRNA and the sheep UHS gene is unknown. In scanning models of translation initiation, it is the small subunit which interacts with the upstream untranslated regions of the message. The large

subunit which contains the 28S RNA doesn't interact until the small subunit has moved to the initiation codon and translation is ready to commence (for reviews see Kozak, 1983; Moldave, 1985)

5.4 Summary

This chapter describes the isolation and characterization of a sheep UHS gene. This gene encodes a protein 181 residues in length and containing 31% cysteine. The *in vitro* translation product of this gene when fractionated by two-dimensional PAGE migrates on the diagonal on which it is believed authentic UHS protein also migrates.

The sheep UHS gene coding region is equivalent to the human UHS gene coding region when they are used as probes for genomic Southern experiments. A probe from the 5' noncoding region of the sheep UHS gene revealed that probably all sheep presumptive UHS genes are homologous in this region. A probe from the 3' noncoding region hybridized to more than one genomic restriction fragment and this may be indicative of recent gene duplication events.

Use of the same probes for northern transfer experiments demonstrated that the sheep UHS gene is expressed in the wool follicle and the size of the message is estimated at 1.5 kb.

Chapter 6

Comparison of UHS Gene Sequences

6.1 Introduction

In this chapter the human and sheep UHS genes that have been described in chapters 4 and 5 are compared to each other, to a mouse presumptive UHS gene (Dr. A. McNab, personal communication), to the sheep UHS cDNA clone K4 (K. Ward, personal communication), and to the sheep B2 high-sulphur protein genes (Powell *et al.*, 1983).

These comparisons are of importance because they may reveal homologies which are important for gene expression, and for protein structure and function. Comparisons may also reveal contrasting differences between the various genes, which may also give insight into gene expression and protein function.

The physical and chemical characteristics of the conceptual protein products of the human and sheep UHS genes were analysed using computer programmes, in order to gain an insight into their structure and function.

6.2 5' Noncoding/Flanking Regions

The human and sheep UHS genes have homologous 5' untranslated regions. Over a length of approximately 50bp contiguous with the initiation codon there is over 90% homology between the two genes (figure 6.1).

When the mouse presumptive UHS gene was compared to the sheep gene, it was found to be more divergent in this region, but does contain some homology to the sheep and human UHS genes (figure 6.1).

It is possible that the homology in the 5' noncoding regions observed between the human and sheep UHS genes extends to all sheep UHS genes, as it was observed experimentally by genomic Southern analysis that a probe from the 5' noncoding region of the sheep UHS protein gene cross-hybridized to all the restriction fragments to which a coding region probe hybridized (see 5.2.4 and 5.3.5). As the probe from the 5' noncoding region also contained 5' flanking sequence it is also possible that homology between the sheep UHS protein genes occurs further upstream, in the promoter region.

Strong conservation between the 5' noncoding regions of genes with closely related coding regions is not a general rule (Kozak, 1983). As strong homology occurs between species for this region of the UHS protein genes, it is possible that it may have some

Figure 6.1

5'Noncoding Region Homology

An alignment of the 5'noncoding regions of the sheep, human and mouse UHS genes. Homology is strongest between the human and sheep UHS genes.

At the top is the sequence of the sheep UHS gene 5' noncoding region. Bases which match with the sheep sequence in the human and mouse sequence are indicated by a star (*). The identity of mismatched bases are indicated. Insertions (-) and deletions (subscript bases) have been introduced into the human and mouse sequences to facilitate alignment.

The presumed CAP sites of the three genes is indicated by a bold letter.

Homology in the 5' Noncoding Region

Sheep Gene: ACACTGAGCCGCTTCTCTCTCTCCACCTGCTCCTCTGACCTACTCCACCCTCAACCCACCAGAACC**ATG**

Human Gene: CTC*AC*C*T**AC***CT***-*****-*****G*****T*******ATG**

Mouse Gene: **TGAA*CAGTTG**AAAA***AGAAAA***G***^{AA}A*****T****-CCA*AT***AC*****ATG**

functional significance, perhaps in the regulation of translation initiation. Such a function has been proposed for the 5' noncoding region of ferritin mRNA, where it is believed that translation of this message is regulated through the binding of a regulatory protein to this region of the mRNA (Casey *et al.*, 1988), and recently such a protein has been found to bind to this region (Leibold and Munro, 1988; Rouault *et al.*, 1988).

As discussed in the introductory chapter (section 1.6) most of the high-sulphur and high-glycine/tyrosine genes have an 18 bp conserved element in the 5' untranslated region, which is contiguous with the initiation codon. When the 5' untranslated regions of the human and sheep UHS genes were compared to that of the sheep high-sulphur B2C gene no strong homology was observed. Hence the 18 bp conserved element is not present in either of the human or sheep UHS genes. The lack of this conserved element in these genes may be significant as it may indicate a difference in the mechanism of translation regulation. Such possibilities are of interest since only the expression of the UHS proteins is affected significantly by the supply of cysteine to the follicle.

Comparison of the sheep, human and mouse UHS genes in the 5' flanking region failed to reveal any extensive homology, as was the case when the high-sulphur B2 and BIIIA genes were compared to the sheep UHS gene.

6.3 3'Flanking/Noncoding Regions

Comparison of the 3' flanking regions of the sheep and human UHS genes revealed a number of short regions of homology with a continuous match seldom longer than 6 bases. In general these were not positionally conserved with respect to the termination codons of these genes. No extensive homology was observed between the sheep B2 and BIIIA high-sulphur genes and the sheep UHS gene in this region.

6.4 Coding Regions

6.4.1 Conceptual Protein Size and Amino Acid Content

Table 6.1 summarizes the amino acid content, chain length and empirical molecular weight of the conceptual translation products of the human and sheep UHS genes, the mouse presumptive UHS gene and that of the sheep UHS cDNA clone, K4.

Table 6.1

Amino Acid Content of UHS Proteins

A summary of the amino acid content, peptide chain length and empirical MW of the conceptual proteins of the three UHS genes and the sheep UHS cDNA clone, K4, which are described in the text. The amino acid analysis of a fraction of UHS protein is presented and was obtained from Gillespie and Broad (1972).

Amino Acid	Protein	Human Gene	Sheep Gene	K4	Mouse Gene
Ala	2.0	1.2	1.7	2.7	1.6
Arg	6.9	2.4	1.1	8.0	4.8
Asn	-	0	0	0	1.6
Asp	0.6	0.6	0	1.3	0.54
Cys	29.9	35.5	31.0	37.3	37.6
Gln	-	4.1	1.1	6.7	9.14
Glu	7.9	0	0	0	1.08
Gly	4.2	16.0	28.0	5.3	2.7
His	1.3	0	0	0	0
Ile	1.7	0.6	0.6	0	0.54
Leu	1.4	0	0	1.3	3.2
Lys	0.9	5.0	5.5	1.3	0.54
Met	0	0	0	0	1.08
Phe	0.5	0	0	0	1.08
Pro	12.8	6.0	4.4	6.7	11.3
Ser	12.7	23.1	20.0	21.3	13.4
Thr	11.1	0	0	1.3	7.0
Trp	-	0	0	1.3	0.54
Tyr	1.8	1.8	0	0	0
Val	4.3	3.0	6.1	5.3	2.1
Chain length	-	168	181	-	168
MW	-	16276	15969	-	19626

As discussed previously (chapter 5, 5.3.3) the human and sheep UHS gene products have an amino acid content which is divergent from that obtained from a purified fraction of wool UHS protein (Gillespie and Broad, 1972).

K4 has a coding region which is also divergent from the protein fraction amino acid analysis; however, care should be taken in drawing conclusions from this clone because it is incomplete.

With the exception of elevated levels of cysteine, the conceptual mouse UHS protein has an amino acid content which approximates the wool UHS protein amino acid analysis much more closely than do the conceptual human and sheep UHS proteins (table 6.1).

Evidence exists at the protein level which indicates that UHS protein with amino acid content divergent from that presented in Table 6.1 is present in hair and wool. Marshall and Gillespie (1976) investigated the high-sulphur proteins of mouse hair and found that there was considerable heterogeneity in the amino acid content of the UHS proteins of this species, which included fractions containing higher levels of glycine than occurs in the fraction of wool UHS protein presented in table 6.1. Also, sheep UHS protein fractions have been observed with higher glycine levels than the fraction presented in table 6.1 (Dr. J.M. Gillespie, personal communication). Thus, the unusual amino acid content of the human and sheep conceptual UHS proteins appears to be in accord with heterogeneity observed directly in the UHS protein of hair and wool.

The coding regions of the various UHS protein genes indicate that a large proportion of the variability in amino acid content may simply reflect variation in the number and type of reiterated amino acid sequences that compose a particular protein.

The chain lengths and empirical molecular weights of the three conceptual UHS proteins are presented in table 6.1. Despite having a longer chain length than the human UHS protein, the sheep UHS protein has a lower empirical molecular weight due to the higher content of glycine of the latter.

6.4.2 Codon Usage

The proportions of the codons encoding cysteine, glycine, proline, serine, and valine for the various UHS genes and also for the high-sulphur BIIIA and B2C genes, are set out in table 6.2. Overall the pattern of codon usage for these amino acids is similar in all of the genes, with there being a marked preference for C in the redundant third codon position. The similarity in codon usage exhibited by the UHS and high-sulphur genes is not unexpected considering the observation that codon usage is similar for genes within a particular genome, and between genomes of related species (Grantham *et al.*, 1980 a, b).

6.5 Comparison of Coding Region Sequences

The previous chapters revealed that the sheep and human conceptual UHS proteins were composed of a number of different repeats. In the following sections (6.5.1 and 6.5.2) the repeats of the various UHS proteins and also the high-sulphur B2 proteins are compared. These comparisons demonstrate that there are significant similarities between the repeats of the various proteins. Also compared (6.5.3, 6.5.4 and 6.5.5) are the way in which these repeats are organised to form the various complete UHS proteins.

6.5.1 Comparison of Cysteine-Rich Repeats

Cysteine-rich repeats of varying length are observed in the conceptual proteins of all the UHS genes and also the B2 and BIIIA high-sulphur proteins. Comparisons of the cysteine-rich repeats at both the protein and nucleotide levels of the various UHS genes, K4, and the B2 genes reveal that these sequences are all related to varying degrees. The comparisons were conducted by aligning nucleotide sequences of representative cysteine-rich repeated sequences from each gene using the human gene cysteine-rich decapeptide repeat as a benchmark. Consensus sequences were determined for the repeats present in the human and sheep UHS genes and K4. For the sheep high-sulphur B2 and mouse UHS genes where more variation was observed between iterations, comparisons were made to all iterations in that gene, and an average homology determined. The results of these comparisons are summarized in figure 6.2 and 6.3.

Table 6.2

Codon Usage

A summary of the codon usage for five amino acids encoded in the mouse, human and sheep UHS genes, the sheep high-sulphur B2C and BIII A genes, and the sheep presumptive UHS cDNA clone, K4. The proportion of the amino acid type encoded by a particular codon is expressed as a percentage.

Amino Acid	Mouse	Human	Sheep	Gene BIIIA	B2C	K4
Cysteine						
TGT	29	35	32	18	3	18
TGC	71	65	68	82	97	82
Glycine						
GGT	0	7	0	15	0	0
GGC	60	63	61	62	83	75
GGA	20	11	6	15	17	0
GGG	20	19	33	8	0	25
Proline						
CCT	24	10	0	18	0	20
CCC	57	80	75	36	76	60
CCA	19	10	25	36	10	20
CCG	0	0	0	9	14	0
Serine						
TCT	12	10	14	0	7	0
TCC	32	46	54	38	40	63
TCA	0	15	3	5	0	0
TCG	0	3	3	0	13	0
AGT	0	3	0	10	0	0
AGC	56	23	27	48	40	38
Valine						
GTT	0	20	0	0	0	0
GTC	0	51	18	17	20	25
GTA	0	0	0	0	0	0
GTG	100	51	82	83	80	75

Figure 6.2

Cysteine-Rich Deca-and Nona-Peptide Repeats

An alignment of the consensus cysteine-rich decapeptide repeat of the human UHS gene with:

- A The consensus decapeptide repeat of the sheep cDNA clone, K4. 87% nucleotide homology was observed.
- B The nonapeptide repeat of the mouse presumptive UHS gene which shared greatest nucleotide homology (85%) to the human cysteine-rich decapeptide repeat. Average homology to all nonapeptide repeats was about 60%.
- C The most similar high-sulphur B2A repeat (73% nucleotide homology). Average homology to all repeats in the B2A gene was 52%.

A star indicates that there is a nucleotide match between the compared sequences.

A

<u>Human gene</u>									
Q	S	S	C	C	K	P	C	C	S
CAG	TCC	AGC	TGC	TGT	AAG	CCC	TGT	TGC	TCC
***	***	***	***	**	*	***	**	***	***
CAG	TCC	AGC	TGC	TGC	AGA	CCC	TGC	TGC	TCC
Q	S	S	C	C	R	P	C	C	S

K4

B

<u>Human gene</u>									
Q	S	S	C	C	K	P	C	C	S
CAG	TCC	AGC	TGC	TGT	AAG	CCC	TGT	TGC	TCC
***	**	***	***	**	* *	***	**	***	
CAG	TCT	AGC	TGC	TGC	AGG	CCC	TGC	TGC	
Q	S	S	C	C	R	P	C	C	

Mouse Gene

C

<u>Human Gene</u>									
Q	S	S	C	C	K	P	C	C	S
CAG	TCC	AGC	TGC	TGT	AAG	CCC	TGT	TGC	TCC
***	**	***	***	**	**	**	*	***	*
CAG	ACC	AGC	TGC	TGC	CAG	CCA	ACC	TGC	CTC
Q	T	S	C	C	Q	P	T	C	L

B2A High-Sulphur Gene

Figure 6.3

21 Residue Cysteine-Rich Repeat

A match of the human UHS gene consensus cysteine-rich decapeptide repeat to the sheep UHS gene consensus cysteine-rich 21 residue repeat. Best homology was observed when the decapeptide was aligned with the first 10 residues of the 21 residue repeat. Nucleotide homology in this region between the two kinds of cysteine-rich repeat is about 65%.

Human Gene Consensus Decapeptide Repeat

Q	S	S	C	C	K	P	C	C	S
CAG	TCC	AGC	TGC	TGT	AAG	CCC	TGT	TGC	TCC
	***	***	***	***	*	***		***	* *

GGC	TCC	AGC	TGC	TGT	GTG	CCC	GTG	TGC	TGC	TGT	GTG	CCA	GCC	TGC	TCC	TGC	TCC	AGC	TGT	GGC
G	S	S	C	C	V	P	V	C	C	C	V	P	A	C	S	C	S	S	C	G

Sheep Gene Consensus 21 Residue Repeat

The cysteine-rich decapeptide repeat of the human gene contained strong homology to that of K4 (figure 6.2). At the nucleotide level the homology was observed to be in excess of 80% (figure 6.2).

The coding region of the mouse presumptive UHS gene, which has not been discussed previously, was found to contain cysteine-rich nonapeptide repeats (figure 6.7). The mouse nonapeptide repeat resembles the human UHS gene decapeptide cysteine-rich repeat. When pairwise comparisons are made between the human gene consensus cysteine-rich decapeptide and the mouse nonapeptide repeats, an average of 60% nucleotide homology was observed (figure 6.2).

The repeat which has been observed in the B2 proteins (Elleman *et al.*, 1973) was found to have an average of 52% homology when pairwise comparisons were made to the human gene decapeptide repeat (figure 6.2).

The cysteine-rich decapeptide repeats present in the human UHS gene and K4 appear to be absent from the sheep UHS gene. Instead this gene contains a longer 21 residue cysteine-rich repeat which, apart from length, differs from the shorter cysteine-rich repeats of the other genes in at least two other respects: (a) the sheep repeat contains the hydrophobic amino acids valine and alanine; and (b) the sheep repeat contains a triplet of cysteine residues in addition to the pair of cysteine residues which are observed in the decapeptide repeats. Maximum homology of 65% is observed if the nucleotide sequence corresponding to the first 10 residues of the sheep gene 21 residue repeat are aligned with the human decapeptide repeat (figure 6.3).

An alignment of the coding regions of the human and sheep UHS genes which is described below (6.5.3) revealed that there is a single iteration of the longer cysteine-rich repeat of the sheep gene within the first half of the human UHS gene coding region (figure 6.5).

It was also found that a sequence homologous to the 21 residue cysteine-rich repeat occurs at the 5' (N-terminal) end of the coding region of K4. When this sequence was compared to the consensus sheep gene repeat, 94% nucleotide homology was observed (figure 6.4).

Figure 6.4

K4 21 Residue Cysteine and Glycine-Rich Repeat Homologies

- A** An alignment of the sheep consensus 21 residue cysteine-rich repeat with a homologous sequence observed in the coding region of K4. The homology in K4 is at the 5' end of the incomplete coding region. 98% nucleotide homology is observed in this region.
- B** An alignment of the sheep gene consensus glycine-rich repeat with the single iteration of this sequence in the coding region of K4. About 77% nucleotide homology is observed in this region.

A nucleotide match between the compared sequences is indicated by a star (*). The positions of these homologies in the K4 coding region is presented in figure 6.10.

																			<u>K4</u>																					
				C	C	V	T		V	C	C	C	V	P	A	C	S	C	S	S	C	G																		
				TGC	TGT	GTG	ACC		GTG	TGC	TGC	TGT	GTG	CCA	GCC	TGC	TCC	TGC	TCC	AGC	TGT	GGC																		
				***	***	***	**		***	***	***	***	***	***	***	***	***	***	***	***	***	***																		
GGC	TCC	AGC		TGC	TGT	GTG	CCC		GTG	TGC	TGC	TGT	GTG	CCA	GCC	TGC	TCC	TGC	TCC	AGC	TGT	GGC																		
G	S	S		C	C	V	P		V	C	C	C	V	P	A	C	S	C	S	S	C	G																		
<u>Sheep Cysteine-Rich Repeat</u>																																								

<u>K4 Glycine-Rich Sequence</u>									
K	G	G	C	S	W	C	G	C	S
AAA	GGG	GGC	TGC	AGC	TGG	TGT	GGC	TGC	TCC
**	***	***	***	**	*	***	**	**	**
AAG	GGG	GGC	TGC	GGC	TCC	TGT	GGA	GGC	TCT
K	G	G	C	G	S	C	G	G	S
<u>Sheep Consensus Glycine-Rich Decapeptide Repeat</u>									

Both the human UHS and K4 proteins contain a single iteration of a sequence homologous to the partial iteration of the 21 residue cysteine-rich repeat observed in the sheep gene (see chapter 5, 5.3.3). These were observed by alignment of the C-terminal regions of the three proteins discussed in sections 6.5.3 and 6.5.4 (see figure 6.6).

6.5.2 Glycine-Rich and Serine-Rich Repeats

Glycine-rich decapeptide repeats are observed in both the human and sheep UHS proteins (see 4.3.5 and 5.3.3 respectively). The alignment of the human and sheep coding regions which is described below (6.5.3) revealed that there is a high degree of homology between the glycine-rich repeats of the human gene and those of the sheep gene, which at the nucleotide level is about 90% (figure 6.5).

A single iteration of the glycine decapeptide repeat occurs in K4 (figure 6.4). When the consensus sequence for the sheep glycine-rich decapeptide was compared to the glycine-rich sequence in K4 about 77% nucleotide homology was observed.

No evidence could be found for the presence of the glycine-rich repeat in the coding region of the mouse UHS gene or in the B2 genes. However, it is of interest to note that most of the glycine residues occur in the cysteine-poor non-repetitive regions of the B2 proteins (Crewther, 1976).

A serine-rich repeat is observed in the coding region of the human UHS gene (4.3.5) (see figure 6.8). Whilst some short regions of homology to this repeat are observed in the sheep UHS gene coding region, there do not appear to be any full length iterations of this repeat in either the sheep or mouse UHS genes, K4, or the high-sulphur B2 genes.

6.5.3 An Alignment of the Human and Sheep UHS gene coding regions

Of the three complete UHS gene sequences that are available, the human and sheep UHS genes contain the most striking positionally conserved homology. This is best demonstrated by aligning the coding regions of these two genes. When this was conducted (figures 6.5 and 6.6) it revealed that homology is most extensive in the N-terminal and C-terminal regions, with the central regions of both coding regions being more divergent.

Figure 6.5

Alignment of N-Terminal (5') Halves of the Sheep and Human UHS Gene Coding Regions

Alignment commences at the N-terminal residue of the human and sheep UHS proteins. A deletion of 7 amino acid residues (21 nucleotides) was introduced to the human gene to maximize the alignment of the N-terminal (5') end of the coding regions. A match between the respective nucleotide sequences is indicated by a star (*). The glycine-rich repeats present in both sequences are boxed in blue, and the sheep type cysteine-rich repeat is outlined in brown. The cysteine-rich decapeptide and serine-rich nonapeptide repeats present only in the human UHS gene are outlined in red and green respectively. The central regions of the coding regions where homology is not positionally conserved with respect to the beginnings and ends of the coding regions have not been included.

N-terminal Region

Sheep Gene:

	10		20		30
G C S G C	S G G C G S S C G G C	G S R C G G C	S S S C C V P		
GGCTGCTCTGGCTGCT	CCGGAGGCTGCGGCTCCAGCTGTGGGGGCTGCGGCTCCAGGTGTGGGGGCTGCGAGCTCCAGCTGCTGTGTGCCC				
*****	*****	*****	*****	*****	*****
GGCTGCTGTGGCTGCT	CCGGCGGCTGTGGCTCCAGCTGTGGAGGCTGTGACTCCAGCTGTGGGAGCTGTGGCCCCAGCTGCTGTGCACCC				
G C C G C	S G G C G S S C G G C	D S S C G S C	G P S C C A P		

Human Gene:

	40		50		60
V C C C K P V C C C V P A C S C S S C G	K G G C G S S C G G				
GTCTGCTGCTGCAAGCCCGTGTGCTGCTGTGTGCCAGCCTGCTCCTGCTCCAGCTGTGGCAAGGGGGGCTGTGGCTCCAGCTGTGGGGGC					
*****	*****	*****	*****	*****	*****
GTCTACTGCTGCAAGCCCGTGTGCTGCTGTGTTCCAGCCTGTTCTGCTCTAGCTGTGGCAAGCGGGGCTGTGGCTCC---	TGTGGGGGC				
V Y C C K P V C C C V P A C S C S S C G	K R G C G S				

GGCTGCAGGGGCTGTGGCTCT
G C R G C G S

	70		80		90
S K G G C G S C G G S	K G G C G S C G G C	G S S C C K P V C			
TCCAAGGGGGGCTGCGGCTCCTGTGGGGGGTCTAAGGGGGGCTGCGGCTCATGTGGAGGCTGCGGCTCCAGCTGCTGCAAGCCCGTGTGC					
*****	*****	*****	*****	*****	*****
TCCAAGGGAGGCTGTGGTTCTTGTGGCTGCTCCAGTGCAGTTGCTGCAAGCCCTGCTGTGTGTTCTTCAGGCTGTGGGTGTCATCCTGCTGC					
S K G G C G S C G C S	Q C S C C K P C C C	S S G C G S S C C			

Figure 6.6

Alignment of the C-terminal Regions

An alignment of the C-terminal regions of the sheep, human and K4 UHS proteins. The C-terminus of the protein is indicated by a dash (-). A match of either the human or K4 proteins to the sheep protein is indicated by a star (*). The identities of mismatched residues are as indicated. The sequences are outlined as follows:

- Red: Cysteine-rich decapeptide repeats
- Blue: Glycine-rich decapeptide repeats.
- Brown: The partial iteration of the 21-residue cysteine-rich repeat. These sequences in the three proteins can be aligned with the first ten residues of the 21-residue repeat.
- Black: Non-repetitive region of homology shared between the three proteins.

Below are the nucleotide sequences encoding the above protein sequences.

C-Terminal Region

Peptide Sequence

<u>Sheep:</u>	K G G C G S C G C S	Q S S C C V P V C C	Q R K I	-
<u>Human:</u>	Q S S C C K P C * *	* * R * * * * *	* * * *	-
<u>K4:</u>	* * * * * R * * * *	* * * * * A * *	* * R D L R C	-

Nucleotide Sequence

<u>Sheep:</u>	AAAGGGGGCTGCGGCTCCTGTGGCTGCTCC	CAGTCCAGCTGCTGTGTCCCCGTGTGCTGCC	CAGCGCAAGATCTGAGGCTCTGCC
<u>Human:</u>	CAGTCCAGCTGCTGCAAGCCCTGCTGCTCC	*****A*****T*****A***T*****AGT	
<u>K4:</u>	*****GA*****	*****C*****T*CT*****T*****	GAGATCTGAGATGCTGA

Within the first 70 amino acid residues 58 are identical and at the nucleotide level 87% homology occurs (figure 6.5). Two features of interest were found in this region. Firstly, the three glycine-rich repeats found in the human UHS gene (4.3.5) were aligned with the first three glycine-rich repeats of the sheep gene (5.3.3), indicating that the glycine-rich repeats of both genes are almost identical. Secondly, a single 21 residue type cysteine-rich repeat was found in the human UHS protein (figure 6.5). This corresponds to the first cysteine-rich repeat observed in the sheep UHS protein and includes a 6 amino acid (18 bp) internal duplication (5.3.3).

Downstream from the first N-terminal 70 residues, homology between the conceptual proteins breaks down and in the sheep protein the 21 residue cysteine-rich repeats are interspersed with glycine-rich repeats. In the divergent part of the human protein, decapeptide cysteine-rich repeats are present and are interspersed with serine-rich nonapeptide repeats. These variations account for most of the difference in the lengths of the sheep and human UHS conceptual proteins which are 181 and 168 residues respectively.

Another region of strong homology is observed towards the C-terminus of the conceptual proteins, with 13 of the last 16 residues of both being identical (figure 6.6). Nucleotide homology in this region is about 95%, and extends 9 bp beyond the end of the coding regions (figure 6.6). Within this region a single incomplete iteration of the 21 residue cysteine-rich repeat occurs in both proteins (figure 6.6). These iterations correspond to the first ten residues of the full length repeat.

Of particular note is that the last amino acid residue of both the human and sheep gene conceptual protein is isoleucine. In this respect these two proteins are different to all the high-sulphur proteins (Gillespie, 1983) and also the mouse UHS and K4 conceptual proteins, which have a cysteine residue at the C-terminus.

6.5.4 K4 Coding Region Structure

The incomplete coding region of K4 contains sequences that are similar to those observed in both the human and sheep UHS genes. Upstream from the cysteine-rich decapeptide repeats observed in the coding region of K4, single iterations of sequences

homologous to both the 21 residue cysteine-rich and glycine-rich repeats occur (see figure 6.10). These have been discussed in sections 6.5.1 and 6.5.2

The strong homology observed between the C-terminal regions of the human and sheep UHS proteins is also present in K4 (figure 6.6). This includes the homology to the first 10 residues of the 21 residue repeat observed in both the sheep and human UHS proteins. Homology to the human UHS gene in this region is more extensive than to the sheep UHS gene, with both K4 and the human UHS gene containing a cysteine-rich decapeptide repeat in similar positions (figure 6.6). The sheep UHS gene contains a glycine-rich decapeptide repeat in the equivalent position.

K4 differs from the sheep and human UHS proteins in that its last residue is cysteine not isoleucine.

The overall structure of the K4 coding region is presented in figure 6.10.

6.5.5 Mouse UHS gene

The mouse UHS gene coding region differs from the sheep and human UHS proteins and K4 in that its coding region is composed almost entirely of short nona- and pentapeptide cysteine-rich repeats (see figure 6.7). These repeats are homologous to the decapeptide cysteine-rich repeats of the human UHS gene and K4 (see section 6.5.1). The mouse gene does not appear to contain either the decapeptide glycine-rich or 21 residue cysteine-rich repeats that are present in the coding regions of the other UHS genes. Hence the structure of the coding region of the mouse UHS gene is quite different to those of the other two UHS genes and K4.

6.5.6 Summary of Coding Region Sequence Comparisons

Of the four UHS sequences available, the human and sheep genes and K4 have the most similar coding regions. These similarities may be summarized as follows:

- (A): All contain a decapeptide glycine-rich sequence, and this sequence is reiterated in the human and sheep genes.
- (B): Both the human gene and K4 contain sequences which are homologous to the sheep gene 21 residue cysteine-rich repeat.

(C): The human and sheep UHS gene coding regions are almost identical for the first 70 codons. Strong homology is also observed towards the 3' (C-terminal) ends of the coding regions of these genes, and this is shared by K4.

However, some differences were also observed between these genes:

(A): The sheep UHS gene does not appear to contain the cysteine-rich decapeptide repeats that are present in both the human UHS gene and K4.

(B): The human gene contains a serine-rich repeat which does not appear to be present in any of the other UHS genes.

The structures of the human and sheep UHS genes and K4 are summarized in figures 6.8-6.10.

Since the sheep and human UHS genes and also K4 contain extensive homology, it is possible that they evolved from a common ancestral gene. Rearrangements in the coding regions of each may have occurred through homologous recombination between repeated elements in different genes and/or in different regions of the same gene (for review of possible mechanisms for duplication and rearrangement see Maeda and Smithies, 1986). The presence of extensive homology between these genes is also indicative of there being selective pressure to maintain these regions, perhaps being essential for the successful functioning of the gene product.

The mouse UHS gene coding region (figure 6.7) is quite dissimilar to the coding regions of the other two UHS genes and that of K4. This may result from specific differences between mice, humans and sheep. Alternatively, there may be more than one family of UHS proteins in both wool and hair. The mouse UHS gene may encode a protein from a different family of UHS proteins to those encoded by the human and sheep UHS genes and K4. The possibility of more than one UHS protein family will be examined in more detail in the next chapter.

6.6 Repeats and Protein Function

The presence of repeats in the conceptual UHS proteins is a feature that appears to be common amongst keratin proteins. Examples include the inner root sheath protein

Figure 6.7

Mouse UHS gene

The nucleotide sequence of the presumptive UHS gene coding region and some 5' and 3' noncoding sequence. The predicted translation product is written above the nucleotide sequence and contains about 37% cysteine (table 6.1). Both penta- and nonapeptide repeats were observed in this gene. The pentapeptide repeats could be aligned with the first five residues of the nonapeptide repeats. Both types of repeat have been outlined. Only the nonapeptide repeats were used for intergenic comparisons. This sequence was obtained from Dr. A. McNab (unpublished results).

TTTCCTCCTGAAAAACAAAACAAAAACAAACAAATTAAGGAAAAGTCCTGCTTATTGGTCCCTTCATGAGATGTCAAGGGTATAAAAGCCCAAGGTCTGAAACAGTTGTCAAAATCTCAGAA 120

AACTCCTGTGAACACAACCTCCTCCCTCCACATCCAACACCATGACCAACTCCTGCTGCTCCCTTGCTGCAGCCACCTGCTGCAGGACCACCTGCTGCAGGACCACCTGCTGGAGACCA 240

M	T	N	S	C	C	S	P	C	C	Q	P	T	C	C	R	T	T	C	C	R	T	T	C	W	R	P
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

AAGCTGTGTGACCAGCTGTTGCCAGCCCTGCTGCAGCCAGCTGCTGTGGCTCTAGCTGCTGTCAGCCTTGCTGTCAAACCACTTGCTGCAGGACCTGCTTCAGCCATGCTGTGTGAGC 360

S	C	V	T	S	C	C	Q	P	C	C	Q	P	S	C	C	G	S	S	C	C	Q	P	C	C	Q	T	T	C	C	R	T	C	F	Q	P	C	C	V	S
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

AAGCTGCTGCCGACACCCCTGCTGCCAGCCCTGCTGCTGTGTGTCCAGCTGCTGCCAGCCCTGCTGCAGCCAGCTGTTGTGAGTCCAGCTGCTGCCAGCCTAGGTGCTGTGAGTCCAGC 480

S	C	C	R	T	P	C	C	Q	P	C	C	C	V	S	S	C	C	Q	P	C	C	Q	P	S	C	C	Q	S	S	C	C	Q	P	R	C	C	E	S	S
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

TTGCTGCCAGCCAGGTGCTGTATCTCCAGCTGCTGCCAGCCCTGCTGCAGGCCTAGCTGCTGTGAGTCTAGCTGCTGCAGGCCCTGCTGCAGCCCTTCTGCCTCAACCTGTGCTGCCAG 600

C	C	Q	P	R	C	C	I	S	S	C	C	Q	P	C	C	R	P	S	C	C	Q	S	S	C	C	R	P	C	C	Q	P	F	C	L	N	L	C	C	Q
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

CCCAGCCTGCTCTGGACCTGTGACCTGCACCAAAGACTTGCTACCAGCCAACATGTGTTTGTGTCCTGTTGCCTGTGCCAAGGCTGTGGGTCCAGCTGTTGTGAGCCCTGTGGCTGTTGA 720

P	A	C	S	G	P	V	T	C	T	K	D	L	L	P	A	N	M	C	L	C	A	C	C	L	S	Q	G	C	G	S	S	C	C	E	P	C	G	C	*
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

TTGAATTATTCTAATGGAATAGAATAATTTAGTCTCAGACCTGCTGGAGTTTTATTTTACCCAAACTCCCTGCCCTTTCTCTGCCTCTAATACACACACAATCATCAAATTCTCAGTCCT 840

CCCTTTAAGAGATCTGTGAACCAACTTACCAATTTTGATGTTTTCTCCATTTCTCTCCTGTGAGTTTCATCTATTAGAACTGGATTTCATGCTCTCAGCCTTGCCAGACACCATTCCTTA 960

Figure 6.8

Human UHS Gene

A summary of the repeat sequences present in the human UHS gene. The cysteine-rich decapeptide repeats are boxed in red, the glycine-rich repeats in blue. Sequences homologous to the sheep UHS gene 21 residue cysteine-rich repeat are outlined in brown.

The sequence in the 5' noncoding region which is homologous to the 5' noncoding region of the sheep UHS gene (figure 6.9) is outlined in black.

AATTCCACAAGGAAATCATCTCAGGAGGAAGGGCTCATACTTGGATCCAGAAAATATCAACATAGCCAAAGAAAAACAATGACATACCTCCAGGAGCTGTGTAAACAGCAACCGGAAAGAG 120

AAACAATGGTGTGTTCTATGTGGATATAAAGAGCCGGGGCTCAGGGGGCTCCACACCTGCACCTCTTCTCACCTGCTCTACCTGCTCCACCCTCAATCCACCAGAACCAATGGGC 240

C C G C S G G C G S S C G G C D S S C G S C G S G C R G C G P S C C A P V Y C C 360
TGCTGTGGCTGCTCCGGAGGCTGTGGCTCCAGCTGTGGAGGCTGTGACTCCAGCTGTGGGAGCTGTGGCTCTGGCTGCAGGGGCTGTGGCCCCAGCTGCTGTGCACCCGTCTACTGCTGC

K P V C C C V P A C S C S S C G K R G C G S C G G S K G G C G S C G C S Q C S C 480
AAGCCCGTGTGCTGCTGTGTTCCAGCCTGTTCTGCTCTAGCTGTGGCAAGCGGGGCTGTGGCTCCTGTGGGGGCTCCAAGGGAGGCTGTGGTTCTTGTGGGCTGCTCCAGTGCAGTTGC

C K P C C C S S G C G S S C C Q C S C C K P Y C S Q C S C C K P C C S S S G R G 600
TGCAAGCCCTGCTGTTGCTCTTCAGGCTGTGGGTCATCCTGCTGCCAGTGCAGCTGTGCAAGCCCTACTGCTCCAGTGCAGCTGCTGTAAGCCCTGTTGCTCCTCCTCGGGCTGTGGG

S S C C Q S S C C K P C C S S S G C G S S C C Q S S C C K P C C S Q S R C C V P 720
TCATCCTGCTGCCAATCCAGCTGCTGCAAGCCCTGCTGCTCATCCTCAGGCTGTGGGTCATCCTGCTGCCAGTCCAGCTGCTGCAAGCCCTGCTGCTCCAGTCCAGATGCTGTGTCCCT

V C Y Q C K I * 840
GTGTGCTACAGTGAAGATCTGAGGCTCTAGTGGGAAACCTCAGGTAGCTCCCGAAGATCTGTGCTTTCCAACAAGTACTACCTTGAAGCACATCCCCTTCTGAATCTGAAAAGAGC

CTTGGCTCAGGGCGTCTTTTCCAGCCCCTGAGGAAATGGAATGAACCACTCCCTGCCCATTCCTATAAGAATATCCAAGACCCAGGCAATTTTGGCCCTCTTCCACATGCCCCCA 960

TATGTCTGAGCCAACTGCACTGGGGGCTGCCTAGTCCAAGCAAGAGCCTGGGAATT 1015

Figure 6.9

Sheep UHS Gene

The repeats present in the coding region of this gene have been outlined as follows:

- Brown: Cysteine-rich repeats approximately 21 amino acid residues in length. There is some variation in the length of the iterations of this repeat (see chapter 5).
- Blue: Glycine-rich decapeptide repeats.

The sequence in the 5' noncoding region which is homologous to the 5' noncoding region of the human UHS gene (figure 6.8) is outlined in black.

AGGTTGCTAACTACGTCATGAGTAGTGATGCATCACCATCTCTCTCTGTACCGCCTTCTCTCTTGCCTAGTTTTCCCCAGCATCAGTCCATTTACCAATGAGTCAGATGAGGGGATT 120
 GGGGCTAGATCCTTCTTATCTGGATATTAAAGGAACTCGCAGAGGCTCGAGTAAAAAGAGGCAAAATCGCCGACATACTCGTTCAACGCCTAGGAGCCAAGCAGAAACCACCTGCAGGTG 240
 ACAGGCTCACAGGCAACCTGCTTTCTTGGTTTTTTTGTTTTTTTTTTAACTGTTCCAGTAAAGACCAAGAATGATCTTTAACTCTACTCATGTAAGAGTGACACCATCGTGGAGAC 360
 GCCAGCACAGCGAAGGCAACAGCGCCGCTGGGACTCCGCCATCTGTAGGTGATCTGAGGTGAGCAGAGAGGCCAGGGGACCAGTGCCCTGGAGGCCCGGGGAAGCGCCAAAGCCTGAG 480
 GTCCTAAGTACAGGGGCAAGAAAAGGGGAAGAGGACAAAGAGGGGACAAAGCAAGCGGATCTTCTCTCTGAGCCAGGAGGAAGGCAGCAGCGAAAGGGAAAAACAGGGCTGGAGAAATCA 600
 AAACCAGTCACGAGGAAGTGAGCATGTGCCATCACCACACCTGACCCGGCACAGGTATATAGAGGCTCCCGGCCCGGAGGCCACACTGAGCCGCTTCTCTCTCTCCACCTGCTCT 720
 CAP Site
 M G C S G C S G G C G S S C G G C G S R C G G C S S S C C V
 CTGACCTACTCCACCTCAACCCACCAGAACCATGGGCTGCTGGCTGCTCCGGAGGCTGCGGCTCCAGCTGTGGGGGCTGCGGCTCCAGGTGTGGGGGCTGCACTCCAGCTGCTGTG 840
 P V C C C K P V C C C V P A C S C S S C G K G G C G S S C G G S K G G C G S C G
 TGCCCGTCTGCTGCTGCAAGCCCGTGTGCTGCTGTGTGCCAGCCTGCTCTGCTCCAGCTGTGGCAAGGGGGCTGTGGCTCCAGCTGTGGGGGCTCAAGGGGGGCTGCGGCTCTGTG 960
 G S K G G C G S C G G C G S S C C K P V C C C V P A C S C S S C G K G G C G S C
 GGGGGTCTAAGGGGGGCTGCGGCTCATGTGGAGGCTGCGGCTCCAGCTGCTGCAAGCCCGTGTGCTGCTGTGTGCCAGCCTGCTCTGCTCCAGCTGTGGCAAGGGGGGCTGCGGCTCT 1080
 G G S K G G C G S C G G S K G G C G S C G G C G S G C G S S C C V P V C C C V P
 GCGGGGGTCTAAGGGGGGCTGCGGCTCTGCGGGGGTCTAAGGGGGGCTGCGGCTGCTGTGGGGGCTGTGGCTCTGGCTGCGGCTCCAGCTGCTGTGTGCCGTGTGCTGTGTGTC 1200
 A C S C S S C G K G G C G S C G C S Q S S C C V P V C C Q R K I *
 CCGCTGCTCTGCTCCAGCTGTGGAAGGGGGGCTGCGGCTCTGTGGCTGCTCCAGTCCAGCTGCTGTGTCCCGTGTGCTGCGCAGCGCAAGATCTGAGGCTCTGCCCAGACCTTG 1320
 GGTCCCTCATGTTTGGAGTCTTATCGTCTCATGTTCTCCGGCATCAGTCTCTTCCACACACCCTCCCTGATTTACCCACAGCCTCCCTCCTCTCGCTCAGGTACACGTCTCCCTCCA 1440
 GGTTACAGCACTGGGGCCTGTGCTCACCTGCTCTCCAGGGAGATGTGTATTTCTGGTCACAGGAGACACATCTCCTTAGAATCAGGTGGCCCTCCAGAGTCACGTGTGGGTACCTTCTG 1560
 CTCTTCTGGGCCAGCTGGGACCCTGCAAGCCAGGGCACGTTCTGAGTGATGGCGAAGGACCCTCTTCCCTGTGCCCTGCTCCCTACGAGAGCTTCTTCTTCTCTCTGCTGCT 1680
 CCAGCTCCCTCTGTTATAAAAAATAACGACTGAATCCCTCTAATAAACATGTCCCATCTCTGACGCCGACTGTGACTGCCTTCTTTCACCCTGTTTCTTGGGTGGTGACAATTGGTGC 1800
 Poly(A) Site
 TTCTTTGTAACGTGCTCTGGGATCC 1824

Figure 6.10

K4

A summary of the repeats present in the sheep UHS cDNA clone K4. The cysteine-rich decapeptide repeats already observed (see chapter 1, 1.6) in the coding region of this clone are boxed in red. The regions homologous to the sheep UHS gene 21 residue cysteine-rich and decapeptide glycine-rich repeats are boxed in brown and blue respectively.

C C V T V C C C V P A C S C S S C G K G G C S W C G C S Q S S C C R P C C S Q S
CTGCTGTGTGACCGTGTGCTGTGTGTGCCAGCCTGCTCCTGCTCCAGCTGTGGCAAAGGGGGCTGCAGCTGGTGTGGCTGCTCCAGTCCAGCTGCTGCAGACCCTGCTGCTCCAGT 120

S C C R P C C S Q S S C C R P C C S Q S S C C V P A C C Q R R D L R C *
CCAGCTGCTGCAGACCCTGCTGCTCCAGTCCAGCTGCTGCAGACCCTGCTGCTCCAGTCCAGCTGCTGCGTCCCTGCTTGTGTGTCAGCGCAGAGATCTGAGATGCTGAGCTCAGACCT 240

CAGGCGGTTCCAGCCAGTCTCAGCTCCCTGACTGTGAGCAGTGACACGTCCAGACCCCAAGGGCACACTCCATTCTCTCCAAGGTAGTGAAGTCTGAATCTTCGGGACCAGAGATCCTGT 360

TCCACACTTCTGATGGAGAACCCCTTCCCTTGTACTGCCTGAGCCATCTGAGGATTGACTGCCACTTCAATCCAAGAGAAGTTCCTTCTTGGCAGCTGAACTCCACTGCAGACGC 480

CTTAGGGGCTAAACCACAGAAACCACTCAACCTCTGATGGGCACAGAGCTGTTCCAACCATTTCTTGGCTTCTTGCTGCCTAGCCTTGGTTCTCACATAACCTCAGGGCTGGTGTGACC 600

TCTCGGGCTGGGCCCCAAATAATCCTCCCTCTTCTGTGACTCCACCTCCTTCTGATCTCGGTCTAGCAAATCTACTCTGGGCTTGTGAAGCCGCTAAATAAATAGACTGTCTTCAAC 720

trichohyalin (R.B. Presland and M.J. Fietz, personal communication), and the epidermal keratins filagrin (Lonsdale-Eccles *et al.*, 1984), and involucrin (Eckert and Green, 1986). Many other proteins, both structural and catalytic, contain repeats (for examples see CS protein, Dame *et al.*, 1984; yeast adenylate cyclase, Karatako *et al.*, 1985; eukaryotic RNA polymerases, Allison *et al.*, 1985 and Corden *et al.*, 1985; apolipoproteins, McLean *et al.*, 1987).

Ohno (1984) suggests that all coding sequences evolved from repeats of base oligomers, which he believes confers upon the protein product the greater likelihood of assuming a periodical secondary structure, such as α -helices or β -sheets. This is very likely to be the case with structural proteins such as the UHS proteins. The next section presents the results of computer investigation of the possible structure of the human and sheep conceptual UHS proteins.

6.7 Protein Structure

At present there is no direct evidence bearing on the structure of any of the high-sulphur proteins (Gillespie, 1983). The frequent occurrence of cysteine residues in these proteins has rendered structural studies difficult, since the necessary alkylation of these residues may have a disordering effect (Gillespie, 1983). Following is a summary of theoretical predictions of possible structures that the human and sheep conceptual UHS proteins may form. These predictions may be of use in determining what features of the proteins should be investigated further by direct experimentation.

Presented in figures 6.11 and 6.12 are plots of hydrophobicity and charge distribution, on the conceptual sheep and human UHS proteins generated using the "Analsep" programme from the Standen package. Also, these sequences were analysed using the "Predict" computer programme to predict any possible secondary structure and the results of these are presented in figures 6.11 and 6.12.

In the sheep conceptual protein, the region surrounding the cysteine triplet of the sheep cysteine-rich repeat is strongly predicted to form a β -strand, and is also hydrophobic in nature. The remainder of the cysteine-rich repeat, and the glycine-rich repeat, are

Figure 6.11

Computer Analysis of the Sheep UHS Protein

The results of computer analysis of the conceptual protein of the sheep UHS gene. Top are histograms depicting the probability that the protein will form either: (a) turns or coils, or (b) β -strands, as determined by the "Predict" programme (2.2.4 B). Numbers on the Y axis are the number of programmes in the suite that predict the formation of the structure type. Lines corresponding to J on the Y axis are where the Joint programme predicts the formation of a particular type of structure. The coloured lines below the histograms indicate where the predictions are made in the conceptual protein; glycine-rich repeats are marked in blue, and cysteine-rich repeats are marked in brown (see figure 6.9). The N-terminus is oriented to the left-hand side.

Below are plots of hydrophobicity and charge on this protein. Both plots were generated using "Analysep" from the Staden package using a sliding span of 11 residues. The programme determined hydrophobicity by the method of Kyte and Doolittle (1982).

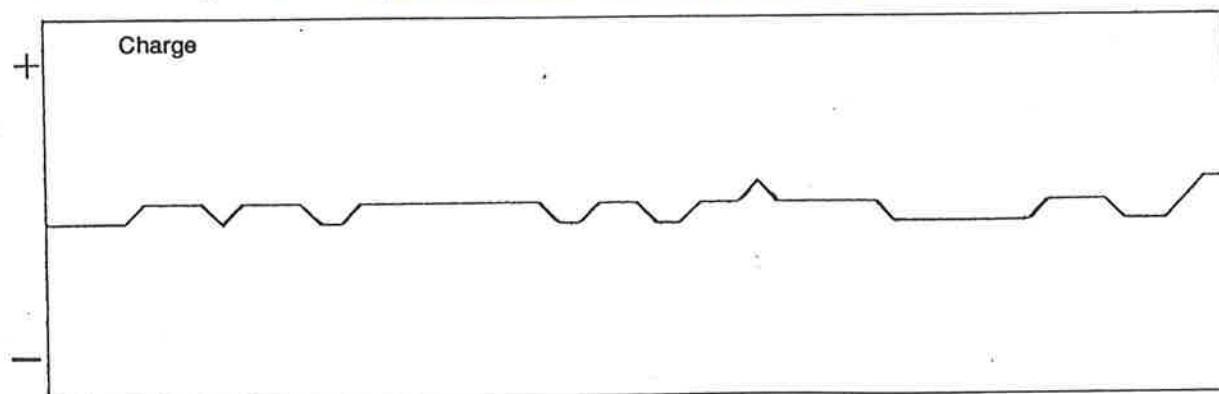
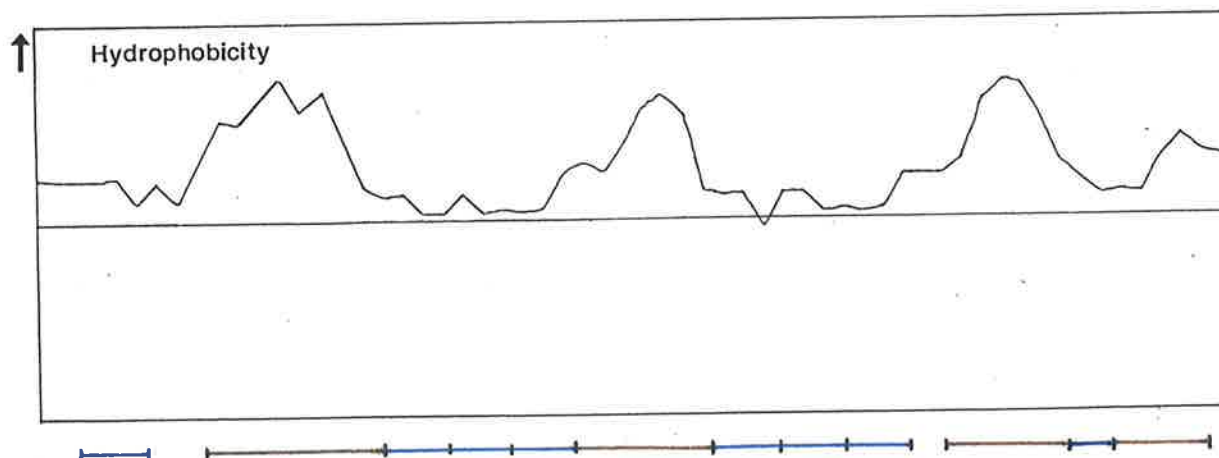
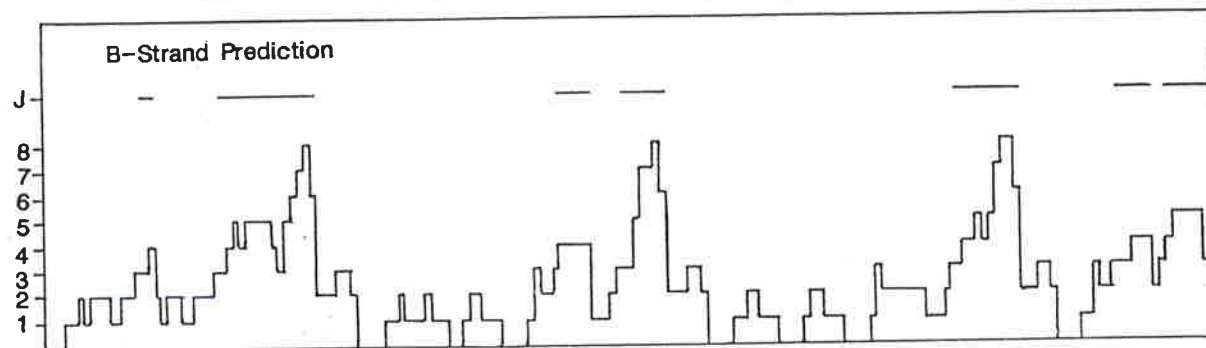
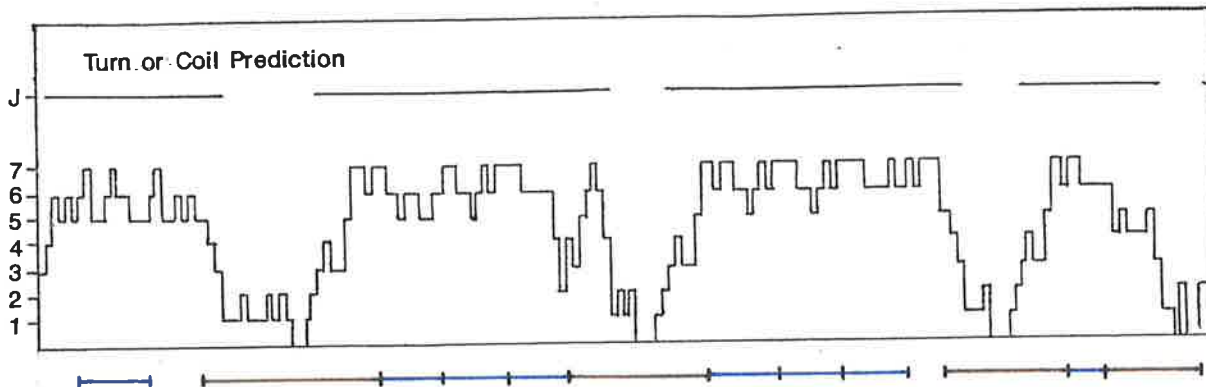
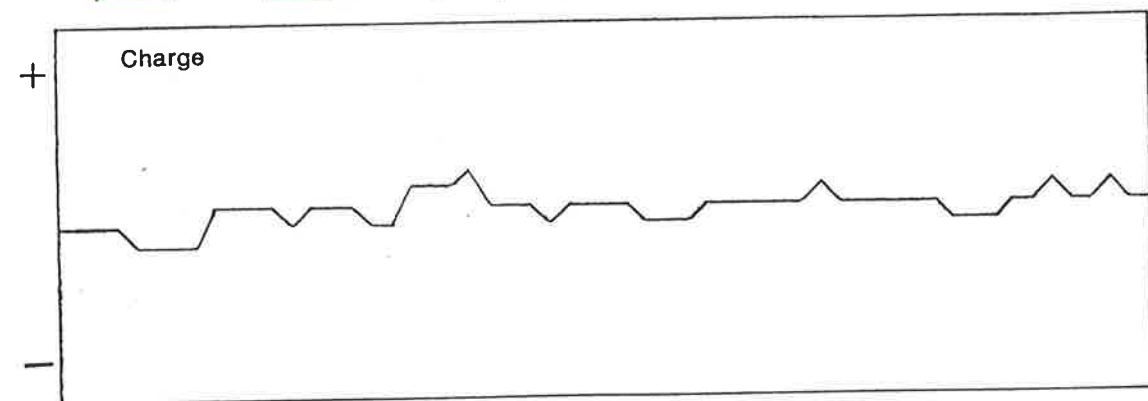
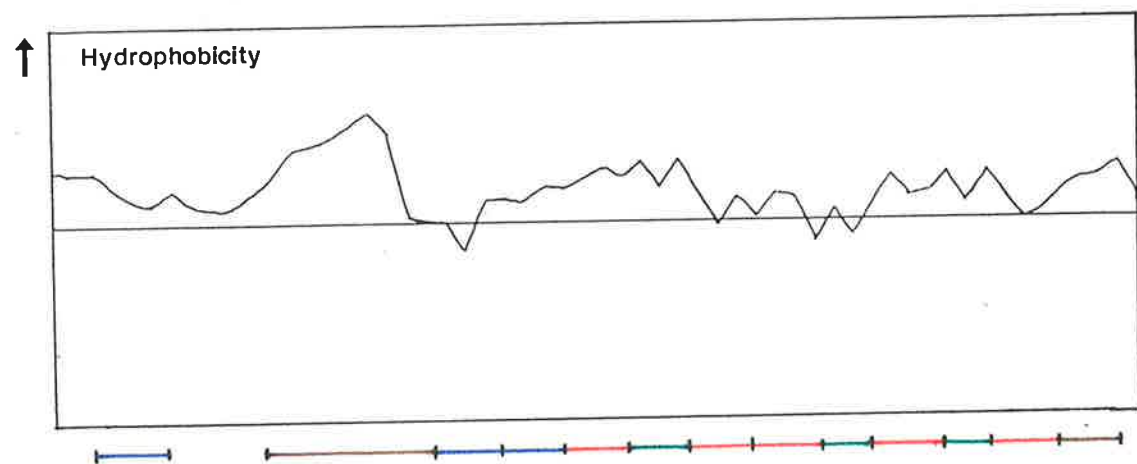
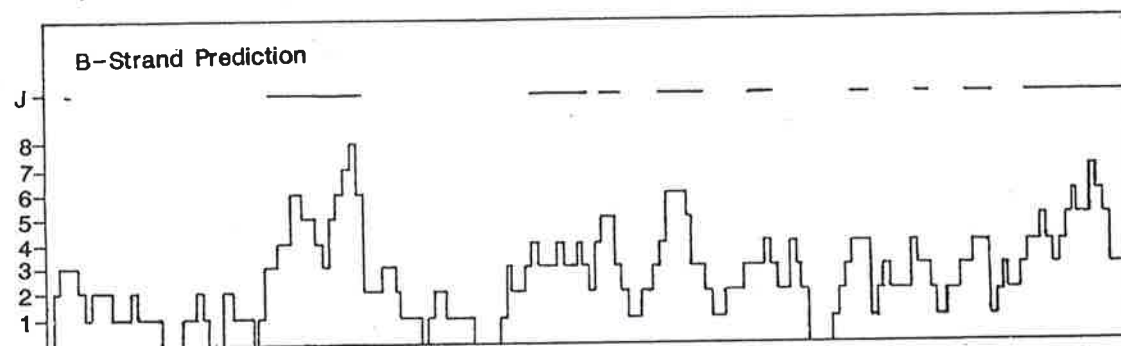
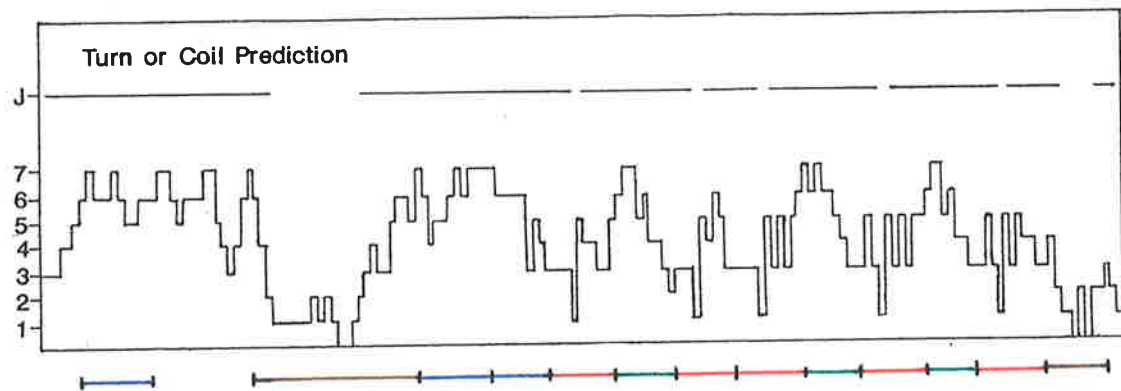


Figure 6.12

Computer Analysis of the Human UHS Protein

Summary of computer analysis of the conceptual protein product of the human UHS gene. The plots types are as described for the sheep UHS gene (figure 6.11). The position of predictions in the conceptual protein are indicated by the coloured lines below the histograms where red indicates the presence of a cysteine-rich decapeptide repeat, blue a glycine-rich repeat, and brown the sheep UHS gene type cysteine-rich repeat (see figure 6.8).



predicted to form turns or coils. Figure 6.11 revealed that there is a relatively even distribution of charge density over the sheep UHS protein.

The significance of these predictions is uncertain, but one possibility is that either inter- or intramolecular disulphide bonds may form between the cysteine triplets, and the formation of these may be facilitated by hydrophobic interaction of the two participating peptide strands. The remainder of the protein may form loops so that the final structure is highly convoluted, and cross-linked.

Results of computer prediction of secondary structure for the human UHS protein are less definite, but moderate predictions of β -strand are made for part of the cysteine-rich decapeptide present in this gene (figure 6.12). The bulk of the protein, including the serine-rich repeats, is predicted to form turns or coils. Since it is observed that homology occurs between the decapeptide repeat of the human gene and that observed in the B2 and BIIIA high-sulphur proteins, it is possible that they are functionally analogous. Hence the human UHS protein decapeptide repeat may form a cross-linked, highly convoluted structure not dissimilar to that suggested by Parry *et al.* (1979).

In the human cysteine-rich repeated sequence, a lysine residue flanks the C-terminal side of the cysteine pair, and in the K4 decapeptide repeat an arginine residue resides in the equivalent position (figure 6.13). In the repeat present in the B2 and BIIIA high-sulphur proteins, usually either an arginine or glutamine residue occurs in this position (Gillespie, 1983). It is also observed that an amino acid substitution of lysine for valine occurs at this position of one of the sheep UHS gene cysteine-rich repeats (chapter 5, 5.3.3). Thus the presence of a positively charged or polar amino acid in this position may have some function in the formation of the secondary structure of these proteins, perhaps through altering the hydrophilicity or polarity of the local environment. Such localized variation may help to determine which cysteine residues are going to pair to form disulphide linkages.

Another possibility has already been discussed in chapter 4 (4.3.6) for the lysine residues of the cysteine-rich decapeptide repeat of the human UHS gene. It was suggested that since both lysine and glutamine residues occur in the cysteine-rich repeat of this gene, it was possible that they form either intra- or intermolecular ϵ -glutamyl- γ -lysine cross-

Figure 6.13

The amino acid sequence surrounding the cysteine pairs in the cysteine-rich decapeptide repeats of the proteins is indicated. Also included are the first 10 residues of one of the sheep cysteine-rich repeats in which a lysine residue has been substituted for the valine residue present in the other iterations of this repeat (see chapter 5, 5.3.3). The amino acid residue flanking the C-terminal side of the cysteine-pair may have a positively charged (Arg and Lys) or polar (Gln) side group.

Sheep Gene: Gly Ser Ser Cys Cys **Lys** Pro Val Cys Cys

Human Gene: Gln Cys Ser Cys Cys **Lys** Pro Cys Cys Ser

K4: Gln Ser Ser Cys Cys **Arg** Pro Cys Cys Ser

B2A: Gln Thr Ser Cys Cys **Gln** Pro Thr Cys Leu

Mouse Gene: Gln Ser Ser Cys Cys **Arg** Pro Cys Cys

links. The presence of such cross-links has been established for proteins occurring in the medulla and inner root sheath (Rogers and Harding, 1976). From a structural viewpoint, the presence of both cystine and ϵ -glutamyl- γ -lysine cross-links would add considerable strength to the fibre.

Experimentation will be required to determine the structure of the UHS proteins. Rather than using UHS protein extracted from wool or hair, chemically synthesized analogue peptides may be of use for *in vitro* studies of UHS protein structure and function. Synthetic peptides would not have to be S-carboxymethylated as is necessary for native UHS protein to allow their purification. Synthetic peptides would have the advantage of allowing the investigation of the effect of systematic alterations in protein sequence upon the structure formed. Possibly, it could be determined using synthetic peptides, whether specific disulphide linkages are formed between cysteine residues in the repeats, and if so which residues interact.

6.8 Summary

In this chapter the human and sheep UHS genes have been compared to each other and also to a mouse presumptive UHS gene, a sheep UHS cDNA clone (K4), and to the B2 high-sulphur genes. The comparisons revealed that there are two distinct cysteine-rich repeats, one of which is a decapeptide, and that variants of this are present in the human UHS gene, K4, and the mouse UHS gene. This repeat appears to be related to cysteine-rich repeats observed in the B2 high-sulphur genes. The longer cysteine-rich, and the glycine-rich repeats, first observed in the sheep coding region, are also present in the human gene coding region, and that of K4. Neither are present in the high-sulphur B2 proteins, or the mouse presumptive UHS gene coding region.

It was found that extensive homology occurred between the human and sheep UHS gene coding regions. This was strongest in the regions corresponding to the N- and C-terminal regions of the UHS proteins encoded by these genes. The homology observed in the C-terminal regions of these proteins was also shared by the K4 protein.

The human and sheep conceptual UHS proteins are unique amongst the high-sulphur proteins in that the C-terminal amino acid is isoleucine and not cysteine.

Structural predictions have been made for the human and sheep UHS proteins and the analyses suggest that β -strands may be formed near the cysteine pairs and triplets observed in these proteins. The remainder of both proteins are predicted to form turns and/or coils, hence it is possible that the proteins form highly convoluted cross-linked structures. The hydrophobicity and polarity of surrounding amino acids may help to determine which cysteine residues form disulphide links. Experimentation is required in order to determine the native structure of these proteins.

The human and sheep UHS genes contain strong homology in the 5' noncoding region, which on the basis of Southern blot analysis may also extend to all related sheep UHS genes. The mouse presumptive UHS gene has some homology to the human and sheep UHS genes in this region. The 18 bp matrix contiguous with the initiation codon of most other wool IFAP genes appears to be lacking from the sheep and human UHS genes.

Chapter 7

UHS Gene Expression

7.1 Introduction

The isolation of the human and sheep UHS genes has enabled an investigation of two ill-defined aspects of UHS gene expression. Firstly, the mechanism by which UHS protein synthesis is elevated in response to increased follicular concentrations of the S-amino acids is unknown. Amongst other possibilities the S-amino acids may act at either the level of UHS gene transcription or at the level of UHS mRNA translation. To investigate these possibilities, preliminary northern transfer and RNA protection experiments were conducted using the human and sheep UHS genes and the sheep UHS cDNA clone, K4, as probes to determine whether or not there were any qualitative or quantitative changes in the UHS encoding mRNA population in response to sulphur-enrichment of the follicle.

Secondly, the histological location of the UHS proteins in the wool and hair fibre and the position in the follicle where UHS protein synthesis takes place is uncertain. Evidence presented in chapter 1 indicates that the UHS proteins may be constituents of the cortex and/or cuticle (1.4.4) and that along with the other high-sulphur proteins, they are synthesized in the more distal regions of the follicle (1.5). In this chapter *in situ* hybridization experiments are described in which the human and sheep UHS genes and K4 were used as probes to locate the corresponding mRNA in human and sheep follicular tissue.

7.2 Results

7.2.1 Sheep Infusion

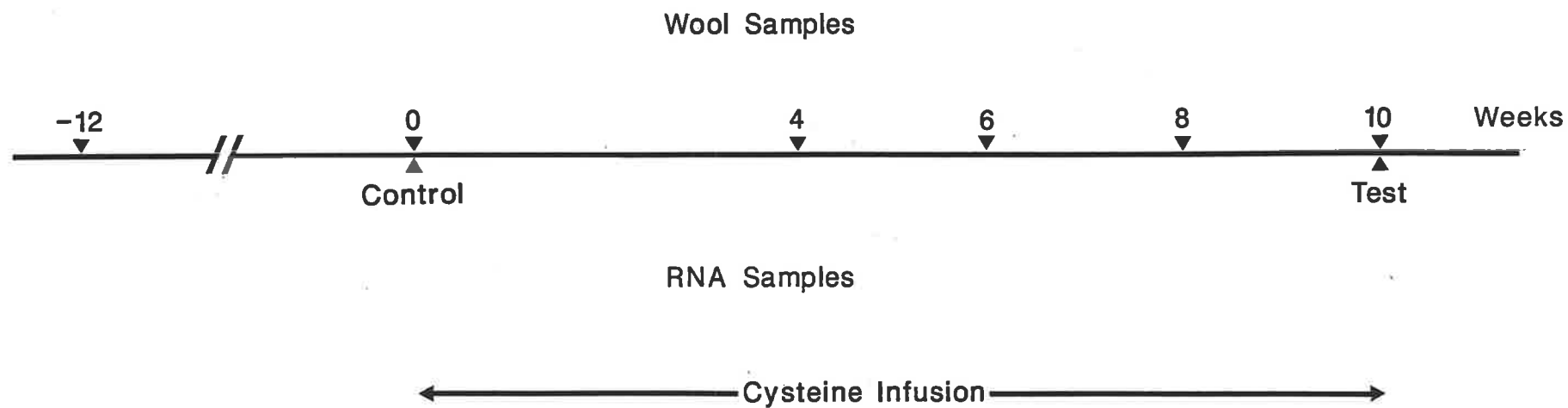
Four sheep were infused *per abomasum* with the equivalent of 3 g/day L-cysteine for a period of 10 weeks. Administration of cysteine at this rate has previously been demonstrated to exert maximum effect on wool sulphur content and growth rates (Reis *et al.*, 1973). The design of this experiment is presented in figure 7.1.

The sheep were held in metabolism cages for the duration of the experiment and were fed on a basal diet of wheaten and oaten chaff. Cysteine-HCl was continuously infused as an aqueous solution for the duration of the experiment. Follicles were taken from one side of each sheep prior to commencement of infusion and from the remaining

Figure 7.1

Design of the Sheep Cysteine Infusion Experiment

Four sheep were infused with cysteine-HCl as described in the text (7.2.1). Wool samples were collected from same position on the side of each sheep at the times indicated. Wool follicles were collected prior to commencement of infusion (week 0) and from the other side of each sheep after 10 weeks of cysteine infusion.



side of each sheep at the termination of infusion. RNA was prepared from both lots of follicles (2.2.3 A). Wool samples were taken as indicated in figure 7.1.

Elemental sulphur analysis was conducted on wool samples taken from two sheep, and the results of these are presented in table 7.1. These indicate that a substantial sulphur-enrichment of the wool of both sheep occurred in response to cysteine infusion.

Two-dimensional PAGE was conducted on the wool samples taken immediately prior to the commencement of cysteine infusion, and on samples taken at the termination of infusion. The results for one sheep are presented in figure 7.2. The remaining three sheep had very similar two-dimensional patterns to this sheep. Quantitative estimates were made on various proteins in the two-dimensional pattern by scanning sections of the fluorograph with a laser densitometer (Zeinen Scanning Densitometer linked to a Varian Vista CDS 401) and the results of these are presented in table 7.2. Exact quantification of individual components in the two-dimensional array was difficult, especially for the high-sulphur proteins because of the complex array of proteins which often overlapped when densitometric scans were undertaken. The intermediate filament proteins as a whole have been shown previously not to vary in response to cysteine infusion (Gillespie *et al.*, 1964). Hence the difference in the amount of IF protein between the pre- and post-infusion two-dimensional gels (ratio 0.77) was taken as a measure of the difference in the total amount of protein loaded onto each gel. Therefore, the estimations of the remaining components including the high-sulphur, UHS and high-glycine/tyrosine proteins in the pre-infusion (control) two-dimensional array were decreased to take the heavier loading on this gel into account.

Three observations of interest may be made on these scans: (a) there is a general increase in the proportions of the proteins lying on the 45° diagonal of between 1.3 (scan c) and 5.0 fold (scan f) in the sulphur-enriched wool; (b) the protein component that comigrates with the sheep UHS gene *in vitro* translation product increases by about 1.4-1.6 fold (scan d and e); and (c) there is an overall decrease in the HGT proteins in the sulphur-enriched wool (scan f). Also, by visual inspection, three components (labelled w) were found to decrease in the sulphur-enriched wool.

Table 7.1

Elemental Sulphur Analysis

Elemental sulphur analysis of selected wool samples from sheep A and C. The results are expressed as percentage sulphur in clean dry wool.

The percentage increase in the amount of sulphur in wool samples taken at week 10 as compared to week 0 is presented at the bottom of the table.

Sulphur analysis was conducted by the Canadian Microanalytical Service.

Sulphur Analysis

Sample	Sheep A	Sheep C
-12	2.58	2.85
0	2.60	2.74
4	3.52	3.56
6	3.65	3.28
8	3.56	3.54
10	3.85	3.42
Increase	48%	25%

Figure 7.2

Two-Dimensional PAGE of Pre- and Post- Cysteine Infusion Wool Samples

At top is the two-dimensional PAGE (2.2.1 F) profile of protein from wool taken from sheep A immediately prior to commencement of cysteine-infusion. The wool was reduced and S-carboxymethylated with [^{14}C]IAA prior to fractionation (2.2.1 E).

At bottom is the profile of wool taken from the same sheep after 10 weeks of cysteine infusion *per abomasum*.

The positions of densitometric scans conducted on these fluorographs are as indicated:

- | | |
|-----------|--|
| Scan a | the type I intermediate filament proteins and transverse scan of UHS diagonal. |
| Scan b | the type II intermediate filament proteins. |
| Scan c | the UHS protein diagonal (almost certainly contains high-sulphur protein also). |
| Scans d&e | replicate scans at slightly different angles to estimate changes in the component that co-migrates with the sheep UHS gene <i>in vitro</i> translation product (indicated by a triangle (Δ)). |
| Scan f | HGT proteins and also a component on the UHS protein diagonal. |

The results of these scans are presented in table 7.2.

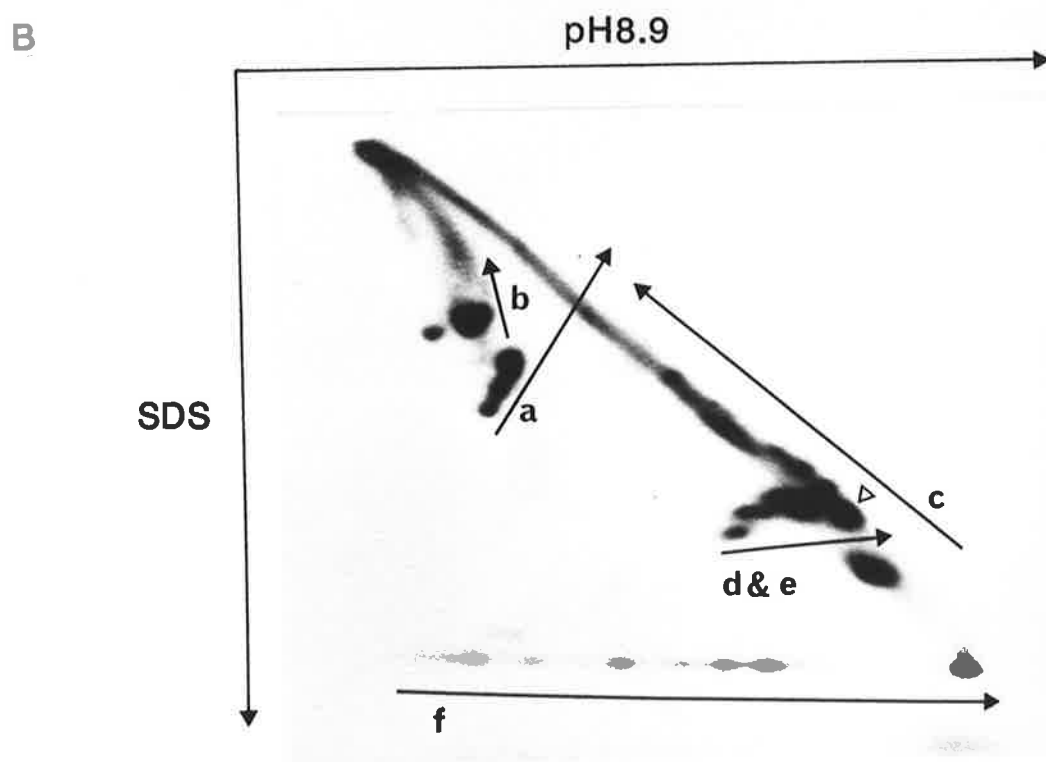
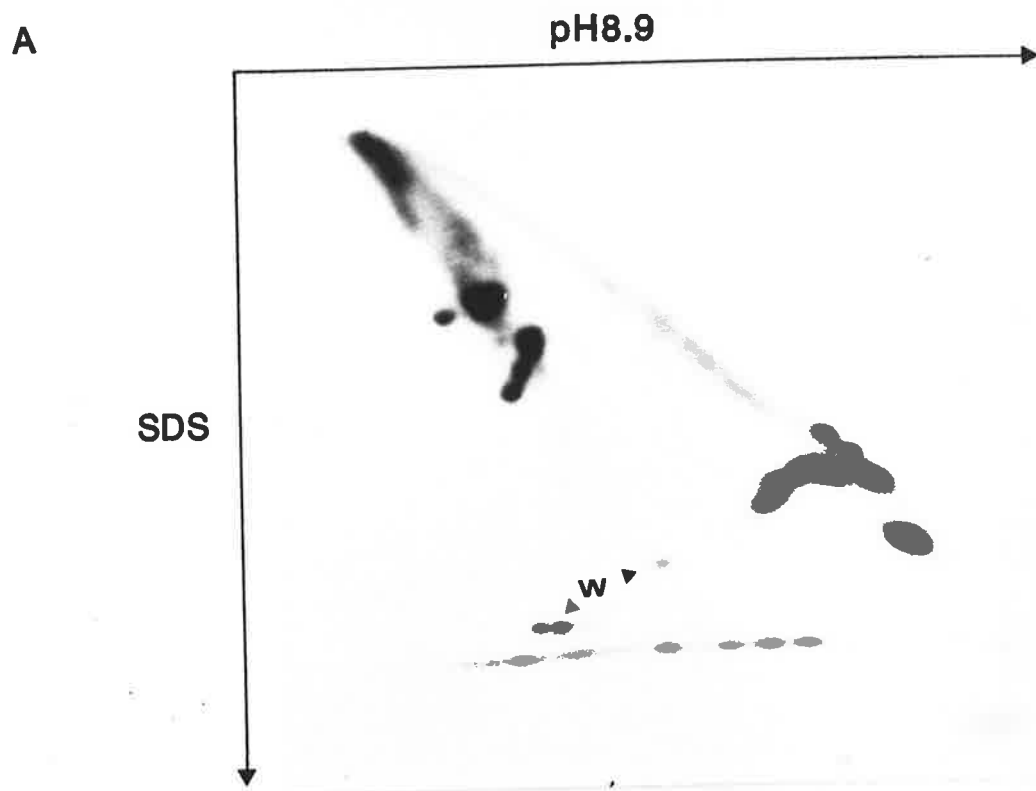


Table 7.2

Densitometric Scans of the Two-Dimensional Gels

The results of estimates of the amount of protein in various regions of the two-dimensional gels on the following three pages. The positions of the scans are indicated in figure 7.2. Individual components were frequently fused with other components into a single peak by the densitometer, making identification of the absorbance of individual components difficult. In scans c, d, e, and f peaks are numbered in the direction that the particular scan was conducted.

As described in the text, the estimates of the proteins on the control two-dimensional scan were adjusted downwards to take into account the greater amount of total protein loaded on this gel. The lower figure for each of the control components is the value corrected for the greater amount of protein loaded onto this gel. Ratios of the components before and at the end of infusion were calculated using the corrected estimates.

Scans d and e are replicates with slightly altered directions to estimate change in the component which comigrates with the sheep UHS gene *in vitro* translation product. Peak 3 in scans d and e correspond to the component which comigrates with the *in vitro* translation product of the sheep UHS gene.

Intermediate Filament Region

Position	Control	Test	Ratio
<u>Scan a</u> Type I I.F.	1467080	1123634	0.77
UHS Smear	157267 120450	441644	3.67
<u>Scan b</u> Type II I.F.	1110060	853062	0.77

High-Sulphur Protein Region

Position	Control	Test	(T/C)
<u>Scan c</u>			
1	4993355 3844883	7682334	2.0
2	1877170 1445421	1916400	1.3
3	156271 120328	468183	3.9
4	127407 98103	153346	1.6
<u>Scan d</u>			
1	893832 688250	691103	1.0
2	462713 356289	495781	1.4
3	732673 564158	776641	1.4
<u>Scan e</u>			
1	954283 734797	777734	1.06
2	447703 344731	413773	1.2
3	659269 507637	809488	1.6

High-Glycine/Tyrosine Region

Position	Control	Test	Ratio
<u>Scan f</u>			
1	169146 129548	88920	0.69
2	203377 155766	44912	0.22
3	141911 108689	49384	0.45
4	125023 95755	12658	0.13
5	139641 106950	56015	0.52
6	176056 134840	109877	0.81
HS/UHS Component	78339 60321	287732	5.0

7.2.2 Northern Transfer Analysis

Northern transfer analysis of follicle RNA isolated before and at the end of the infusion regimen was conducted using coding region probes from both the human and sheep UHS genes. Coding region probes from both genes were identical to those used for genomic Southern and northern transfer experiments described in chapters 4 and 5. Northern transfer experiments in chapter 5 indicated that both probes hybridized only to mRNA of about 1.5 kb in length. In both cases the same filters were probed with a 3'noncoding region probe from the intermediate filament type II gene B (Powell *et al.*, in preparation). Assuming that, like the intermediate filament family as a whole (Gillespie *et al.*, 1964), the expression of this gene does not vary in response to cysteine infusion, it may be used as a control against which the levels of UHS encoding mRNA can be measured.

The levels of hybridization were very similar when the human UHS gene probe was hybridized to equal amounts of pre- and post-infusion RNA from sheep A and B (figure 7.3). More variation was observed in the hybridization levels of the IF probe. This however is almost certainly experimental in origin, as uneven hybridization across the tracks was observed with this probe. Hence UHS probe hybridization levels were not adjusted to take into account the variation observed with the IF probe.

More variation was observed when the sheep UHS gene was used as a probe on RNA isolated from the remaining two sheep (figure 7.4). There was not a consistent trend in this variation, with an increase in the level of hybridization with RNA from one infused sheep and a decrease in hybridization being observed in RNA from the other sheep. Changes in the two-dimensional protein patterns for both sheep were essentially the same, hence it is most likely that this variation is random and does not reflect a genuine change in the mRNA levels to which this probe hybridizes. There was relatively even hybridization of the IF probe to pre- and post-infusion RNA for both sheep in this experiment (figure 7.4).

Figure 7.3

Northern Transfer Analysis, Sheep A and B

Northern transfer analysis was conducted as described in section 2.2.3 B. The following tracks were hybridized to both the IF type II gene B 3' noncoding probe (labelled by oligo labelling, see 2.2.2 D) and the human UHS gene coding probe at the same time (7.2.2):

C contained approximately 10 µg of total follicle RNA taken immediately prior to cysteine infusion.

T 10 µg total follicle RNA isolated after 10 weeks of cysteine infusion.

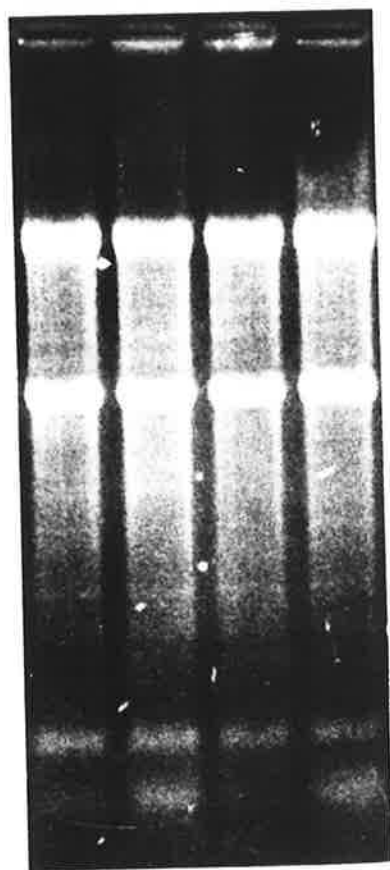
Equivalent tracks were hybridized to these probes individually (labelled IF and UHS). After hybridization the filters were washed in 2 X SSPE, 0.1% SDS; 65°C.

At top is a photograph of the agarose gel after staining with ethidium bromide prior to transfer to Zetaprobe.

The intensity of hybridization was determined by densitometric scanning, and the results of this are presented below.

Probe	Control	Test	Ratio (T/C)
<u>Sheep A</u>			
IF	459164	325696	0.71
UHS	439604	445109	1.01
<u>Sheep B</u>			
IF	157483	156824	1.00
UHS	547828	509653	0.93

A B
C T C T



A B
IF UHS C T C T



Figure 7.4

Northern Transfer Analysis, Sheep C and D

Northern transfer was conducted as described in figure 7.3. Initially the filter was hybridized to both the sheep UHS gene coding region probe and the IF type II gene B 3' noncoding region probe. The UHS probe used in this hybridization proved to have a low S.A.. Hence the filter was rehybridized to this probe alone after it had been freshly prepared. An autoradiograph of the UHS probe alone is presented at bottom.

The results of densitometric scans to determine the level of densitometric scanning are presented below.

Probe	Control	Test	Ratio (T/C)
<u>IF probe</u>			
Sheep C	1456760	1620490	1.11
Sheep D	1418630	1572970	1.11
<u>UHS probe</u>			
Sheep B	2998350	3839790	1.3
Sheep C	4180370	2513660	0.6

IF
C D
C T C T



UHS
C T C T



7.2.3 RNA Protection Experiments

As no gross changes in UHS encoding mRNA were observed using northern transfer analysis, it was decided to use the more sensitive and quantitative method of RNA protection to assay RNA levels.

RNA protection experiments were conducted using a 3' noncoding probe from the sheep UHS cDNA clone, K4. An *Mbo*II 300 bp fragment from the 3' noncoding region of K4 was subcloned into a pGEM-2 *Sma*I vector (figure 7.5). The orientation of the insert was such that antisense RNA would be produced by transcription from the SP6 promoter (figure 7.5). Run-off transcripts labelled to high specific activity with [³²P]UTP were produced using SP6 RNA polymerase (2.2.3 D) after the clone had been linearized with *Hind*III. In a similar fashion reverse transcripts were prepared from the 3' noncoding region of an intermediate filament type II gene (Powell *et al.*, in preparation).

RNA protection was conducted on similar amounts of total RNA isolated before and after cysteine infusion from two of the sheep used in these experiments. Both probes were included in a single reaction so that the intermediate filament probe could act as a direct internal control against which variation in the levels of the K4 mRNA could be measured. The probes were also allowed to hybridize individually so that the protected fragments corresponding to each gene could be differentiated.

The results of this experiment are presented in figure 7.6. The densitometric scans indicate that for sheep A, relative to the levels of IF-protected mRNA, the levels of K4-protected mRNA are almost identical between pre- and post-infusion RNA samples. For sheep B an average increase of 1.35 fold in the relative levels of K4-protected mRNA is observed. This may reflect a genuine increase in the levels of the K4 coding mRNA, but given the lack of a response in the other sheep may simply be the result of experimental variation.

7.2.4 Follicle Tissue *in situ* Hybridization

In situ hybridization was conducted on human and sheep follicular tissue using cRNA probes constructed from the human and sheep UHS genes and also the sheep UHS cDNA clone K4. Sheep follicular tissue used throughout these experiments was taken

Figure 7.5

K4 3' Noncoding pGEM-2 Clone

The construction of this clone is described in section 7.2.3. The direction of *in vivo* mRNA transcription is indicated by the arrow in the insert (thin line). pGEM-2 DNA is represented by the thick line. The orientation of the fragment was determined using an internal *Pst*I. Antisense cRNA was produced using SP6 RNA polymerase after the clone had been linearized using *Hind*III (2.2.3 D).

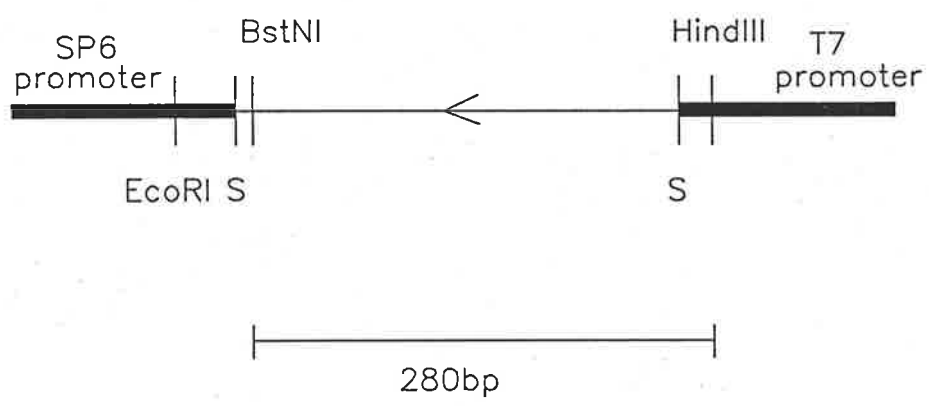


Figure 7.6

RNA Protection, Sheep A and B

RNA protection was conducted as described in section 2.2.3 C. 10 µg of total follicle RNA was used in each protection reaction. The tracks are as follows:

C contain RNA isolated from wool follicles before cysteine infusion. Both the IF and K4 3' noncoding probes were used in these reactions.

T contain RNA isolated after 10 weeks of cysteine infusion and also contain both probes.

Control tracks in which the IF and K4 probes were hybridized individually in the absence of follicle RNA (the pair of tracks entitled "1") and in the presence of 10 µg of total follicle RNA (the pair of tracks entitled "2").

RNA protection reactions were replicated for sheep B (B and B'), and the replicates were fractionated on a different polyacrylamide gel. The results of densitometric scans are presented below.

Probe	Control	Test	Ratio (T/C)
<u>Sheep A</u>			
IF	1590142	3302316	2.07
UHS	184241 382621 ¹	361449	0.95 ²
<u>Sheep B</u>			
IF	2831064	1647644	0.59
UHS	434384 252806 ¹	376921	1.42
<u>Sheep B'</u>			
IF	587458	590030	1.00
UHS	41139	54865	1.30

¹ Estimate has been corrected to take into account variation in the IF protection levels.

² Ratio calculated using adjusted estimates.

No RNase

A

B

B'

IF UHS

¹IF UHS

²IF UHS

C

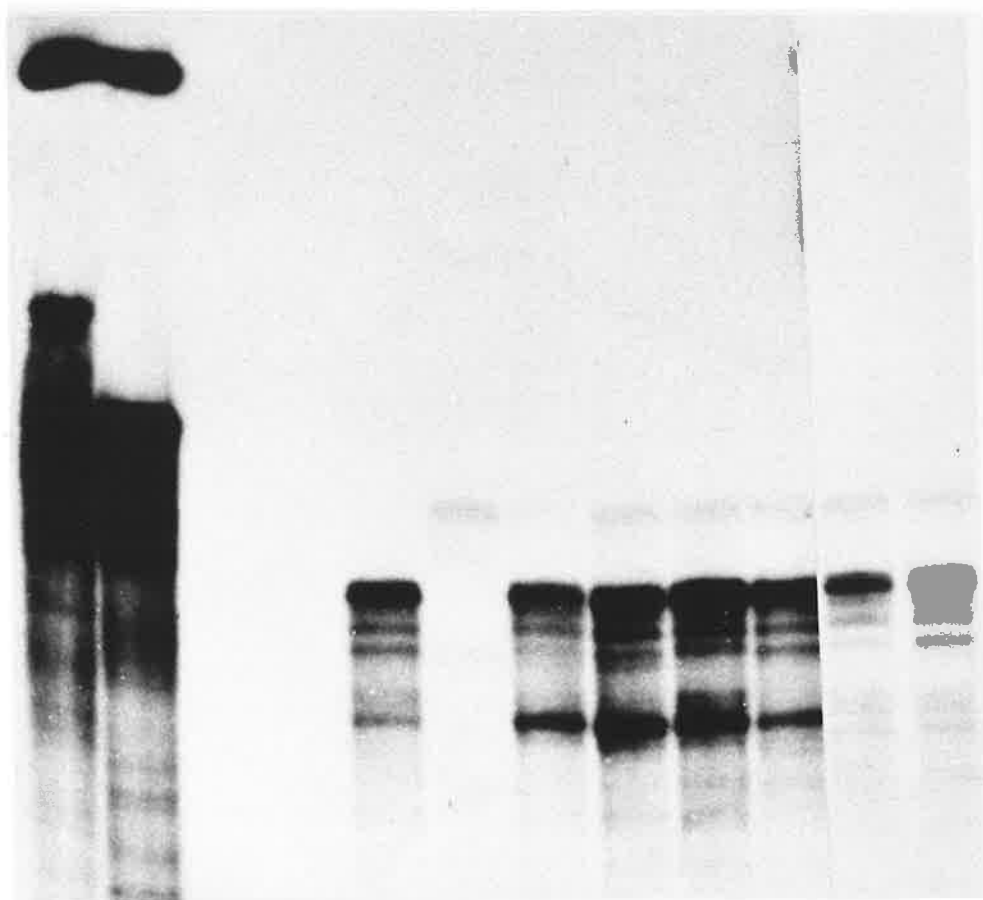
T

C

T

C

T



from sheep that had not been infused with S-amino acids. The maps of the pGEM-2 constructions from which the cRNA probes were synthesized are presented in figures 7.5 and 7.7. For comparative purposes sheep follicular material was also hybridized to a cRNA probe from the high-sulphur BIIIA gene coding region (Powell *et al.*, in preparation). The probes were synthesized using either SP6 or T7 RNA polymerase (2.2.3 D) as indicated in figures 7.5 and 7.7 and were radioactively labelled by incorporation of [α -³⁵S]UTP. Tissue *in situ* hybridization was conducted as described in chapter 2 (2.2.3 F). After hybridization, washing and autoradiography, the sections were stained by the SACPIC staining procedure (Auber, 1950). The results of *in situ* hybridization are presented in the following figures:

Figure 7.8: the human UHS gene coding region hybridized to transverse sections of human hair follicles.

Figure 7.9: transverse and oblique sections of sheep wool follicles hybridized to the sheep UHS gene coding region probe.

Figure 7.10: transverse sections of wool follicles hybridized to coding region probes from the sheep UHS and high-sulphur BIIIA genes.

Figure 7.11: transverse sections of wool follicles hybridized to the K4 3' noncoding probe and also the sheep 3' noncoding probe.

Figure 7.12: longitudinal sections of sheep wool follicles hybridized to the sheep UHS gene coding region probe.

Figure 7.13: longitudinal sections of sheep wool follicles hybridized to the sheep high-sulphur BIIIA gene coding region probe.

Figure 7.14: control sections pretreated with RNase prior to hybridization to the sheep UHS coding region probe.

Two general observations were made with the UHS gene probes. Firstly, with all the UHS probes used and with both sheep and human follicle tissue, silver grains were mostly localized to the periphery of the nascent fibre over the cuticle and perhaps also inner-root sheath. Hybridization to the cuticle was most obvious in the human tissue

Figure 7.7

pGEM-2 Clones for *In Situ* Hybridization

A Human UHS Gene Coding Probe

The construction of this clone (pGEME-HUHS) is described in Chapter 4, section 4.2.3. The cRNA probe covered the entire coding region and was synthesized using T7 RNA polymerase (2.2.3 D) after the clone had been linearized using *EcoRI*.

B Sheep UHS Gene Coding and 3' Noncoding Region Probes

The coding region pGEM-2 construct (pGEM-SUHS) been described in Chapter 5, section 5.2.3. The probe covered the entire coding region and was synthesized using T7 RNA polymerase (2.2.3 D) after the clone had been linearized using *EcoRI*.

The 530 bp *BglII/BamHI* fragment from the 3' noncoding region which was used for Southern and Northern transfer (5.2.4) was cloned into a pGEM-2 *SmaI* vector. The orientation of the insert was determined using the internal *EcoRI* site and the direction of *in vivo* mRNA transcription is indicated by an arrow in the insert. The antisense probe was produced using T7 RNA polymerase after the clone had been linearized using *SacI*.

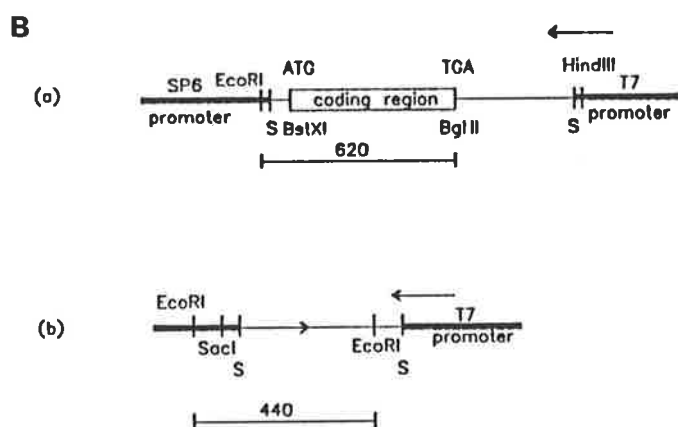
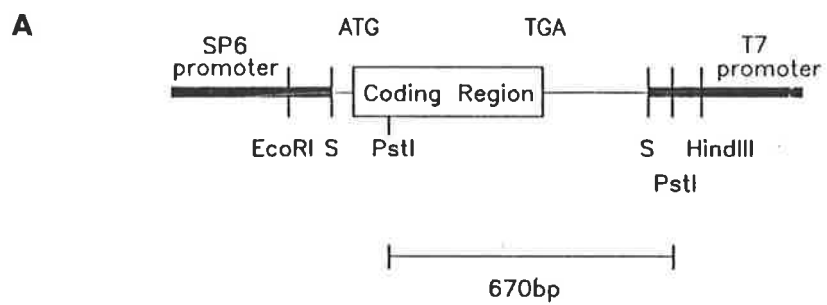


Figure 7.8

Tissue *In Situ* Hybridization of Transverse Sections of Human Scalp Hair Follicles

A This section was treated with RNase prior to hybridization to the human coding region probe (section 7.2.4). The structures of the hair follicle are indicated as follows:

ORS the outer root sheath.

IRS the inner root sheath.

CUT the fibre cuticle.

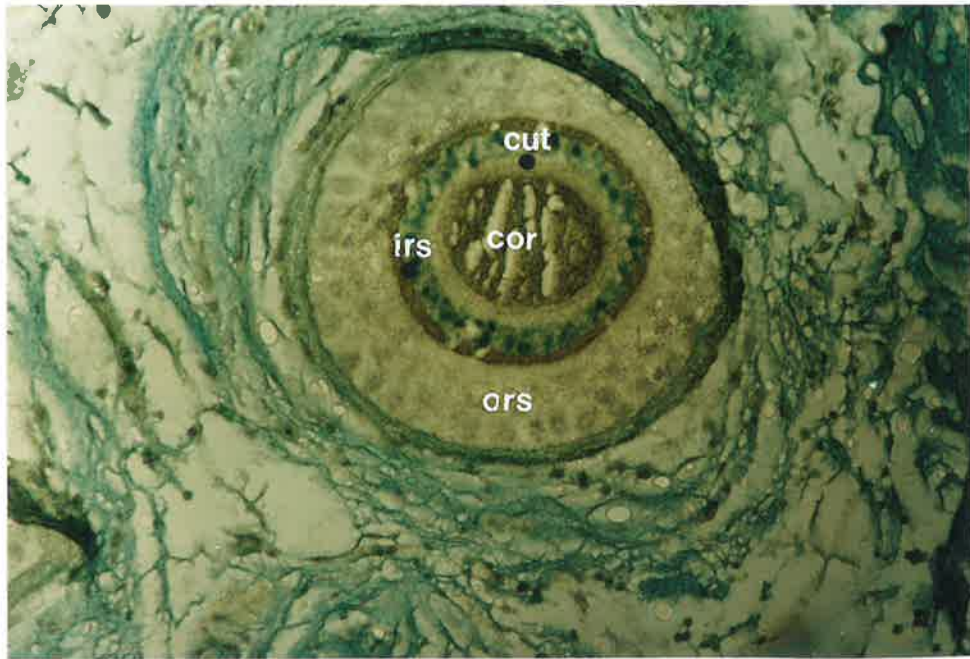
COR the fibre cortex.

B A similar section hybridized to the human coding region probe (section 7.2.4). Silver grains can be clearly observed over the periphery of the fibre in the region of the cuticle.

Both sections were photographed at 250X magnification after a 32 day exposure.

These are preliminary and are to be repeated when more human scalp tissue becomes available.

RNase Pretreated



Human UHS Gene Coding Probe

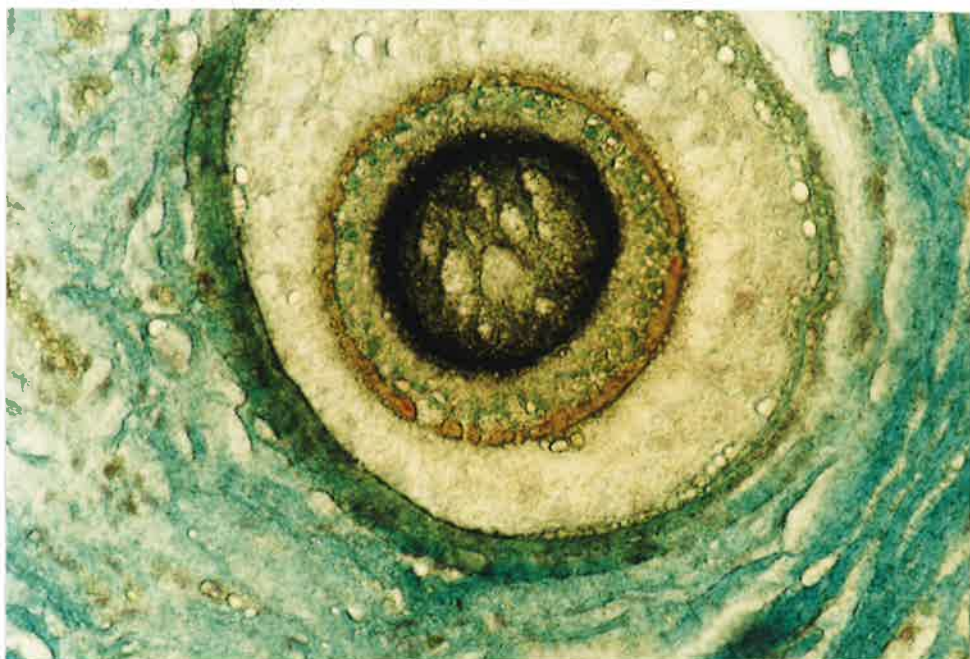


Figure 7.9

Tissue *In Situ* Hybridization of Merino Wool Follicles to the Sheep UHS Gene Coding Probe

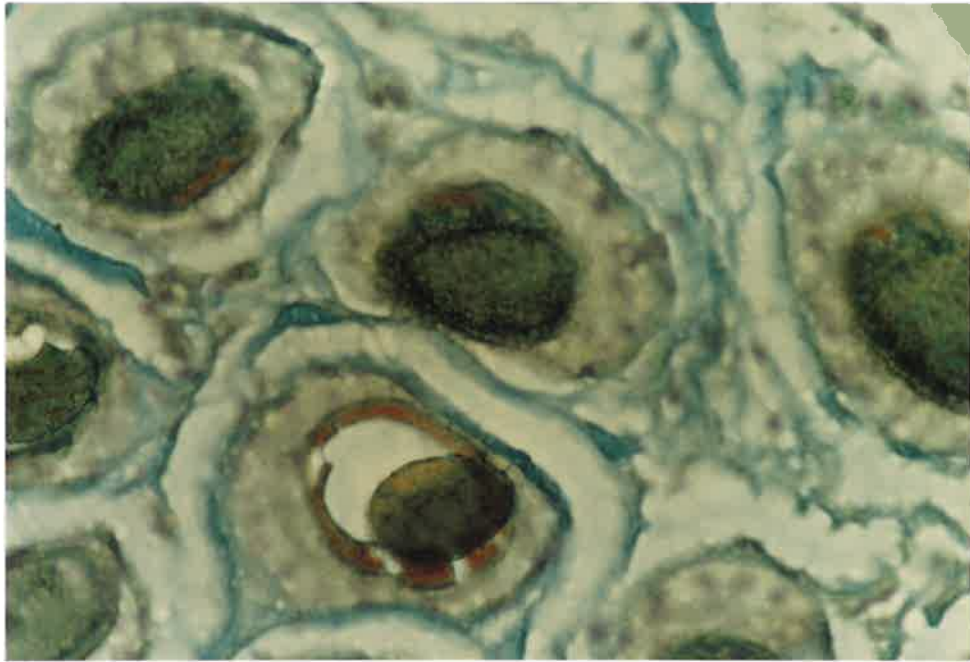
- A A transverse section photographed at 250X magnification. Silver grains can be observed forming a ring at the periphery of the nascent wool fibres.

- B An oblique section photographed at 400X magnification. A similar pattern of hybridization to A can be observed in this section.

The histological components of the wool follicle in both transverse and longitudinal section are as indicated in figure 7.14.

Both sections were exposed for 31 days.

Sheep UHS Gene Coding Probe



Sheep UHS Gene Coding Probe



Figure 7.10

In Situ Hybridization of Transverse Sections of Merino Wool Follicles.

- A The same section as figure 7.9 A photographed at 400X magnification. The section was hybridized to the sheep UHS gene coding region probe.
- B A consecutive section hybridized to the high-sulphur BIIIA gene coding region probe and photographed at 400X magnification. Silver grains are observed over the whole cortex which is consistent with this gene being expressed predominantly in the cortex.

The histological components of the wool fibre in both transverse and longitudinal section are as indicated in figure 7.14.

Both sections were exposed for 31 days.

Sheep UHS Gene Coding Probe



BIIIA Coding Probe

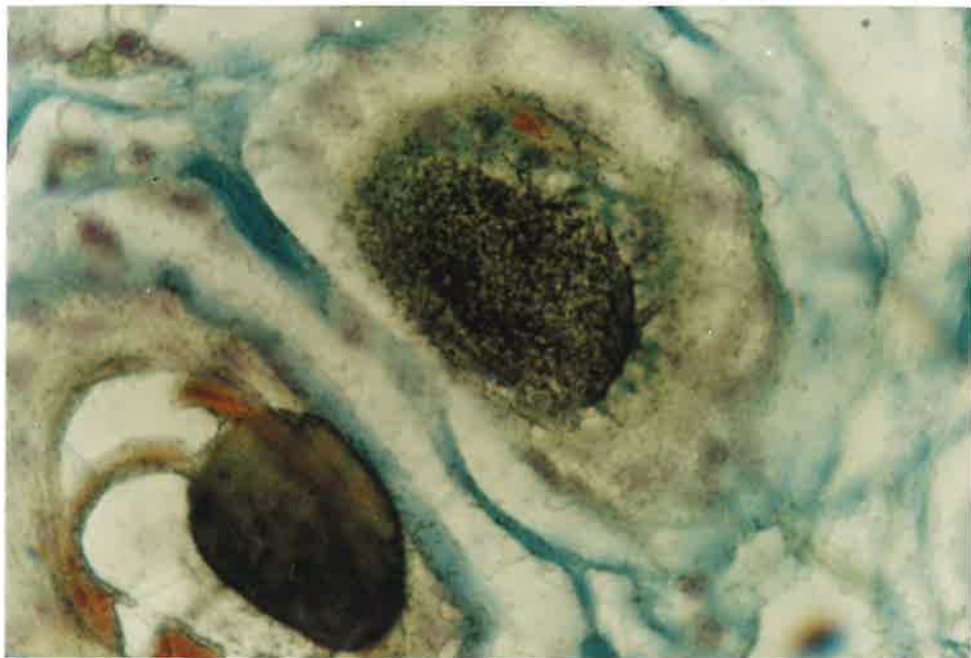


Figure 7.11

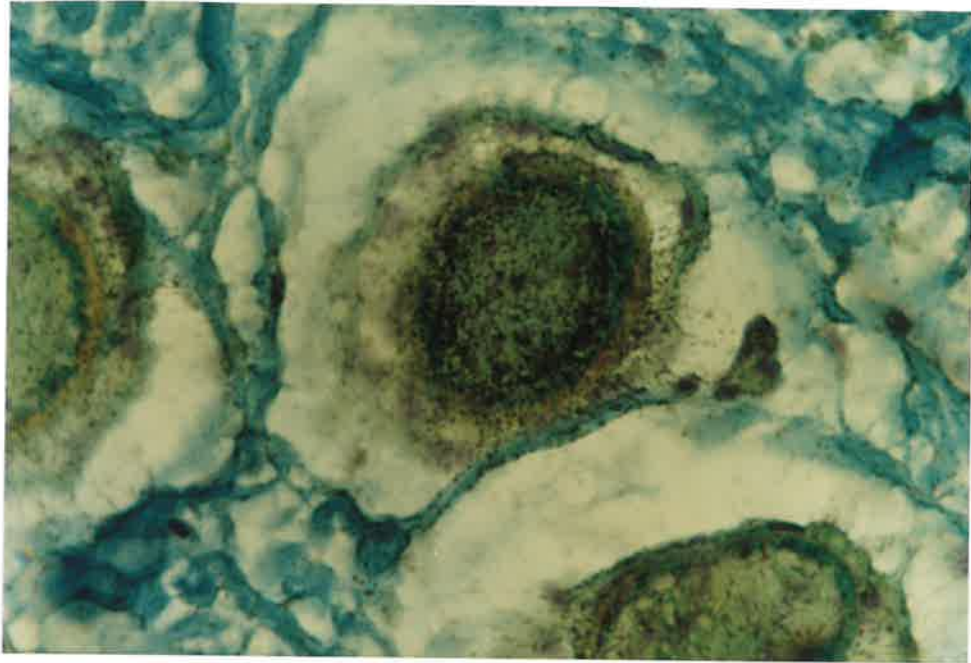
In Situ Hybridization of Transverse Sections of Merino Wool Follicles

- A** The section was hybridized to the sheep UHS gene 3' noncoding region probe and photographed at 400X magnification. The hybridization pattern with this probe was not as obvious as with the coding region probe, but silver grains still appear to be localized predominantly as a ring around the fibre.
- B** The section was hybridized to the K4 3' noncoding region probe and photographed at 400X magnification. The K4 probe has a similar pattern of hybridization to the sheep UHS gene.

The histological components of the wool fibre in both transverse and longitudinal section are as indicated in figure 7.14.

Both sections were exposed for 31 days.

Sheep UHS Gene 3'NC Probe



K4 3'Noncoding Probe

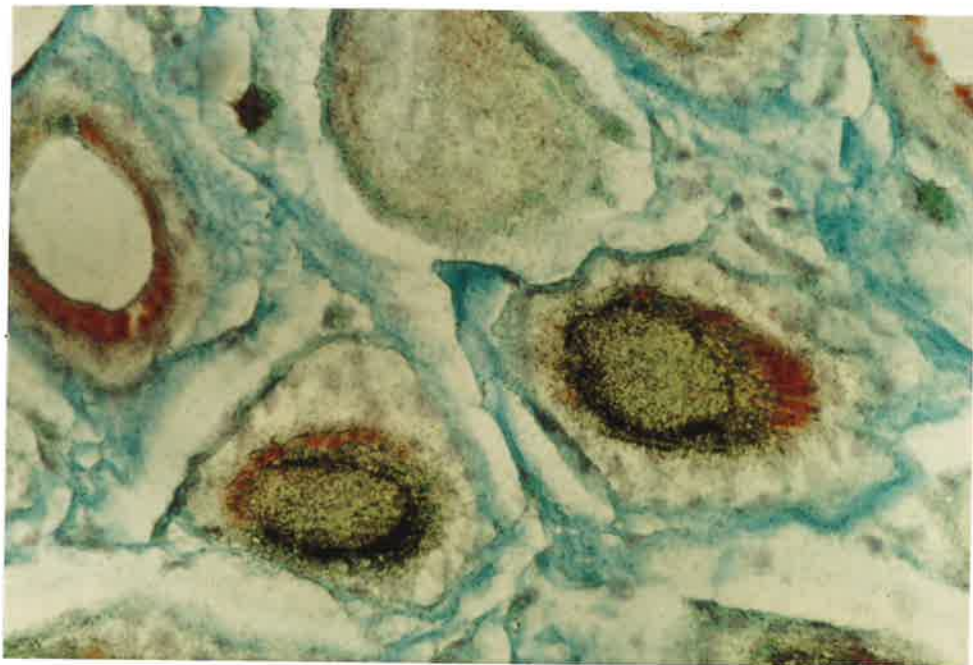


Figure 7.12

A Wool Follicle Longitudinal Section Hybridized to the Sheep UHS Gene Coding

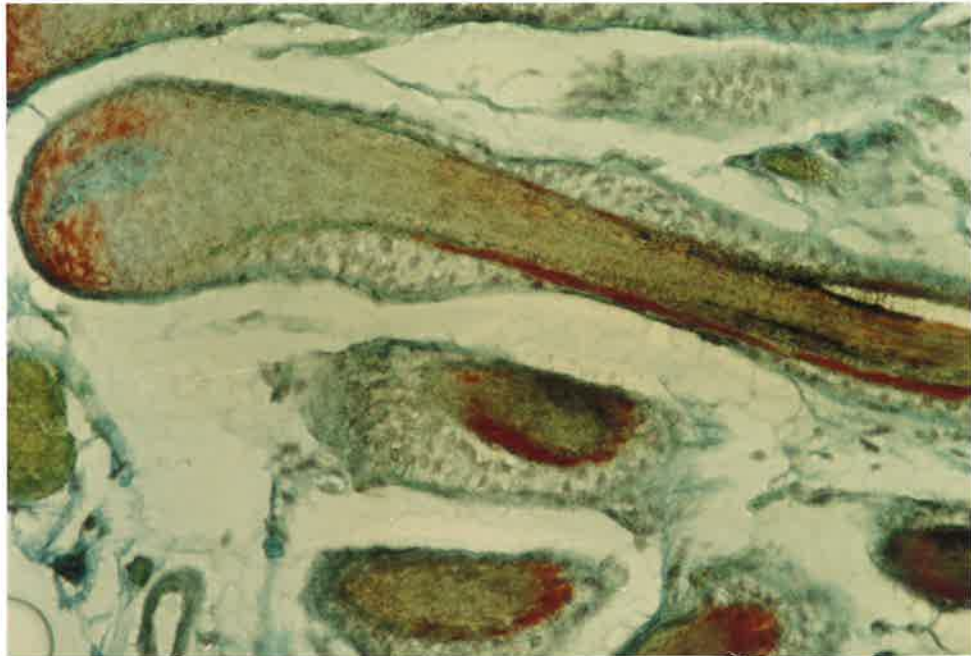
Probe

- A** This section was hybridized to the sheep UHS gene coding region probe photographed at 250X magnification. Silver grains can be observed in two parallel lines along the edge of the fibre. The region to which the probe hybridizes is high above the bulb, immediately below where the fibre is fully keratinized (orange colour).
- B** The same section photographed at 400X magnification. It will be noted that where the inner root sheath is detached from the fibre the silver grains are located predominantly over the edge of the fibre and indicates that the gene is expressed in the fibre and not the inner root sheath.

The histological components of the wool fibre in both transverse and longitudinal section are as indicated in figure 7.14.

The section was exposed for 31 days.

Sheep UHS Gene Coding Probe



Sheep UHS Gene Coding Probe

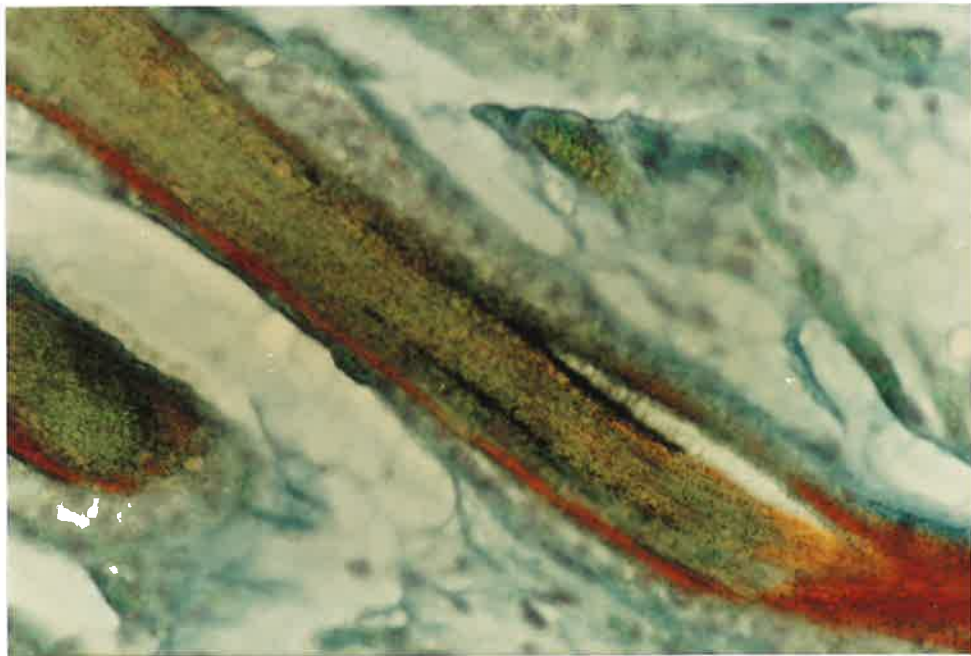


Figure 7.13

Longitudinal Section Hybridized to the BIIIA Coding Probe

A section hybridized to the BIIIA coding region probe photographed at 250x magnification. Silver grains are located over the whole fibre in the more distal region of the follicle. This *in situ* was conducted and photographed by Miss R. Keough.

BIIA Coding Probe



Figure 7.14

RNase pretreated Wool Follicle Sections Hybridized to the Sheep UHS Gene Coding Probe

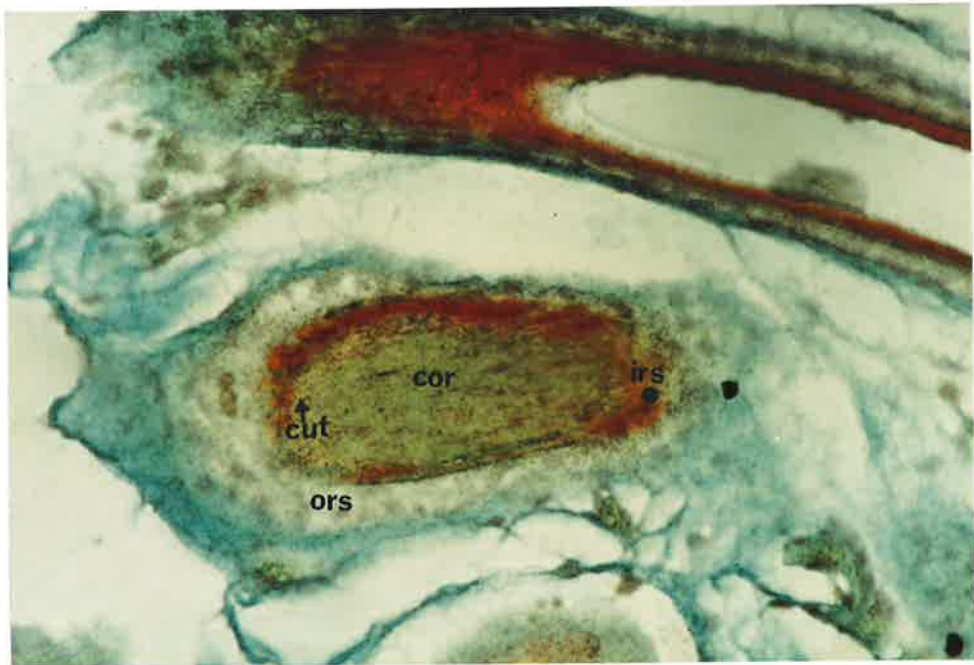
- A A transverse section photographed at 250X magnification.
- B A longitudinal section photographed at 250X magnification.

Hybridization of the other probes described in this chapter to wool sections pretreated with RNase resulted in similar observations to those presented in this figure.

The structures have been labelled as follows:

- ORS Outer root sheath.
- IRS Inner root sheath.
- CUT Fibre cuticle
- COR Fibre cortex.

RNase Pretreated



RNase Pretreated



where this structure is 8-10 cell layers thick (Swift, 1976). In contrast the high-sulphur BIIIA gene probe hybridized evenly over the whole wool fibre.

Secondly, silver grains were mostly localized over the more distal regions of the follicle when the UHS gene probes were hybridized to longitudinal sections of wool follicles, well above the follicle bulb and dermal papilla. The BIIIA gene also hybridizes to the more distal regions of the fibre.

7.3 Discussion

7.3.1 Sheep Infusion Experiment

The elemental sulphur analysis (table 7.1) and two-dimensional PAGE analysis of the wool (figure 7.2; table 7.2) clearly indicate that there has been an increase in the levels of the UHS proteins in response to infusion *per abomasum* in the sheep.

Scans along the diagonal in which the UHS proteins and also some high-sulphur protein components are believed to reside (Gillespie *et al.*, 1982; see also chapter 4, 4.3.6 and chapter 5, 5.3.4) indicate that there are increases of between 1.3 and 5.0 fold in these components relative to the IF proteins.

These scans also indicated that there was an increase of 1.4-1.6 fold in the particular wool protein component that comigrates with the *in vitro* translation product of the sheep UHS gene. Hence if the authentic product of the sheep UHS gene does actually migrate in this region, there is a small increase in the proportion of the component upon sulphur-enrichment of the wool. This increase is substantially less than is observed for some other components on the diagonal.

It is of interest that scans of the regions of the two-dimensional gels in which the high-glycine/tyrosine proteins lie reveal that there was an overall decrease in the levels of these proteins in the sulphur-enriched wool. Previously it was believed that sulphur-enrichment through S-amino acid supplementation did not inhibit synthesis of the high-glycine/tyrosine proteins (Frenkel *et al.*, 1975; Gillespie, 1983). The overall decrease of about 40% in the high-glycine/tyrosine proteins is comparable to that observed when epidermal growth factor was administered to sheep (Gillespie *et al.*, 1982).

Also observed was a decrease in three unknown components (labelled w in figure 7.2). Interestingly these components have been observed to decrease in the regrowth wool after administration of epidermal growth factor (Gillespie *et al.*, 1982).

7.3.2 Northern Transfer and RNA Protection Analysis

Northern transfer analysis using coding region probes from the sheep and human UHS genes did not reveal any consistent changes in the follicle RNA populations in response to cysteine infusion. Similar observations were made when the more specific method of RNA protection was employed. Hence, overall, there is no evidence to suggest that there was a substantial change in the levels of mRNA to which the UHS gene probes hybridized in response to sulphur-enrichment.

The above data is only preliminary and further information is required to establish the significance of the apparent lack of any change in the abundance of UHS protein encoding mRNA in response to cysteine infusion. However given the clear increase observed in the UHS proteins by two-dimensional PAGE and elemental sulphur analysis, an obvious change would have been expected in the levels of the UHS encoding mRNA if these genes were under transcriptional regulation. Hence these results cast doubt on the possibility that the increased synthesis of the UHS proteins is mediated by a transcriptional mechanism and rather, may indicate that enhanced translation of existing UHS coding mRNA may occur in the sulphur-enriched follicles.

An alternative explanation for the northern transfer and RNA protection results is that the mRNA species to which the human and sheep UHS genes and K4 hybridized are encoded by UHS genes that are expressed constitutively and do not respond to increased cysteine levels. This hypothesis will be elaborated upon below.

In addition to repetition of the northern transfer and RNA protection experiments to substantiate the preliminary findings presented in this chapter, some additional experiments are suggested to further investigate the effect of cysteine-infusion on UHS gene expression.

In situ hybridization experiments using the UHS gene probes should be conducted in which follicles taken before and after cysteine-infusion are compared to determine if an

alteration in the pattern of UHS gene expression occurs in response to sulphur-enrichment. These experiments should include protein, histological and ultra-structural analysis, so that changes in the UHS encoding mRNA (or lack thereof) can be compared to changes in the levels of the UHS proteins and to changes in the ultra-structure of the nascent fibre.

Experiments could also be conducted in which mRNA transcribed *in vitro* from the sheep UHS gene is translated *in vitro* in a wool follicle homogenate in the presence of varying concentrations of cysteine. A message synthesized *in vitro* from another gene from either the high-sulphur or intermediate filament families could be included in these reactions as an internal control. The relative levels of translation of the control protein (IF or HS) and the UHS protein could be determined by electrophoresis followed by fluorography. The effect of cysteine concentration on the synthesis of the UHS protein could then be determined. Such an experiment may indicate *in vitro* whether it is possible to increase the synthesis of UHS protein from a given amount of mRNA merely by increasing the concentration of cysteine available for protein synthesis.

At present there are no antibodies specific for the UHS proteins. The possession of an antibody specific for the UHS protein synthesized from the sheep UHS gene would be a great advantage. The possession of such an antibody would enable the correlation of the effect of sulphur-enrichment on the protein with that on the mRNA transcribed from the sheep UHS gene. An antibody may also be used for *in situ* UHS protein detection in conjunction with *in situ* mRNA hybridization. It may be possible to produce such an antibody from synthetic analogue peptides made to part of the conceptual UHS protein sequence.

7.3.3 Tissue *in situ* Hybridization

When the human UHS gene was used as a probe for *in situ* hybridization to human hair follicles it was found that silver grains were localized predominantly to the periphery of the nascent hair fibre. Similar observations were made using the sheep UHS gene and K4 as probes for *in situ* hybridization to wool follicles. The wool and human hair cuticles have been observed to contain cysteine-rich protein (Swift, 1976; Ley and Crewther, 1980) and the wool cuticle contains protein with a two-dimensional PAGE mobility similar to

that of the UHS proteins (Ley and Crewther, 1980; Marshall and Ley, 1986). Hence it is most likely that the hybridization observed in the human and sheep tissue results from expression of UHS genes in the fibre cuticle.

γ -glutamyl- ϵ -lysine cross-links are believed to occur between some proteins in the fibre cuticle (Ley and Crewther, 1980). Hence the localization of the human and sheep UHS proteins to the cuticle increases the possibility that the glutamine and lysine residues of these proteins are involved in such cross-links. This is particularly true for the human UHS protein which has almost equimolar amounts of these two residues (see chapter 4).

The number of grains over the wool cortex appeared to be greater than the background level. Hence it is possible that some hybridization of the UHS probes to mRNA occurs in this region. This raises the possibility that UHS proteins also occur in the cortex. The EM observations of Rogers (1964) indicate that apparent large increases occur in the amount of cortical matrix protein when the wool fibre is sulphur-enriched, suggesting that synthesis of the UHS proteins occurs in the cortex. The wool follicular tissue used for *in situ* hybridization was not sulphur-enriched, hence the possibility remains that the sheep UHS gene and/or K4 will cross-hybridize to UHS-encoding mRNA present in the cortex of sulphur-enriched wool fibres. As discussed in section 7.3.2, tissue *in situ* hybridization together with ultra-structural analysis of sulphur-enriched wool follicles should help determine if UHS protein expression occurs in the fibre cortex.

One alternative possibility which may be considered is that there are two families of UHS protein. One family may reside in the cuticle and be synthesized constitutively. That is, the expression of the genes encoding this family are unaffected by cysteine supply. The second family could reside in the cortex and is only produced when the supply of cysteine to the follicle is elevated. If the two families were sufficiently distant for nucleic acid probes not to cross-hybridize then this would explain the apparent lack of a change in mRNA levels during sulphur-enrichment observed using the cuticle UHS genes described in this thesis. The observation that the authentic keratin component that comigrates with the *in vitro* translation product of the sheep UHS gene increased by a smaller margin than many other components on the two-dimensional diagonal is also consistent with this hypothesis.

If a cortical UHS protein family exists, then the mouse UHS gene (see chapter 6) which has a coding region quite different to the human and sheep UHS genes and K4, may belong to this family. The amino acid content of the protein encoded by this gene agrees more closely the amino acid analysis of UHS protein fractions isolated from sulphur-enriched wool than do the cuticle UHS conceptual proteins.

The second observation of interest that was made with the wool follicle tissue was that the sheep UHS gene probe hybridized to the distal regions of the follicle. These results confirm the belief that the UHS proteins are amongst the last proteins to be synthesized prior to the completion of keratinization of the fibre (Dr J.M. Gillespie, personal communication; Marshall and Gillespie, 1988). The BIIIA probe also hybridizes to the distal region of the nascent fibre (figure 7.13). The hybridization patterns of the IF probes indicate that these proteins are synthesized predominantly in the proximal regions of the follicle (Powell *et al.*, in preparation). Thus, overall the *in situ* hybridization patterns support the two-stage hypothesis of keratin synthesis in which it is believed that the intermediate filament proteins are the first to be synthesized, in the proximal regions, and are followed by the high-sulphur proteins which are synthesized in the more distal regions of the follicle (see chapter 1, 1.5).

7.4 Summary

In this chapter two aspects of UHS gene expression have been investigated. Firstly the effect of sulphur-enrichment on the UHS-encoding mRNA populations was investigated. These preliminary experiments suggest that there was not a significant alteration in the mRNA populations to which the human and sheep UHS genes and the sheep UHS cDNA clone K4 hybridized in northern transfer and RNA protection experiments. If this observation is substantiated then it may indicate that sulphur-enrichment occurs through an increase in the rate of protein synthesis from an invariant amount of mRNA rather than from an increase in the amount of UHS encoding mRNA.

Alternatively, there may be two families of UHS proteins, one which is not affected by cysteine supply and is a constituent of the cuticle, the other is synthesized only in the presence of high levels of cysteine and is a constituent of the cortex.

It has been determined using tissue *in situ* hybridization that the sheep and human UHS genes and the K4 gene are expressed in the wool and hair cuticles of their respective species. Also determined by this method was that these genes are expressed in the more distal regions of the wool follicle near where keratinization is completed. They are almost certainly amongst the last major fibre keratin groups to be synthesized.

Chapter 8

Final Discussion

This thesis has described the isolation and characterization of two UHS genes, one from the human genome and one from the sheep genome. The salient features of the sheep and human UHS genes may be summarized as follows:

- (a) The amino acid contents of both genes deduced from the coding sequences are substantially different from those published by other workers and also observed in the present study for fractions of wool UHS proteins.
- (b) The coding regions of both genes contain a number of different types of repeats, each with distinctive amino acid sequences. Most of the repeats are shared between the two UHS genes and also with the cDNA clone K4. One of the repeats present in the human UHS gene and also in K4 resembles the repeat found in the sheep high-sulphur B2 and BIIIA proteins.
- (c) strong conservation is observed between the 5' untranslated regions of both genes.
- (d) both genes have coding regions which are uninterrupted by introns.
- (e) genomic Southern experiments indicated that there are a number of related UHS genes in both the human and sheep genomes.
- (f) the genomic clones from which the human and sheep UHS genes were isolated each contained at least one more possible UHS gene, indicating that there is close linkage of at least some of the UHS genes in both the human and sheep genomes. Chromosomal *in situ* hybridization indicated that two distant UHS gene clusters occur in the human genome.
- (g) Northern transfer experiments demonstrated that the sheep UHS gene is expressed in the wool follicle and that the human UHS gene cross-hybridizes to wool follicle mRNA. Tissue *in situ* hybridization showed that the sheep and human UHS genes are: (a) expressed in the cuticles of wool and hair; and (b) expressed in the more distal regions of the nascent fibre and are probably one of the last groups of proteins to be synthesized in the keratinizing fibre.
- (h) preliminary northern transfer and RNA protection experiments were conducted to determine if there was a modulation of the UHS coding mRNA populations in response to cysteine infusion *per abomasum* of sheep. These experiments provided

no evidence for any qualitative or quantitative changes in the mRNA to which the human and sheep UHS genes and also K4 hybridized.

Whilst it can be seen from the above summary that significant progress has been made in the analysis of the UHS proteins and the genes that encode them, there are several major points of interest yet to be understood. Four of these are discussed below.

A Protein Structure

A brief investigation of the possible structures of the proteins encoded by the human and sheep UHS genes has been discussed in chapter 6. Also discussed were possible approaches for elucidating the three-dimensional structure of these proteins. Unlike the intermediate filament proteins, very little is known of the structures of any of the wool IFAPs. It is important to overcome this deficiency in our knowledge so that the interactions of all of these proteins can be elucidated in relation to the mechanical and chemical properties of the wool fibre. In the long term a knowledge of the structure of the component proteins will be essential if engineering of the wool fibre to alter its properties is to take place.

B Expression

Very little is known of the mechanism by which the expression of any of the wool keratin protein genes are regulated. The UHS proteins are of particular interest because of the inductive effect the sulphur-containing amino acids have on the synthesis rates of these proteins. Apart from the intellectual satisfaction of elucidating the mechanism of regulation of the UHS genes, there is also a commercial interest. Experimental programmes are underway to improve wool growth rates through increasing the supply of cysteine to the follicle (for examples see "The Biology of Wool and Hair", Rogers G.E., Reis P.J., Ward K.A. and Marshall R.C. eds., Chapman and Hall, London 1988 in press). As a consequence the rapidly growing wool will have elevated levels of the UHS proteins. The effect of these proteins on the commercial qualities of the fibre are unknown, but it may be necessary to alter the expression of the UHS genes. A knowledge of the mechanism of regulation of UHS protein synthesis will be essential for this.

The strong conservation in the 5' untranslated regions of the human and sheep UHS genes is of interest (see chapter 6). The possibility that this region has a function in the regulation of translation should be investigated. This could be accomplished through alteration of the sequence in this region and observing the effect upon expression *in vitro* in follicle homogenates and *in vivo* after transgenesis. Also the 5' untranslated region of the sheep or human UHS gene could be added to another gene and the effect upon the expression of this gene observed *in vitro* and *in vivo*.

C Possible Cortical UHS Protein Components

As discussed in chapter 7 it is possible that as well as being present in the cuticle the UHS proteins may also reside in the cortex of the wool and hair fibres.

Experimental approaches have been suggested in chapter 7 which may help to determine if this is the case.

D Isolation of more sheep UHS genes

One of the most important steps which must be undertaken is the isolation and characterization of more UHS protein encoding genes. Genomic Southern analysis indicates the presence of a number of genes that cross-hybridize to the UHS genes already characterized. Mapping of the lambda clone from which this gene was isolated indicated the presence of another UHS gene in this clone. Hence a second gene can be easily characterized. As mentioned in chapter 5 some difficulty has been encountered in isolating UHS genes from the sheep genome. This problem is not limited to the UHS genes as difficulty was also encountered in isolating the high-sulphur BIIIB genes from random libraries (Frenkel *et al.*, 1988) and is being encountered with the isolation of the trichohyalin gene (Dr. R.B. Presland and Mr. M.J. Fietz, personal communication).

Alternative procedures for isolating these genes may need to be developed. One possibility is to use long range mapping methods such as pulsed field gel electrophoresis (PFGE) (Schwatz and Cantor, 1984). This would allow mapping of the UHS genes, and the restriction fragments upon which they reside could be isolated by preparative PFGE and subcloned in a similar manner to that described by Anand *et al.* (1988).

Once isolated and characterized the additional UHS encoding genes will undoubtedly help in answering the outstanding questions which remain on the UHS proteins, their structure, and their expression.

Appendix A

Appendix A

Sequencing the 1.0 kb *Eco*RI Fragment Containing the Human UHS Gene

The strategy used to determine the sequence of the 1.0 kb *Eco*RI fragment is summarized in figure A.1. Initial sequencing was conducted on opposite orientations of the fragment in M13mp18. Further sequence was obtained from three *Pvu*II subfragments which had been cloned into a M13mp18 *Sma*I vector.

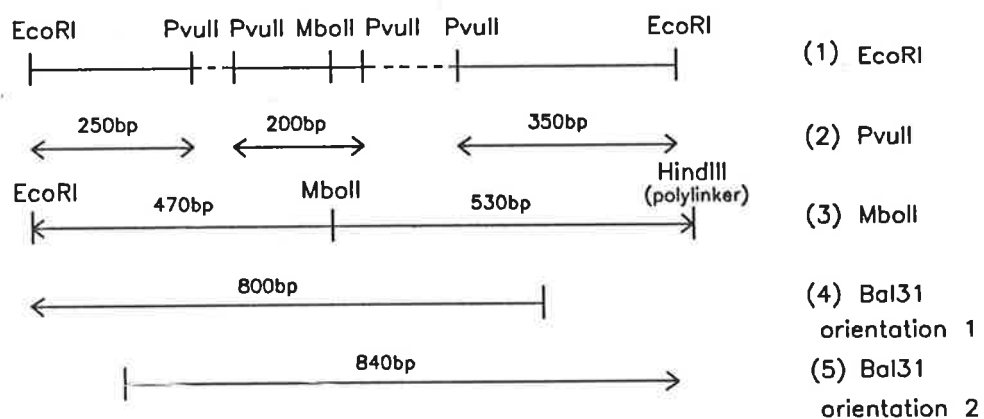
By inspection of the partial sequence an *Mbo*II site was discovered midway through the fragment. To obtain sequence overlapping the *Pvu*II sites used for initial subcloning and sequencing, an *Mbo*II/*Hind*III and an *Mbo*II/*Eco*RI fragment were resected from the pTZ18 clone and subcloned into M13mp18 *Sma*I/*Hind*III and M13mp19 *Sma*I/*Eco*RI vectors respectively. Sequencing (2.2.2 N) of these clones revealed the presence of regions of the 1.0 kb *Eco*RI fragment not covered by the three *Pvu*II subfragments. These regions contained multiple *Pvu*II sites which when cleaved with this enzyme would result in a number of very short (approximately 30 bp) fragments.

To complete the sequence of the 1.0 kb *Eco*RI fragment BAL31 exonuclease was used to create deletion fragments from the 1.0 kb *Eco*RI fragment. Deletions were made in one direction by cleavage of pTZ18 clone containing the 1.0 kb *Eco*RI fragment with *Hind*III, followed by deletion with BAL31 exonuclease under conditions described in section 2.2.2 L. The progressively deleted inserts were excised from the pTZ18 vector by cleavage with *Eco*RI and recloned into a M13mp18 *Eco*RI/*Sma*I vector. Deletions were conducted in the opposite orientation in a similar manner, using a clone in which the 1.0 kb *Eco*RI fragment was inserted into the vector in the opposite orientation. A suitable deletion fragment in each orientation was discovered to complete the sequence of the 1.0 kb *Eco*RI fragment in both orientations.

The complete sequence of the 1.0 kb *Eco*RI fragment is presented in chapter 4; figure 4.4.

A.

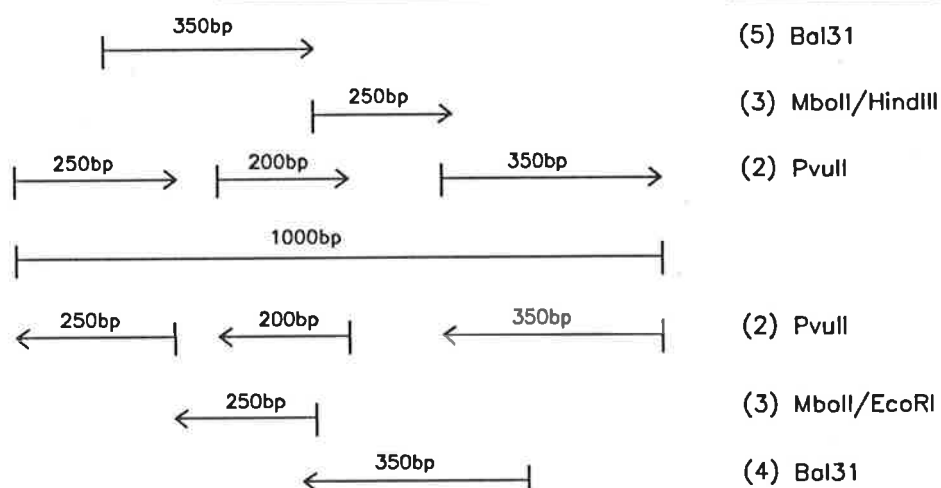
SUBCLONES



B.

SEQUENCING STRATEGY

SUBCLONES USED



Appendix B

Appendix B

Sequencing Strategy for the Sheep UHS Gene

Initially, to determine if the gene coding region straddles any of the restriction sites used for subcloning, both ends of the 2.3 *EcoRI/HindIII* and 2.8 *BamHI* fragments (see chapter 5, section 5.2.1) were sequenced. The sequences in these regions were inspected for the presence of a cysteine-rich open reading frame. No such reading frame was observed, hence it was concluded that the coding region was contained within the 2.3 kb *EcoRI/HindIII* fragment.

Because of the observation with the human UHS gene that repeated sequences occur in the coding region, it was decided to avoid the use of restriction enzymes for the construction of subclones for initial sequence analysis of the sheep gene. Rather, BAL31 exonuclease was used to construct deletion fragments from the *EcoRI/HindIII* 2.3 kb fragment. To facilitate isolation of the deleted inserts, deletions were conducted on this fragment subcloned into M13mp19. Plasmid vectors such as pTZ18R/19R, pGEM1/2 and pUC8/9 are too similar in size to the 2.3 kb *EcoRI/HindIII* fragment to allow specific isolation of the progressively deleted insert.

The strategy for creating BAL31 deletion fragments is presented in figure B.1. BAL31 deletion fragments were created in one direction by firstly cleaving the M13mp19 vector containing the *EcoRI/HindIII* fragment in the polylinker using *EcoRI*. For the opposite orientation the same clone was cleaved using *HindIII* which dissects the vector within the polylinker on the opposite side of the insertion to *EcoRI* (figure B.1). Clones so linearized were subjected to the action of BAL31. The progressively deleted inserts were resected from the vector by cleavage with *HindIII* for the one orientation and *EcoRI* for the opposite orientation (figure B.1). They were recloned into M13mp19 *HindIII/SmaI* and M13mp18 *EcoRI/SmaI* vectors respectively. Single stranded DNA was prepared from representative subclones from each of the deletion intervals and sequenced.

Some of the intervals proved to be too great to allow overlapping sequence. By inspection of the partial sequence a presumptive UHS coding region was discovered which

was flanked by *Bst*XI and *Bg*III sites. These sites enabled the construction of BAL31 deletion subclones commencing close to the coding region (figure B.1). BAL31 deletions were conducted on the 2.3 kb *Eco*RI/*Hind*III fragment cloned into M13mp19 after cleavage with either *Bst*XI or *Bg*III. Progressively deleted inserts were isolated from the vector with either *Eco*RI for clones cleaved initially with *Bst*XI, or *Hind*III for those cleaved with *Bg*III, and recloned as described above (figure B.1). Sequence analysis of some of these clones proved difficult because of GC rich sequence and in particular the presence of multiple strings of six G residues. Ambiguous regions were resequenced using a modified form of the T7 DNA polymerase called "Sequenase" (Tabor and Richardson, 1987) and the dGTP analogue dITP (2.2.2 N).

To complete the sequence of the presumptive coding region and flanking regions, three restriction fragments were subcloned from the *Bam*HI fragment which covered the smaller *Eco*RI/*Hind*III fragment containing the gene. These included a 300 bp *Bst*XI, a 120 bp *Pvu*II and a 250 bp *Bam*HI/*Pvu*II fragment. The position of these fragments within the 2.8 kb *Bam*HI fragment is depicted in figure B.1. The 300 bp *Bst*XI fragment was cloned initially into a pGEM-2 *Sma*I vector and resected from this clone using *Eco*RI/*Hind*III. This fragment was subcloned into M13mp18/19 *Eco*RI/*Hind*III vectors. The 250 bp *Pvu*II/*Bam*HI fragment was isolated from the 2.8 kb *Bam*HI fragment which had been subcloned into M13mp18 *Bam*HI. It was isolated by cleavage with *Pvu*II and *Hind*III (in M13mp18 polylinker) and subcloned into an M13mp19 *Hind*III/*Sma*I vector. The sequencing strategy using these clones is presented in figure B.2.

The complete sequence obtained from these clones is presented in chapter 5, figure 5.4. The coding region conceptual amino acid sequence, TATA box and polyadenylation signal of the presumptive UHS gene contained in this sequence are delineated in this figure.

Figure B.1

Generation of Clones for Sequence Analysis of the Sheep UHS Gene

A BAL31 Deletion Clones

The central region of this figure is a map of the 2.3 kb *EcoRI/HindIII* fragment endfilled and cloned into a M13mp19 *SmaI* vector (see text). The hatched box indicates the position of the coding region in this fragment.

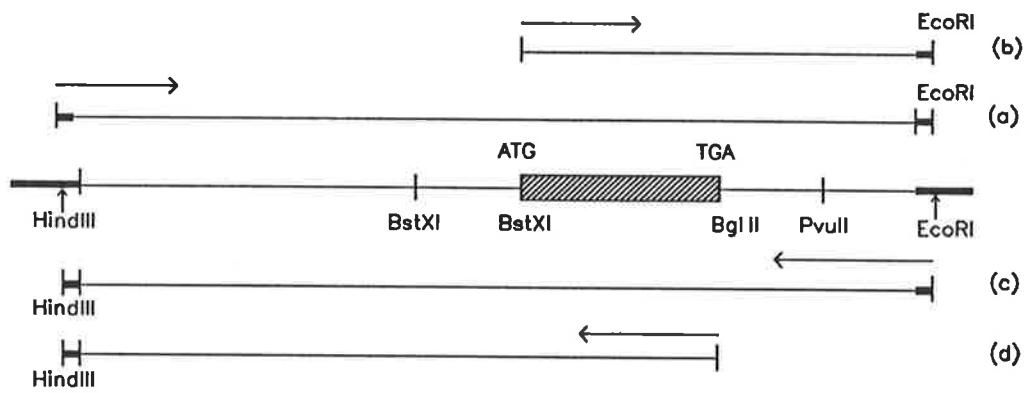
Initial BAL31 deletions were conducted from the *HindIII* (a) and *EcoRI* (c) sites in the M13mp19 polylinker. After progressive deletion of the fragments in the direction indicated by arrows they were excised from the vector using *EcoRI* (a) and *HindIII* (c). The excised fragments were subcloned as described in the text.

More BAL31 deletions were conducted from the *BstXI* (b) *BglIII* (d) sites in the insert. The progressively deleted fragments were excised with *EcoRI* (b) and *HindIII* (d) and subcloned as described in the text.

B Restriction Fragments

The positions of the 300 bp *BstXI*, 120 bp *PvuII* and 250 bp *PvuII/BamHI* fragments in the 2.8 kb *BamHI* fragment. These were subcloned as described in the text (5.2.2) and used to complete the sequence of the sheep UHS gene.

(A)



(B)

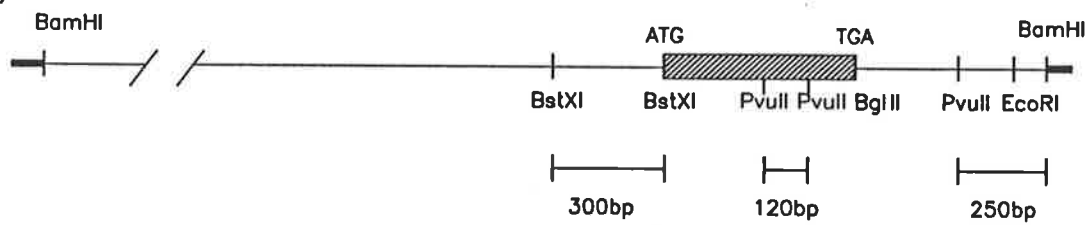
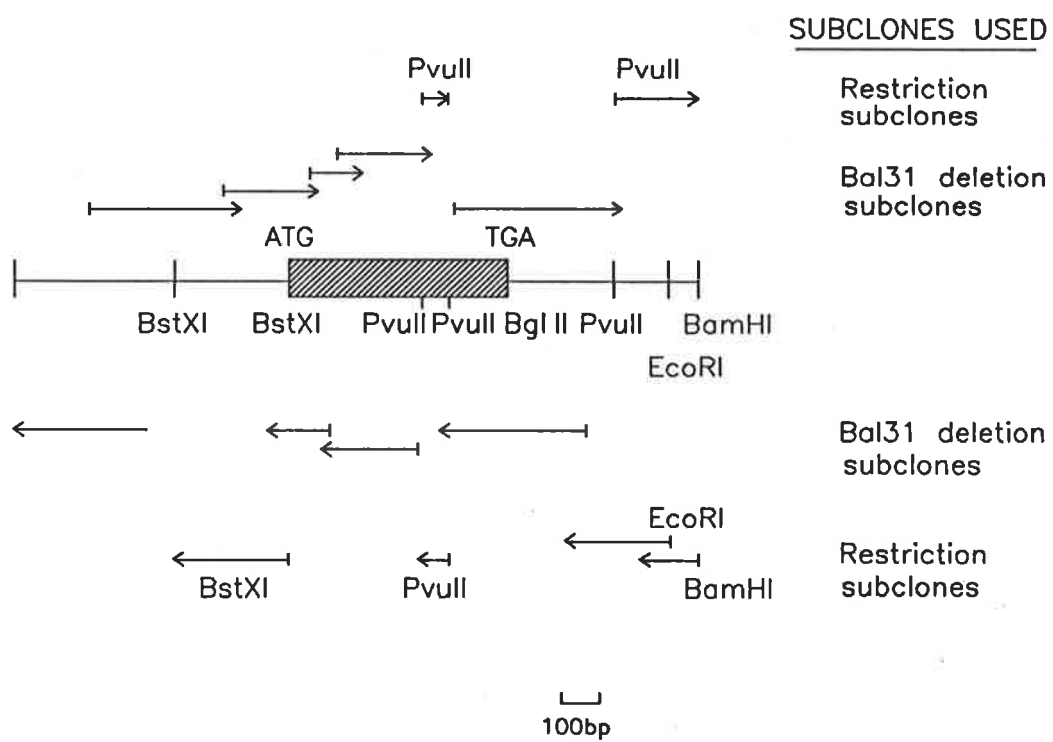


Figure B.2

Sequencing Strategy for the Sheep UHS Gene

This figure summarizes the position of the sequence obtained from the clones constructed from the 2.8 *Bam*HI and 2.3 *Eco*RI/*Hind*III fragments. Arrows indicate the direction of sequencing and the clones used to generate the sequence are indicated in the right margin. The presumptive transcription unit of this gene was sequenced in both orientations (see chapter 5, figure 5.4). The 5' noncoding/flanking regions were sequenced in both orientations for 300 bp upstream from the methionine initiation codon. The 3' noncoding/flanking regions were sequenced in both orientations up to the *Bam*HI site which is approximately 100 bp downstream from the presumptive polyadenylation signal.



SUBCLONES USED

Restriction
subclones

Bal31 deletion
subclones

Bal31 deletion
subclones

Restriction
subclones

Bibliography

- Ahmadi B. and Speakman P.T. (1978) *FEBS Lett.* **94**, 365-367.
- Akrigg A. and Cook P.R., (1980), *Nucleic Acids Res.* **8**, 845-854.
- Allison L.A., Moyle M., Shales M. and Ingles C.J., (1985), *Cell*, **42**, 599-610.
- Anand R., Honeycombe J., Whittaker P.A., Elder J.K. and Southern E.M. (1988) *Genomics* **3**, 177-186.
- Auber L. (1950) *Trans. Roy. Soc. Edin.* Vol. LXII, part I pp 191-254.
- Benoist C. and Chambon P., (1981), *Nature*, **290**, 394-310.
- Benoist C, O'Hare K. Breathnach R, Chambon P, (1980) *Nucleic Acids Res.* **8**, 127-142.
- Benton W.D. and Davies R.W. (1977) *Science (Wash. DC)* **196**, 180-182.
- Bertolino A.P., Gibbs P.E.M. and Freedberg I.M. (1982) *J. Invest. Dermatol* **79**, 173-177
- Birnboim H.C. and Doly J. (1979) *Nucl. Acids. Res.* **7**, 1513-1523.
- Birnstiel M.L., Busslinger M. and Strub K., (1985), *Cell*, **41**, 349-359.
- Bradford M. (1976) *Anal. Biochem.* **72**, 248-256.
- Brawerman G. (1981). *CRC Crit. Rev. Biochem.* **10**, 1-38.
- Breathnach R. and Chambon P. (1981) *Ann. Rev. Biochem.* **50**, 349-383.
- Briggs M.R., Kadondaga J.T., Bell S.P., Tjian R., (1986), *Science*, **234**, 47-52.
- Brooker J.D., May B.K. and Elliott W.H. (1980) *Eur. J. Biochem.* **296**, 17-24.
- Buchanan J.H., (1977), *Biochem. J.*, **167**, 489-481.
- Burgess A.W., Ponnuswamy P.K. and Scheraga H.A. (1974) *Israel J. Chem.* **12**, 239-286.
- Casadaban M.J. and Cohen S.N. (1980), *J. Mol. Biol.*, **138**, 179-207
- Casey J.L., Hentze M.W., Koeller D.M., Caughman S. W., Roualt T.A., Klauser R.D. and Harford J.B., (1988), *Science*, **240**, 924-928.
- Cavener D.R., (1987) *Nucleic Acids Res.* **15**, 1353-1361.
- Chapman R.E., (1976), *Proc. 5th Int. Wool Text. Conf. 1975*, Aachen.
- Chapman R.E. and Gemmell R.T., (1971), *J. Ultrastruct. Research* **36**, 342-354.
- Chapman R.E. and Gemmell R.T., (1973), *J. Cell Sci.* **13**, 811-819.
- Choo K.H., Folby G., Greco S., Lau Y.F. and Kan Y.W., (1986) *Gene*, **46**, 277-286.
- Chou P.Y. and Fasman G.D. (1974) *Biochemistry* **13**, 222-245.
- Corden J., Wasylyk B., Buchwaslder A., Sassone-Corsi T., Keding C., Chambon P. (1980) *Science*, **209**, 1406-1414.
- Corden J.L., Cadena D.L., Ahearn J.M. and Dahmus M.E., (1985), *Proc. Natl. Acad. Sci. USA*, **82**, 7934-7938.

- Cox K.H., Delean D.V., Angener L.M. and Angener R.C. (1984) *Dev. Biol.* **101**, 485-502.
- Crewther W.G., (1976), *Proc. 5th Int. Wool Text. Res. Conf. Aachen 1975* vol. II, 1-103.
- Crewther W.G., Dowling L.M., Steinert P.M. and Parry D.A.D. (1983) *Int. J. Biol. Macromol.* **5**, 267-274.
- Cutting J.A. (1984) *Methods in Enzymology* **104**, 451-455.
- Dagert M. and Ehrlich S.D. (1979) *Gene*, **6**, 23-28.
- Dale B.A., Resing K.A., Haydock P.V. Fleckman P., Fisher C. and Holbrook K.A. (1988), in *The Biol. of Wool and Hair* (Rogers G.E., Reis P.J., Ward K.A., Marshall R.C. eds), Chapman and Hall in press.
- Dame J.B., Williams J.L., McCutchan T.F., Webber J.L. Wirtz R.A., Hockmeyer W.T., Haynes J.D., Schneider I., Roberts D., Sanders G.S., Reddy P., Diggs C.L. and Miller L.H., (1984), *Science*, **225**, 593.
- Davis L.G., Dibner M.D. and Battey J.F. (1986) *Basic Methods in Molecular Biology* (Elsevier).
- Diaz-Ruis J.R. and Kaper J.M. (1978) *Prep. Biochem.* **8**, 1-17.
- Downes A.M., Sharry L.F. and Rogers G.E., (1963). *Nature (London)* **199**, 1059.
- Downes A.M., Sharry, L.F., and Rill, A.R. (1964). *Aust. J. Biol. Sci.* **17**, 945.
- Downes A.M., Reis P.J., Sharry L.F., and Tunks D.A. (1970) *Aust. J. Biol. Sci.* **23**, 1077-1088.
- Dopheide T.A.A. (1973) *Eur. J. Biochem.* **34**, 120-124.
- Drapeau G.R. (1976) *Methods in Enzymology* **45**, 469-475.
- Dufton M.J. and Hider R.C. (1977) *J. Mol. Biol.* **115**, 177-193.
- Dynan W.S. and Tjian R., (1985), *Nature*, **316**, 774-778.
- Echert R.L. and Green H., (1986), *Cell*, **46**, 583-589.
- Efstratiadis A., Posakony J.W., Maniatis T, Lawn T.M., O'Connell C., Spritz R.A., DeRiel J.K., Forget B.G., Weissman S.M., Slightom J.L., Blechl A.E., Smithies O., Baralle F.E., Shoulders C.C., Proudfoot N.J. (1980) *Cell*, **21**, 653-668.
- Eliopoulos E.E., Geddes A.J., Brett M., Pappin D.J.C. and Findlay, J.B.C. (1982) *Int. J. Biol. Macromol* **4**, 263-268.
- Elleman T.C. Lindley H. and Rowlands R.J. (1973) *Nature* **246**, 530-531.
- Epstein W.L. and Maibach H.I. (1969) in *Hair Growth* (eds. Dobson R.L. and Montagna W.), Oxford University Press 83-97.
- Feinberg A.D. and Vogelstein B. (1983), *Anal. Biochem.*, **132**, 6-13.
- Fraser I.E.B. (1969a) *Aust. J. Biol. Sci.* **22**, 213-229.
- Fraser I.E.B. (1969b) *Aust. J. Biol. Sci.* **22**, 231-238.

- Fraser R.D.B., Jones L.N., Macrae T.P., Suzuki E. and Tulloch P.A. (1980) *Proc. 6th Int. Wool Text. Res. Conf., Pretoria* (1980) pp 1-33.
- Frenkel M.J., Gillespie J.M. and Reis P.J. (1974) *Aust. J. Biol. Sci.* **27**, 31-38.
- Frenkel M.J., Gillespie J.M. and Reis P.J. (1975) *Aust. J. Biol. Sci.* **28**, 331-338.
- Frenkel M.J., Sleight M.J., Ward K.A., Powell B.C. and Rogers G.E. (1988) *Genomics*, in press.
- Friedman M. and Krull L.H. (1969) *Biochem. Biophys. Res. Comm.* **37**, 630-633.
- Friedman M., Krull L.H. and Cavins J.F. (1970) *J. Biol. Chem.* **254**, 3868-3871.
- Garnier J., Ostguthorpe D.J. and Robson B. (1978) *J. Mol. Biol.* **120**, 97-120.
- Ghosh P.K., Levowitz P., Frisque R.J., and Gluzman Y., (1981), *Proc. Nat. Acad. Sci.*, **78**, 100-104.
- Gillespie J.M., (1962), *Aust. J. Biol. Sci.*, **15**, 262-277.
- Gillespie J.M. (1964) *Aust. J. Biol. Sci.* **17**, 282-300.
- Gillespie J.M., (1968), in *Symposium on Fibrous Proteins* (1967), Butterworths. pp 362-363.
- Gillespie J.M. (1983) in *Biochem. and Physiol. of the Skin* 1st edn (Goldsmith L.A. ed.) Ox. Univ. Press pp 475-510.
- Gillespie J.M. and Inglis A.S. (1965) *Comp. Biochem. Physiol.* **15**, 175-185.
- Gillespie J.M. and Reis P.J. (1966) *Biochem. J.* **98**, 669-677.
- Gillespie J.M. and Broad A. (1972) *Aust. J. Biol. Sci.* **25**, 139-145.
- Gillespie J.M. and Frenkel M.J. (1974) *Comp. Biochem. Physiol.* **47B**, 339.
- Gillespie J.M., and Marshall, R.C., (1980) *Proc. 6th Int. Wool Text. Conf., Pretoria*, pp 67-77
- Gillespie J.M., and Marshall, R.C., (1988), in *The Biol. of Wool and Hair* (Rogers G.E., Reis P.J., Ward K.A., Marshall R.C. eds), Chapman and Hall in press.
- Gillespie J.M., Reis P.J., and Schinckel P.G. (1964) *Aust. J. Biol. Sci.* **17**, 548-560.
- Gillespie J.M., Broad A. and Reis P.J. (1969) *Biochem. J.* **112**, 41-49.
- Gillespie J.M., Frenkel M.J., and Reis P.J., (1980) *Aust. J. Biol. Sci.* **33**, 125-136.
- Gillespie J.M., Marshall R.C., Moore P.M., Panaretto B.A. and Robertson D.M., (1982) *Text. Res. J.* **79**, 197-200.
- Goldberg M.L., (1979) Ph.D. thesis. Stanford Univ.
- Gonzalez I.L, Gorski J.L., Campen T.J., Dorney D.J., Erickson J.M., Sylvester J.E. and Schmickel R.D., (1985), *Proc. Natl. Acad. Sci. USA*, **82**, 7666-7670.
- Grantham R., Gautier C., Gouy M., Mercier R. and Pave, A. (1980 a) *Nucl. Acids Res.*, **8**, r49-r62.

- Grantham R., Gautier C. and Gouy M., (1980 b) *Nucl. Acids Res.*, **8**, 1893-1915.
- Green M.R., Pastewka J.V. and Peacock A.C. (1973) *Anal. Biochem.* **56**, 43-51.
- Grosschedl R. and Birnstiel M.L., (1980), *Proc. Natl. Acad. Sci.*, **77**, 1432-1436.
- Hanahan D. and Meselson M. (1980) *Gene* **10**, 63-67.
- Harrap B.S. and Gillespie J.M. (1963) *Aust. J. Biol. Sci.* **16**, 542-556.
- Hassouna N., Michot B. and Bachellerie J.P., (1984), *Nucl. Acids Res.*, **12**, 3563-3583.
- Holbrook K.A., Fisher C., Daoe B.A. and Hartley R. (1988) in *The Biology of Wool and Hair*, (Rogers G.E., Reis P.J., Ward K.A., Marshall R.C. eds), Chapman and Hall in press.
- Hynd P.I., Schlink A.C., Philips P.M. and Scobie, D.R. (1986) *Aust. J. Biol. Sci.*, **39**, 329-339.
- Hynd P.I. (1988) in *The Biol. of Wool and Hair* (Rogers G.E., Reis P.J., Ward K.A., Marshall R.C. eds), Chapman and Hall pp: 169-184 in press.
- Ish-Horowicz D. and Burke J.F. (1981) *Nucl. Acids Res.* **9**, 2989-2998.
- Jones L.N. (1975) *Biochim. Biophys. Acta* **412**, 91-98.
- Jones L.N. (1976) *Biochim. Biophys. Acta* **446**, 515-524.
- Kabat E.A. (1973) *Proc. Nat Acad. Sci.* **70** (5), 1473-1477.
- Kaplin I.J., and Whitely K.J., (1978) *Aust. J. Biol. Sci.* **31**, 231-240.
- Kataoka T., Broek D. and Wigler M., (1985), *Cell*, **43**, 493-505.
- Khoury G. and Gruss P., (1983), *Cell*, **33**, 313-314.
- Konaska M.M., Padgett R.A. and Sharp P.A. (1984) *Cell* **38**, 731-736.
- Kozak M., (1981) *Nucleic Acids Res.* **9**, 5233-5252.
- Kozak M., (1983), *Microbiol. Rev.*, **47**, 1-45.
- Kozak M., (1984) *Nucleic Acids Res.* **12**, 857-871.
- Kozak M., (1986) *Cell*, **44**, 283-292.
- Kreig P.A. and Melton D.A. (1987) *Methods in Enzymology* **155**, 397-415.
- Kuczek E.S. and Rogers G.E. (1985) *Eur. J. Biochem.* **146**, 89-93
- Kuczek E.S. and Rogers G.E. (1987) *Eur. J. Biochem.* **166**, 79-85
- Kyte J. and Doolittle R.F. (1982) *J. Mol. Biol.* **157**, 105-132.)
- Laemmli U.K. (1970) *Nature* **227**, 680-685.
- Lee L.D., Ludwig K. and Baden H.P. (1978) *Forensic Sci.* **11**, 115-121.
- Leibold E.A. and Munro H.N., (1988) *Proc. Nat. Acad. Sci. USA*, **85**, 2171-2175.

- Ley K.F., Crewther W.G. (1980) *Proc. 6th Int. Wool Text. Res. Conf.* (1980) Pretoria vol. II, 13-27
- Lilley D.M., (1980), *Proc. Natl. Acad. Sci. USA*, **77**, 6468-6472.
- Lilley D.M., (1988), *Trends in Genet.*, **4**, 111-114
- Lim V.I. (1974) *J. Mol. Biol.* **88**, 857-872 and 873-894.
- Lindley H., Gillespie J.M. and Haylett T. (1968), *Symposium on Fibrous Proteins 1967*, (Ed. W.G. Crewther.), 353-361. (Butterworths (Australia): Sydney)
- Lonsdale-Eccles J.D., Resing K.A., Meek R.L. and Dale B.A., (1984), *Biochemistry*, **23**, 1239-1245.
- Lynch M.H., O'Guin W.M., Hardy C., Mak L. and Tung-Tein Sun (1986). *J. Cell Biol.* **103**, 2593-2626.
- Maeda N. and Smitheis O., (1986), *Ann. Rev. Genet.*, **20**, 81-108.
- Mandel M. and Higa (1970) *J. Mol. Biol.* **53**, 154.
- Maniatis T, Fritsch E.F., and Sambrook J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Press.
- Marshall R.C., (1981 a), *Textile Res. J.*, **51**, 106-108.
- Marshall, R.C. (1981 b) *Text. Res. J.* **51**, 384-388.
- Marshall R.C., (1983), *J. Invest. Dermatol.* **80**, 519-524.(check)
- Marshall R.C. and Gillespie J.M. (1976 a) *Aust. J. Biol. Sci.*, **29**, 1-10
- Marshall R.C. and Gillespie J.M. (1976 b) *Aust. J. Biol. Sci.*, **29**, 11-20.
- Marshall R.C. and Ley K.F. (1986) *Text. Res. J.* **56**, 772-774
- Marshall R.C., Gillespie J.M., Inglis A.S. and Frenkel M.J. (1980) *Proc. 6th Int. Wool Text. Conf.* Pretoria Vol. II pp 147-158.
- Marshall, R.C. and Gillispie, J.M. (1988), in *The Biol. of Wool and Hair* (Rogers G.E., Reis P.J., Ward K.A. and Marshall R.C. eds) Chapman and Hall in press.
- McKnight S.L., and Kingsbury R., (1982), *Science*, **217**, 316-324.
- McLauchlan J., Gaffney D., Whitton J.L., and Clements J.B., (1985). *Nucl. Acids Res.* **13**, 1347-1368.
- McLean J.W., Tomlinson J.E., Kuang W-J., Eaton D.L., Chen E.Y., Fless G.M., Scanu A.M. and Lawn R.M., (1987), *Nature*, **300**, 132-137.
- Melton D.A., Krieg P.A., Rebalati M.R., Maniatis T., Zinn K. and Green M.R. (1984), *Nucl. Acids. Res.*, **12**, 7035-7056
- Mercer E. H. (1961) *Keratin and Keratinization* (P. Alexander P. and Bacq Z. M eds) Pergamon.
- Merril G.R., Goldman D. and Keuren M.L.V. (1984) *Methods in Enzymology* **104**, 441-447.

- Messing J., (1979), *Recombinant DNA Technical Bulletin*, **2**, 43-48.
- Messing J. (1983) *Methods in Enzymology* **101**, 20-78.
- Miller J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Moldave (1985), *Ann. Rev. Biochem.* **54**, 1109-1149.
- Moreau D., Hen R., Wasylyk B., Everett R., Gaub M.P. and Chambon P. (1981) *Nucl. Acids Res.* **9**, 6047-6068.
- Mount S.M. (1982) *Nucleic Acids Res.* **10**, 459-472.
- Nagano K. (1973) *J. Mol. Biol.* **75**, 401-420.
- Nevins J.R., (1983) *Ann. Rev. Biochem.* **52**, 441-466.
- Norlander J., Kempe T. and Messing J. (1983) *Gene* **26**, 101-106.
- Ohno S., (1984), *J. Mol. Evol.*, **20**, 313-321.
- Orwin D.F.G. (1976) *J. Ultrastruct. Res.* **55**, 312.
- Orwin D.F.G. (1979) *Int. Rev. Cytol.* **60**, 331-374.
- Orwin D.F.G. and Thomson R.W. (1972) *J. Cell Biol.*, **11**, 205-219.
- Parry D.A.D., Fraser R.D.B. and MacRae T.P. (1979) *Int. J. Biol. Macromol.* **1**, 17-22.
- Pelham H.R.B. and Jackson R.J. (1976) *European J. Biochem.* **67**, 247-256.
- Platt T., (1986), *Ann. Rev. Biochem.*, **55**, 339-372.
- Powell B.C. and Rogers G.E. (1986) in *Biology of the Integument vol. 2 Vertebrates* (J. Bereiter-Hahn A.G Matoltsy K.S. Richards Eds) Springer-Verlag **2**, 695-721.
- Powell B.C., Sleigh M.J., Ward K.A. and Rogers G.E. (1983) *Nucleic Acids Res.* **11**, 5327-5346.
- Powell B.C., Cam G.R., Fietz M.J. and Rogers G.E. (1986) *Proc. Natl. Acad. Sci.* **83**, 5048-5052
- Powell B.C., Kuczek E.S., Crocker L., O'Donnell M. and Rogers G.E. (1988) in *The Biol. of Wool and Hair* (Rogers G.E., Reis P.J., Ward K.A. and Marshall R.C. eds) Chapman and Hall in press.
- Proudfoot N.J., Brownlee G.G., (1976) *Nature*, **263**, 211-214.
- Radloff R., Baver W. and Vinograd J. (1967) *Proc. Natl. Acad. Sci.* **57**, 1514.
- Reed K.C. and Mann D.A. (1985) *Nucl. Acids Res.* **13**, 7207-7221.
- Reis P.J. (1965) In *Biology of the Skin and Hair Growth* (Lyne A.G. and Short B.F., eds.) pp 365-375.
- Reis P.J. (1967) *Aust. J. Biol. Sci.* **20**, 809-825.
- Reis P.J. (1979) in *Physiological and Environmental Limitations of Wool Growth* (Black J.L. and Reis, P.J. eds.) University of New England Publishing Unit pp. 223-242.

- Reis P.J. (1982), *Proc. Aust. Soc. Anim. Prod.* pp 479-482.
- Reis P.J. (1988) in *The Biol. of Wool and Hair* (Rogers G.E., Reis P.J., Ward K.A., Marshall R.C. eds), Chapman and Hall pp. 185-203 in press.
- Reis P.J. and Schinckel P.G. (1963) *Aust. J. Biol. Sci.* **16**, 218.
- Reis P.J. and Schinckel P.G. (1964) *Aust. J. Biol. Sci.* **17**, 532.
- Reis P.J., and Tunks D.A. (1974) *Aust. J. Agric. Res.*, **25**, 919.
- Reis P.J. and Tunks D.A. (1982) *Aust. J. Biol. Sci.* **35**, 49-62.
- Reis P.J. and Gillespie J.M. (1985) *Aust. J. Biol. Sci.* **38**, 151-163.
- Reis P.J., Tunks D.A. and Downes A.M. (1973) *Aust. J. Biol. Sci.* **26**, 249-258.
- Rigby P.W.J., Dieckmann M., Rhodes C. and Berg P. (1977) *J. Mol. Biol.*, **133**, 237-251.
- Rogers A.W. (1979) in *Techniques of Autoradiography* (Elsevier/North-Holland Biomedical Press, Amsterdam).
- Rogers G.E. (1959) *Ann. N.Y. Acad. Sci.* **83**, 378-399.
- Rogers G.E. (1964) in *The Epidermis* (Montagna, W. and Lobitz W.C., eds.) Academic Press.
- Rogers G.E. (1963) *J. Histochem. Cytochem.* **11**, 700.
- Rogers G.E. and Harding H.W.J., (1971) *Biochemistry*, **10**, 624-630
- Rogers G.E., Kuczek E.S., MacKinnon P.J., Presland R.B., Fietz M.J., (1988), in *The Biol. of Wool and Hair* (Rogers G.E., Reis P.J., Ward K.A., Marshall R.C. eds), Chapman and Hall in press.
- Rouault T.A., Hentze M.W., Caughman H.S., Hartford J.B. and Klausner R.D. (1988) *Science*, **241**, 1207-2120.
- Rylatt D.B. and Parish C.R. (1982) *Anal. Biochem.* **121**, 213-214.
- Sanger F., Nicklen S. and Coulson A.R. (1977) *Proc. Natl. Acad. Sci.* **74**, 5463-5467.
- Sanger F., Coulson A.R., Barnell B.G., Smith A.J.H. and Roe B.A. (1980) *J. Mol. Biol.* **143**, 161-178.
- Schartz D.C. and Cantor C.R. (1984) *Cell* **37**, 67-75.
- Shlomai J. and Kornberg A., (1980), *Proc. Natl. Acad. Sci. USA*, **77**, 799-803.
- Shenk T. (1981) *Curr. Top. Microbiol. Immunol.* **93**, 25-46.
- Southern E.M. (1977) *Methods in Enzymology* **68**, 152-176.
- Sullivan K.M. and Lilley D.M., (1986), **47**, 817-827.
- Swart L.S. (1973) *Nature* **243**, 27-29
- Swart L.S., Jourbert F.J. and Parris D. (1976) *Proc. 5th. Int. Wool Text. Conf. 1975 Aachen Vol II*, 254-263.

- Swift J.A. (1976) *Proc. 5th Int. Wool Text. Conf.* 1975 Aachen vol II, 12-22.
- Steinert P.M. and Rogers G.E. (1971) *Biochim. Biophys. Acta.* **238**, 150-155.
- Steinert P.M. and Rogers G.E. (1973) *Biochim. Biophys. Acta.* **312**, 403-412.
- Tabor S. and Richardson C.C. (1987) *Proc, Nat. Acad. Sci. USA*, **84**, 4767-4771.
- Tenehouse H.S. and Gold R.J.M. (1976) *Biochem. J.* **159**, 149-160.
- Thomas P.S. (1983) *Methods in Enzymology* **100B**, 255-266.
- Thompson E.O.P. and O'Donnell I.J. (1962) *Aust. J. Biol. Sci.* **15**, 757-68.
- Vogelstein B, and Gillespie D. (1979) *Proc. Natl. Acad. Sci., USA*, **76**, 615.
- Wallace, A.L.C., (1979), in *Physiological and Environmental limitations to wool Growth* (Black, J.L. and Reis, P.J., eds) University of New England Publishing Unit.
- Ward K.A., *et al.* (1988) *Gene*, in press.
- Ward K.A. and Kasmarik S.E. (1980) *J. Invest. Dermatol.* **75**, 244-248
- Weiher H., Konig M. and Gruss P. (1983) *Science* **219**, 626-631.
- Weintraub H., Pei Feng Cheng and Conrad K., (1986), *Cell*, **47**, 817-827.
- Wilkinson E.F. (1971) *Biochem. J.* **125**, 371-373.
- Wilson P.A., Henrikson R.C. and Downes A.M. (1971) *J. Cell Sci.* **8**, 489-512.
- Winter G. and Fields S. (1980) *Nucl. Acids Res.* **8**, 1965-1974.
- Woods E.F. and Gruen L.C. (1981) *Aust. J. Biol. Sci.* **34**, 515-526.
- Zeugin J.A. and Hartley J.L. (1985) *Focus* **7**, no. 4, 1-2.