HIGH-GLYCINE/TYROSINE KERATIN GENES OF WOOL

Elizabeth Salome Kuczek B.Sc. (Hons.)

A thesis submitted for the degree of
Doctor of Philosophy in the University of Adelaide
Department of Biochemistry
March, 1985

CONTENTS

THESIS SUMMARY

The investigations presented in this thesis were aimed at examining the structure and complexity of the high-glycine/tyrosine (HGT) keratin genes expressed in the wool fibre of the domestic sheep, *Ovis aries*.

These studies fell into two categories: (i) construction and characterisation of nucleic acid probes encoding HGT proteins of the wool fibre and, (ii) use of such probes to isolate corresponding genes from the sheep genome and to examine their distribution by Southern blot analysis.

Total poly(A)$^+$ RNA isolated from wool follicles and purified by affinity chromatography was found to direct the synthesis of wool keratin polypeptides in a wheat germ cell-free system. A cDNA library was constructed from total wool follicle mRNA and screened by two different approaches. The first of these involved the use of oligo(dT)-primed cDNA synthesised to partially enriched-HGT RNA prepared by sucrose gradient fractionation. A clone was isolated, which on sequence analysis could not be identified as containing HGT sequences. A more specific screening approach, using synthetic oligonucleotides as probes, was then adopted which selected cDNA clones for the two major HGT components present in the wool fibre. These clones, although not containing full-length mRNA sequences, encoded protein components F and C2, with only a few minor discrepancies with the available protein sequence data. The two cloned sequences did not cross-hybridize with each other due to limited (22%) sequence homology in their coding regions, with no homology between the 3' non-coding regions. These sheep cDNA clones, pSHGT-F and pSHGT-C2, are the first coding sequences cloned and characterised for proteins rich in glycine and the aromatic residues, tyrosine and phenylalanine. Proteins with similar compositions have been found in a wide variety of mammalian species, as well as in certain structures of birds, insects and reptiles.

Using pSHGT-F as a probe, the HGT-F gene was isolated from a sheep genomic library and sequenced in its entirety together with about 1.5 kb of flanking region DNA. The HGT-F gene shows the conventional eukaryotic gene consensus signals in addition to an 18-base sequence element, 5'-CCGCCCAACCCAGACACC-3', preceding the initiation codon, ATG. This sequence is conserved in most high-sulphur keratin genes (Powell *et al.*, 1983) and may constitute a wool matrix-gene specific control element.

Genomic Southern blot analysis indicated that HGT-F was an

unique gene and that in addition to the HGT-C2 gene, there were at least five other C2-related sequences in the sheep (Merino) genome. At present, the full extent of the heterogeneity of the HGT gene family, in both size and sequence, has not been determined. Additional protein or gene sequence data is required to establish whether there is further HGT protein sequence dinergence (as seen for component F) or whether there exists a common structural homology as implied by the HGT-C2 Southern blot data.

This thesis also reported the occurrence of HGT keratin and related genes in the human, mouse and phenotypically HGT-deficient sheep genomes. Using the isolated HGT-F and C2 cDNA clones as probes, cross-hybridization was observed in the human and mouse genomes, however the nature of these sequences was not examined. Similarly, HGT-F and/or C2 sequences were detected in the DNA from Lincoln and Felting lustre-mutant sheep, both of which have a wool HGT protein content of less than 1%. Interestingly, the presence of the structural genes for HGT components F and C2 in the genome of the Felting lustre-mutant Merino suggested that the negligible level of HGT proteins in this mutant fleece-type was presumably due to a defect at the level of HGT gene transcription or mRNA translation.

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.

ELIZABETH SALOME KUCZEK

ACKNOWLEDGEMENTS

I wish to thank *Professor W. H. Elliott* for the opportunity to undertake these studies in the department.

I sincerely thank my supervisor, *Professor G. E. Rogers* for his advice, support and encouragement during the course of this work. In addition, I would like to express my appreciation to the following people:

Dr. Adrian Egan and *Mr. Ian Ridgway* at the Waite Institute for their technical aid concerning the animal work.

Past and present members of *'Keratin Korna'* for providing an enjoyable atmosphere in which to work. Special thanks to *Lesley Crocker*, for her technical assistance and in the preparation of photographic diagrams, to *Dr. Barry Powell* and *Dr. Keith Gregg*, for their critical reading of this thesis.

Dr. Michael Bawden, Bill Kalionis and other members of the department for their help and friendly discussions.

My family, for their understanding and support.

Mrs. To, for her care in typing this thesis.

I wish to acknowledge that during the course of this work I was supported by a Commonwealth Postgraduate Award.

ABBREVIATIONS

A_n:	absorbance measured at wavelength, n
bp:	base pair kb: kilobase
BSA:	bovine serum albumin
cDNA:	DNA complementary to mRNA
Ci:	curie
cpm:	counts per minute
ddNTP:	dideoxynucleoside triphosphate
DMSO:	dimethylsulphoxide
DNA:	deoxyribonucleic acid
dNTP:	deoxynucleoside triphosphate
DTT:	dithiothreitol
Hepes:	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
oligo(dT):	oligodeoxythymedylic acid
pfu:	plaque forming units
poly(A):	polyadenylic acid
POPOP:	1,4-bis-[2-(5-phenyloxazolyl)]-benzene
PPO:	2,5-diphenyloxazole
RNA:	ribonucleic acid mRNA: messenger RNA
	rRNA: ribosomal RNA
	tRNA: transfer RNA
rpm:	revolutions per minute
S:	Svedberg unit of sedimentation coefficient
S-23:	supernatant after centrifugation at 23,000g
TCA:	trichloroacetic acid
Tris:	2-amino-2(hydroxymethyl)-1,3-propandiol
UV:	ultraviolet

CHAPTER 1

INTRODUCTION

CHAPTER 1

1.1 GENERAL INTRODUCTION

The wool follicle is an example of a terminally differentiating tissue which produces a structure consisting of distinctive and co-ordinately expressed groups of proteins, the wool keratins. Although numerous studies have been carried out on the constituent proteins and histology of the wool follicle and fibre, little is known about the molecular events leading to follicle development and the control of specific keratin gene expression. Clearly many processes are involved in follicle differentiation and fibre development, the most basic of which is the selective expression of genes in particular cell types and the suppression of other genetic material. The fundamental question which needs to be resolved is: how are specific gene products produced in the different cell types of the fibre, all of which arose from pluripotent cells at the base of the follicle?

The development of recombinant DNA techniques enables the purification and preparation of individual keratin genes, which will allow an investigation into the selective co-ordinate expression of the different keratin gene families. Furthermore, there exists the possibility of applying this type of information to the economically important wool industry and to related medical aspects of diseases and abnormalities in the keratins of hair and skin. The remainder of this chapter is aimed at providing a background to the research presented in this thesis and consequently reviews the morphology and formation of the wool follicle, the protein composition of the wool fibre, in particular the high-glycine/tyrosine wool proteins, and finally the current status of keratin gene structure and organization.

1.2 WOOL STRUCTURE AND BIOSYNTHESIS

1.2.A Follicle Morphology and Development

Hair and wool are keratinized fibres whose development and structures are essentially similar. The mature mammalian hair (and wool fibre) consists of a column of dead epidermal cells and is produced in follicles formed by a downgrowth of the epidermis (Mercer, 1961). At the base or bulb of the follicle there is an invagination of the surrounding connective tissue called the dermal papilla which is separated by a membrane from the pluripotent bulb cells destined to form the fibre and its associated structures. These early steps in the path of follicle development have not been elucidated but are probably hormone-responsive and their effect is undoubtedly exerted through the interaction of the dermal papilla with immature follicle bulb cells.

The histological structures of the follicle and fibre and their differentiation has been described in detail elsewhere (for reviews see Rogers, 1964; Epstein and Maibach, 1969; Orwin, 1979) and will only be discussed briefly here. Figure 1.1 shows a diagrammatic representation of the hair (wool) follicle and indicates the positions of the five developmental stages:

(1) Undifferentiated zone:

In this region around the papilla at the base of the follicle, cells are undifferentiated, spherical in shape. In active follicles, cells are rapidly dividing and constantly forcing columns of epithelial cells further upwards.

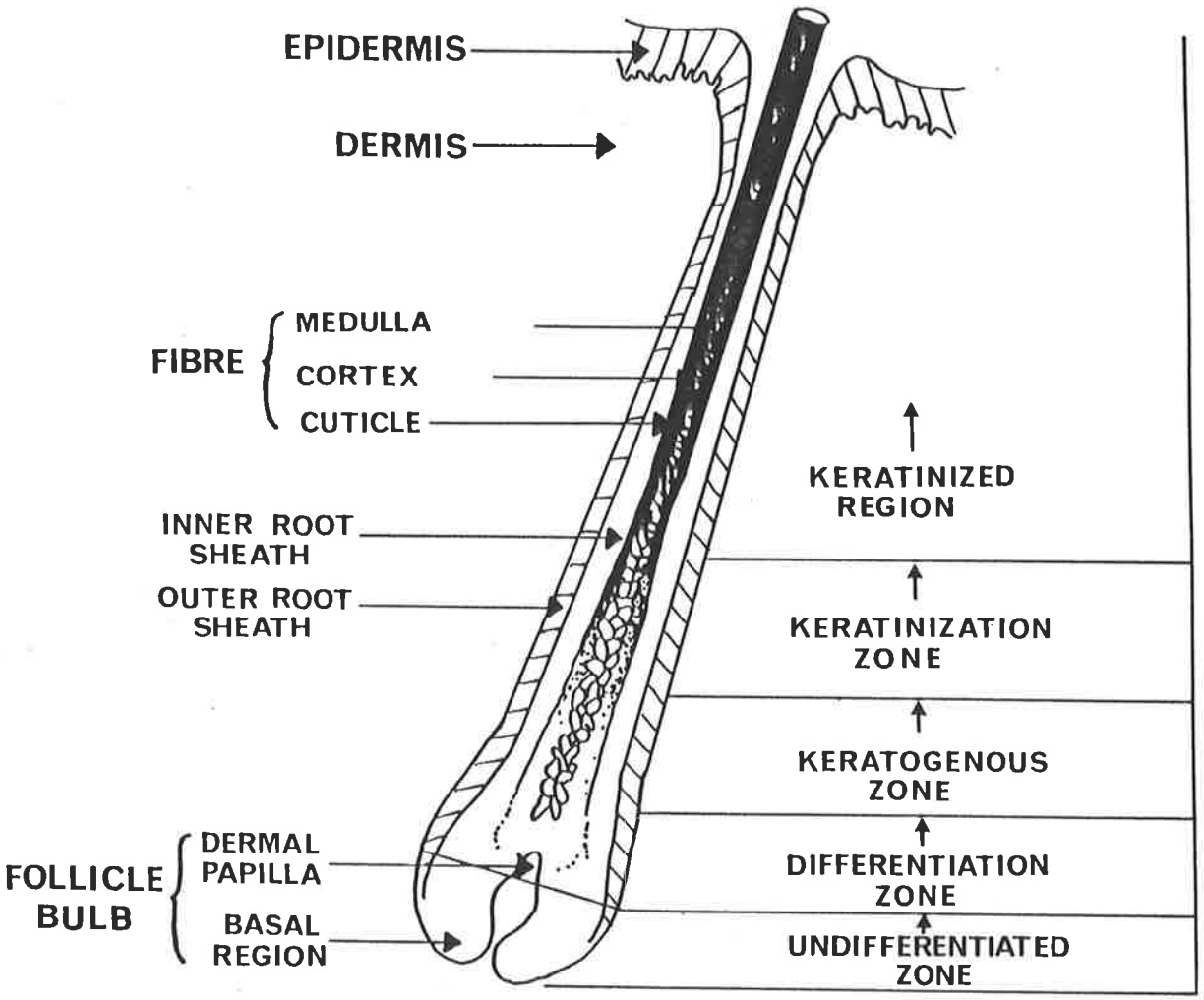
(2) Differentiation zone:

As the cells rise they differentiate to form the characteristic cell layers of the follicle, which arranged concentrically and proceeding inwards are (a) the outer root sheath (b) the inner root sheath (c) the hair fibre which consists of a cuticle, cortex and medulla, although in some fibres the latter is absent. These cell layers all originate in the follicle bulb, with the addition that the outer root sheath also contains cells of epidermal origin. Differentiation is observed as a change in the shape of cells with alterations in the cellular components. In this region large numbers of ribosomes, polysomes and mitochondria appear in the cell cytoplasm.

FIGURE 1.1

WOOL FOLLICLE MORPHOLOGY

Diagram of a longitudinal section of a medullated wool follicle showing the developmental zones and major histological regions discussed in the text.
Adapted from Montagna and Parakkal (1974).


EPIDERMIS
DERMIS
FIBRE
MEDULLA
CORTEX
CUTICLE
INNER ROOT SHEATH
OUTER ROOT SHEATH
FOLLICLE BULB
DERMAL PAPILLA
BASAL REGION
KERATINIZED REGION
KERATINIZATION ZONE
KERATOGENOUS ZONE
DIFFERENTIATION ZONE
UNDIFFERENTIATED ZONE

(3) Keratogenous zone:

Here the cell layers are fully committed and show distinct morphologies and protein products. Keratin proteins reside in the cuticle and cortex and are distinctly different from the proteins present in the other cell layers. An additional structural feature of the cortex of wool fibres is its asymmetry in cross-section. Approximately one half of the cortex (orthocortex) has cells in which the composition and arrangement of keratin proteins is different to that seen in the other half or paracortex (Rogers, 1959 a, b; Kaplin and Whiteley, 1978).

(4) Keratinization zone:

In this region cells dehydrate, nuclei and cytoplasmic organelles disintegrate.

(5) Keratinized region:

The cells here are dead and packed with protein. Organelle structures have disappeared.

The pathway of cyto-differentiation as studied by various workers (Short et al., 1965; Sims, 1967; Wilkinson, 1969; Epstein and Maibach, 1969) can be viewed in relation to the above description of wool fibre development as follows.

Initially each cell destined to form part of the fibre is undifferentiated, synthesizing DNA and dividing (undifferentiated zone), mitosis only occurring in the bulb of the follicle. RNA synthesis begins (differentiation zone) and ribosomal and messenger RNA species move into the cytoplasm. The synthesis of specific structural proteins is initiated (keratogenous zone), then DNA synthesis ceases and the nucleus becomes permanently inactive while protein synthesis continues (keratinization zone).

It is obvious from this brief summary of the differentiation process that a number of co-ordinated changes must occur within each cell as it moves up the follicle. Differential gene expression is indicated by the fact that different components are found in the various cell types, for eg., the bilateral (ortho/para) structure of the cortex; however control at the translational and post-translational levels may also be involved.

1.2.B Protein Composition

The keratin proteins of a hair (or wool) fibre reside in the cortex and the thin outer covering of cuticle. Most of the knowledge concerning the chemical and physical properties of these proteins has been obtained from numerous studies carried out on wool (for reviews see Crewther, 1976, also Ley and Crewther, 1980). However, there is sufficient data available for other mammalian hairs to indicate that the general protein composition observed in wool also applies to hair structures (see Gillespie, 1983 for review). Necessarily then, most of the present discussion will draw on the wool model and specifically the cortical keratins, since the cortex is the predominant structure in wool and the greater proportion of extractable proteins derive from it.

The mandatory first stage in the study of the wool keratins requires protein disaggregation and solubilization of the three-dimensional disulphide-bonded polymeric structure. The best procedure appears to be the reductive one using 2-mercaptoethanol and the blocking of -SH groups with iodoacetate as originally used by Goddard and Michaelis (1934). The derivatized soluble proteins are referred to as S-carboxymethylkerateines or SCMK and in this form have been readily fractionated by chemical and gel electrophoretic methods into three major classes; the low-sulphur proteins - referred to collectively as SCMKA, the high-sulphur proteins - SCMKB, and the high-glycine/tyrosine proteins. Evidence from X-ray and electron microscopic studies (Rogers 1959a, b; Fraser *et al.*, 1972) have also shown that the keratin of the cortex is made up of 8 nm diameter filaments embedded in a matrix and these same studies indicated that the low-sulphur proteins are the subunits of the microfibrils, and the high-sulphur and high glycine/tyrosine proteins reside in the inter-microfibrillar matrix. Details on these proteins will now be discussed, the majority of the information being obtained from the reviews by Crewther (1976), Gillespie (1983) and Powell and Rogers (1985).

1. Low-sulphur proteins

The low-sulphur proteins constitute about 60% of the proteins in a fibre and form the filamentous structures of mammalian keratins, the microfibrils. They are composed of relatively low levels of SCM-cysteine, and high proportions of glutamic and aspartic acid, leucine, lysine and arginine. SDS-polyacrylamide gel electrophoresis and isoelectric-focusing data reveal two major families of proteins known as components 7 and 8, and a minor family, component 5 (see Table 1.1) each with a number of individual chains with molecular weights in the range 40,000-60,000. Protein sequence data has recently indicated that component 5 belongs to the component 7 family (L. Sparrow, M. Frenkel personal communications). Partial amino acid sequences are known for components 5 and 7c (Dowling et al., 1979; Sparrow and Inglis, 1980) and one protein, component 8-c1, of the other low-sulphur group has been fully-sequenced (Crewther et al., 1982). The low sulphur proteins are distinguished from other keratin proteins by the presence of α-helix. The complete sequence of the α-helical rich fragments showed that sequence homology (about 30%) was present but that the degree of homology enabled two sequence types to be distinguished, which were referred to as Type I and Type II. They are known to originate respectively from low-sulphur proteins, component 8 and component 7 (Crewther, 1976).

The most interesting advance in the knowledge of hair keratin structure has come about from the study of the genes for cyto-keratins (Hanukoglu and Fuchs, 1982, 1983; Steinert et al., 1983) which has been correlated with the protein chemical studies of wool keratins (Dowling et al., 1983; Crewther et al., 1983). It was realized that the low-sulphur proteins are a subclass of cyto-keratins which are members of the larger group of structural proteins that constitute the 8-10 nm intermediate filaments or IF, generally found in the cytoplasm of cells (for reviews see Fuchs and Hanukoglu, 1983; Steinert et al., 1984). Comparison of the

TABLE 1.1

WOOL KERATIN PROTEINS

The characteristic features of the various keratin protein families present in the wool fibre are indicated in the table opposite. Protein data has been compiled from Crewther (1976); Swart et al., (1976) and Gillespie (1983).

a The prefix SCMK indicates that the proteins are the S-carboxymethylated derivatives. The numbers in parentheses represent estimates for the number of individual chains in the family.

b Recent amino acid data (L. Sparrow, personal communication) has shown that SCMKA-5 and SCMKA-7 are highly homologous and as such belong to the same and not separate sub-families as previously indicated (Crewther, 1976).

c The proteins of this group have not been characterized.

d There is some protein-chemical data which suggests that the Type I and Type II high-glycine/tyrosine (HGT) proteins should be considered as subclasses, each containing several families of proteins. However, the numbers of different polypeptide chains may be a conformational anomaly.

PROTEIN CLASS AND LOCATION IN FIBRE	CHARACTERISTIC AMINO ACID CONTENT	PROTEIN FAMILIES WITHIN THE CLASS	FAMILY NOTATION[a]	MOLECULAR WEIGHT
LOW-SULPHUR (MICROFIBRILS)	Cysteine (1-3 moles %)	2	SCMKA - 5/7[b] (1)/(3) SCMKA - 8 (4)	56-58,000 38-43,000
HIGH-SULPHUR (MATRIX)	Cysteine (20-30 moles %)	4	SCMKB - 1 (?)[c] SCMKB - 2 (7) SCMKB - IIIA (11) SCMKB - IIIB (4)	23-26,000 19,000 16,000 11,000
ULTRA-HIGH SULPHUR (MATRIX AND CUTICLE)	Cysteine (> 30 moles %)	?	-	> 20,000
HIGH-GLYCINE/TYROSINE (MATRIX)	Glycine (20-40 moles %) Tyrosine (12-21 moles %)	2	Type I (10)[d] Type II (5)	6-9,000 6-9,000

amino acid sequence of IF proteins showed common structural features including similar peptide domains and extensive regions of coiled-coil α-helix (Steinert et al., 1980; Geisler et al., 1982; Geisler and Weber, 1982). It is now apparent that the structural pattern of the low sulphur components 7 and 8 is homologous to that of other IF proteins in which helical and non-helical domains are present.

2. High-sulphur proteins

Every hard α-keratin (wool, hair, hoof, horn, quill) which has been examined has been found to contain sulphur-rich proteins in amounts wich vary from 7 to 10% for some horns and up to 50% in some hairs (Gillespie and Frenkel, 1974). In wool the high-sulphur protein content varies from 20-30% and is located mainly in the matrix surrounding the microfibrils. These proteins have a high content of SCM-cysteine (about 20-30%), are rich in proline, serine and threonine but contain little histidine, lysine and no methionine. These proteins, collectively designated as SCMKB, have a molecular weight range from 10,000-30,000 (Table 1.1). The high-sulphur proteins are highly heterogenous and their exact number is not known. Their fractionation by chromatography on DEAE-cellulose and subsequent gel electrophoresis suggested at least 35 separate proteins and a similar number can be distinguished by two-dimensional electrophoresis (Darskus, 1972). The high-sulphur proteins are grouped into four sub-families on the basis of their molecular weights (Table 1.1) with each group comprising a number of closely related components. More wool high-sulphur proteins have been sequenced than any other type of keratin protein, with 22 different amino acid chains known. Of these, 21 were directly sequenced (Swart et al., 1976; Crewther, 1976) and one SCMKB-2 sequence was deduced from gene cloning and sequencing (Powell et al., 1983). All of the sequenced proteins have a C-terminal SCM-cysteine residue and both the SCMKB-IIIA

and SCMKB-2 families contain considerable proportions of a repetitive cystine-rich sequence which has been interpreted as either a pentapeptide (Swart, 1973; Parry et al., 1979) or decapeptide repeat (Elleman et al., 1973). The pentapeptide repeating unit takes the form, Cys-Cys-$\frac{\text{Arg}}{\text{Gln}}$-Pro-X where X can be serine, threonine or valine. Elleman et al., (1973) has suggested that these proteins may have arisen from a primordial sulphur-rich polypeptide by a series of duplications of the gene coding for the repeating sequence. The apparent continuous distribution of molecular sizes in the ultra-high-sulphur proteins (first described by Gillespie et al., 1964), a family which is particularly rich in the amino acids of the repeating pentapeptide, may provide support for Elleman's suggestion. Very little is known about this group of proteins other than their high SCM-cysteine content (in excess of 30%) and their molecular weights which appear to be greater than that seen for the high-sulphur proteins. An equivalent group of proteins appear to be present in mouse and human hair which in contrast to the wool proteins, are synthesized constitutively. No single species has been isolated and sequenced for this group of proteins (Powell and Rogers, 1985).

At present the nature of the high-sulphur protein structures has not been clearly established. Protein sequence data would seem to suggest that α-helix and β-structures do not exist, possibly due to the disordering effect of the large numbers of negatively-charged SCM-cysteine groups. Studies on the solutions of SCMKB components isolated from wool have provided no evidence for either of these two types of structures.

3. High-glycine/tyrosine proteins

The high-glycine/tyrosine proteins together with the high-sulphur proteins form the intermicrofibrillar matrix of the cortex. This conclusion is based on the indirect evidence that the proportion of matrix space calculated from X-ray diffraction studies equates closely

with the total proportion of high-sulphur and high-glycine/tyrosine proteins estimated by isolation (Fraser et al., 1973). In wool, this heterogenous group of small proteins with molecular weights in the range 5,000-10,000 (see Table 1.1), is found in varying amounts up to 13% of the total fibre protein content. They have a very unusual amino acid composition, with high contents of glycine (25-40%) and tyrosine (15-20%), moderately high contents of phenylalanine and serine, negligible levels of lysine, isoleucine, histidine, glutamic acid, alanine and no methionine.

There are two major families of high-glycine/tyrosine proteins. Type I is poor in SCM-cysteine and rich in phenylalanine while the reverse is true of Type II components (not to be confused with Type I and Type II low-sulphur filament proteins). These proteins are heterogenous in number (Zahn and Biela, 1968; Gillespie and Frenkel, 1974), although the extent of heterogeneity has not been fully determined and appears to be in part, due to an artefact of the protein extraction procedure (Frenkel, 1977), and/or conformational anomalies occurring during the separation procedures as a result of the high content of aromatic residues. It appears that these proteins may not be as heterogenous as first thought (Marshall et al., 1980). Complete amino acid sequence data has been obtained for three Type I components from wool (Dopheide, 1973; Marshall et al., 1980). Two of the proteins sequenced are members of the one sub-family and there is no difference in the sequences to indicate why they are isolated as separate components by ion-exchange chromatography (Marshall et al., 1980). There is also no apparent homology between these two components and the protein sequenced by Dopheide (1973). A more detailed discussion on the high-glycine/tyrosine proteins relevant to this thesis is given below in section 1.3.

1.2.C Protein Synthesis

The mechanism of synthesis of the different keratin protein families within the cortex of the wool fibre has been the subject of debate over many years. Rudall (1956) reported that the sulphur content of the horn keratin protein increased with age of the keratinocyte, possibly reflecting the synthesis of low-sulphur proteins prior to that of high-sulphur proteins. The early radioisotope work of Downes et al., (1963) and Rogers (1964) also supported this 'two-stage' process. However, subsequent studies by Fraser (1969a, b), in which keratins synthesized in three succeeding levels of the fibre were examined, suggested that concurrent synthesis occurred, although at different rates, with a linear synthesis of the low-sulphur proteins but an exponential synthesis of the high-sulphur proteins. Chapman and Gemmell (1971) refined this interpretation of protein synthesis through further electron microscopic work and suggested different patterns of synthesis for the ortho- and paracortex, the two distinct cellular regions responsible for the bilateral nature of the wool fibre (Horio and Kondo, 1953). Microfibrils (low-sulphur proteins) and matrix (high sulphur and high-glycine/tyrosine proteins) are synthesized simultaneously in the orthocortex but sequentially in the paracortex, with the production of matrix initially lagging behind the microfibrils then increasing sharply until there is dual synthesis of both. The relative rates of synthesis of the different classes of keratin during fibre formation remains inadequately answered, particularly when one considers that there is no data available on the biosynthesis of the high-glycine/tyrosine proteins.

Very little is known of the process of aggregation and stabilization of the microfibril-matrix complex. Recent physiochemical and cross-linking studies have shown that the low-sulphur helical chains have a preferred mode of association to dimers and tetramers, where the subunit polypeptide chains form pairs of double-stranded coiled coils composed of one member of each low-sulphur protein family, one Type I and

one Type II (Woods and Gruen, 1981). In the case of the high-sulphur protein structure, Lindley (1977) and Parry et al., (1979) examined features of these amino acid sequences and in particular the repeating sequences and suggested theoretical evidence for the possible existence in the native state of highly convoluted, looped, folded structures stabilized by intrachain disulphide bonds. A similar structural arrangement may also exist for the high-glycine/tyrosine sequences (see section 1.3.C). Since each class of matrix protein has a clear demarcation into two subgroups, in the high-glycine/tyrosine proteins based on differences in amino acid composition (Type I and Type II), and in the high-sulphur proteins on the presence or absence of the pentapeptide repeating sequence, it might be expected that subgroups within each class will have differences in peptide chain folding and perhaps of location and function in the matrix (Gillespie, 1983). Information concerning this proposition is lacking nor is anything known on the method by which the matrix proteins infiltrate the microfibrils, the disposition of disulphide bonds within the matrix proteins and the final process in the stabilization of the microfibril-matrix protein structure by copper-catalyzed disulphide bond formation, resulting in a three-dimensional network of protein chains.

1.2.D Factors Affecting Wool Composition and Growth

The keratin proteins in the fibre are all subject to variation through genetic, dietary and physiological factors (Frenkel et al., 1974; Reis, 1979; Gillespie et al., 1982). The least variable are the low-sulphur and a proportion of the high-sulphur proteins, thought to be essential for the development of the basic fibre structure (Gillespie and Reis, 1966; Fraser and MacRae, 1980; Gillespie and Marshall, 1983), while the most profound changes occur in the ultra-high-sulphur and high-glycine/tyrosine protein families (Frenkel et al., 1974; Marshall

and Gillespie, 1981, Gillespie and Marshall, 1983). Some of the factors affecting wool composition and growth will now be discussed.

1. Genetic control

Comparative studies on wool from different breeds of sheep has shown a wide range in keratin composition due to differing ratios of matrix to microfibrils, but also of changes in the relative proportions of the two matrix protein classes (Gillespie, 1983). For example, Lincoln sheep normally produce wool which contains less than 1% high-glycine/tyrosine protein and approximately 20% high-sulphur protein in contrast to that grown by the fine wool of Merino, which contains about 13% and 30% of these two protein classes respectively (see Figure 1.2). It is interesting to note that the hard keratins produced in the various keratinized structures of one animal appear to contain identical groups of proteins, although the relative proportions may differ. This has been shown for horn, hoof and wool (Gillespie, 1968; Marshall and Gillespie, 1977) with the variability presumably reflecting the different structure-function requirements of each structure.

Variations in the wool protein content from the same breed of sheep have been observed (Frenkel et al., 1974) and these presumably are due to other contributing factors as discussed below.

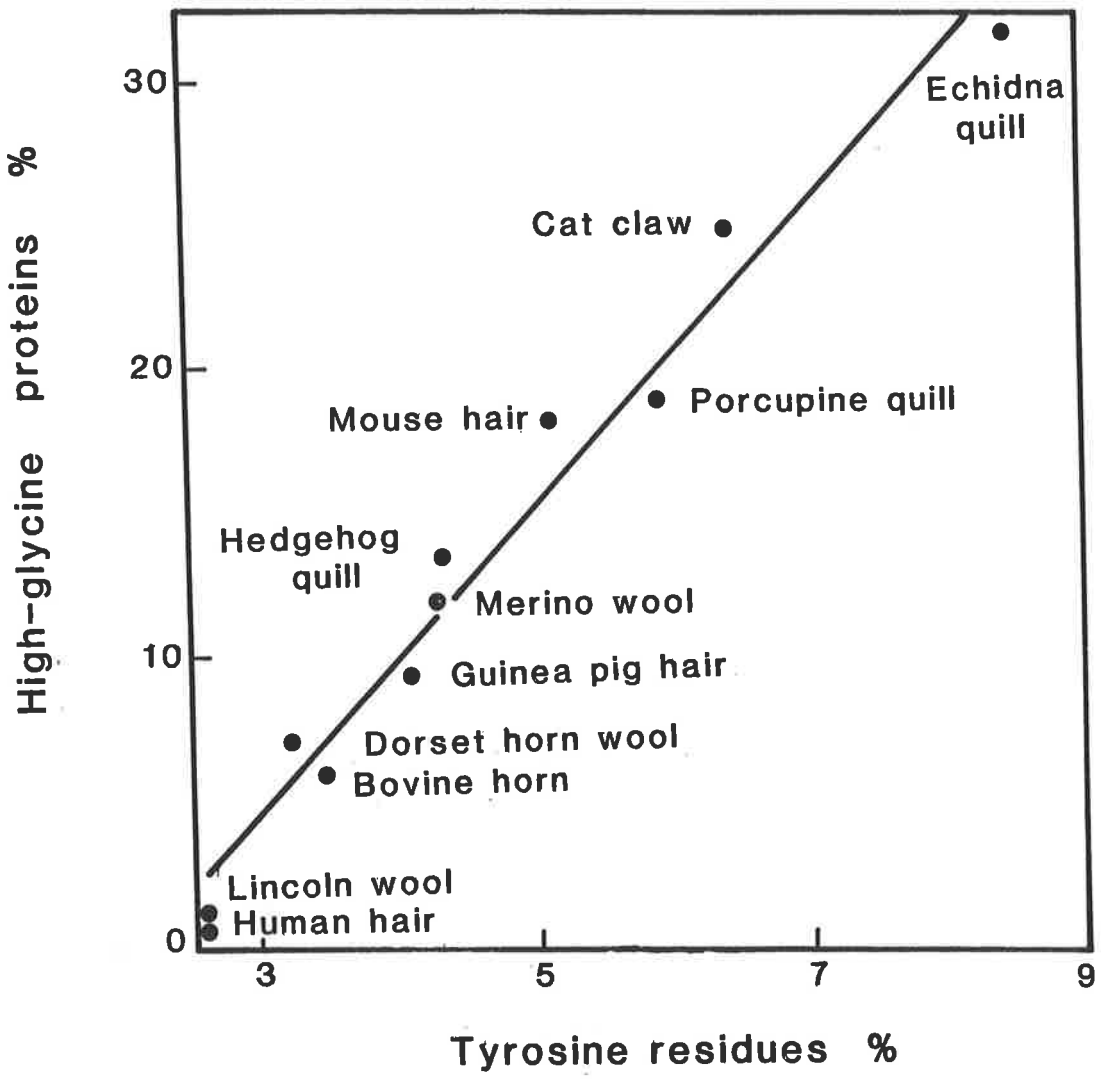
2. Nutritional control

A great deal of research has been directed towards the dietary control of wool growth and the manipulation of its composition. When cysteine or methionine (sulphur-containing amino acids) were infused into the abomasum of sheep, there was a very substantial increase in the growth rate of wool, and furthermore, the cysteine content of the wool was increased by as much as 45% (Gillespie and Reis, 1966). This change in composition was due to the synthesis of the ultra-high-sulphur (UHS) proteins, an extremely heterogenous family of polypeptides which differ

FIGURE 1.2

VARIABILITY IN TYROSINE AND HIGH-GLYCINE/TYROSINE PROTEIN CONTENTS OF VARIOUS KERATINS

A graph showing the relationship between the tyrosine content of a number of keratins (hairs, unless otherwise stated) and their respective content of HGT proteins.
Adapted from Gillespie and Frenkel (1974).


High-glycine proteins %
30
20
10
0
3
5
7
9
Tyrosine residues %
Echidna quill
Cat claw
Porcupine quill
Mouse hair
Hedgehog quill
Merino wool
Guinea pig hair
Dorset horn wool
Bovine horn
Lincoln wool
Human hair

markedly from the high-sulphur keratin proteins.

The proportion of high-glycine/tyrosine proteins, can be substantially decreased by various dietary treatments applied to sheep. These treatments usually involve dietary imbalances such as sheep on a diet of wheat receiving an abomasal infusion of methionine, or when sheep on a basal diet are infused with wheat gluten, zein or amino acid mixtures lacking lysine or phenylalanine (Frenkel et al., 1974, 1975; Reis, 1979). There is often an initial increase in the level of UHS proteins which accompanies the decrease in high-glycine/tyrosine protein synthesis. It has been suggested that the increase in the UHS proteins results from greater availability of sulphur-containing amino acids as a consequence of the reduced wool growth rate caused by these treatments (Gillespie et al., 1980). Supplements of phenylalanine and tyrosine do not stimulate the synthesis of HGT proteins so that the importance of aromatic acids in their regulation is uncertain. No mechanism has been proposed for the dietary regulation of the high-glycine/tyrosine proteins and as yet no way has been found to increase their synthesis.

The ability of the sheep to vary the type of wool produced according to its nutritional status probably reflects a complicated control mechanism which optimizes the properties of the wool coat for the animal's needs according to the available resources.

3. Physiological control

Numerous variations in the composition of wool unrelated to dietary changes have been observed which suggests that there may also be physiological control over the synthetic activities of the follicle (Gillespie et al., 1980). Extreme reductions in wool growth rate can be induced by adminstering chemical defleecing agents to sheep. Regrowth wool following mimosine or cyclophosphamide-induced defleecing showed major changes in composition over a prolonged period of time, such as a

substantial decrease in high-glycine/tyrosine protein level and an initial increase in the high-sulphur protein content. The high-glycine/tyrosine protein level returned to pretreatment values after 10 weeks, whereas the high-sulphur proteins decreased and remained below pretreatment levels for at least 12 weeks (Frenkel et al., 1975; Gillespie et al., 1980).

In another study, the first wool growth or regrowth following plucking, was observed to contain a reduced content of high-glycine/tyrosine proteins, possibly reflecting a foetal-type protein profile (Gillespie et al., 1980). These workers suggested that the changes in keratin protein proportions following defleecing may be characteristic of fibres synthesized by new or regenerating follicles rather than attributed to a specific effect of the chemical agent.

4. Hormonal control

Adrenal, gonadal, thyroid and pituitary hormones all appear to effect hair and wool growth (see review, Ebling and Hale, 1983). Thyroidectomy causes a reduction of 40% in wool growth while the administration of thyroxine produces a stimulation of wool growth rate. Adrenocorticotrophic hormone or adrenal corticosteroids cause a suppression of wool growth, while hypophysectomy results in complete cessation. Wool growth has also been shown to be inhibited by the administration of mouse epidermal growth factor (Moore et al., 1982).

These control factors acting within the follicle, discussed in the sections above, are as yet poorly understood with respect to their modes of action at the molecular level.

1.3 HIGH-GLYCINE/TYROSINE (HGT) PROTEINS

1.3.A Occurrence and Variability

Proteins rich in glycine and tyrosine occur in all the mammalian keratins studied (Gillespie and Frenkel, 1974). The content of

these proteins varies from 1-30% by weight across the species (for eg., negligible levels in human hair, 20% in mouse hair and 30% in echidna quill) and is primarily responsible for the differences in tyrosine content observed between the different species (see Figure 1.2). The level of these proteins can vary within a species, eg., sheep wool (1-13%) and this variability is determined by genetic and dietary factors (see 1.2.D)

1.3.B Structure and Heterogeneity

Most of the data concerning this family of proteins has come from studies on wool and mouse hair (Gillespie and Frenkel, 1974; Marshall et al., 1980). Extraction under identical conditions has shown that the HGT protein families in these keratin structures have similar degrees of heterogeneity. However, protein sequence data is available only for the wool keratins and this in itself is very limited so that the similarities of the wool and mouse proteins have yet to be determined.

The extent of heterogeneity of the high-glycine/tyrosine wool proteins remains unresolved at present. Early work using various techniques to examine these proteins, such as QAE-cellulose chromatography (pH 10.5), cellulose-acetate electrophoresis (pH 8.9), DEAE-cellulose chromatography (pH 2.9), and phosphate-cellulose chromatography (pH 2.9) (Gillespie and Frenkel, 1976), indicated that the Type I family consisted of at least 10 subgroups of one or more polypeptide species of varying amino acid composition (see Figure 7.6), and that the Type II family of Merino wool consisted of one family of five components which were almost identical in amino acid composition. Ultracentrifugation data (Gillespie, 1972) showed that all components were of similar size and suggested that the heterogeneity was related to charge. When Type I proteins were prepared by extraction at pH 7.7, each subgroup contained only one member, which could be converted to the larger number of components by subsequent

treatment at alkaline pH (Marshall et al., 1980). In view of the harsh procedure routinely used for extraction of the high-glycine/tyrosine proteins from wool (urea-thioglycolate solutions at pH 10.5-11), the heterogeneity of these keratins may be attributable to the extraction procedure. Furthermore, Gillespie and Frenkel (1976) and Frenkel (1977) have suggested that the heterogeneity observed for the 10 QAE-cellulose subgroups following fractionation of the Type I proteins may be due to differences in the degree of amidation and similarly, there is growing evidence that at least part of the heterogeneity revealed by electrophoresis of DEAE-cellulose fractions at pH 2.9 is due to variations in the relative proportions of free and blocked N-terminal amino groups. The unusual amino acid composition of the high-glycine/tyrosine proteins may contribute to their apparent heterogeneity. The insolubility of these proteins below pH 10 is caused by their high tyrosine content, and because many of the chromatographic properties of the high-glycine/tyrosine proteins are determined by interactions between the aromatic residues and the chromatographic matrix, it is possible that the apparent heterogeneity arises from different conformational states of aromatic residues (Marshall et al., 1980). With current refined techniques for protein purification and separation it may be realized that the heterogeneity of the high-glycine/tyrosine proteins is not as extensive as first thought.

Two complete and different HGT protein sequences are known (Dopheide, 1973; Gillespie et al., 1980). Both belong to the Type I sub-family and differ significantly in size and amino acid sequence. Component C2 (84 amino acids) and component F (61 amino acids) share similarities such as (1) glycine-rich and glycine-poor sections in the molecules (2) glycine is often followed by a hydroxyl-containing residue such as tyrosine and serine giving rise to -Gly-X- repeat sequences (3) the relative infrequency of Gly-Gly sequences (two in C2 and one in F) (4) and the longest stretch of amino acids in common being the

pentapeptide, Gly-Cys-Gly-Tyr-Gly (5) when the two sequences were aligned for maximum sequence homology (see Figure 5.1) stretches of amino acids up to 3 residues in length were contained within two sections common to both proteins and separated by a non-homologous region. This latter region in the component C2 sequence corresponds to a glycine-poor region which is also devoid of the other abundant amino acids, tyrosine and serine and contains the only histidine, alanine and valine residues in the molecule. The relevance, if any, of this sequence arrangement to the function of the HGT proteins is not known.

1.3.C Location and Function

It now appears evident that the HGT proteins form part of the intermicrofibrillar matrix of the cortex (Fraser et al., 1973) since the amount of these proteins in some keratins, 13% in Merino wool, as much as 30-35% in echidna quill, is far too much to be accomodated other than in the matrix.

Semi-quantitative measurements of the different protein families in various mammalian keratins (Gillespie, 1972) have shown large differences in the proportion of HGT proteins present in these structures. It has been suggested that such variations are related to specific structural requirements for each keratin but at the present time no consistent examples are apparent to support such an hypothesis. For example, in sheep hoof, horn and wool keratins the constituent proteins are similar but their relative proportions are seen to differ and presumably are controlled at the level of gene expression in each tissue.

The fact that individual keratins (eg. wool) do not have unique compositions suggests that their actual composition is not of paramount importance but that the properties of each particular structure may be reflected in the way in which the constituent proteins are arranged and in the cross-links between them (Gillespie and Frenkel,

1974). In the case of wool, it is apparent that the matrix is stabilized not only through its disulphide (hence high-sulphur protein) content, but that the HGT proteins are also involved via non-covalent interactions (Bendit and Gillespie, 1978). Furthermore, mechanico-chemical data has shown that stress-breakage of the wool fibre usually occurs in the matrix and it was conceivable that the weakening of wool caused by chemical defleecing agents resulted from a suppression of HGT proteins synthesis (Gillespie and Frenkel, 1976). Again, this obervation is inconsistent with the fact that Lincoln wool (HGT content < 1%) is of normal strength.

Recently, comparison of the HGT protein sequences with scale keratin and epidermal (mouse and human) keratin sequences has suggested (D. Parry, personal communication) a role for HGT proteins or segments in protein chain/molecular aggregation. Some interesting points of note from this comparison were: (1) striking similarities between the carboxyl terminal region of the scale keratin sequence and the HGT proteins. This is shown in Figure 1.3A and the conformation of the HGT segment (predicted by the Chou-Fasman and Robson predictive schemes) is shown in part B of this figure. This structure can be seen to lend itself to a highly flexible molecule capable of many interactions, as might be required for the HGT sequences in the three-dimensional wool structure (2) the probability that the glycine-rich terminal regions of the epidermal keratins (Fuchs and Hanukoglu, 1983; Steinert et al., 1984) arose independently from those of the scale and HGT keratins is quite high. However there is a significant probability ($\approx 2.6 \times 10^{-2}$) that the HGT and scale kerains have common origins (3) it seems that hard α-keratins (wool, horn, hair) contain HGT proteins which were originally part of the low-sulphur proteins but were cleaved to form a separate molecule; whereas the HGT segments were maintained in the soft α-keratins such that even scale requires the same type of HGT segment.

One final point on the localization and a possible function of

FIGURE 1.3

COMPARISON OF HGT-COMPONENT C2 AND THE GLYCINE-RICH SEGMENT OF SCALE KERATIN AND THE PROPOSED STRUCTURE FOR THE GLYCINE-RICH REGION

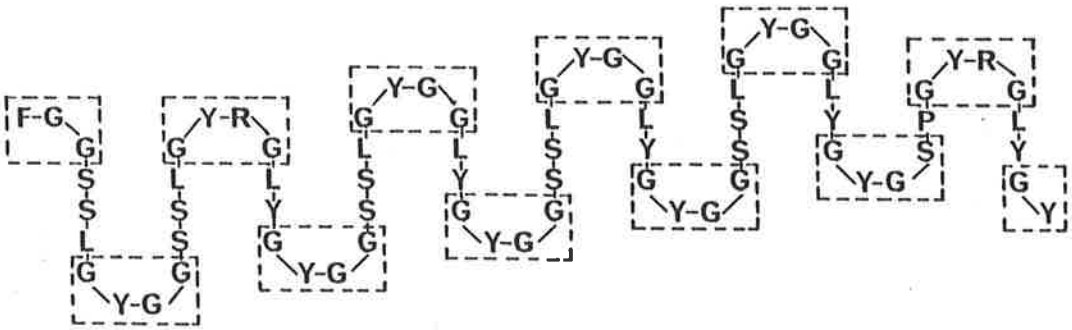
PART A shows the short-range amino acid sequence homologies that exist between a glycine-rich repeat segment of scale keratin (S. Wilton, personal communication) and the HGT protein component C2 (Marshall et al., 1980). Only the relevant regions indicated by their amino acid position are shown. The one-letter amino acid code is used.

PART B shows the proposed conformation of the glycine-rich repeat in the scale keratin sequence (D. Parry, personal communication). β-turns are formed by Gly-Tyr-Gly-Gly, with two lines of highly reactive tyrosines formed by the β-bends. Leu-Tyr and Ser-Ser-Leu form β-strands which are short and alternate in size giving rise to the structure illustrated and described as a multistranded twisted antiparallel β-pleated sheet.

A.

Scale repeat		G	Y	G	G	S	S	L	G	Y	G	G	L	Y
C2	36 – 42				G	S	P	L	G	Y	G			
C2	45 – 52			G	Y	S	S	L	G	Y	G			
C2	53 – 57		F	G	G	S	S							
C2	64 – 68	G	Y	G	G	S								

B.



the HGT proteins in wool, deals with the natural wave or crimp of the fibre. Horio and Kondo (1953), Mercer (1954) and Fraser and Rogers (1955) showed that the cortex of wool fibres is divided into two parts, the orthocortex and paracortex, with the orthocortex always on the convex side of the cortical filament. This is illustrated in Figure 1.4. Orwin (1979) and others have shown that these two regions differ in the proportion of microfibril and matrix with greater amounts of matrix in the paracortical cells. This observation supports the statement of Mercer, Golden and Jeffries (1954) that "the ortho-para differentiation appears to be related to the unsymmetrical keratinization occurring in the 'curved' follicle of the sheep skin". Recent immunohistochemical studies suggest that the HGT Type II proteins (Hewish and French, personal communication) are apparently specifically located in the orthocortex of the wool fibre, but care must be taken in interpreting these results since the highly cross-linked nature of the keratinized fibre can present difficulties for antibody penetration. Although much has been written about the orthocortex - paracortex structure it must be remarked that a clear understanding of its relation to fibre crimp has not so far been obtained, or whether the HGT proteins have a functional role in this phenomenon.

The data presented above shows the limited extent of knowledge concerning the structure, arrangement and synthesis of the HGT proteins in the wool fibre and that further studies are required to establish the function of these proteins in the keratin complex.

1.4 KERATIN GENE STUDIES

The following section deals with current investigations of gene structure and organization for the epidermal and wool keratins.

1.4.A Epidermal Keratin Genes

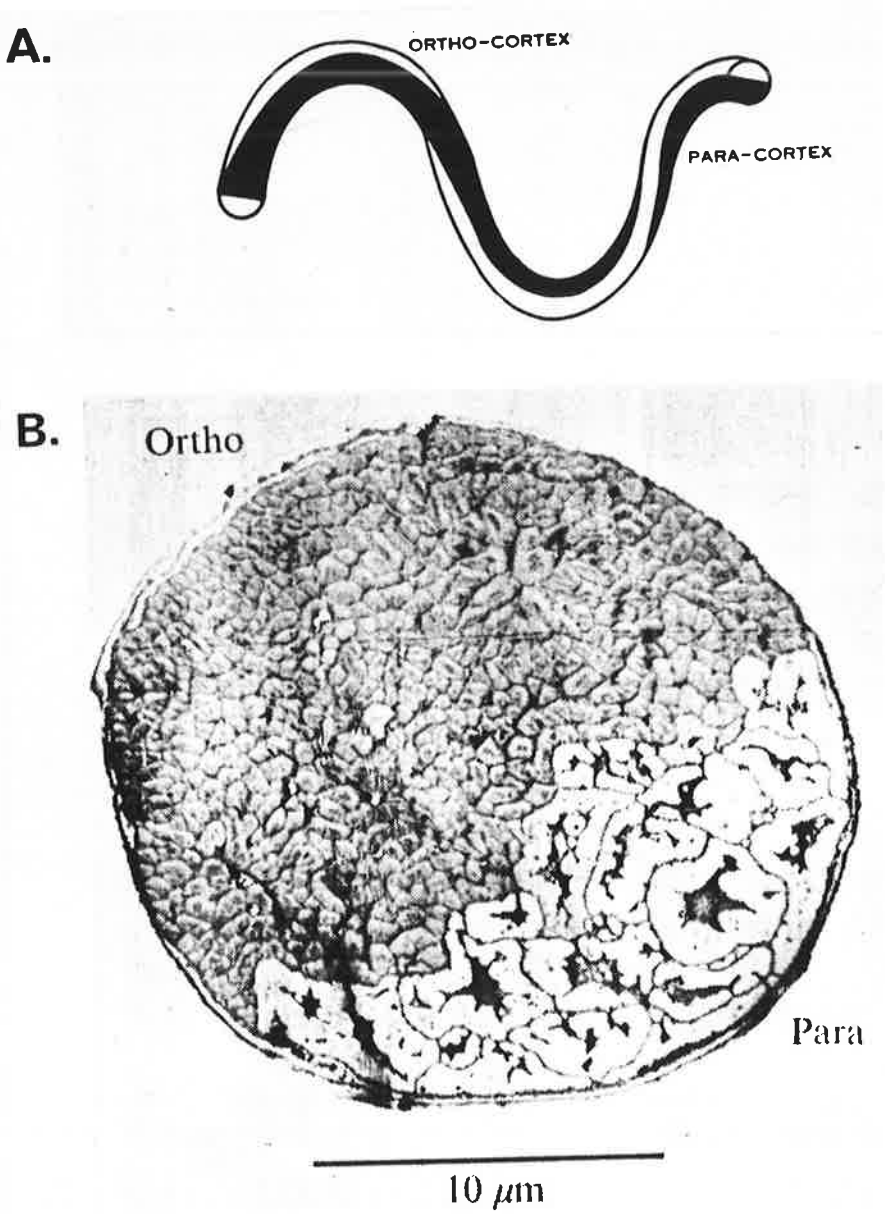
The epidermal keratins are a family of at least 10 related

FIGURE 1.4

DISPOSITION OF THE ORTHOCORTEX AND PARACORTEX IN A WOOL FIBRE

PART A shows the disposition of orthocortex and paracortex relative to the crimp of a fine wool fibre (from Mercer, 1954). The orthocortex is always found on the convex side of the cortical filament, i.e., it is on the outside face at the crests and passes from one side of the fibre to the other.

PART B shows an electron-micrograph of a cross-section of a fine (unmedullated) wool fibre, with the thin outer cuticle, and orthocortex and paracortex. The distinct textures of these two regions are due to differences in the arrangement and composition of the microfibril and matrix components (Rogers 1959; Kaplin and Whiteley, 1978).


A.
ORTHO-CORTEX
PARA-CORTEX
B.
Ortho
Para
10 μm

proteins of M_r 40,000-70,000 which form cytoplasmic filaments in most vertebrate epithelial cells (Fuchs et al., 1981 and references therein). There are two classes of genes, distinguished on the basis of their nucleic acid sequences and which do not cross-hybridize. In addition, the two classes appear to be maintained as separate unlinked sub-families throughout vertebrate evolution (Fuchs et al., 1981). The analysis of nucleic acid sequences of human epidermal cDNA clones (Hanukoglu and Fuchs, 1982; 1983) allowed direct comparisons to be made with protein sequences of other intermediate filament types. The comparisons (Steinert et al., 1983; Hanukoglu and Fuchs, 1983) suggested that : (1) intermediate filaments contain an internal core of four helical domains separated by three regions of β-turns (2) the coiled-coil α-helical regions are structurally homologous and of the same size (3) the amino and carboxyl terminal portions are non-helical and variable both in size and amino acid sequence (although the three epidermal sequences do contain at least one region extremely rich in glycine). The variability of the non-helical regions probably accounts for the size heterogeneity among epidermal keratins.

Quax et al., (1983) have reported the gene structure of the intermediate filament protein, vimentin. This gene, isolated from a hamster genomic library, is present as a single copy, comprises about 10 kb of DNA and contains more than 8 kb of intron sequences. A recent comparison of the vimentin and a human epidermal keratin gene sequence (Marchuk et al., 1984) has indicated, that although there does not appear to be a correlation between exon-intron organization and protein domains, the positions of the introns within these two genes are highly conserved. Surprisingly, these same exon-intron boundaries are maintained in a Type I low-sulphur wool keratin gene (B. Powell, personal communication), which is related to the cyto-keratin group of intermediate filaments (see 1.2.B.1).

1.4.B Wool Keratin Genes

Sequence data for the Type I and Type II low-sulphur wool proteins has indicated that these microfibrillar components are members of the intermediate filament class of structures (Weber and Geisler, 1982; Dowling et al., 1983; Crewther et al., 1983). Sheep genomic clones, have been isolated (K. Ward, personal communication) corresponding to components 7 (Type II sequence) and 8 (Type I sequence) respectively. Sequence analysis to date has shown that in the case of component 8 four introns are located in the coding region. A further three introns have been approximately positioned by R-looping studies. The most interesting feature of the intron structure is that some are close to the boundaries of domains which have been recognized as α-helical from secondary structure analysis using the Chou and Fasman rules (Crewther et al., 1983). It is not known yet whether there are additional introns in the 5' and 3' non-coding regions.

Two sheep genomic clones containing three B2 high-sulphur keratin genes have been isolated (Powell et al., 1983). Two genes in one clone, encode the B2A and B2D proteins and are closely linked in the genome, separated by 1.9 kb and are transcribed in the same direction. The single gene in the other clone codes for the B2C member of this high-sulphur sub-family. Although there is extensive sequence conservation in the 5' non-coding regions, the 3' non-coding regions diverge both in length and sequence. Within the 5' non-coding region adjacent to the initiation codon of the three B2 high-sulphur keratin genes, there is a highly conserved 18 bp sequence which is also present in another gene (data obtained from a cDNA clone sequence, K. Ward and M. Sleigh, unpublished data) coding for a member of a different, unrelated high-sulphur keratin sub-family, the BIIIB proteins. Recently, it has been established that the genomic clone containing the B2C gene, also contains a BIIIA gene positioned approximately 5 kb downstream from the

B2C gene and transcribed in the opposite direction (B. Powell, personal communication). Two other genomic clones have been isolated (M. Frenkel, personal communication) which contain two different members of the BIIIB high-sulphur keratin family. However, a gene within one of these genomic clones is non-functional and represents a pseudogene (M. Frenkel, personal communication). One of the interesting features of the high-sulphur keratin genes is the absence of introns both in the coding and non-coding regions.

1.5 AIMS OF THIS THESIS

Genes for the low-sulphur wool proteins (a subclass of keratin intermediate filament proteins, Crewther *et al.*, 1983) and the high-sulphur wool proteins (matrix components, Powell *et al.*, 1983) have been isolated and are currently under examination. The main objective of the work described in this thesis was to extend the knowledge of wool keratin gene structure and organization, by initiating an analysis of the high-glycine/tyrosine (HGT) keratin genes. The selection of pure HGT cDNA species, derived from mRNAs expressed in the wool follicle, was adopted as the first step in the characterization of HGT genes from the sheep genome. Knowledge of the gene structures would have several ramifications including: (1) confirmation of the limited HGT protein sequence data and substantiation, or otherwise, of the observed complexity of these proteins in the wool fibre, and (2) identification of possible structural relationships between the various keratin gene families. Hopefully regions essential for selective and co-ordinate expression of these genes in the wool follicle cells would be suggested by the presence of conserved nucleotide sequences and/or common features in the overall organization of the keratin gene families.

CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2

2.1 MATERIALS

2.1.A TISSUE

In the experiments reported in this thesis, RNA, DNA and wool samples were collected from the following domestic sheep Ovis aries : Merino, Merino-Dorset Horn X Border Leicester, Lincoln and Felting lustre Merino. The sheep were maintained indoors in individual pens, being fed 800g of lucerne chaff per day and water ad libitum.

2.1.B BACTERIAL STRAINS

The E. coli K12 strains used in this study were:

MC1061 : ara D139, Δ(ara, leu)7697, ΔlacX74, gal U^-, gal K^-, hsr^-, hsm^+, str A (Casadaban and Cohen, 1980). Host for pBR322 transformation.

JM101 : lac, pro, sup E, thi, F'traD36, pro AB, lac I^q, ZΔM15 (Messing, 1979). Host for M13 transfection.

K802 : hsd R^+, hsd M^+, gal^-, met^-, sup E (Wood, 1966). Host for λ Charon 28 propagation.

2.1.C BACTERIOPHAGE STRAINS

M13mp83, M13mp93 (Messing and Vieira, 1982), M13mp19 (Norrander et al., 1983). Vectors used for cloning restriction fragments prior to sequencing.

λCharon 28 : b1007, KH54, NIN5 (Rimm et al., 1980). Vector used for cloning sheep genomic DNA.

2.1.D ENZYMES

All restriction endonucleases were purchased from New England Biolabs.

Calf-thymus terminal deoxynucleotidyl transferase, T_4 polynucleotide kinase, S_1 nuclease were from Boehringer-Mannheim.

E. coli DNA polymerase I (Klenow fragment) and T_4 DNA ligase were initially purchased from Boehringer-Mannheim but later from Biotechnology Research Enterprise of South Australia (BRESA).

E. coli deoxyribonuclease I (DNase I), Lysozyme and Ribonuclease A (RNase A) were from Sigma Chemical Co.

AMV reverse transcriptase (avian myeloblastosis virus RNA-dependant DNA polymerase) was initially a gift of Dr. J.R.E. Wells but later purchased from Molecular Genetic Resources Inc.

2.1.E RADIOCHEMICALS

[α-^{32}P]dATP, [α-^{32}P] dCTP, [α-^{32}P] dGTP (specific activity, >1800 Ci/mmol) and [γ-^{32}P]ATP (specific activity, > 2000 Ci/mmol) were initially gifts of Dr. R.H. Symons but later obtained from BRESA.

The remaining radiochemicals listed below, were purchased from the Radiochemical Centre, Amersham.

[5-^{3}H]dCTP (specific activity, 21 Ci/mmol), [8-^{3}H]dGTP (specific activity, 9.5 Ci/mmol).

iodo-[2-^{14}C] acetic acid (specific activity 54 mCi/mmol).

L-[4,5-^{3}H]Leucine (specific activity, 71-130 Ci/mmol).

L-[3-^{3}H]Serine (specific activity, 18-28 Ci/mmol).

L-[2,3,4,5-^{3}H]Tyrosine (specific activity, 76-94 Ci/mmol).

2.1.F SYNTHETIC DNA OLIGONUCLEOTIDES

The various synthetic oligonucleotides (14-mer, 17-mer and 25-mer) used as probes in Chapters 5 and 7, were obtained from BRESA. Synthesis was performed by Dr. D. Skingle and Mr. S. Rogers using the solid-phase phosphite method (Beaucage and Caruthers, 1981) with morphilinoamidites and then purified by reverse-phase high performance liquid chromatography.

2.1.G CHEMICALS

The following chemicals were purchased from the Sigma Chemical Co., : Acrylamide and bis-acrylamide (N,N'-methylene-bis-acrylamide), ampicillin, BCIG (5-bromo-4-chloro-3-indolyl-3-D-β-galactopyranoside), deoxynucleoside triphosphates (dNTPs), ethidium bromide, EDTA (ethylenediaminetetracetic acid), guanidine hydrochloride (Grade 1), iodoacetic acid, IPTG (isopropylthiogalactoside), salmon sperm DNA, SDS (sodium dodecyl sulphate), Trizma base and E. coli tRNA.

Agarose (Type 1), low-melting point agarose, dideoxynucleoside triphosphates (ddNTPs), oligo(dT)-cellulose and ultrapure sucrose were from BRL.

Polyethylene glycol (PEG) and phenol were from BDH laboratories.

Sequencing primer (17mer : 5'-GTAAAACGACGGCCAGT-3') was purchased initially from New England Biolabs but later from BRESA.

Tetracycline was a gift from Upjoin Pty., Ltd.

TEMED (N,N,N,N'-tetramethylethylenediamine) was obtained from Eastern Kodak Co.

Vertex denture material, liquid and powder : Dentimex Zeist, Holland.

Other chemicals were routinely obtained from Sigma Chemical Co., BDH Chemicals Ltd., Ajax Chemicals Ltd., and May and Baker Ltd., and were either analytical grade or the highest available purity.

2.1.H MEDIA AND BUFFERS

Double-distilled water was used throughout all preparations and solutions and media all sterilized by autoclaving.

1. Growth media for E. coli MC1061

Luria broth (L-broth) contained : 1% Bacto-tryptone (Difco), 0.5% yeast extract (Difco), 1% NaCl and the pH was adjusted to 7.2 prior to autoclaving. Agar plates were prepared by adding 1.5% Bacto-agar (Difco) to the above media. Where appropriate, the media and plates

(after cooling to 45°C) were supplemented with the relevant antibiotics, ampicillin at 50μg/ml or tetracycline at 15μg/ml.

2. Growth media for E. coli JM101

M13 minimal media contained : 1.05% K_2HPO_4, 0.45% KH_2PO_4, 0.1% $(NH_4)_2SO_4$, 0.05% Na citrate and was supplemented after autoclaving with 0.02% $MgSO_4$, 0.0005% Thiamine-HCl and 0.2% glucose. M13 minimal plates were prepared with M13 minimal media containing 1.5% Bacto-agar.

2xYT broth contained : 1.6% Bacto-tryptone, 1% yeast extract, 0.5% NaCl and the pH was adjusted to pH 7.0. Soft YT agar overlay contained 2xYT broth supplemented with 0.7% Bacto-agar.

3. Growth media for E. coli K802

Broth used for the growth of E. coli K802 consisted of L-broth supplemented with 0.2% maltose. Z plates contained 1% amine A, 0.5% NaCl and 1.2% Bacto-agar with the pH adjusted to 7.2. The soft agarose overlay consisted of 0.7% agarose in water with the pH adjusted to 7.0.

NZCYM broth for the propagation of Charon 28 bacteriophage contained 1% NZamine A, 0.5% NaCl, 0.5% yeast extract, 0.1% casamino acids, 0.25% $MgSO_4$ with the pH adjusted to 7.5.

4. Buffers

Buffers commonly used in this study are listed below.

NET : 100mM NaCl, 1mM EDTA, 10mM Tris-HCl pH 7.5.

PSB : 100mM NaCl, 10mM $MgCl_2$, 0.05% gelatine, 10mM Tris-HCl pH 7.4.

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

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

TBE : 130mM Tris, 45mM Boric acid, 2.5mM EDTA, pH 8.8.

TE : 10mM Tris-HCl pH 7.5, 0.1mM EDTA.

2.1.I MISCELLANEOUS

Cellophane dialysis tubing : BDH Chemicals Ltd.

Glass fibre (GF/A) filter discs : Whatman Ltd.

Nitrocellulose filter paper : Sartorius; Schleicher and Schuell.

Sephacryl S-1000 : Pharmacia.

2.1.J GLASSWARE AND EQUIPMENT

All glassware and equipment where necessary was alkali washed (in 1N KOH to minimize RNase contamination), rinsed well with double-distilled sterile water and where possible sterilized by dry heat or autoclaving.

2.2 METHODS

2.2.A ETHANOL PRECIPITATION

Unless otherwise stated, all samples were made 0.3M with respect to Na acetate using a 3M stock solution at pH 5.2. About two to three volumes of redistilled ethanol were added, and after thorough mixing the samples were chilled for either 20 minutes in an ethanol/dry ice bath or left at -20°C for at least 5 hours. Precipitates were collected by centrifugation at 12,000 rpm for 15 minutes at 4°C in an Eppendorf centrifuge (for small volumes) or at 15,000 rpm for 20 minutes at 4°C in a Beckman centrifuge (for large volumes). The supernatant was removed and the pellet washed with 70% (v/v) ethanol to remove residual salt, and then dried under vacuum before resuspension in the appropriate solution.

2.2.B PHENOL EXTRACTION

Before use phenol (redistilled) was equilibrated with buffer, 0.1M Tris-HCl pH 8.0, 0.5mM EDTA. The extraction was carried out by mixing the aqueous sample with an equal volume of phenol/chloroform (1:1). After centrifugation, residual phenol was removed by extraction with

diethylether, which was evaporated prior to ethanol precipitation of the aqueous phase. For all extractions, aqueous samples were made 5-10mM with respect to EDTA, to ensure no loss of material to the organic phase.

2.2.C mRNA PREPARATION

1. Extraction of total RNA from wool follicles

Sheep were immobilized on a table and the area of the sheeps' side from which follicles were to be collected was sheared close to the skin using Oster clippers with a size 40 head and a local anaesthetic (Xylocaine, 10ml in 0.5ml aliquots) was injected subcutaneously. A 1:1 mixture of Vertex denture powder and liquid was spread in small regular patches over the sheared area and strips of fibreglass tape were embedded in the quick-setting resin to provide a strengthening backbone and a gripping edge for each patch. Once set (about 5 minutes), the strips were carefully pulled away from the skin and quickly immersed in liquid nitrogen in which they were stored until the extraction procedure was begun.

RNA was routinely prepared using an extraction procedure employing guanidine hydrochloride as detailed below. The method was essentially that of Brooker et al., (1980), but with some modifications.

The frozen follicles were initially scraped from their backing strips into a small volume of liquid nitrogen and then transferred to a sterile tube containing a freshly prepared solution of 6M guanidine hydrochloride, 0.2M Na acetate pH 5.5, and 0.1M 2-mercaptoethanol (10ml per gram of follicle material). Using an Ultraturrax homogenizer (small probe) at high speed, the mixture was homogenized at 0°C then ethanol precipitated with two volumes of ethanol. After standing at -20°C for at least 4 hours, the precipitate was collected by centrifugation at 15,000 rpm for 20 minutes at 4°C and resuspended in 10ml of freshly-made 7M urea, 0.1M Tris-HCl pH 8.5, 5mM EDTA and 0.1% SDS. This aqueous solution was

extracted twice with phenol/chloroform (1:1), once with ether and ethanol precipitated.

To remove high molecular weight DNA from the preparation, a selective high salt precipitation was performed according to a procedure reported by Diaz-Ruis and Kaper (1978). An equal volume of 4M LiCl was added to the nucleic acid solution (resuspended in water) and the mixture left at 4°C overnight. The high molecular weight RNA (greater than 4S) becomes insoluble and can be recovered by centrifugation (14,000 rpm for 15 minutes at 4°C). The pellet was then washed with 70% ethanol, before a final ethanol precipitation to remove excess salt.

2. Isolation of poly(A)$^+$ RNA

Total RNA extracted from wool follicles was further purified by two rounds of chromatography on oligo(dT)-cellulose as described below.

The RNA, resuspended in 10mM Tris-HCl pH 7.5, 1mm EDTA, 0.1% SDS, was dissociated by heating at 65°C for 3 minutes and then immediately chilled on ice to prevent reaggregation of RNA chains. The sample was adjusted to 0.5M KCl and layered onto a 1.5ml (total capacity, 64 OD units/gram) oligo(dT)-cellulose column equilibrated with binding buffer (0.5M KCl, 10mM Tris-HCl pH 7.5, 1mM EDTA). The column was washed with a further 7ml of binding buffer and the combined initial eluate was passed once more through the column. The elution of bound material was carried out with elution buffer (10mM Tris-HCl pH 7.5, 1mM EDTA). Occasionally this low salt eluate was readjusted to 0.5M KCl, redissociated and passaged through the column once more. All buffers were kept on ice during the above procedure. The poly(A)$^+$ and poly(A)$^-$ fractions (as judged by the absorbance at 260nm) were pooled and precipitated with ethanol. RNA samples were resuspended in water and stored until use at -80°C.

3. Sucrose gradient centrifugation of RNA

Linear density gradients of 10-40% (w/v) and 5-20% (w/v) sucrose in NET buffer were prepared using a gradient former. The RNA, resuspended in 0.1M Tris-HCl pH 7.5, 0.1% SDS, 1mM EDTA was heated at 65°C for 3 minutes and snap-chilled on ice, before samples were layered onto the 11.5ml gradients. After centrifugation in a Beckman SW41 rotor at 37,000 rpm for 16 hours at 4°C, appropriate fractions, as judged by the A_{260} profiles, were collected manually by upward displacement using an ISCO model 640 flow cell. Where required RNA fractions were pooled first and then ethanol precipitated twice to ensure complete removal of all traces of sucrose.

2.2.D CELL-FREE TRANSLATION OF RNA

The cell-free wheat germ extract used in translation experiments was a gift of Dr. J. Brooker and was prepared as described by Roman et al., (1976), except that Tris-acetate buffer was replaced by 20mM Hepes-KOH pH 7.6.

The complete system contained in a final volume of 25µl:5µl of wheat germ (S23) extract, 20mM Hepes pH 7.6 (adjusted with KOH), 2mM DTT, 1mM ATP (neutralized with KOH), 20µM GTP, 8mM creatine phosphate, 40µg/ml creatine phosphokinase, 2mM Mg acetate, 60mM KCl, 5-10µCi of ^{3}H-labelled amino acid (see below) and 25µM of all other unlabelled amino acids. The amount of the various ^{3}H-labelled amino acids used in these assays were as follows :

5µCi L-[4,5-^{3}H] Leucine (71-130 Ci/mmol)

5µCi L-[3-^{3}H] Serine (18-28 Ci/mmol)

10µCi L-[2,3,4,5-^{3}H] Tyrosine (76-94 Ci/mmol)

Both total (5 to 10µg) and poly(A)$^{+}$ (0.5 to 2µg) RNA samples were incubated for 1 hour at 26-27°C. Rabbit globin mRNA (0.25µg) obtained from Amersham was translated under the same reaction conditions, as an

indication of the endogenous activity of the S23 wheat germ extract.

The incorporation of ^{3}H-labelled amino acid(s) into protein during cell-free translation was determined by estimating the total incorporation by TCA precipitation. Duplicate samples (1.5µl) were spotted onto Whatman GF/A filters and air-dried. A batch-wash procedure was used as described by Bollum (1968), in which the filters were initially placed in ice-cold 10% TCA (to remove aminoacyl-tRNAs), then briefly rinsed with 10% TCA, ethanol and ether. After drying, toluene scintillant (0.35% PPO, 0.035% POPOP in toluene) was added to the filters and the incorporation of radioactivity determined in a Packard liquid scintillation spectrometer.

The cell-free translation reactions were terminated by the addition of gel loading buffer (see E.3 below) or by immediately carrying out the S-carboxymethylation modification as described below.

2.2.E POLYACRYLAMIDE GEL ANALYSIS OF WOOL PROTEINS AND IN VITRO TRANSLATION PRODUCTS

1. Extraction and ^{14}C-labelling of wool proteins

Wool, collected from the various breeds of sheep used in this work, was washed sequentially in petroleum (BP 40°C-80°C), filtered ethanol, water and finally rinsed with diethylether and allowed to dry.

Labelling of wool proteins by S-carboxymethylation was by the procedure reported by Marshall and Gillespie (1982). Wool (2-3mg) was extracted overnight at room temperature under nitrogen in a reducing solution (300µl) containing 0.05M Tris, 0.05M DTT, 8M urea, pH 9.3. A 50µl aliquot (free of unsolubilized material) was radiolabelled by S-carboxymethylation of a proportion of the cysteine residues with 5µl of iodo-[2-^{14}C]acetic acid (54 mCi/mmol) and 0.5µl of 3M Tris (added to maintain the pH at about 8). After thorough mixing and allowing the

mixture to stand at room temperature for 5 minutes, excess unlabelled 1M iodoacetic acid (25µl) containing 2.3M Tris was added to block the remaining sulphydryl groups. After a further 5 minute incubation at room temperature, the alkylation was terminated by the addition of 5µl of 14M 2-mercaptoethanol. Bromophenol blue was then added as a tracker dye and this mixture could be either loaded directly onto a polyacrylamide gel (see below) or stored frozen at -20°C until required.

2. S-carboxymethylation of cell-free translation products

To allow a comparison of in vitro synthesized translation products with ^{14}C-labelled wool keratins on polyacrylamide gels, the translation products were reduced and alkylated in the following manner.

The reaction sample (5µl) was mixed with 50µl of reducing solution (as above in E.1) and incubated at 37°C for one hour. After adding 30µl of 3M Tris-HCl pH 8.0, 10µl of 30% iodoacetic acid, the mixture was left at room temperature for 20 minutes before terminating the reaction with 2µl of 2-mercaptoethanol. The volume of the S-carboxymethylated translation product sample was concentrated by acetone precipitation (9 volumes) at -20°C overnight. The pellets were washed with ether, dried, then resuspended in 25µl of gel loading buffer.

3. SDS-urea polyacrylamide gel electrophoresis

The various protein samples were analyzed in an SDS-polyacrylamide gel system based on that of Laemmli (1970), with a number of modifications. The slab gel (14cmx14cmx1mm) consisted of a 15% (bis : acrylamide ratio of 1:100) separating gel in 0.375M Tris-HCl pH 8.8, 0.1% SDS with the addition of urea to 4M. Protein samples were dissociated by heating at 100°C for 3 minutes in loading buffer (0.0625M Tris-HCl pH 6.8, 1% SDS, 5% 2-mercaptoethanol, 4M urea, 0.001% bromophenol blue) prior to loading onto gels. Electrophoresis was carried out in electrode buffer (0.05M Tris-HCl pH 8.8, 0.375M glycine, 0.1% SDS) at 15mA

until the tracker dye had migrated to the bottom of the gel (usually about 6 hours). The gel was either fixed and simultaneously stained followed by destaining according to Swank and Munkres (1971) or processed immediately for fluorography.

4. Fluorography

A fluorographic procedure as described by Laskey and Mills (1975) was employed. However, the initial DMSO wash and subsequent DMSO/PPO/naphthalene treatment were carried out at 50°C since this elevated temperature (as compared to room temperature) enabled a two-thirds reduction in the processing time. Gels were dried down onto 3MM Whatman paper using a slab gel dryer (Hoefer Scientific Instruments) and set up for autoradiography as described below.

2.2.F AUTORADIOGRAPHY

Gels and nitrocellulose filters were autoradiographed using Fuji X-ray film at room temperature for up to 24 hours or at -80°C, with the addition of a Cronex Lightning Plus intensifying screen (Dupont), for periods longer than 24 hours.

2.2.G CONSTRUCTION OF A WOOL FOLLICLE cDNA LIBRARY

1. Synthesis of first strand cDNA

Total wool follicle poly(A)$^+$ RNA (as prepared above in C.2) was used as the template for synthesis of the first strand cDNA. Oligo(dT)-primed reverse transcription of poly(A)$^+$ RNA (4 g) was carried out in a 25µl reaction mix containing 100mM Tris-HCl pH 8.3, 50mM KCl, 6mM $MgCl_2$, 30mM 2-mercaptoethanol, 25µg/ml oligo(dT)$_{10-12}$, 1mM each of dATP, dTTP, dGTP, 0.25mM dCTP, 6 M [α-^{32}P] dCTP (480 Ci/mmol) and 950 units/ml AMV reverse transcriptase. The reaction was incubated for 90 minutes at 42°C before a 1µl aliquot was removed and assayed for cDNA synthesis, i.e.,

to determine the number of ^{32}P-nucleotides of known specific activity incorporated. The aliquot was chromatographed through a Sephadex G-50 medium column (14cmx0.4cm) in TE buffer and the amount of $[\alpha\text{-}^{32}P]$ dCTP added to the reaction was such that the cDNA synthesis could be monitored, but that the proportion of labelled nucleotides incorporated which would introduce single-stranded breaks upon decay was minimal.

2. Synthesis of second strand cDNA

In order to minimize any loss of cDNA and for ease of manipulation, the second strand synthesis was performed in the same reaction vessel as the first strand, essentially as described by Wickens et al., (1978).

The first strand reaction was stopped and mRNA-cDNA hybrids simultaneously denatured by heating at 100°C for 3 minutes, then quickly chilled on ice to faciliate self-priming of the first strand. Second strand synthesis was initiated by adjusting to final concentrations of 100mM Tris-HCl pH 8.3, 6mM $MgCl_2$, 30mM 2-mercaptoethanol, 50mM KCl, 1mM each of dATP, dCTP, dGTP, dTTP, 1500 units/ml AMV reverse transcriptase in a volume of 50µl. The reaction was kept for 6 hours at 37°C, with fresh enzyme added after 3 hours incubation. Double-stranded cDNA (ds cDNA) was purified by phenol/chloroform (1:1) extraction, ether washing, Sephadex G-50 chromatography (to remove unincorporated nucleotides) and finally ethanol precipitated.

3. S_1 nuclease digestion

To generate ds cDNA as a suitable substrate for terminal transferase, the ds cDNA was incubated (total volume 200µl) at 37°C for 30 minutes in a buffer containing 30mM Na acetate pH 4.6, 0.3M NaCl, 4.5mM $ZnSO_4$ and 500 units/ml of S_1 nuclease.

4. Size-selection of ds cDNA

The S_1 nuclease-treated ds cDNA sample was neutralized by adjusting to 0.1M Tris-HCl pH 8.0, 5mM EDTA before phenol/chloroform (1:1) extraction, ether washing and ethanol precipitation. The ds cDNA to be used in subsequent cloning into E. coli was fractionated through a Sephacryl S-1000 column (30cmx1cm) equilibrated with TE buffer and the elution profile compared with that of known molecular weight markers (Sau 96-digested pBR322 DNA) chromatographed under the same conditions. Ds cDNA material greater than approximately 200 base pairs in length was pooled and ethanol precipitated.

5. Homopolymeric tailing

Homopolymeric oligo(dC) tails were added to the ds cDNA using calf-thymus terminal deoxynucleotidyl transferase. 500 pmoles of [5-^{3}H] dCTP (21 Ci/mmol) was dried down and resuspended in a solution of 25µl containing 140mM Na cacodylate, 30mM Tris-HCl pH 7.6, 0.2mM DTT, 12.5µg/ml BSA, 2mM $CoCl_2$, 0.5 pmoles of ds cDNA and 8.5 units of terminal transferase. The reaction was incubated at 37°C and monitored by the conversion of ^{3}H-dCTP into a TCA-insoluble precipitate. When an average of 20 nucleotides per end were added, the reaction was stopped by the addition of EDTA to 5mM, followed by phenol/chloroform (1:1) extraction, ether washing and ethanol precipitation.

Vector pBR322 DNA was linearized by cleavage with PstI restriction endonuclease and tailed with [8-^{3}H] dGTP (9.5 Ci/mmol) using the same procedure as used for the ds cDNA material.

6. Annealing of tailed cDNA and vector DNA and transformation of E. coli

Equimolar amounts of dC-tailed ds cDNA and dG-tailed pBR322 vector DNA were annealed in a small volume (20µl) containing 0.2M NaCl by heating the mixture at 65°C for 10 minutes, followed by a one hour

incubation at 45°C, then slow cooling to 4°C at which the mix was stored.

E. coli strain MC1061 cells were made competent and transformed using a modification of the method described by Dagert and Ehrlich (1979). E. coli cells were grown overnight at 37°C in L-broth and then diluted 1 in 50 into fresh L-broth and grown at 37°C to an A_{600} of 0.6 - 0.7. The cells were chilled on ice for 30 minutes, harvested by low speed centrifugation at 4°C and resuspended in a 1/2 volume of ice-cold 0.1M $MgCl_2$. The cells were pelleted immediately, resuspended in a 1/20 volume of ice-cold 0.1M $CaCl_2$, and kept on ice for at least two hours before use.

0.2ml of E. coli competent cells were added to 0.1ml of the annealed DNA in 10mM Tris-HCl pH 7.6 and stirred occasionally at 0°C for 30 minutes. After heat shock at 42°C for 2 minutes, the cells were kept on ice for a further 30 minutes, then warmed to room temperature. 0.5ml of L-broth was added to the pBR322 transformed cells followed by an incubation at room temperature for 20 minutes. The transformed cells were mixed with 3ml of 0.7% L-agar and plated onto L + tet plates and incubated for 24 hours at 37°C.

7. Detection and storage of recombinant colonies

Detection of pBR322 plasmids carrying DNA complementary to wool follicle RNA was carried out using a modification of the procedure of Grunstein and Hogness (1975). Colonies found to be tet^r (tetracyline-resistant) and amp^s(ampicillin-sensitive) were transferred by sterile toothpicks to a nitrocellulose sheet that had been boiled three times in distilled water and overlaid onto a L + tet agar plate. The plates were incubated overnight at 37°C. DNA from the colonies grown on the nitrocellulose filter was immobilized onto this support by transferring the filter sequentially onto 3MM paper saturated with 0.5M NaOH for 7 minutes, 1M Tris-HCl pH 7.5 for 2 minutes, 1M Tris-HCl pH 7.5 for 2 minutes and 1.5M NaCl, 0.5M Tris-HCl pH 7.4 for 4 minutes. The

nitrocellulose filter was washed in ethanol, air dried, then baked at 80°C for 2 hours under vacuum.

The recombinant colonies detected by ^{32}P-labelled wool follicle cDNA (see Chapter 3) and to be used in all subsequent screenings were selected and stored as described below.

For rapid replication of ordered arrays of colonies, for the various screenings discussed in this thesis, the recombinant clones were maintained as glycerol stocks (overnight cultures mixed 1:1 with sterile 80% glycerol) in the wells of microtitre trays (0.6cm diameter, Linbro tissue culture plates). Each set of colonies could be replicated onto nitrocellulose filters by using a metal replicator constructed in such a way that each tooth corresponded to one well. With this equipment, a given set of clones, ordered in a defined way and therefore identified by two coordinates (see Chapter 5, Figure 5.4) could be replicated whenever and as many times as needed. The trays were kept at -20°C for periods of up to 4 months or at -80°C for longer storage.

2.2.H PREPARATION OF ^{32}P-LABELLED RNA AND DNA

1. Oligo(dT)-primed cDNA

Reaction mixtures of 50µl contained 1-2µg of wool follicle poly(A)$^+$ RNA, 50mM Tris-HCl pH 8.3, 20mM KCl, 8mM $MgCl_2$, 5mM DTT, 0.25µg oligo(dT)$_{10-12}$, 1mM each of dATP,dGTP, dTTP, 50µCi of dried down [α-^{32}P] dCTP (>1800 Ci/mmol) and 15 units of AMV reverse transcriptase. After incubation at 42°C for 90 minutes, protein was removed by phenol/chloroform (1:1) extraction in the presence of 5mM EDTA, 0.1% SDS. Alkaline hydrolysis of the mRNA template was carried out at 37°C for at least one hour in 0.3M NaOH. After neutralization, by the addition of HCl to 0.3M and Tris-HCl pH 8.0 to 0.1M, the single-stranded cDNA was purified away from unincorporated nucleotides by chromatography over a Sephadex G-50 (medium) column.

2. Kinased rRNA

Sheep rRNA (28S and 18S, isolated from sucrose gradient fractions, see C.3 above) was hydrolysed by heating to 100°C for 10 minutes in sterile water. Following ethanol precipitation to concentrate the sample, RNA and 100µCi of dried down [γ-^{32}P] ATP (>2000 Ci/mmol) were resuspended in 50mM Tris-HCl pH 9.5, 10mM $MgCl_2$, 5mM DTT, 10 units of T_4 polynucleotide kinase and incubated at 37°C for 30 minutes. ^{32}P-labelled rRNA was separated from unincorporated [γ-^{32}P] ATP by Sephadex G-50 chromatography.

3. Nick-translated DNA

The inserts from cDNA clones used as probes were labelled with [α-^{32}P] dCTP by the nick-translation method of Rigby et al., (1977). Reactions containing 0.1-0.5µg of DNA, 50mM Tris-HCl pH 7.6, 10mM $MgSO_4$, 0.1mM DTT, 50µg/ml gelatin, 25µM each of dATP, dGTP, dTTP, 50µCi dried down [α-^{32}P] dCTP (>1800 Ci/mmol), 5 units of E. coli DNA polymerase I and 40pg DNase I were incubated at 15°C for 90 minutes. Protein and unincorporated nucleotides were removed as in H.1 above.

4. Kinased oligonucleotides

The various synthetic DNA oligonucleotides (14-mer, 17-mer and 25-mer) were supplied free of the 5' terminal phosphate group and therefore could be labelled at the 5' end by transfer of ^{32}P from [γ-^{32}P] ATP using T_4 polynucleotide kinase in a reaction volume of 10-15µl containing 0.5µg of the synthetic oligonucleotide, kinase buffer as in H.1, 50 µCi of [γ-^{32}P] ATP and 4 units of T_4 polynucleotide kinase. After incubation at 37°C for one hour, the ^{32}P-labelled oligonucleotide was purified from the unincorporated label by electrophoresis through a denaturing 20% polyacrylamide gel (see below L.2) and the single band excised and recovered by elution (section N.1).

5. 3' end-labelled DNA

Both restricted pBR322 DNA (used as a molecular weight marker) and recombinant plasmid DNA was end-labelled with Klenow (large fragment of E. coli DNA polymerase I) and [α-^{32}P] dCTP as described below, to faciliate the detection of short fragments, which would not be seen under UV light after ethidium bromide staining of small amounts of DNA.

The 3' ends of restricted DNA (<0.5µg) were labelled immediately following restriction digestion (section K) with the appropriate enzymes, in a reaction mix containing 60mM Tris-HCl pH 7.5, 8mM $MgCl_2$, 10mM DTT, 10µCi dried down [α-^{32}P] dCTP and 1 unit of Klenow. Following incubation at room temperature for 30 minutes, the DNA was ready for loading onto a gel.

6. Primer-extended cDNA

Synthetic oligonucleotides, were annealed to about 2µg of RNA in a volume of 10µl in the presence of 0.1M KCl, by heating at 65°C for 10 minutes, 42°C for 15 minutes then slow cooling to room temperature. This mixture was then diluted to 20µl by the addition of 50mM Tris-HCl pH 8.3, 10mM $MgCl_2$, 10mM DTT, 600µM each of dATP, dGTP, dTTP, 50µCi dried down [α-^{32}P] dCTP and 12 units of AMV reverse transcriptase and incubated at 37°C for 2.5 hours. cDNA extension products were analyzed following electrophoresis on a denaturing 10% polyacrylamide gel (see section L.2).

7. Oligo-labelled DNA

The inserts from cDNA clones used as probes were labelled with [α-^{32}P] dCTP by the 'oligo-labelling' method described by Feinberg and Vogelstein, (1983). 100-200ng of DNA dissolved in 10µl of H_2O was heat-denatured at 100°C for 2 minutes, then quickly chilled on ice. The DNA solution was mixed with 25µCi of dried down [α-^{32}P]dCTP and 12.5µl of a 2x buffer containing 40µM each of dATP, dGTP, dTTP, 400mM Hepes pH 6.6, 10mM $MgCl_2$, 20mM 2-mercaptoethanol, 800µg/ml gelatin and 600µg/ml of

random primer. The reaction was incubated at 40°C for 20-30 minutes following the addition of 2.5µl of Klenow (2.5 units). The reaction mix was then loaded directly onto a Sephadex G-50 column to remove unincorporated nucleotides.

2.2.I TRANSFER HYBRIDIZATIONS AND cDNA LIBRARY SCREENING CONDITIONS

A variety of ^{32}P-labelled probes (see section H) were used in screening recombinant clones (colonies, phage and purified DNA) for specific DNA sequences. Depending on the nature of the probe, two basic procedures were adopted, as detailed below.

1. Synthetic oligonucleotide probes

Prehybridization was carried out in 6xNET, 0.5% NP40, 5xDenhardt's (0.02% (w/v) each of BSA, polyvinylpyrrolidone and Ficoll) 100µg/ml salmon sperm DNA (denatured and sonicated), for at least 4 hours. Hybridizations were performed in the same solution except that Denhardt's was decreased to 1xstrength and 250µg/ml *E. coli* tRNA replaced salmon sperm DNA. ^{32}P-labelled oligonucleotides (see H.4) were present at 1-5x10^{4}cpm/ml in the hybridization mix. All washes were done in 6xSSC.

The temperature of hybridizations and washings, varied according to the length and degeneracy (see relevant Chapters 5 and 7) of each oligonucleotide probe, as follows.

	14-mer	17-mer	25-mer
Prehybridization			
T°	41°C	41°C	41°C
Time		at least 4 hours	
Hybridization			
T°	*RT°	37°C	47°C
Time	2 hour	16-20 hours	16-20 hours
Washing			
	*RT° 2x20min	37°C 2x30min	37°C 2x20min
	37°C 1x10min	45°C 1x10min	50°C 2x20min

* RT° = room temperature

2. Other ^{32}P-labelled probes

Prehybridization and hybridization conditions, for the various other ^{32}P-labelled probes used throughout this thesis, for nitrocellulose filters prepared with RNA, DNA or colonies were essentially those of Wahl et al., (1979). Prehybribization was for at least 2 hours at 41°C, while hybridizations were at 41°C overnight with ^{32}P-labelled probe present at 1-5x10^4 cpm/ml. The filters were washed with increasing conditions of stringency until an acceptable level of radioactivity was detectable on the filters. The washes were initiated in 2xSSC/0.1% SDS at room temperature (twice for 20 minutes) then continued at 1xSSC/0.1% SDS at 65°C (once for 30 minutes), and finally at 0.1xSSC/0.1% SDS at 65°C (once for 30 minutes).

2.2.J PLASMID DNA PREPARATION

1. Preparative scale procedure

Recombinant plasmid DNA was prepared by a modified procedure of Birnboim and Doly, (1980).

A single recombinant colony of E. coli containing the recombinant plasmid was used to inoculate 5ml of L-broth containing the appropriate antibiotic (50µg/ml ampicillin or 15µg/ml tetracycline) and grown overnight at 37°C with aeration. The overnight culture was diluted 100-fold into 500ml of fresh media and grown with aeration to stationary phase. The cells were pelleted by centrifugation and resuspended in 4ml of ice-cold 25mM Tris-HCl pH 8.0, 10mM EDTA, 15% (w/v) sucrose and lysozyme was added to a final concentration of 3mg/ml. After 40 minutes on ice, 8ml of 0.2M NaOH, 1% SDS (made fresh on the day of use) was added and the mixture was left on ice for 10 minutes. 5ml of 3M Na acetate pH 4.6 was mixed gently into the solution, and after a 40 minute incubation on ice, the cellular debris and chromosomal DNA was pelleted by

centrifugation at 18,000 rpm for 20 minutes at 4°C. The supernatant was decanted carefully to avoid disturbing the soft pellet and treated with 50µg of RNase A (heat denatured at 80°C for 20 minutes to inactivate any contaminating DNase) for 2 hours at 37°C to degrade contaminating RNA. After extraction twice with phenol/chloroform (1:1) and ether washing, the nucleic acid was precipitated by the addition of 2 volumes of ethanol, chilling at -20°C for at least 3 hours and centrifugation at 15,000 rpm for 25 minutes at 4°C. The pellet was resuspended in 1.6ml of water, to which 0.4ml of 4M NaCl and 2ml of 13% PEG 6000 was added. The mixture was left on ice for at least one hour and the DNA pelleted by centrifugation at 14,000 rpm for 15 minutes at 4°C, washed with ice-cold 70% ethanol, dried *in vacuo*, then resuspended in 0.5ml of water.

2. Analytical scale procedure - 'minipreps'

A rapid and efficient method for analytical preparations of plasmid DNA was used as reported by Ish-Horowicz and Burke (1981), which combines the methods of Davis *et al.*, (1979) and the alkaline lysis method (Birnboim and Doly, 1980) described above. A final yield of 2-10µg of DNA was obtained and could be used directly for restriction digestion analysis.

2.2.K RESTRICTION DIGESTION OF DNA

All restriction endonuclease digestions of DNA were carried out using the conditions for the appropriate enzymes detailed in the New England Biolabs catalogue. Generally, one unit of enzyme was added for each microgram of DNA and the reaction mix incubated for at least a 2-fold excess of time to ensure complete digestion. When preparative amounts of DNA (>10µg) were to be digested, the reaction volumes were scaled up and digestion continued overnight. Invariably, all digestions were ethanol precipitated prior to analysis by gel electrophoresis.

2.2.L. POLYACRYLAMIDE GEL ELECTROPHORESIS

All polyacrylamide solutions (acrylamide to bis-acrylamide ratio of 20:1) were deionized using Elgastat mixed bed resin (1g for 25ml of solution stirred gently for 30 minutes). The TBE pH 8.8 buffer was used routinely for all polyacrylamide gels. DNA was visualized on the gels by autoradiography or after staining with 0.1% (w/v) ethidium bromide for 10 minutes and then viewed under UV light. Preparative and sequencing gels were pre-electrophoresed for 0.5 hour prior to loading of samples.

1. Non-denaturing gels

Samples were loaded onto gels (6% or 8%) in the following 6x loading buffer : 0.1mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in water. Electrophoresis was at 25mA for small gels (14cmx14cmx1.1mm) and 20mA for large gels (20cmx40cmx0.5mm) until the tracker dye had migrated the required distance.

2. Denaturing gels

Varying percentage gels (6%, 10%, 20%) were prepared as above with the addition of 7M urea. Samples were loaded in the following buffer : formamide, 0.03% bromophenol blue, 0.03% xylene cyanol, 20mM EDTA pH 8.0. The samples were denatured by heating at 100°C for 1 to 3 minutes, briefly chilled on ice and then loaded onto the gels. Sequencing gels (20cmx40cmx0.2mm) and gels(14cmx14cmx1mm) used for isolating ^{32}P-oligonucleotides were electrophoresed at 15-20mA until the tracker dye had migrated the required distance.

2.2.M AGAROSE GEL ELECTROPHORESIS

Agarose gel mixes, 0.8% to 2% (w/v) agarose in TAE buffer were set up in a horizontal gel apparatus (of varying dimensions) for preparative and analytical purposes, while vertical slab gels (14cmx14cmx3mm) were used when transferring DNA to nitrocellulose filters.

All gels were poured once the molten agarose had cooled to 65°C.

1. Non-denaturing gels

Low-melting point agarose gels were poured once the agarose had cooled to 37°C (as recommended by the suppliers). Sample buffer was the same as that used for non-denaturing polyacrylamide gels (see L.1). Electrophoresis was usually at 30-50mA.

2. Denaturing gels

(i) alkaline agarose gels (14cmx14cmx1mm), used for the analysis of cDNA products, consisted of 1% (w/v) agarose in 30mM NaOH, 1mM EDTA as described by McDonell et al., (1976). Electrophoresis and loading buffers were the same as the gel buffer except that bromophenol blue was added as a tracker dye to the sample loading buffer.

(ii) glyoxal gels (14cmx14cmx1mm) used for electrophoresis of RNA prior to transfer to nitrocellulose consisted of 1.5% (w/v) agarose containing 10mM Na phosphate pH 6.5 in a vertical slab gel set-up. RNA samples were dissociated by heating at 50°C for 30 minutes in 1M glyoxal, 10mM Na phosphate pH 6.5. After cooling on ice, loading buffer (as in L.1) was added and electrophoresis carried out at 20-30mA for up to 3 hours with constant recirculation of electrode buffer (10mM Na phosphate pH 6.5) to maintain a constant pH and thus prevent glyoxal dissociation from the RNA. RNA was visualized by staining with 0.1% (w/v) ethidium bromide for 10 minutes and viewed under UV light.

2.2.N ELUTION OF DNA FROM GELS

1. Elution from polyacrylamide gels

Gel slices containing the required DNA fragment were excised either by visualizing the bands under UV light or by superimposing an

autoradiogram from which the position of the bands could be aligned. The DNA was eluted by soaking the gel slices overnight in 300µl of the gel elution solution (0.5M NH_4 acetate, 10mM Mg acetate, 1mM EDTA, 0.1% SDS, pH 7.6) of Maxam and Gilbert (1980), after which the supernatant was ethanol precipitated. Carrier tRNA (5µg) was added to facilitate precipitation of ^{32}P-labelled oligonucleotides (see H.4) eluted from gel slices.

2. Electroelution from agarose gels

The gel slice containing the required DNA fragment was placed floating inside a sterile piece of dialysis tubing (18/32), to which 400µl of TE buffer was added to cover the gel slice. The tubing was placed within a plastic-holder across the path of an electric current in a horizontal gel apparatus and the DNA eluted from the slice by electrophoresis at 100mA for 15 minutes. The TE buffer was removed and the tubing washed several times with a further 300µl of the same buffer, to dislodge any DNA that may have stuck to the tubing. DNA was then ethanol precipitated from the collected buffer.

3. Elution from low-melting point agarose gels

The gel slice containing the required DNA fragment was excised out of low-melting point agarose and placed into an Eppendorf tube to which was added approximately two gel volumes of 50mM Tris-HCl pH 8.0, 0.5mM EDTA. The gel slice was melted and mixed in with the added buffer by holding the tube at 65°C for 30 minutes. After cooling to 37°C, DNA was recovered by two phenol extractions, one ether wash, followed by ethanol precipitation of the aqueous phase.

2.2.0 SOUTHERN TRANSFER PROCEDURE

Restricted DNA fractionated on 1 to 2% agarose gels (see M.1) was transferred to nitrocellulose by the method of Southern (1975), as

modified by Wahl et al., (1979) using 1M NH_4 acetate, 0.02M NaOH for neutralization and transfer as recommended by Smith and Summers, (1980). Transfer of DNA was allowed to proceed overnight before baking the filters for 2 hours at 80°C in vacuo to bind the DNA prior to hybridization. Prehybridization, hybridization and washing conditions varied according to the type of ^{32}P-labelled probe used and are described in section I as well as in the legends to the relevant figures.

2.2.P NORTHERN TRANSFER PROCEDURE

RNA fractionated on 1% agarose gels (M.2.ii) was transferred to nitrocellulose using the procedure described by Thomas (1980), in which after baking at 80°C for 2 hours in vacuo, the filters were washed in boiling 20mM Tris-HCl pH 8.0 for 20 minutes to remove residual glyoxal. Prehybridization, hybridization, and washing conditions were as described in section I.2.

2.2.Q SEQUENCING PROCEDURE FOR OLIGONUCLEOTIDES

Prior to use in hybridization experiments, the synthetic oligonucleotides described in Chapters 5 and 7, were labelled as in section H.4 and the sequences confirmed by the chemical cleavage method of Maxam and Gilbert (1980) as modified by Banaszuk et al., (1983). The cleavage products were electrophoresed on a denaturing 20% polyacrylamide gels (see L.2) and detected by autoradiography.

2.2.R SUBCLONING INTO M13 PHAGE VECTORS AND SEQUENCING OF RECOMBINANT PHAGE DNA

Restriction fragments of DNA to be sequenced were subcloned into M13 phage vectors (Messing and Vieira, 1982; Norrander et al., 1983) and sequenced using the "dideoxy" chain termination method of Sanger et al., (1977) as described below.

1. Preparation of M13 vectors

The replicative form of M13mp93 was prepared by the method of Birnboim and Doly (1980) and was kindly supplied by C.P. Morris. The replicative form (RF) was digested with PvuII and HindIII to generate the desired termini. The linearized RF was fractionated on a 1% agarose gel (see M.1) to remove any intact molecules and extracted as described in section N.3 and resuspended in water to a final concentration of 20ng/µl.

M13mp93SmaI, M13mp93AccI, M13mp93 EcoRI/HindIII and M13mp19 PstI/SmaI were generous gifts of M.J. Bawden.

M13mp93 SmaI was also kindly supplied by C.P. Morris

2. Ligation conditions

The DNA fragment and appropriate M13 vector were combined in a ratio of 3:1 in a 10-15µl reaction mix containing 50mM Tris-HCl pH 7.6, 10mM $MgCl_2$, 10mM DTT, 1mM rATP. All ligations, regardless of the termini (blunt or sticky ends), were incubated with 0.5 unit of T_4 DNA ligase for 4-16 hours at 14°C.

3. M13 transfection of E. coli JM101

Competent cells of E. coli strain JM101 were prepared essentially using the method of Messing et al., (1981). Overnight cultures of JM101 were diluted 1/100 in fresh 2xYT broth (Materials H.2) and grown to an A_{600} of 0.6 - 0.7. The culture was cooled on ice for 20 minutes and the cells sedimented at 5,000 rpm for 5 minutes at 4°C. The pelleted cells were resuspended in 2ml of ice-cold freshly diluted 0.1M $CaCl_2$ and left on ice for at least one hour. 0.2ml of competent cells (prepared and used on the same day) were added to 1-5µl of the ligation mix and incubated on ice for 40 minutes. After heat shock treatment at 42°C for 2 minutes, the cells were mixed with 3ml of molten YT soft agar (Materials H.2) containing 20µl of IPTG (24mg/ml in water) and 20µl of BCIG (20mg/ml in dimethylformamide). The density of the competent cells

were such that no additional bacteria were needed as a feeder lawn. This mixture was poured onto M13 cloning plates (Materials H.2) followed by incubation at 37°C overnight.

4. Preparation of single strand template DNA for sequencing

An overnight culture of JM101 was diluted into 2xYT broth and 1ml aliquots dispensed into tubes. The cultures were infected with recombinant M13 phage (from a fresh plaque) and incubated at 37°C with aeration for 5 - 6 hours. The cells were sedimented at 12,000 rpm for 10 minutes and phage particles precipitated from the supernatant by adding 200µl of 20% PEG, 2.5M NaCl and leaving the tubes at room temperature for 15 minutes. Phage pellets were collected by centrifugation at room temperature for 10 minutes and then resuspended in 100µl of TE buffer. Phage DNA was recovered by phenol extraction of the phage suspension with 50µl of buffer-saturated phenol, followed by ether washing (5 volumes) and finally ethanol precipitation (-20°C, overnight) of the aqueous phase. The DNA to be sequenced was resuspended in 25µl of 5mM Tris-HCl pH 8.0, 0.1mM EDTA.

5. Annealing conditions

The universal primer (17-mer) was annealed to the template M13 DNA by mixing 1µl of primer (2.5ng), 1µl of 10xTM buffer (100mM Tris-HCl pH 8.0, 100mM $MgCl_2$) and 5-8µl of single strand DNA template in a final volume of 10µl. The annealing was initiated by heating at 100°C for 3 minutes followed by slow cooling to room temperature over a period of one hour.

6. Polymerization reactions : "dideoxy" chain termination sequencing procedure

The dideoxynucleotide (ddNTP) solutions and deoxynucleotide or zero mixes (dNTPs or N°) were prepared as shown in the table below and

stored separately at -20°C.

Zero mixes :

	A°	G°	C°	T°
dATP	0.02mM	0.2mM	0.2mM	0.2mM
dGTP	0.2mM	0.02mM	0.2mM	0.2mM
dTTP	0.2mM	0.2mM	0.2mM	0.02mM
Buffer	1xTE	1xTE	1xTE	1xTE

Zero mixes made up from stock solutions of 10xTE (50mM Tris-HCl pH 8.0, 1mM EDTA) and 0.5mM dNTP solutions (in water).

ddNTP solutions :

ddATP : 0.3mM in water
ddGTP : 0.3mM in water
ddcTP : 0.04mM in water
ddTTP : 0.5mM in water
These working solutions were made from 10mM ddNTP stock solutions (in water).

Prior to sequencing 1µl of the appropriate ddNTP and dNTP solutions were dispensed into eppendorf tubes and left on ice. The annealed DNA and primer were mixed with 1µl of 10mM DTT and 3µM (about 16µCi) of dried down [α-^{32}P] dCTP (>1800 Ci/mmol). Sequencing reactions were commenced by the addition of 1µl of Klenow (1 unit/µl) and after mixing, dispensing 2.6µl into each of the four reaction tubes. After 15 minutes at 37°C, 1µl of dCTP chase solution (0.5mM dCTP in 5mM Tris-HCl pH 8.0, 0.1mM EDTA) was added to each reaction tube and incubation continued for a further 10 minutes, before terminating the reactions with the addition of 3µl of formamide loading buffer (see L.2).

7. DNA sequencing gels

Just prior to loading onto denaturing 6% polyacrylamide gels (section L.2), the reaction mixtures (plus loading buffer) were heated at 100 for 3 minutes then chilled on ice. 0.5µl of each sample was loaded onto the gels which were electrophoresed at 18-20mA (for about 1.5 hours)

until the tracker dye had migrated the required distance.

Following electrophoresis the gel, supported by a glass plate, was immediately washed with 12% acetic acid. The acetic acid wash was continued for about 5 minutes until all the urea had been removed. The gel was then baked in an 100°C oven for 45 minutes, followed by autoradiography overnight at room temperature.

2.2.S GENOMIC DNA PREPARATION PROCEDURES

Genomic DNA for use in Southern blot experiments was initially obtained from the high-salt precipitation step employed in the follicle RNA extraction procedure (see section 2.2.C.1). Later DNA was extracted from liver (Merino and Lincoln sheep, mouse) and blood (Felting lustre-mutant Merino) by a procedure essentially that of J. Haley (personal communication) as follows.

The tissue was homogenized in liquid nitrogen then slowly placed into a 50% redistilled phenol/20mM Tris pH 7.8, 0.5% SDS, 1mM EDTA, 1M NaCl solution while stirring. After a centrifugation for 10 minutes at 4000rpm, the aqueous phase was extracted once more with phenol followed by several ether extractions. RNase A (50µg/ml) was added and the solution transferred to a dialysis bag and dialysis continued at the same time as the RNase A digestion at 37°C for 16 hours in a buffer of 10mM Tris pH 7.8, 10mM NaCl, 1mM EDTA. The dialysed solution was then made to 10mM EDTA, 0.5% SDS and 100µg/ml of Proteinase K was added. The digestion was carried out for 3-16 hours at 37°C with gentle swirling after which the mixture was made to 1M NaCl and extracted with phenol as above. The aqueous phase was then either washed with ether and ethanol precipitated or dialyzed against the same buffer as described above.

Human genomic DNA was a gift of B. Powell.

Lincoln and Felting lustre Merino DNA were gifts of G. Cam and J. Forrest.

2.2.T SCREENING OF THE GENOMIC LIBRARY

The sheep genomic library used in Chapter 6 was prepared and kindly donated by Drs. P. Roche, P. Aldred and R. Crawford from the Howard Florey Institute of Experimental Physiology and Medicine, Melbourne. The library was constructed by partial Sau3A digestion of Merino liver DNA, followed by the ligation of 12-20 kb fragments to BamHI- digested λCharon 28 (Rimm et al., 1980) arms. The library consisted of a once amplified stock of 0.75×10^6 unique recombinants.

First round screening was carried out on large (150mm) Z-agar plates (Materials H.3) containing approximately $3\text{-}5 \times 10^4$ plaque forming units (pfu) on a lawn of indicator bacteria E. coli K802 (Wood, 1966). Duplicate nitrocellulose filters were prepared from each plate, essentially as described by Benton and Davis (1977). The phage were adsorbed to the filters (1 minute for the first filter, 2 minutes for the second), denatured by soaking in 0.5M NaOH, 1.5M NaCl for 1 minute, neutralized with two 1 minute washes in 0.5M Tris-HCl pH 7.5, 1.5M NaCl, air dried, then baked at 80°C for 2 hours under vacuum.

The hybridization of radioactive cloned probe to these filters and detection of complementary nucleotide sequences (see Chapter 6) was carried out under conditions as described in I.2 above. Plaques which bound the ^{32}P-labelled keratin probe were picked into 1ml of PSB buffer and purified through two further rounds of screening at low plaque density.

2.2.U PREPARATION OF λ RECOMBINANT PHAGE DNA

The liquid culture method of Kao et al., (1982) was used, with some modifications, to prepare DNA from λ recombinant phage.

Approximately 6×10^4 pfu (plaque forming units) were added to 0.25ml of PSB, 0.25ml of 10mM $MgCl_2$, 10mM $CaCl_2$ and 0.25ml of host indicator bacteria strain E. coli K802 (overnight culture grown in L-broth

plus 0.2% maltose). This mix was incubated at 37°C for 15 minutes then used to inoculate 50ml of pre-warmed NZCYM broth (Materials H.3) and the incubation was continued with aeration at 37°C. Following lysis of the cultures (6 to 7 hours), chloroform (2 to 3 drops) was added and after an incubation at 37°C for 10 minutes, the cellular debris was removed by centrifugation at 10,000 rpm for 10 minutes at 4°C. 50µg of DNase I and RNase A was added to the supernatant, which was kept on ice for one hour, before collecting the phage by centrifugation at 19,000 rpm for 3 hours at 4°C. The phage pellets were resuspended in 0.4ml of 0.3M NaCl, 0.1M Tris-HCl pH 8.0 and allowed to sit overnight at 4°C to allow complete resuspension of the pellet. Phage DNA was released by gentle extraction of the phage suspension with phenol (buffer saturated) until no interface was visible, once with chloroform and twice with diethylether before ethanol precipitation at -20°C. The DNA was resuspended in 10mM Tris-HCl pH 7.6, 0.1mM EDTA and stored at 4°C. This procedure yielded 100-150µg from a 50ml preparation.

2.2.V SUBCLONING OF PHAGE DNA INTO PLASMID VECTOR, pBR322

A HindIII DNA fragment isolated from a recombinant λCharon 28 phage (see Chapter 6) was to be analyzed in great detail and was therefore subcloned into the plasmid pBR322, so that large quantities of this DNA could be harvested.

The ligation conditions were similar to those described in G.6 except that vector pBR322 DNA was linearized with HindIII restriction endonuclease (gift of B. Powell). Competent *E. coli* MC1061 cells and the transformation procedure were essentially as described in section G. 6.

Recombinant colonies were detected by antibiotic screening which showed sensitivity to tetracycline (as a result of insertional activation of the tetracycline-resistance gene) and resistance to ampicillin i.e. tet^s, amp^r. Cloning of the correct HindIII DNA fragment,

subcloned in the above manner, was further checked by screening recombinant colonies (tet^s, amp^r) with the cloned ^{32}P-labelled keratin probe (see Chapter 6).

2.2.W MISCELLANEOUS

Cell densities were determined by measurement of the absorbance of a cell suspension at 600nm in a 1cm cuvette using a Hitachi Model 101 spectrophotometer.

RNA and DNA concentrations were estimated spectrophotometrically (Varian Superscan 3) assuming that one A_{260} unit equals 40µg/ml for RNA and 50µg/ml for DNA.

2.2.X CONTAINMENT FACILITIES

All work involving recombinant DNA was carried out under C1/EK1 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

ISOLATION AND CHARACTERIZATION OF WOOL FOLLICLE RNA

AND

CONSTRUCTION OF A WOOL FOLLICLE cDNA LIBRARY

CHAPTER 3

3.1 INTRODUCTION

Extensive information concerning the structural organization of the HGT keratin genes and their associated regulatory sequences in the sheep genome is necessary before the molecular basis of the synthesis and complexity of the HGT proteins can be understood. If mRNAs for these proteins are isolated and purified, they can be used as probes to identify their respective structural genes. Recombinant DNA technology has provided the means through which single mRNA species can be isolated by the conversion of a complex mRNA population into double-stranded complementary DNA (ds cDNA) with the subsequent production of homogenous probes by cloning of the DNA. Briefly, double-stranded molecules, whose sequences are derived directly from the mRNA, are joined to a vector DNA molecule that can replicate autonomously in a bacterial cell. These molecules are then used to transform bacterial cells which are plated out for single colonies. Since each colony is derived from a single cell and transformation conditions are such that each cell carries only one type of recombinant molecule, each colony carries copies of one recombinant molecule and thus represents a single mRNA species.

Before proceeding to the synthesis of cDNA molecules, it was necessary to isolate and characterize the mRNA from wool follicle cells actively-synthesizing keratin proteins.

Many eukaryotic mRNA molecules have a polyadenylic acid tract at their 3'terminus (Lee *et al.*, 1971; Darnell *et al.*, 1971) and a number of techniques have been developed utilizing the poly(A) segment to specifically select mRNA. The most common include poly(U) bound to filters, cellulose or sepharose (Kates, 1970; Lindberg and Persson, 1972; Kates, 1973), and oligo(dT) bound to cellulose (Aviv and Leder, 1972). These methods depend on the formation of stable base-pairing between the

poly(A) tail of the mRNA and the poly(U) or oligo(dT) attached to an inert support.

It has been shown (Ward and Kashmarik, 1980) that wool follicle mRNA sequences bind oligo(dT) due to the presence of poly(A) tracts at the 3' ends of the RNA and translation of the follicle RNA in a rabbit reticulocyte cell-free translation system (Pelham and Jackson, 1976) led to the production of essentially all three keratin protein classes found in the wool fibre. Ward and co-workers subsequently isolated (unpublished data) cloned sequences for several low-sulphur and high-sulphur wool proteins but no high-glycine/tyrosine (HGT) clones. The exclusion of HGT keratin sequences was probably due to the size-selection performed on the ds cDNA material to eliminate low molecular weight cDNA species prior to cloning (K. Ward, personal communication).

The initial and crucial step towards unravelling the structure and expression of HGT keratin genes was to therefore construct a wool follicle cDNA library from which these particular keratin cDNA clones could be isolated. Although this study was aimed at isolating HGT recombinants, the total poly(A)$^+$ RNA fraction was chosen as starting material since cloning of RNA species other than those encoding HGT proteins may be of interest in further studies on wool follicle mRNAs. For this purpose a cDNA library which would be truly representative of the wool follicle mRNA population would be essential.

A variety of methods for cloning eukaryotic mRNA species (Efstratiadis et al., 1976; Rabbits, 1976; Zain et al., 1979; Land et al., 1981) are well established. All these methods start with the synthesis of a DNA copy (cDNA) of the mRNA using AMV reverse transcriptase. The cDNA serves as a template and primer for synthesis of the second strand which is covalently linked to the first strand. The hairpin loop at the 3' end of the molecule is specifically cleaved with single-strand specific S_1 nuclease (Vogt, 1973) rendering the ds cDNA ends suitable for the

addition of homopolymeric tails or linkers. In the addition of oligo(dG) tails to the vector molecules and oligo(dC) tails to the ds cDNA molecules, hydrogen-bonding between the complementary homopolymeric tails creates an open circular hybrid molecule capable of transforming E. coli. In the second procedure synthetic linkers are added to the ds cDNA molecules which are cleaved with an appropriate restriction endonuclease and then ligated into vector DNA cleaved with compatible enzymes.

The cDNA tailing method was incorporated in the cloning scheme used in this work, since it involved less handling of the DNA than the "linker" method which until recently (Kurtz and Nicodemus, 1981) was considered as the less efficient ligation procedure. The plasmid pBR322 (Bolivar et al., 1977; Sutcliffe, 1978) was chosen as the vector molecule because it is small (about 4.3 kb), well characterised, contains a single PstI recognition site which can be regenerated following insertion and joining of foreign DNA by the addition of dG/dC tails and carries two antibiotic resistance genes which facilitate the selection of recombinant molecules.

The purification of mRNA from wool follicles and the construction of a wool follicle cDNA library is described in this chapter.

3.2 RESULTS

3.2.A Isolation of Wool Follicle Poly(A)$^+$ RNA

A dramatic improvement in the yield of follicles resulted from the adoption of a new fast-setting resin as an adhesive for the wool roots (see 2.2.C.1). This resin set within 5 to 10 minutes (as compared with the Araldite adhesive employed by Ward and Kashmarik (1980) which was left overnight to set in place) and the hardened strips had a smooth flat surface from which the frozen follicle material was easily collected.

A number of points were considered in deciding on the particular breed of sheep to be used in this study. They were as follows : (1) optimal density of fibres (2) avoiding poor elasticity of the skin which invariably caused the skin to tear during the harvesting procedure (3) a tendency of certain wool fibre types to break off above the surface of the skin, leaving the intact follicles below the surface. The response of a variety of sheep breeds (Corriedale, Merino, Border Leicester, Dorset Horn and Merino-Dorset Horn X Border Leicester) to the follicle harvesting technique described above determined the breed of sheep, Merino-Dorset Horn X Border Leicester, as the main source of material for this work.

A number of extraction methods of RNA from wool follicles were attempted. Several of these (Tushinski et al., 1977; Ward and Kashmarik, 1980) were unsatisfactory in that the yield and integrity of RNA isolated were quite variable and it became clear from these initial attempts that isolation of intact RNA required rapid and sustained inactivation of ribonucleases.

Two procedures employing guanidine hydrochloride as a protein denaturant and ribonuclease inhibitor were compared to determine the best method of RNA extraction from wool follicles. The first of these methods (Glisin et al., 1974) employed CsCl centrifugation as a means of separating protein, DNA and RNA and gave essentially undegraded RNA as determined by the ratio of 28S to 18S rRNA of 2:1 (as examined by gel electrophoresis, see 2.2.M) but the yields were very low. The procedure finally adopted for all RNA preparations was a modification of the method described by Brooker et al., (1980). This method is detailed in 2.2.C.1 and employed phenol extractions for protein removal and LiCl-precipitation (Diaz-Ruis and Kaper, 1978) for removal of DNA from the preparations. A comparison of follicle RNA prepared by two different procedures is shown in Figure 3.1. RNA prepared by the guanidine hydrochloride procedure was essentially undegraded (tracks 1 and 2) when compared to chicken feather

FIGURE 3.1

AGAROSE GEL ELECTROPHORETIC ANALYSIS OF WOOL FOLLICLE RNA (COMPARISON OF POLYSOME AND GUANIDINE HYDROCHLORIDE RNA ISOLATION PROCEDURES)

10µg of each RNA prepared by the different procedures described in section 3.2.A, was analyzed by electrophoresis on a 1% non-denaturing agarose gel (2.2.M.1). Samples were loaded onto the gel in the presence of formamide to disaggregate RNA complexes and selected samples were also heated at 65°C for 3 minutes prior to electrophoresis.

TRACKS:

1. Heat denatured sample
2. Unheated sample

} Guanidine-HCl procedure

M. Chicken rRNA 28S and 18S molecular weight markers

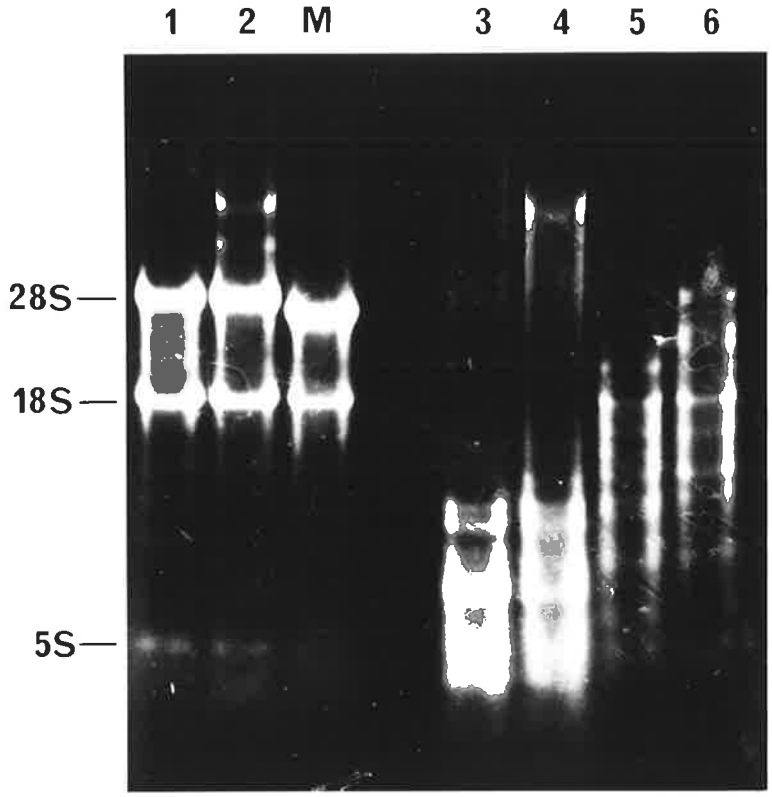
3. Heat denatured sample
4. Unheated sample
5. Heat denatured sample
6. Unheated sample

} Polysome procedure

Preparation 1

Preparation 2

The 28S, 18S and 5S rRNA species are indicated.


1
2
M
3
4
5
6
28S
18S
5S

28S and 18S rRNA markers. However RNA prepared by the polysome method of Ward and Kashmarik (1980) was usually degraded and gave rise to a large number of low molecular weight RNA species (tracks 3, 4, and 5, 6). In all cases, RNA aggregation was observed as high molecular weight material (tracks 2, 4, 6) which was less apparent after heat-dissociation of samples immediately prior to electrophoresis on the non-denaturing gel (tracks 1, 3, 5).

Total RNA samples were further purified by two rounds of oligo (dT)-cellulose chromatography (Aviv and Leder, 1972) with a heat-dissociation step carried out before the second passage through the column, to disrupt aggregated RNA. A yield of 10-15µg of poly(A)$^+$ RNA per gram of starting follicle material was obtained and was free of DNA as judged by gel electrophoresis analysis of RNA samples.

3.2.B In Vitro Translation Products Directed By Follicle RNA

To examine the protein products specified by the isolated RNA, wool follicle poly(A)$^+$ RNA was translated in a wheat germ cell-free system (Roberts and Paterson, 1973), see section 2.2.D and the ^{3}H-labelled protein products analyzed on SDS-urea 15% polyacrylamide gels (2.2.E).

The purpose of these experiments was to identify the presence of mRNA species in the isolated RNA which encoded the constituent keratin protein families within the wool fibre and therefore a detailed optimization of translation conditions for wool follicle mRNA was not carried out. However the following reaction conditions for the wheat germ cell-free system were noted : (1) optimal salt concentrations were found to be 2mM Mg acetate and 60mM KCl (2) 5 to 10µg of total RNA gave a 3-fold increase in TCA-precipitable material over the background level and 0.5 to 2µg of poly(A)$^+$ RNA gave 10 to 18-fold stimulation above background compared to purified rabbit globin mRNA which routinely gave up to

100-fold incorporation.

The fractionation of wool follicle RNA *in vitro* translation products on the basis of molecular weight is shown in Figure 3.2. To overcome the difficulty in matching stained gels with autoradiograms, wool proteins labelled *in vitro* with ^{14}C-iodoacetic acid (tracks marked W) were used to indicate the electrophoretic mobilities of the translation products. The polypeptides synthesized *in vitro* ranged from M_r 5,000 to 70,000 with many prominent products (tracks marked L and T). Due to the distinctive amino acid composition of the three major keratin families in the wool fibre (Crewther, 1976), different polypeptide products could be identified depending on the type of ^{3}H-labelled amino acid used in the translation reactions. For example, the HGT proteins are characterized, as their name suggests, by a high content of tyrosine. Accordingly, when ^{3}H-tyrosine was used the majority of the labelled translation products migrated in the region characterized by these low molecular weight keratins (Figure 3.2C, track T) with some low molecular weight high-sulphur components also evident. When ^{3}H-leucine was used to label the translation products the low-sulphur keratins (leucine content about 11 residues %, ie., 11 leucine residues per 100 residues) were well represented together with some high-sulphur protein components (leucine content about 4 residues %). See Figure 3.2B, track L. Translation of total RNA although less efficient yielded essentially identical products to those of poly(A)$^{+}$ RNA (data not shown).

Due to their high content of cysteine and their propensity to form aggregates via disulphide bridges, the high-sulphur proteins electrophoresed as fuzzy, broad bands on the SDS-urea-polyacrylamide gels (see track L, Figure 3.2B). The disruption of disulphide bonds and stabilization by alkylation through S-carboxymethylation improved the resolution of these proteins but was not always used since the procedure often resulted in loss of more than 50% of the labelled material. As a

FIGURE 3.2

SDS-UREA POLYACRYLAMIDE GEL ELECTROPHORESIS OF WOOL FOLLICLE RNA IN VITRO TRANSLATION PRODUCTS

Wool follicle poly(A)$^+$ RNA, isolated as described in 2.2.C.2, was translated in a wheat germ cell-free system (2.2.D). A one-dimensional SDS-urea 15% polyacrylamide gel system (2.2.E.3) was used to examine and compare the electrophoretic patterns produced by the ^{3}H-labelled translation products and ^{14}C-labelled wool proteins (2.2.E.1). The major wool keratin protein families are indicated as follows,

LS = Low-sulphur
HS = High-sulphur
HGT = High-glycine/tyrosine

PART A TRACKS:

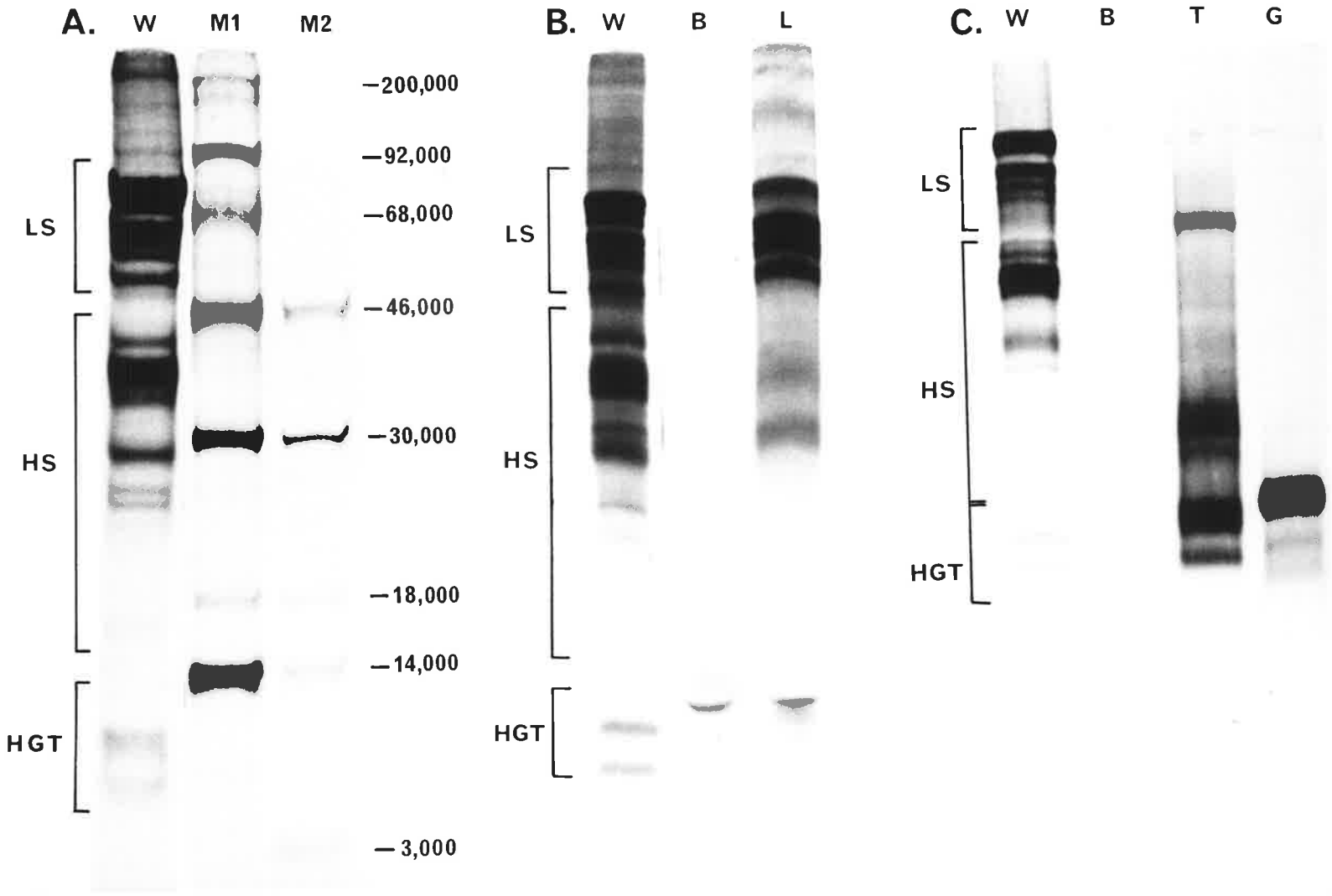
W. ^{14}C-labelled wool proteins
M1. Commercial ^{14}C-labelled protein molecular weight markers M_r range 14,000 - 200,000
M2. Commercial ^{14}C-labelled protein molecular weight markers M_r range 3,000 - 46,000

PART B TRACKS:

W. ^{14}C-labelled wool proteins
B. Endogenous wheat germ activity = background level
L. ^{3}H-leucine labelled *in vitro* translation products

PART C TRACKS:

W. ^{14}C-labelled wool proteins
B. Endogenous wheat germ activity = background level
T ^{3}H-tyrosine labelled *in vitro* translation products
G. Rabbit β-globin mRNA translation product - used as a control for wheat germ activity and as a molecular weight marker ($M_r \simeq 14,000$)


A.
W
M1
M2
LS
HS
HGT
200,000
92,000
68,000
46,000
30,000
18,000
14,000
3,000
B.
W
B
L
LS
HS
HGT
C.
W
B
T
G
LS
HS
HGT

result, *in vitro* keratin translation products did not show exactly the same patterns (unless S-carboxymethylation was carried out prior to gel electrophoresis) as keratins labelled *in vitro* with ^{14}C-iodoacetic acid, as is evident in Figure 3.2. This result is consistent with the observations (Marshall, 1983; B. Powell, P. Morris, personal communications) that S-carboxymethylated keratins migrate anomalously as compared with the unmodified proteins. Non-identical patterns could also be explained by the fact that a bias is created towards the identification of *in vitro* translation products which are rich in the particular ^{3}H-labelled amino acid used to label the synthesized proteins. Differences in *in vitro* translational activities between the various keratin mRNAs may also account for the altered polypeptide pattern when compared with the *in vitro* ^{14}C-labelled wool proteins. This difference in translational efficiencies has been observed with other eukaryotic mRNA molecules that are normally expressed in the same cell for example, rabbit α-globin mRNA translates more efficiently than rabbit β-globin mRNA in a wheat germ cell-free system (Roberts and Paterson, 1973) and ovalbumin always translates more efficiently than conalbumin in the rabbit reticulocyte system (Palmiter and Smith, 1974).

Whilst the polypeptides have not been directly identified all indirect evidence is consistent with the conclusion that the major products directed by the affinity-purified RNA correspond to the three wool keratin protein familes. This RNA was therefore used as the starting material for cloning as described below.

3.2.C Construction of a Wool Follicle cDNA Library

1. Synthesis of ds cDNA

Double-stranded cDNA was prepared by a procedure (Rougeon and Mach, 1976; Wickens *et al.*, 1978) in which the purification of the

single-stranded intermediate was omitted by using AMV reverse transcriptase for the synthesis of both strands of the cDNA. This procedure is detailed in Chapter 2, methods section G.

3µg of total poly(A)$^+$ RNA was used as a template for cDNA synthesis. By the number of ^{32}P-nucleotides of known specific activity incorporated, 12.8% copy of the mRNA was calculated, ie., is 0.384µg of cDNA had been synthesized. A final amount of 0.1µg of ds cDNA was used for cloning.

2. S_1 nuclease digestion and size analysis of ds cDNA

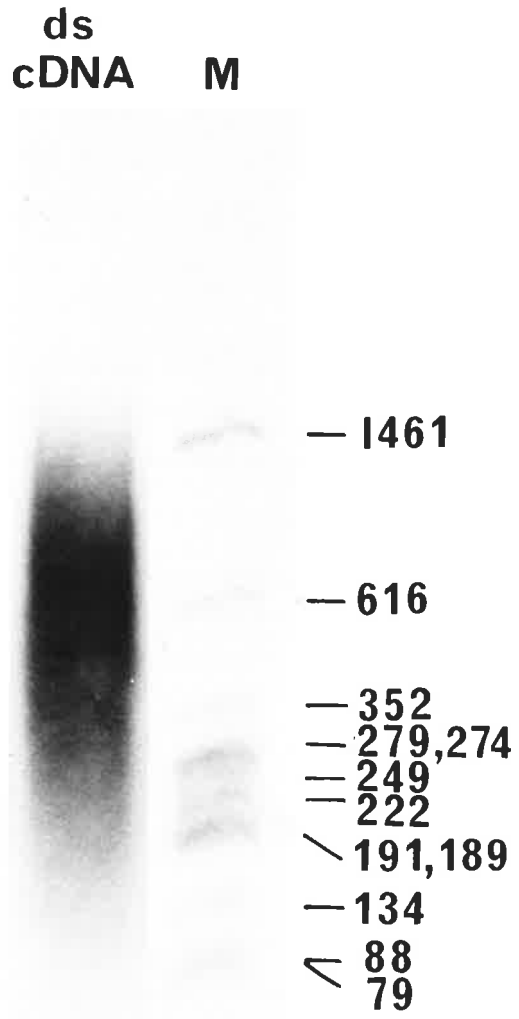
Removal of the hairpin loop structure of the ds cDNA was achieved by digestion with S_1 nuclease. Size-selection of the S_1-treated ds cDNA prior to cloning was carried out by chromatography on a Sephacryl-S1000 column. A known molecular weight marker (Sau96-digested pBR322 DNA end-labelled with [α-^{32}P]dCTP) was passed through the same column under identical conditions and fractions analyzed by gel electrophoresis, as a guide to select a specific size range of ds cDNA molecules. Alkaline agarose gel electrophoresis of S_1-treated ds cDNA prior to size-fractionation is shown in Figure 3.3. Although the average length of the ds cDNA was about 500 bp, it was important to remove as much of the low molecular weight material as possible because the presence of a large number of short ds cDNA molecules results in the cloning of these sequences in preference to the large molecules (Maniatis et al., 1982). However it was important to strike a balance between excluding low molecular weight material but not losing potential HGT cDNAs which would likely be short. The minimum size of HGT mRNAs was calculated at about 300 bases, assuming short 5' and 3' untranslated regions of about 50 bases and a coding region ranging from 180-270 bases as calulated from the molecular weight of the HGT proteins which range from 6,000 to 9,000. Therefore ds cDNA material greater than approximately 200 bp was

FIGURE 3.3

SIZE ANALYSIS OF DS cDNA

A sample of the total wool follicle ds cDNA, prepared by the AMV reverse transcriptase method described in the text (3.2.C.1), was electrophoresed on a 1% alkaline agarose gel (2.2.M.2). The ds cDNA ranged in size from 100 to 1500 bp, as judged by the co-electrophoresis of molecular weight markers (M) generated by end-labelling a Sau 96 digest of pBR322 with $[\alpha\text{-}^{32}P]$dCTP (2.2.H.5). Fragment sizes are given in bp.

Ds cDNA material greater than 200 bp was collected by combining the appropriate fractions following Sephacryl S1000 - chromatography (3.2.C.2), and used for subsequent cloning into E. coli.


ds
cDNA
M
1461
616
352
279,274
249
222
191,189
134
88
79

collected, pooled and used for cloning into the plasmid vector pBR322.

3. Cloning of the ds cDNA

Tailing of the ds cDNA molecules was performed using terminal transferase and dCTP as substrate. Roychoudry et al., (1976) have shown that in the presence of Co^{++} ions terminal transferase will preferentially add nucleotide triphosphates to the 3' end of a double-stranded DNA molecule without prior treatment with exonuclease. The reaction was monitored by the incorporation of ^{3}H-dCTP as described in 2.2.G.5. Vector pBR322 DNA linearized with PstI was tailed in the same manner using ^{3}H-dGTP.

The number of dG:dC residues required for optimal annealing and transformation efficiency has been determined by Peacock et al., (1981). In general it has been found that the number of residues on the plasmid and insert DNA should be equal, with about 20 residues being added to each 3' end of the DNA for maximum annealing efficiency. Accordingly, the tailing reactions were stopped by the addition of the chelating agent EDTA when an estimated 20 residues were added to each 3' end.

Equimolar amounts of dC-tailed ds cDNA and dG-tailed pBR322 DNA were annealed in a high salt solution to stablilize dG/dC base pairing of the complementary homopolymeric tails and this mixture was used to transform E coli. Approximately $3x10^{4}$ transformants per µg of ds cDNA were obtained when the annealed mixture was used to transform E. coli compared to $7x10^{6}$ per µg of intact pBR322 and a background of $1x10^{3}$ per µg for PstI-cleaved dG-tailed vector.

3.2.D Detection of Recombinants

1. Antibiotic screening of transformants

Following the screening of colonies on L + amp and L+ tet phates (2.2.G.7), where insertional inactivation of the ampicillin resistance gene of pBR322 resulting from the insertion of foreign DNA into

the unique PstI site was used as the marker, about 85% of the colonies were found to be recombinants.

2. Hybridization screening with wool follicle ^{32}P-cDNA

Nitrocellulose filters containing recombinant (amp^s, tet^r) colonies prepared by the method of Grunstein and Hogness (1975) were screened with ^{32}P-labelled wool follicle cDNA (2.2.H.1). Figure 3.4A shows the result of screening a small part of the library using this particular probe with individual colonies giving a wide range of hybridization intensities reflecting the abundance of poly(A)$^+$ RNA corresponding to each cDNA clone. Greater than 90% of all colonies demonstrated variable degrees hybridization to the wool follicle ^{32}P-labelled cDNA. The nature of the 'non-hybridizing' clones were not examined further and could represent very low abundance mRNAs.

The HGT proteins constitute the least abundant keratin family in the wool fibre (see 1.2.C.3) and it was assumed that the HGT mRNAs would be present at a correspondingly low level in the wool follicle as compared with the other keratin mRNA species. For this reason clones which showed a range of intensities of hybridization signals with the ^{32}P-labelled follicle cDNA were chosen to represent the wool follicle cloned mRNA population, in order to avoid a bias for the more abundant mRNAs expressed in the wool follicle cells.

About 1400 clones were chosen to constitute the cDNA library to be used in this study and these were maintained as described in 2.2.G.7 and screened with a variety of probes as detailed in Chapters 4, 5 and 7.

3. Screening for rRNA sequences

The cloning of follicle rRNA sequences seemed unlikely since the RNA used for cloning was first selected by oligo(dT)-cellulose chromatography and then primed with oligo(dT)$_{10-12}$ for cDNA synthesis. To confirm this, ^{32}P-labelled follicle rRNA (28S and 18S) was used to probe a

FIGURE 3.4

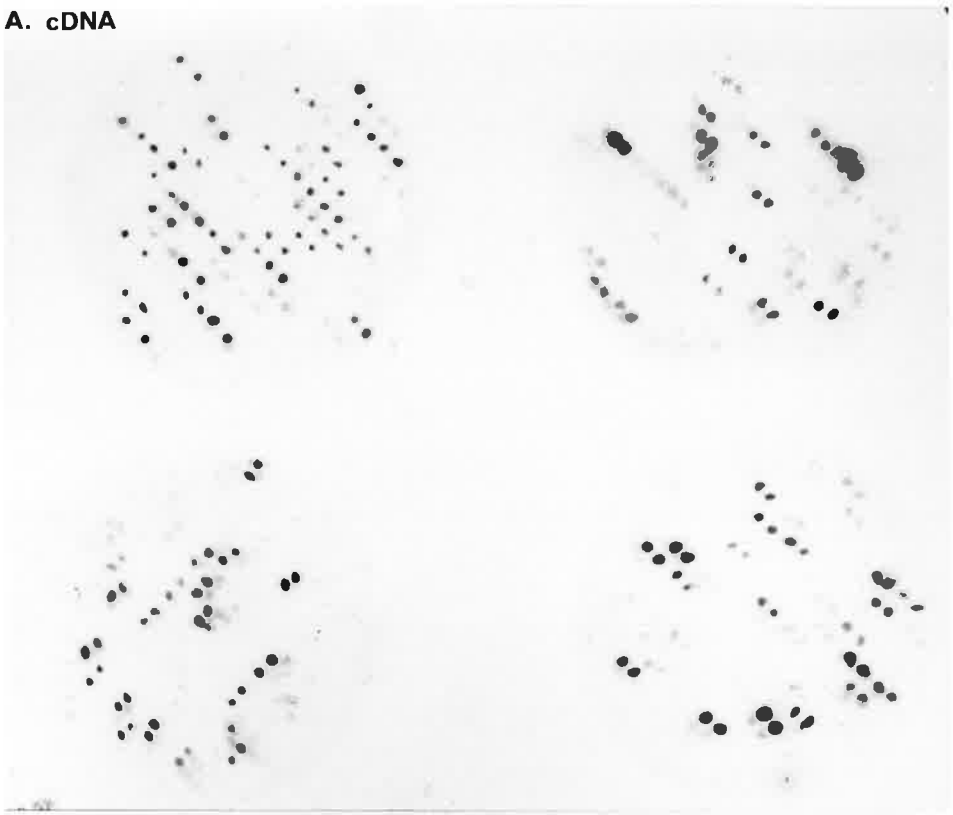
DETECTION OF WOOL FOLLICLE cDNA SEQUENCES BY THE COLONY SCREENING PROCEDURE

Colonies were spotted in duplicate onto nitrocellulose filters, denatured *in situ* (2.2.G.7), then hybridized to total wool follicle ^{32}P-labelled cDNA (2.2.H.1) or kinased wool follicle rRNA (2.2.H.2), washed and autoradiographed.

PART A : ^{32}P-cDNA screening at 1xSSC/0.1% SDS, 65°C

PART B : ^{32}P-rRNA screening at 2xSSC/0.1% SDS, 65°C.

A. cDNA



B. rRNA

portion of the cDNA library as shown in Figure 3.4B and as expected, no hybridization was observed, indicating that no rRNA sequences had been cloned.

3.2.E ANALYSIS OF A RANDOM SET OF RECOMBINANTS

To obtain an estimate of the size of inserted DNA segments, 'minipreps' of DNA from seven recombinant colonies and two non-recombinants were prepared. However, four of the seven recombinants chosen for the analysis could not be resected with PstI so this method could not be used to determine the insert length of these clones. Although the dG/dC tailing procedure used in the cloning scheme described in this chapter was designed to reconstitute PstI sites and thus permit easy excision of the inserted DNA, others workers have similarly found that as few as 40% of the DNA sequences inserted in this way are finally resectable by PstI cleavage (Villa-Komaroff et al., 1978). Instead the DNA was digested with HpaII, which has restriction sites close to and either side of the PstI site, end-labelled with $[\alpha\text{-}^{32}P]$ dCTP and electrophoresed on a 6% polyacrylamide gel (see 2.2.L.1). pBR322 DNA was included to indicate the digestion pattern resulting from vector DNA.

In Figure 3.5, recombinants (tracks 1, 3, 4, 6, 7, 8) were apparent by (a) the presence of one or more additional bands of various lengths as compared with pBR322 vector DNA and (b) the absence of the 110 bp fragment of pBR322 which contains the PstI site and whose size was altered as a result of the insertion of foreign DNA at that site. The two non-recombinants maintained this 110 bp band and had no additional bands (see tracks 2 and 5) and therefore produced the same digestion pattern as pBR322 DNA.

The average insert size of these selected recombinants was estimated by subtracting the contribution of vector DNA from the length of newly generated fragments and found to be about 350 bp which was less than

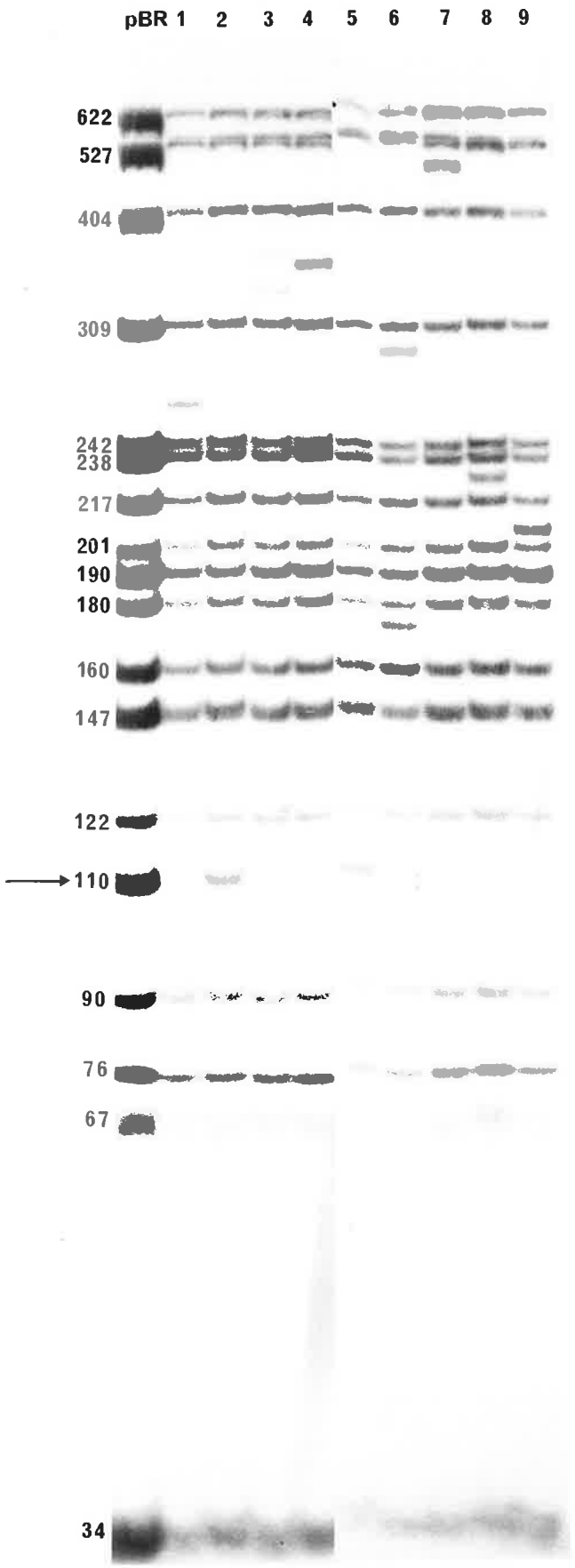
FIGURE 3.5

INSERT SIZE ESTIMATION OF RANDOMLY-SELECTED WOOL FOLLICLE cDNA RECOMBINANTS

Approximately 0.1 g of plasmid DNA was digested with HpaII (3.2.K), end-labelled with $[\alpha-^{32}P]$ dCTP, fractionated on a 6% polyacrylamide gel (2.2.L.1) and autoradiographed. HpaII-digested pBR322 was also included to show which of the fragments were derived from the vector molecule. Plasmid bands are indicated and the 110 bp fragment present in pBR322 and non-recombinant plasmids, but absent in the recombinant plasmids, is shown by the arrow. Fragments sizes are given in bp.

TRACKS :

pBR.	pBR322 DNA
1.	Clone A
2.	Non-recombinant (pBR322) DNA
3.	Clone B
4.	Clone C
5.	Non-recombinant (pBR322) DNA
6.	Clone D
7.	Clone E
8.	Clone F
9.	Clone G


pBR 1 2 3 4 5 6 7 8 9
622
527
404
309
242
238
217
201
190
180
160
147
122
110
90
76
67
34

the average ds cDNA size (500 bp) used for cloning.

3.3 DISCUSSION

This chapter described the isolation of wool follicle mRNA and the cloning of the mRNA population into E. coli.

RNA was extracted from wool follicles by a method which maintained the inactivation of ribonucleases and prevented degradation of the RNA. mRNA was then purified from total RNA by oligo(dT)-cellulose chromatography. Analysis of the in vitro translation products synthesized in a wheat germ cell-free system showed that the selected poly(A)$^+$ RNA faithfully represented mRNA species encoding the three principal wool keratin families present in the wool fibre. This mRNA population was converted into a library of cDNA clones using the dG/dC tailing procedure as described Maniatis et al., (1976) and others. Hybridization of follicle ^{32}P-labelled cDNA to greater than 90% of the colonies confirmed the successful cloning of follicle mRNA species. An average insert size for seven recombinants was estimated at about 350 bp, but because of the small sample size this obviously could not be taken as a true indication of average insert size within the cDNA library.

The availability of cloned mRNAs from the wool follicle was the initial step towards the analysis of mRNA and gene structures for individual wool proteins. Chapter 4 and 5 discuss the approaches used to detect and isolate HGT wool keratin sequences from this cDNA library.

CHAPTER 4

SCREENING FOR SHEEP HGT KERATIN cDNA CLONES:

INDIRECT APPROACH

CHAPTER 4

4.1 INTRODUCTION

The isolation of specific cloned DNA sequences depends, to a great extent, on an available assay to detect those clones containing the desired sequence. Until recently, assays primarily involved the use of specific hybridization probes either purified from abundant mRNAs, such as globin mRNA, or prepared from previously isolated clones. For example, Woo et al., (1983) used a rat phenylalanine hydroxylase cDNA clone to isolate the corresponding sequence from a human liver cDNA library.

It was not possible to employ these approaches to detect HGT cDNA clones for the following reasons. Firstly, no HGT-sequence-related probes have been isolated at all. Secondly, the abundance of HGT mRNA is likely to be very low when the following points are considered. The HGT protein content of the breed of sheep examined in this thesis was calculated to be about 3%. This figure was based on the weight of HGT proteins extracted as a proportion of the initial starting wool weight from a Merino-Dorset Horn cross sheep (M. Calder, personal communication). The contribution of component F (the most abundant protein species) to the total tyrosine-rich fraction of Merino wool was estimated by Dopheide (1973) at 15%, so that if the HGT protein amount in the wool fibre reflects the proportion of HGT mRNA in the follicle RNA population, then the relative abundance of HGT-F mRNA would be about 0.45%. Therefore the isolation of pure HGT-F and other HGT mRNA species in the way that can be done for globin mRNA, would be an extremely difficult task due to their low abundance.

The initial approach used to detect HGT cDNA clones, as described in this chapter, was to obtain a partially purified wool follicle RNA fraction enriched for HGT mRNAs and synthesize ^{32}P-labelled cDNA to this RNA for use as a probe to the follicle cDNA library.

An alternative approach considered in determining the coding capabilities of the cDNA clones was hybrid-selected translation (Parnes et al., 1981). In this procedure, cDNA clones to be screened are bound to nitrocellulose filters and tested for the ability of the filter to select HGT keratin mRNA. The filter-selected mRNA is then assayed for its ability to direct the synthesis of HGT polypeptides in an in vitro translation reaction. However, several difficulties were envisaged in using this approach : (1) it would involve tedious preparations of pooled plasmid DNAs (from 1400 clones) bound to nitrocellulose filters (2) large amounts of wool follicle RNA would be needed to ensure adequate representation of HGT mRNA species present for selection (3) because the proportion of HGT-mRNAs was expected to be very low, resulting in the production of correspondingly small amounts of HGT translation products, specific immuno-detection of HGT proteins would be an asset. Unfortunately, no such antibodies were available at the time this approach was considered.

4.2 RESULTS

4.2.A Detection of Low-sulphur and High-Sulphur Keratin cDNA Sequences

It was of interest to divide the cDNA library into groups of related sequences using the available low-sulphur (LS) and high-sulphur (HS) keratin clones. From the available protein sequence data (Crewther, 1976) it was expected that a single coding-region probe for each keratin sub-family would cross-hybridize with all members of the sub-family under low stringency conditions. This has been shown for the high-sulphur SCMKB-2 gene sequences (Powell et al., 1983). Specific DNA probes for the two HS keratin sub-families, SCMKB-I and SCMKB-IIIA, were not available at the time of these experiments and since cross-hybridization was not likely

to occur between different keratin sub-families, the corresponding cDNA clones would not be detected. The four cDNA clones (gifts of K. Ward and B. Powell) employed as probes are listed in Table 4.1, together with their respective protein counterparts and the proportion of clones in the cDNA library detected by these probes. All four clones contained both coding and non-coding sequences.

A similar hybridization trend was obtained for both the LS and HS keratin probes but for convenience only part of the LS keratin cDNA hybridization results are shown in Figure 4.1. At high stringency washing conditions (Figure 4.1B) colonies presumably containing sequences highly homologous to that of the probe gave strong signals, whereas at low stringency many more clones were seen to hybridize with varying intensities of signals (Figure 4.1A). The clones detected under low stringency conditions probably represent related protein sequences within the same keratin sub-family which are homologous to the probe but are not identical to it (see Figure 4.1A).

Analysis of the combined results for the screening experiments showed a lower than expected number of LS and HS cDNA sequences, bearing in mind that these keratins constitute the bulk of the proteins synthesized in the wool follicle. Only 35% of the clones within the cDNA library hybridized to the LS and HS keratin probes. The remaining clones (about 65% of the cDNA library) which did not show significant hybridization to the cloned keratin probes discussed above, were used in subsequent screening for HGT keratin cDNA sequences.

4.2.B Enrichment for HGT keratin mRNA

1. Sucrose gradient fractionation of wool follicle RNA

The molecular weight range of each of the three main classes of keratins in the wool fibre (see Table 4.1) differ sufficiently so that

TABLE 4.1

LOW-SULPHUR AND HIGH-SULPHUR KERATIN PROBES
USED FOR SCREENING THE cDNA LIBRARY

The low-sulphur and high-sulphur wool keratin cDNA clones used as probes to the wool follicle cDNA library, constructed as described in Chapter 3, are listed in the table opposite.

Protein data has been compiled from Crewther (1976), Swart *et al.*, (1976) and Gillespie (1983).

Explanations of superscripts a, b and c are given in Table 1.1

d The cDNA clones listed were gifts of B. Powell and K. Ward. A dash indicates that cloned sequences for these sub-families have not been isolated or were not available at the time of screening.

e The percentage figures refer to the total number of clones detected by the low-sulphur and high-sulphur keratin probes in screening of the cDNA library.

PROTEIN CLASS	CHARACTERISTIC AMINO ACID CONTENT	PROTEIN FAMILIES WITHIN THE CLASS	FAMILY NOTATION[a]	MOLECULAR WEIGHT	AVAILABLE PROBES[d]	% cDNA CLONES DETECTED[e]
LOW-SULPHUR KERATINS	Cysteine (1-3 moles %)	2	SCMKA - 5/7 (1)/(3)[b] SCMKA - 8 (4)	56-58,000 38-43,000	p09 pV15	9.5%
HIGH-SULPHUR KERATINS	Cysteine (20-30 moles %)	4	SCMKA - 1 (?)[c] SCMKB - 2 (7) SCMKA-IIIA (11) SCMKA-IIIB (4)	23-26,000 19,000 16,000 11,000	- pSK10 - pKS18	24%

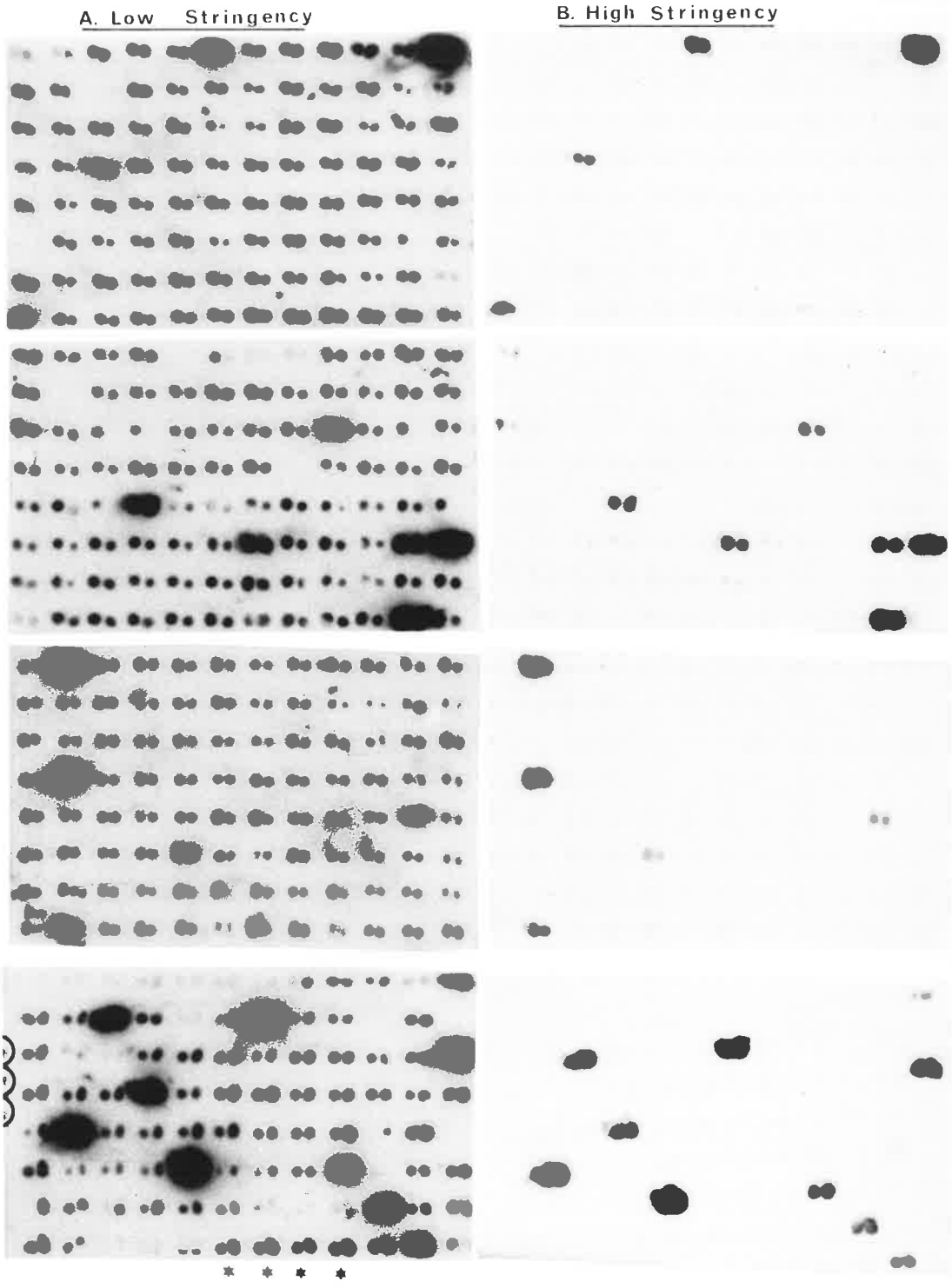
FIGURE 4.1

SCREENING FOR LOW-SULPHUR KERATIN SEQUENCES
IN THE WOOL FOLLICLE cDNA LIBRARY

The inserts of the low-sulphur cDNA clones, p09 and pV15 (see Table 4.1), were ^{32}P-labelled by nick-translation (2.2.H.3), then combined and used to probe the wool follicle cDNA library. Both low and high stringency conditions were used, as indicated below, and only part of the screening results are shown in this figure.

PART A : Low stringency (2xSSC/0.1% SDS, 65°C) wash results. The circled colonies on the left-hand side of the bottom filter indicate pBR322 control colonies. The colonies indicated by asterisks are examples of the weakly-hybridizing clones discussed in the text (4.2.A).

PART B : High stringency wash (0.1xSSC/0.1% SDS, 65°C) results of the same filters as shown in Part A.


A. Low Stringency
B. High Stringency

a partial size fractionation of their corresponding mRNAs seemed feasible. If the 5' and 3' non-coding regions of all follicle keratin mRNAs are not too dissimilar in length, then the mRNAs encoding the HGT proteins (M_r 6,000-9,000) should be smaller than those of the HS proteins (M_r 11,000-26,000) and LS proteins (M_r 38,000-58,000) and possibly be readily fractionated from them. Therefore, size-selection for a wool follicle RNA fraction encoding primarily HGT proteins was attempted.

Total wool follicle RNA was centrifuged through a 10-40% linear sucrose gradient. The RNA absorbance profile in Figure 4.2A is dominated by the characteristic 28S, 18S and 5S eukaryotic ribosomal RNA peaks although other RNAs are evident sedimenting between them. The RNA material sedimenting between the 5S and 18S rRNA peaks was collected and rerun on a shallow gradient (5-20%) to allow better resolution of the low molecular weight RNA species which should include the HGT mRNAs. Twenty-seven fractions were collected across this gradient (Figure 4.2B) and analyzed for their coding capacity as described below.

2. Analysis of the fractionated RNA by cell-free translation

To determine whether any region of the RNA profile was enriched for HGT mRNA, approximately 0.5μg of each collected fraction was translated in a wheat germ cell-free system (2.2.D). ^{3}H-tyrosine and ^{3}H-serine were used concomitantly in all the translation reactions to label the proteins synthesized *in vitro*, since these amino acids are present in characteristically high proportions in the three wool keratin protein families (Crewther, 1976). For example, HGT proteins have a tyrosine content of 12-21 residues % (residues per 100 residues), LS proteins have a serine content of 7-8.5 residues %, while the serine content of the HS proteins is 10-16 residues %.

The ^{3}H-labelled protein products were examined on the basis of molecular weight by electrophoresis on acrylamide gels in the presence of

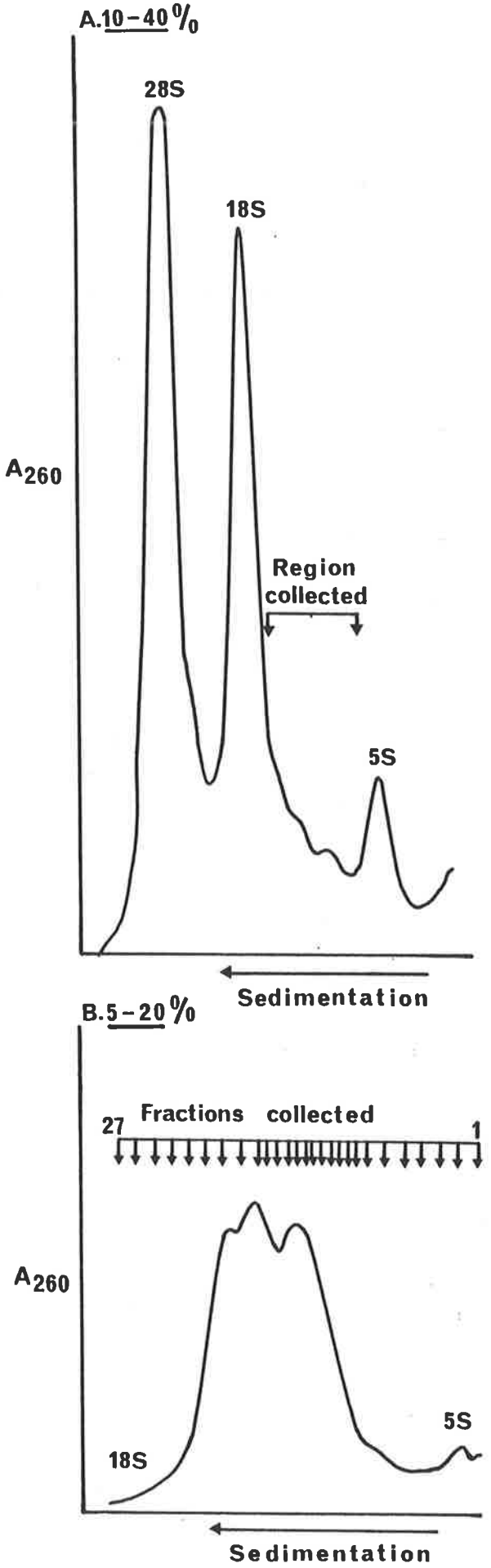
FIGURE 4.2

SUCROSE GRADIENT FRACTIONATION OF WOOL FOLLICLE RNA

PART A : Total wool follicle RNA (0.5mg) was centrifuged through a linear 10-40% sucrose gradient (2.2.C.3) and then fractionated according to the absorbance profile indicated by a UV spectrophotometer. The RNA sedimenting between the 5S and 18S rRNA peaks was collected and precipitated with ethanol.

PART B : The 5S-18S RNA collected as above in Part A, was centrifuged through a linear 5-20% sucrose gradient and twenty-seven fractions, spanning the absorbance profile as shown, were collected and separately precipitated with ethanol.

Note : the absorbance peaks in both profiles are not to scale.


A. 10-40 %
28S
18S
A260
Region collected
5S
Sedimentation
B. 5-20 %
27
Fractions collected
1
A260
5S
18S
Sedimentation

SDS and urea followed by fluorography (Figure 4.3). Translation products across the gradient profile are shown in tracks 1 through to 27. The first twelve fractions (see tracks 1 to 12) gave negligible translational activity above the background level (track B) and so it was presumed that the absorbance profile in this region was due to degraded 28S and 18S or aggregated 5S rRNA species. In the next five fractions (tracks 13 to 17) the HGT proteins were the predominant products, after which increasing amounts of HS proteins were visible, with the appearance in fraction 24 of a major LS protein which persisted in the last collected fractions (tracks 25 to 27) but at a level just above background. Ward and Kashmarik (1980) have shown that 21S wool follicle RNA translates to give LS protein products. Since only 5S to 18S wool follicle RNA was collected, fractionated and translated in the experiments discussed here, the low level of LS in vitro translation products was therefore not surprising. It was presumed that the last fractions (25 to 27) of the gradient contained predominantly 18S rRNA and a small amount, if any at all, of LS keratin mRNA. Furthermore, the observation that HGT protein products appeared over a wide range of the gradient fractions suggested that RNA aggregation had occurred and that an improvement in the RNA fractionation would require inclusion of a disaggregating agent such as formamide in subsequent sucrose gradients.

The first RNA fraction showing HGT protein activity (fraction 13), and the one least likely to contain the mRNAs encoding the smallest HS proteins, the SCMKB-IIIB family (Mr ≈ 11,000), was chosen for synthesis of ^{32}P-labelled 'enriched-cDNA'.

4.2.C Synthesis and Screening with 'HGT-enriched' cDNA

Approximately 1µg of the HGT-enriched RNA fraction 13 was used as template for the synthesis of oligo(dT)-primed cDNA (2.2.H.1). Filters containing wool follicle cDNA recombinant colonies were hybridized with

FIGURE 4.3

IN VITRO TRANSLATION PRODUCTS DIRECTED BY WOOL FOLLICLE 5S-18S RNA

Wool follicle 5S-18S RNA, fractionated as described in Figure 4.2B, was translated in a wheat germ cell-free system (2.2.D). Samples were electrophoresed on SDS-urea 15% polyacrylamide gels (along with ^{14}C-labelled wool protein markers) and the gels (2.2.E) autoradiographed after fluorography.

TRACKS :

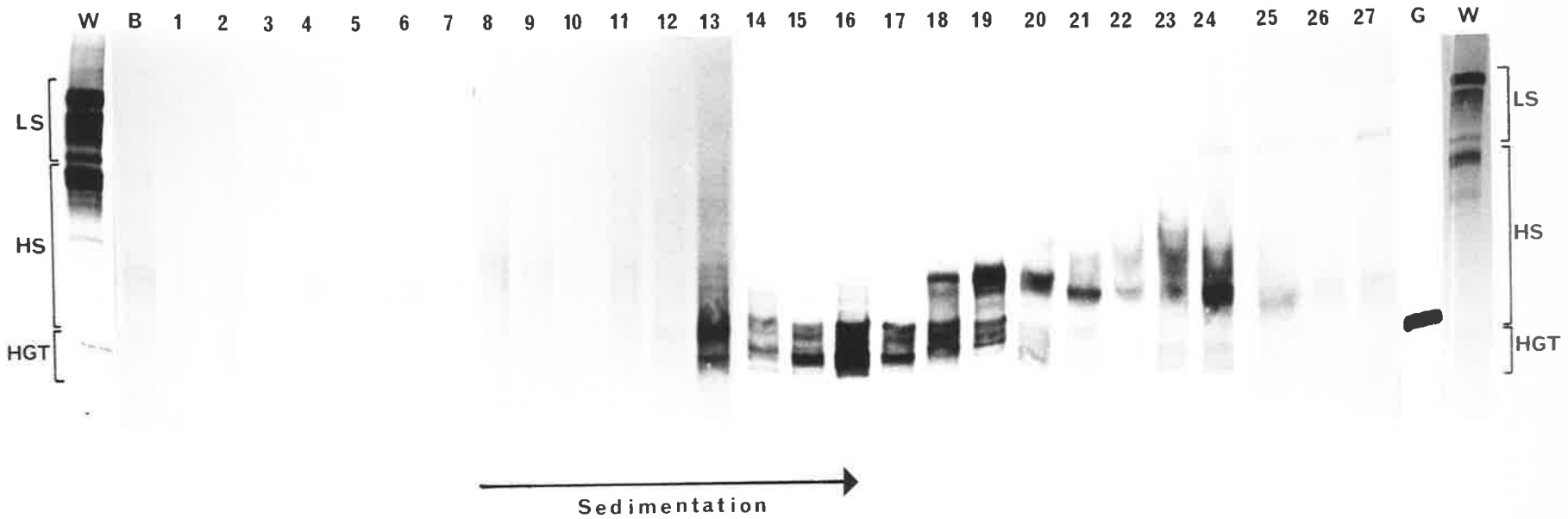
W.	^{14}C-labelled wool proteins
B.	Endogenous wheat germ activity = background level
G.	Rabbit β-globin mRNA translation product, $M_r \simeq 14,000$
1 → 27	Translation products labelled with ^{3}H-tyrosine and ^{3}H-serine and directed by RNA fractions of increasing sedimentation values

The three major wool keratin protein families are indicated as,

LS = Low-sulphur

HS = High-sulphur

HGT = High-glycine/tyrosine


W
B
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
G
W
LS
HS
HGT
LS
HS
HGT
Sedimentation

the ^{32}P-labelled cDNA probe under standard conditions (2.2.I.2). After stringent washes (0.5xSSC/0.1% SDS at 65°C) were carried out, a small number of colonies showed various degrees of hybridization while one colony gave a strong signal. Figure 4.4 shows the hybridization of the 'HGT-enriched cDNA' probe to clone p11E6. This clone was therefore chosen for further analysis by sequencing.

4.2.D Sequence Analysis of Clone p11E6

DNA was prepared (2.2.J.1) from plasmid p11E6 and restricted with PstI and a selection of four-base-specific endonucleases e.g. HpaII, HaeIII. PstI resection of the insert indicated a total length of about 300 bp with an internal PstI site, hence generating three fragments; one fragment containing solely pBR322 DNA and two fragments of insert DNA only. Due to its small insert size, the sequence of this clone was rapidly determined by isolating the two PstI and two HpaII fragments containing insert DNA (2.2.N.1), followed by sequencing according to the 'dideoxy' chain termination method (2.2.R).

The sequence of the insert from p11E6, shown in Figure 4.5, consists of 268 bp of a cloned DNA fragment which was bounded by homopolymeric tails, 25-30 residues in length added to the insert DNA during the cloning procedure. The remnant of a poly(A) tail, 12 nucleotides long is located at one end of the insert sequence, preceded at a distance of 17 bases by AATAAA, thought to be the putative eukaryotic polyadenylation signal (Proudfoot and Brownlee, 1976; Baralle, 1983). No translation termination codons (TGA,TAA,TAG) are present in the sequence upstream from this hexanucleotide, indicating that the cloned DNA in plasmid p11E6 does not extend into a protein-coding region.

The cDNA library was then rescreened with the nick-translated (2.2.H.3) insert of p11E6 in an attempt to isolate a longer clone which included the protein-coding sequence. However, only the same p11E6 clone

FIGURE 4.4

SCREENING FOR HGT KERATIN SEQUENCES USING AN 'ENRICHED cDNA' PROBE

RNA from fraction 13 (see 4.2.B.2 and Figure 4.3) was used as template for the synthesis of ^{32}P-labelled cDNA (2.2.H.1). This probe was used to screen the cDNA library under standard conditions (2.2.I.2) with a final wash at 1xSSC/0.1% SDS, 65°C.

Clone, p11E6 gave a strong hybridization signal with this cDNA probe and is indicated by the arrow on the autoradiogram opposite.

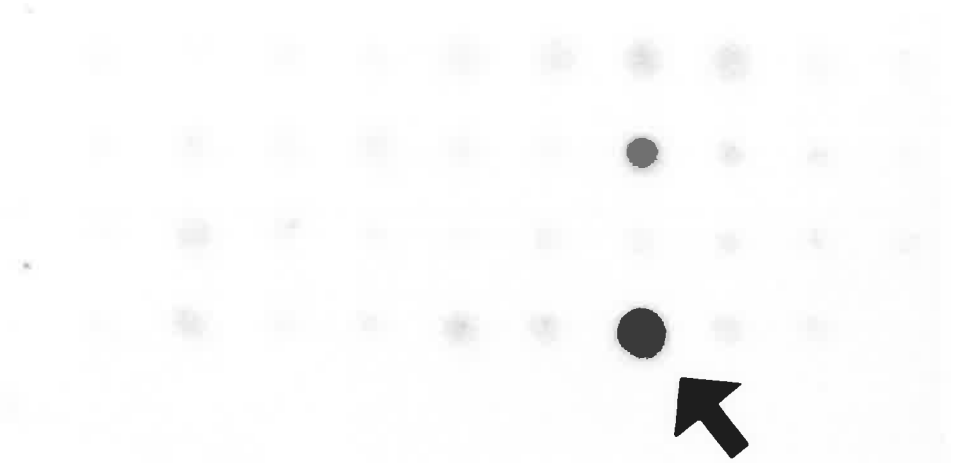
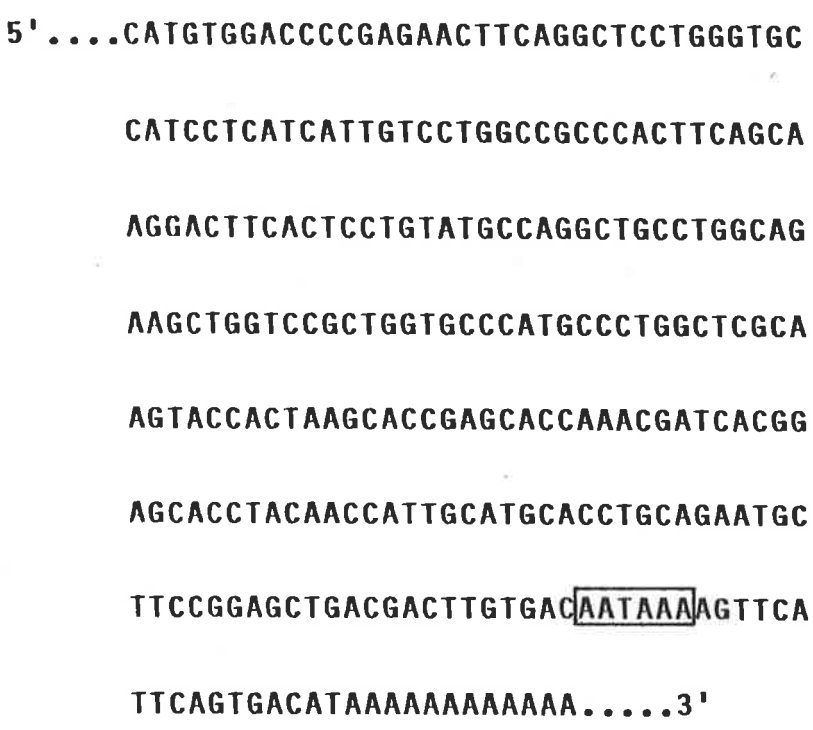


FIGURE 4.5

DNA SEQUENCE OF CLONE, p11E6

The insert from clone p11E6 (isolated as described in Figure 4.4), was sequenced by the"dideoxy" chain termination method (2.2.R), utilizing the internal PstI site as discussed in the text (4.2.D).

The putative polyadenylation signal, AATAAA is boxed and occurs 17 bases upstream to the remnants of a poly(A) tail consisting of 12 A residues.


5'....CATGTGGACCCCGAGAACTTCAGGCTCCTGGGTGC
CATCCTCATCATTGTCCTGGCCGCCCACTTCAGCA
AGGACTTCACTCCTGTATGCCAGGCTGCCTGGCAG
AAGCTGGTCCGCTGGTGCCCATGCCCTGGCTCGCA
AGTACCACTAAGCACCGAGCACCAAACGATCACGG
AGCACCTACAACCATTGCATGCACCTGCAGAATGC
TTCCGGAGCTGACGACTTGTGACAATAAAAGTTCA
TTCAGTGACATAAAAAAAAAAA.....3'

gave a strong hybridization signal on rescreening, with a few other colonies giving weaker signals. This result indicated that only one copy of this particular mRNA was present in the cDNA library. Therefore the cloned DNA in this recombinant plasmid contained the 3' untranslated region of an mRNA molecule which could not be identified due to the absence of protein-coding sequence.

4.3 DISCUSSION

The work described in this chapter involved the use of a partially purified mRNA population in the screening of a wool follicle cDNA library with the aim of isolating cDNA sequences for HGT wool keratins.

Initially, the cDNA library was screened to detect LS and HS keratin clones using the available probes. The low number of clones (35%) detected was surprising since Gillespie and Frenkel (1974) showed that the HGT proteins constitute no more than 13% of the wool fibre, so that the remainder would be composed primarily of the LS and HS keratins. One point of note is that HS probes were available for only two of the four HS keratin families, and if as shown in Table 4.1 the number of protein chains reflects the relative abundance of each sub-family then only half of the HS keratin sequences would be detected. It was possible that errors were made in the scoring of positive colonies, since the wide gradations in hybridization signals at low stringency conditions (see Figure 4.1A) made it difficult to distinguish between the non-hybridizing colonies and weakly-hybridizing colonies, and thus a number of clones may have been dimissed as background hybridization. Since LS and HS (and HGT) keratin cell-free translation products were synthesized from the RNA used for cloning (see 3.2.B), the under-representation of these sequences may reflect differences in the efficiency with which the keratin mRNAs are copied into ds cDNA. It has been reported (Rougeon and Mach, 1976) that

AMV reverse transcriptase synthesizes the second strand of different mRNA species with different efficiencies, but it is presently unknown whether this explains the observed screening results. Extensive analysis of the cDNA clones by DNA restriction and sequence analysis would be needed to clearly establish the similarities of the clones selected in this initial screening experiment and to then assign them to the various protein families.

Fractionation of wool follicle RNA on sucrose gradients yielded a fraction which encoded predominantly HGT protein products. RNA from this fraction was then used as template for cDNA synthesis and the resultant cDNA probe was used to screen for HGT keratin sequences in the wool follicle cDNA library. As a result of this work a single clone (p11E6) was isolated which apparently contained only 3' non-coding sequence and as such could not be directly identified; this sequence also showed no homology with the corresponding regions of the HGT cDNA clones characterized later in this thesis (5.2.E and 7.2.C). On rescreening of the cDNA library, p11E6 gave a strong hybridization signal with itself only. This hybridization result was somewhat puzzling, since the HGT proteins are small (M_r 6,000-9,000) and the size-range of the ds cDNA used for cloning would have allowed for almost full-length HGT cDNA species, even with long 3' non-coding regions as found in p11E6. HGT cDNA clones might be under-represented in the library if, firstly the mRNA concentration is less than that estimated from the HGT protein content of the breed of sheep used in this study or secondly, if the presence of secondary structures in the mRNAs (due to the unusual amino acid composition of these proteins) caused a less efficient transcription of HGT sequences as compared to the other keratin mRNA sequences. So even though it was difficult to isolate large amounts of wool follicle mRNA, sucrose gradient fractionation yielded several RNA fractions which directed the synthesis of HGT protein (although not exclusively), but for

unknown reasons cDNA made from this 'enriched' RNA did not detect cloned HGT sequences. The nature of the cDNA probe or the identity of p11E6 was not established for reasons discussed below.

Hybrid-selected translation (Parnes et al., 1981) was considered as a means of identyfing the protein product encoded by the truncated cDNA clone p11E6, but before this work was initiated the technique of custom-made oligonucleotide synthesis became available in our department.

This technique allowed the preparation of highly specific DNA probes, derived from the HGT protein sequences, to be synthesized and used for the detection of complementary sequences in the cDNA library. So rather than pursue the identification of clone p11E6 or the other colonies which showed weak hybridization signals with ^{32}P-labelled cDNA probe, which may or may not have represented HGT cDNA species, it was decided to utilize synthetic probes as a more specific screening approach for HGT cDNA clones. The construction of these HGT-specific oligonucleotide probes is described in detail in Chapter 5, prior to the isolation and characterization of cloned wool HGT mRNA sequences.

CHAPTER 5

IDENTIFICATION AND CHARACTERIZATION OF HGT-F cDNA CLONES: SYNTHETIC PROBE APPROACH

CHAPTER 5

5.1 INTRODUCTION

Synthetic oligonucleotides have become extremely valuable in recombinant DNA work as detection probes for specific mRNAs (Legon et al., 1982; Wallace et al., 1981) cDNAs (Sood et al., 1981; Suggs et al., 1981) and genes (Nickoloff and Hallick, 1982) by transfer hybridization techniques, and for high density screening of bacterial colonies (Colantuoni et al., 1983; Knott et al., 1984) and plaques (Montgomery et al., 1978) for specific DNA sequences. The relative ease with which large numbers of clones can be screened has enabled the isolation of cloned sequences for low abundance mRNAs such as human anti-haemophilic Factor IX (Choo et al., 1982) and human X-linked 3-phosphoglycerate kinase (Singer-Sam et al., 1983). Synthetic oligonucleotides therefore provide a general procedure for detecting cloned DNA sequences for any protein whose amino acid sequence (or part thereof) is known.

The chemically synthesized oligomers have been employed in the isolation of cloned sequences in basically two ways, either directly as hybridization probes or as primers for the synthesis of radiolabelled cDNA that is used as a probe. The use of short, synthetic ^{32}P-labelled oligomer probes has become the more widely accepted approach for two main reasons. Firstly, greater specificity can be obtained by the hybridization approach than by using the priming approach. A reluctance in using the latter approach presumably arose from the reported (Chan et al., 1979; Noyes et al., 1979) detection of cDNA species unrelated to the desired cloned mRNA, where it was speculated that these prominent cDNA species probably arose from an interaction between a non-specific RNA and the synthetic oligonucleotide primer, that was stabilized in the presence of reverse transcriptase. With regards to the direct hybridization approach, Wallace et al., (1981) have shown that under certain conditions

a single mismatched base-pair can be distinguished between oligonucleotide : polynucleotide duplexes. Secondly, with the priming approach, the amount of probe obtained is dependant on the amount of RNA available as template. In isolating cloned sequences for very low abundance mRNAs, use of the primer approach would necessitate large quantities of RNA to obtain sufficient amounts of cDNA probe.

In view of the above observations, the direct hybridization probe approach, as described in more detail later in this chapter, was employed to isolate wool HGT keratin cDNA clones. The identity of these oligomer-selected clones was determined by nucleotide sequence analysis of the cloned DNA using the "dideoxy" chain termination method of Sanger et al., (1977).

5.2 RESULTS

5.2.A Design of Synthetic Oligonucleotides Corresponding to HGT-F Keratin

Two wool proteins, components C2 and F, belonging to the HGT keratin class have been fully sequenced (Dopheide, 1973; Marshall et al., 1980). These two protein sequences, designated HGT-F and HGT-C2, are illustrated in Figure 5.1 and are characterized by their difference in size, amino acid composition and a low sequence homology despite both having a high glycine and tyrosine content. Even though a small number of gaps were introduced to optimize the alignment between the protein sequences shown in Figure 5.1, direct homology between components C2 and F is limited to small blocks of at most three residues in length.

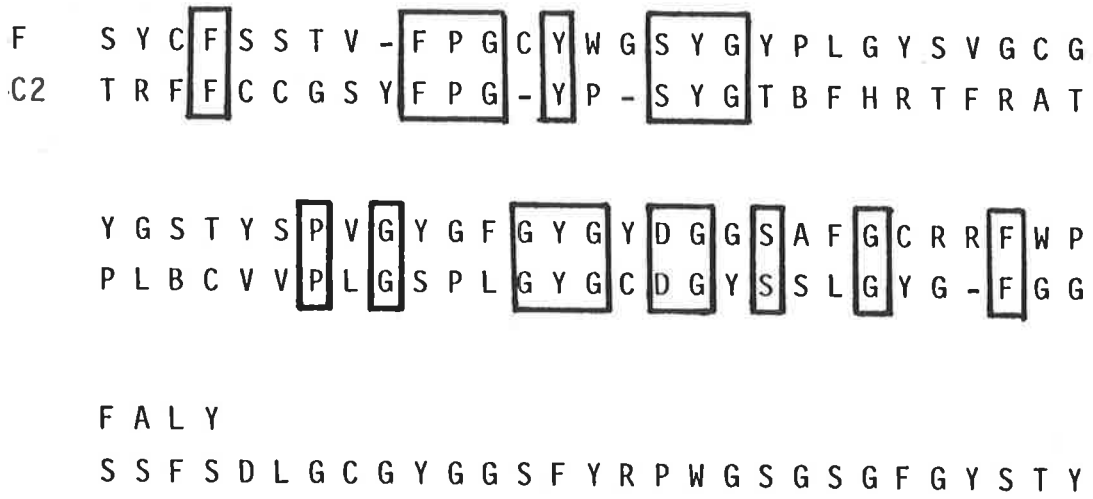
For the purpose of synthesizing a HGT keratin-specific oligonucleotide probe, and in view of the limited homology between the two known protein sequences, it was decided to concentrate efforts on component F which, as reported by Frenkel et al., (1973), constitutes the

FIGURE 5.1

AMINO ACID SEQUENCE HOMOLOGY BETWEEN THE TWO HGT WOOL PROTEINS

Complete amino acid sequences of two distinct HGT wool proteins are known (see 1.3.B) and are presented in the diagram opposite in the one-letter amino acid code. Component F (61 amino acids) is given on the top line and component C2 (84 amino acids) is given on the bottom line. Both end with a C-terminal tyrosine but each has a different N-terminal residue (F has a serine, while C2 a threonine), neither of which is acetylated.

In this figure, the two protein sequences have been aligned to give maximum homology with the minimum number of introduced gaps (indicated by a dash). The residues in common are boxed.


F S Y C F S S T V - F P G C Y W G S Y G Y P L G Y S V G C G
C2 T R F F C C G S Y F P G - Y P - S Y G T B F H R T F R A T
Y G S T Y S P V G Y G F G Y G Y D G G S A F G C R R F W P
P L B C V V P L G S P L G Y G C D G Y S S L G Y G - F G G
F A L Y
S S F S D L G C G Y G G S F Y R P W G S G S G F G Y S T Y

most abundant HGT protein in wool. Two synthetic oligonucleotide probes specific for the HGT-F protein were constructed as described below.

1. 14-mer mix oligonucleotide probe

This probe involved the synthesis of a set of oligomers complementary to all the possible coding sequences for a small portion of the HGT-F protein. As with most oligonucleotide probes, regions of the protein containing the least codon ambiguity were examined, with particular emphasis on methionine and tryptophan residues both of which are encoded by a single codon. Since the HGT proteins do not contain methionine residues, the stretch of amino acids surrounding the tryptophan residue at postion 56 was used for constructing a mixed-probe containing sixteen different sequences, each 14 nucleotides in length. See Figure 5.2. Within this mixture must be one sequence complementary to the DNA coding for that part of the protein, and which under stringent hybridization conditions will form a perfectly matched duplex whereas the other fifteen oligonucleotides will form mismatched duplexes. This mixed-probe method as developed by Wallace and co-workers (see Wallace et al., 1981; Reyes et al., 1981) has emerged over the last few years as a reliable technique for the detection of specific cloned sequences by colony hybridization (for example, Suggs et al., 1981).

2. 25-mer unique oligonucleotide probe

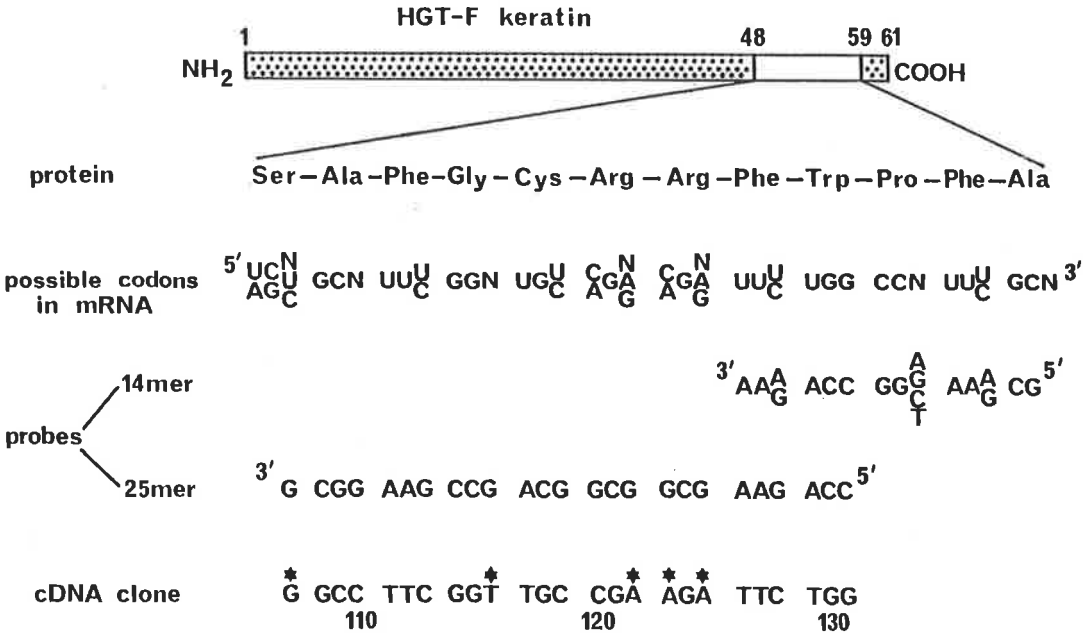
The second HGT-F specific probe extended from the trytophan residue at position 56 towards the amino terminal of component F and rather than contain a mixture of shorter oligonucleotides, which cover all possible arrangements of codons (as for the 14-mer mixed probe), it consisted of one long oligonucleotide sequence 25 bases in length. See Figure 5.2. By increasing the probe length it was hoped to increase specificity of the hybridization to HGT sequences and avoid unrelated hybridizations as might be seen with the 14-mer mixed probe. In designing

FIGURE 5.2

DESIGN OF THE HGT-F SPECIFIC OLIGONUCLEOTIDE PROBES

The sequence of the two oligonucleotide probes, designed and used as described in 5.2.A and 5.2.C, are presented. The 14-mer probe was a mixture of 16 oligonucleotides, while the 25-mer was an unique oligonucleotide sequence. N = A, G, C or U.

Mismatches between the 25-mer probe and the cloned HGT cDNA sequence are indicated by an asterisk and are discussed in section 5.3.


HGT-F keratin
1
48
59 61
NH2
COOH
protein
Ser–Ala–Phe–Gly–Cys–Arg–Arg–Phe–Trp–Pro–Phe–Ala
possible codons in mRNA
5′ UCN AGU C GCN UUU C GGN UGU C CGN AGA G CGN AGA G UUU C UGG CCN UUU C GCN 3′
probes
14mer
3′ AAA G ACC GGA G C T AAA G CG 5′
25mer
3′ G CGG AAG CCG ACG GCG GCG AAG ACC 5′
cDNA clone
G GCC TTC GGT TGC CGA AGA TTC TGG
110
120
130

this 25-base oligonucleotide probe the following parameters were considered :

(1) utilization of the unique tryptophan codon (2) codon usage in sequenced wool keratin genes (Powell et al., 1983) and (c) the relative stability of the mismatched G:T base pair. The codons UGC, GGC, GCC and CGC encoding respectively amino acids cysteine, glycine, alanine and arginine were selected in constructing the 25-mer, since they were used more frequently than their alternative degenerate codons in wool keratin genes (Powell et al., 1983). No clear preference was obvious for the tyrosine and phenylalanine codons, but it has been shown (Szostak et al., 1979) that the effects of possible mismatches were minimized by selecting G whenever there is a choice between A and G in the oligonucleotide probe; thus G was chosen in the probe sequence for the positions complementary to the third base of the tyrosine and phenylalanine codons, so the result will be correct base-pairing or a G:T mismatch.

5.2.B Characterization of Labelled Oligonucleotide Probes

Since the oligonucleotides were synthesized without 5' terminal phosphate groups, preparation of ^{32}P-labelled probes was a relatively simple task. The single-stranded oligomers were labelled at their 5' ends by transfer of ^{32}P from $[\gamma\text{-}^{32}P]$ ATP using polynucleotide kinase (see 2.2.H.4) in a small reaction volume (15µl) and unincorporated nucleotides removed by electrophoresing the complete reaction mix on high-percentage denaturing polyacrylamide gels (Figure 5.3). Each was sequenced according to the method of Banaszuk et al., (1982) and the sequences were confirmed on those shown in Figure 5.2.

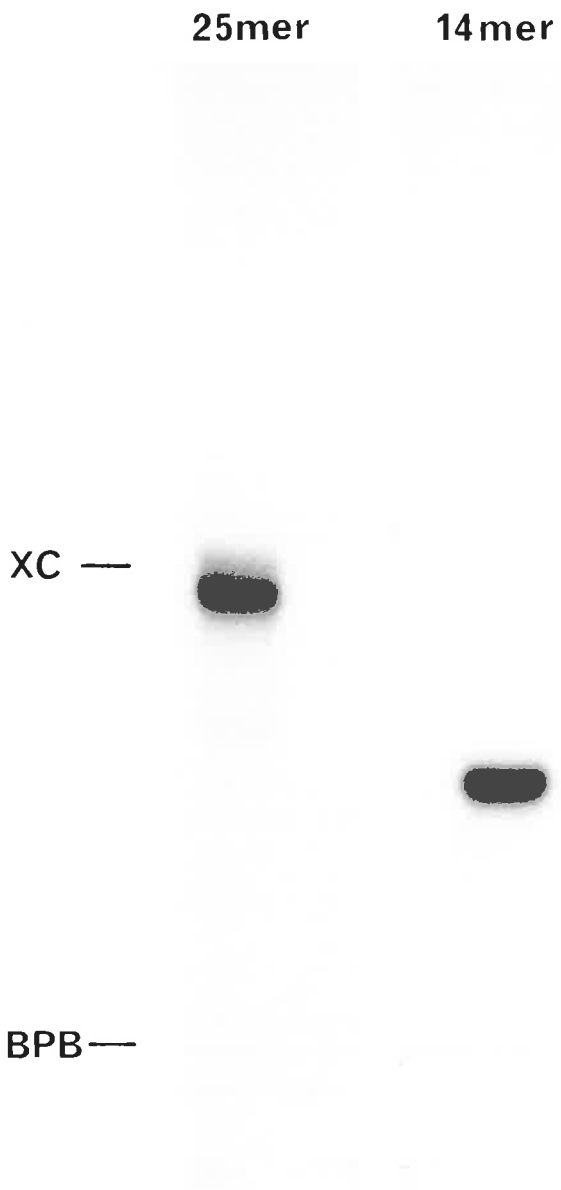
5.2.C Screening of the cDNA Library with the 14-mer and 25-mer Probes

The wool follicle cDNA library described in Chapter 3 was

FIGURE 5.3

POLYACRYLAMIDE GEL ANALYSIS AND ISOLATION OF ^{32}P-LABELLED OLIGONUCLEOTIDE PROBES

The oligonucleotides, whose sequences were intially confirmed by the sequencing procedure described in 2.2.Q, were prepared for use as probes in the following manner. The 14-mer and 25-mer were ^{32}P-labelled by kinasing (2.2.H.4) and unincorporated nucleotides removed by electrophoresis through a denaturing 20% polyacrylamide gel (2.2.L.2). Following very short exposure times (less than 10 seconds), the single bands as shown in the autoradiogram opposite, were excised and eluted (2.2.N.1), then concentrated by precipitation with carrier tRNA and ethanol.


25mer
14mer
XC
BPB

screened with the 14-mer mixed probe using hybridization and washing conditions as detailed in section 2.2.I.1. The initial result showed high background hybridization to all colonies caused by an unexpected match-up between pBR322 (the cloning vector) DNA and 12 consecutive nucleotides of one of the 14-mer sequences in the probe mix. In order to obtain a more selective hybridization and to avoid the type of background problem encountered with the 14-mer mix, the second probe, a 25-mer of potentially greater specificity due to its increased length and unique sequence was hybridized to the cDNA library.

Following screening of the cDNA clones with the ^{32}P-labelled 25-mer (hybridization and washing conditions as in section 2.2.I.1), four colonies gave strong signals, even after additional washes at 50°C. These colonies are shown in Figure 5.4 and were picked for further analysis of the cloned insert DNA.

Interestingly, on re-examination of the cDNA library screening results with the 14-mer mixed probe, these four clones were amongst about 20 colonies which gave hybridization signals above the high backgound level, suggesting that more stringent washing conditions may have distinguished between 14-mer positive clones and the 12-base pBR322 hybridization.

5.2.D Characterization of the Cloned cDNA for HGT-F Keratin by Southern Blot Analysis

Plasmid DNA was prepared (2.2.J.1) from the presumptive bacterial clones for HGT-F cDNA, designated pSHGT-1,2,3 and 4 (plasmid sheep high-glycine/tyrosine clone -1,2,3 and 4). The various DNAs were digested with HpaII and the Southern blot hybridized with the ^{32}P-labelled 25-mer (Figure 5.5). Only the insert fragments of plasmids pSHGT-1,2,3 and 4 hybridized with the probe (lanes 1,2,3,4). No hybridization was observed to the DNA from two 25-mer negative clones (lanes 5 and 6) or

FIGURE 5.4

SCREENING OF cDNA COLONY FILTERS USING THE ^{32}P-LABELLED 25-MER PROBE

The ^{32}P-labelled 25-mer was hybridized to nitrocellulose filters containing duplicate colonies of cDNA clones representing the whole cDNA library. Hybridization and washing conditions were as described in 2.2.I.1.

Only the filters showing the four positive colonies detected with this probe are given in the figure. These four clones are specified by their positions on each of the four filters A, B, C and D.

25-mer positive clones: A7H
B9D
C7A
D9C

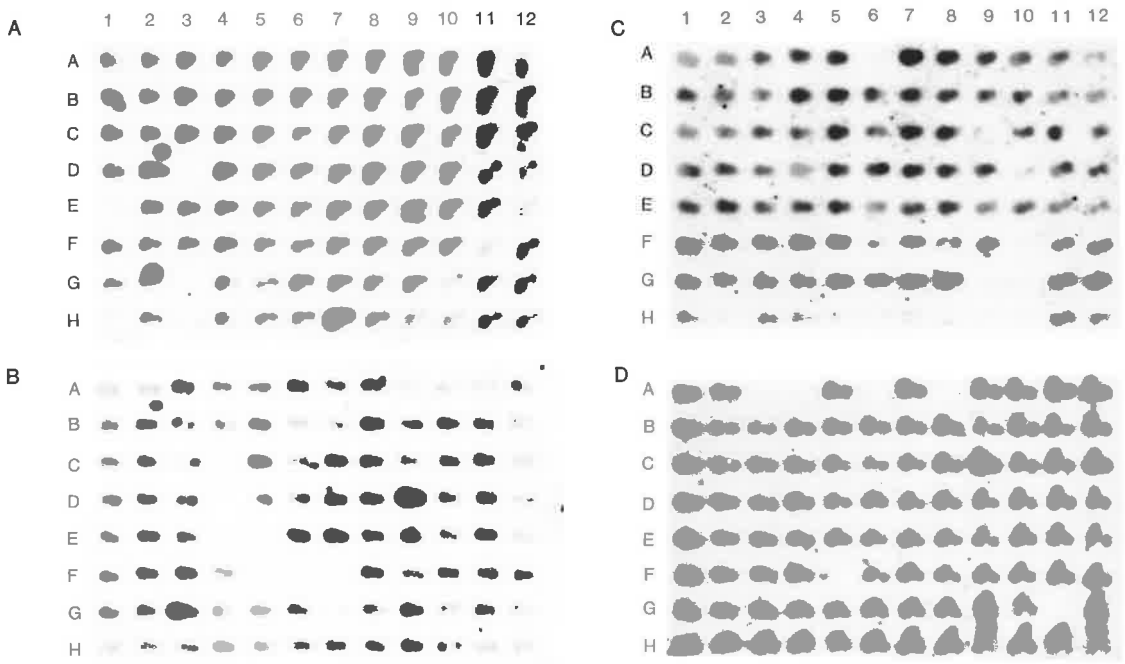

A
1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
B
A
B
C
D
E
F
G
H
C
1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
D
A
B
C
D
E
F
G
H

FIGURE 5.5

SOUTHERN BLOT ANALYSIS OF SELECTED cDNA CLONES

Plasmid DNA was purified (2.2.J) from the four 25-mer positive clones (see Figure 5.4) and digested with HpaII. The digestion products were electrophoresed on a 2% agarose gel (2.2.M.1), alongside HpaII digests of two 25-mer negative clones and pBR322 DNA. Following transfer to nitrocellulose (2.2.0), hybridization with the 25-mer probe was essentially as described in 2.2.I.1, except that it was carried out at 42°C overnight in the presence of 10% dextran sulphate. The blot was washed at room temperature in two changes of 6xSSC.

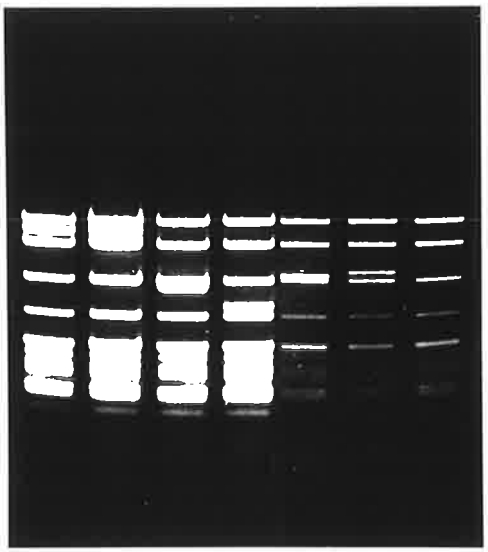
PART A : shows the ethidium bromide-stained gel of the HpaII-digested plasmid DNAs

PART B : shows the corresponding autoradiogram following hybridization to the 25-mer probe.

TRACKS :

1. pSHGT-1
2. pSHGT-2
3. pSHGT-3
4. pSHGT-4
5. clone p1C10
6. clone p7G10
7. pBR322 DNA

A. 1 2 3 4 5 6 7



B. 1 2 3 4 5 6 7

to vector pBR322 DNA (lane 7). These results confirmed the preliminary data obtained at the colony screen level.

5.2.E Sequence Analysis of Clones pSHGT-1,3 and 4

DNA from each of the four plasmid clones detected by the 25-mer probe was restricted with a variety of 4-base-specific endonucleases such as HpaII, RsaI, HaeIII and Sau3A, to generate suitable fragments for M13 cloning and "dideoxy" sequencing. Since the restriction patterns of pSHGT-1 and pSHGT-2 were identical in all cases (for example, see Figure 5.5A, lanes 1 and 2), these plasmids were judged to contain the same cloned cDNA sequence and therefore only pSHGT-1 was chosen for sequence analysis. The cDNA inserts in pSHGT-1,3 and 4 were approximately 425,220 and 154 bp long respectively and each contained a single RsaI site and two HaeIII sites (see Figure 5.6A). The two RsaI and the two large HaeIII cDNA-containing fragments were subcloned into M13mp19 SmaI-digested vector by blunt-end ligation (2.2.R.2) and sequenced by the "dideoxy" chain termination method of Sanger (1977).

All three clones contained overlapping sequences, which were identical, with 70 bp of 3' non-coding sequence common between pSHGT-1 and pSHGT-3. Due to the common identity of 3' non-coding regions, it was presumed that the sequences were derived from reverse transcription of the same mRNA, so the data has been combined and presented as one sequence, pSHGT-F (Figure 5.6B). The amino acid sequence derived from the DNA sequence in Figure 5.6B, corresponded to the sheep wool HGT keratin, component F (Dopheide, 1973). This combined data provided nucleotide sequence for 77% of the protein coding region and all of the 3' non-coding region of the HGT-F mRNA molecule.

5.2.F Size Estimation of HGT-F mRNA

1. Northern blot analysis

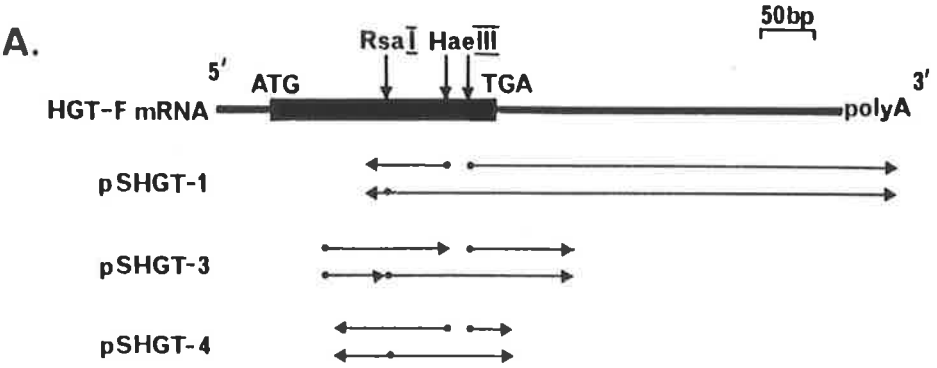
FIGURE 5.6.

SEQUENCING STRATEGY AND THE COMBINED SEQUENCE DATA FOR THE cDNA CLONES, pSHGT-1, 3 AND 4

PART A shows the combined map of the HGT-F cDNA as derived from the analysis of clones pSHGT-1, 3 and 4, as well as the sequencing strategy for each insert. The inserts of these clones were sequenced by the chain-termination method of Sanger et al., (1977), after subcloning into phage M13mp93 (Messing et al., 1981) of fragments produced by cleavage with HaeIII and RsaI restriction endonucleases. The direction and extent of sequencing from relevant restriction sites are indicated by the arrows.

PART B presents the combined sequence data from clones pSHGT-1, 3 and 4. The sequence complementary to the 25-mer probe is underlined, the termination and polyadenylation signals are boxed and discrepancies between the published protein and cloned cDNA sequence (discussed in 5.3) are indicated by the asterisks.

A.



B.

```
5'.....AC TGG GGC AGC TAT GGC TAC CCG CTG GGC TAC AGT GTG GGC TGT
         Trp-Gly-Ser-Tyr-Gly-Tyr-Pro-Leu-Gly-Tyr-Ser-Val-Gly-Cys
      αα14

GGC TAC GGT AGT ACC TAC TCC CCA GTG GGC TAT GGC TTC GGC
Gly-Tyr-Gly-Ser-Thr-Tyr-Ser-Pro-Val-Gly-Tyr-Gly-Phe-Gly

TAT GGC TAC AAC GGC TCT GGG GCC TTC GGT TGC CGA AGA TTC
Tyr-Gly-Tyr-Asn-Gly-Ser-Gly-Ala-Phe-Gly-Cys-Arg-Arg-Phe
           *       *  *

TGG CCA TTT GCT CTC TAC TGA TTTGCTGAAATACCAGAGGCATGGAAT
Trp-Pro-Phe-Ala-Leu-Tyr-Stop
      αα61

CTTCTCCCCAAACCCACGAGGCGGACTTCCAGGTCCTCAGAGACTCATCAGCCTC

CAGTTAGCTGCTTTTCACATCGGCACAGAGTCTCAAGGGAAGAAGATGAAAAACC

ACTTGCCTCCAGCTGCCTTCCTGCATGATGTTTGTTGGGACATTTTGAGAAACTT

GACACCCAAACACGTTTTACGTTTGAATTTTTCACCATGCTCATGACTCTTGTCA

TTATCAAGTTGTGGGATGTGTCTGTCAAATTCTCAATAAACTTGTCTCAACCGCA

TATTATAAAAAAAAAAAAAAAA.....3'
```

The size of the mRNA for the HGT-F protein encoded in pSHGT-1 was determined by Northern blot hybridization. When poly(A)$^+$ wool follicle RNA was denatured in glyoxal, electrophoresed on a 1% agarose gel, transferred to nitrocellulose then hybridized with the nick-translated insert from pSHGT-1, one discrete band was visible on the resultant autoradiogram (Figure 5.7). No other hybridizing bands were detected, even when low stringency washing conditions (2xSSC/0.1% SDS at 65°C) were employed, which indicated that no other closely-related RNA species were present in the wool follicle RNA population. The single RNA species detected was about 600 bases long. Allowing for a protein-coding region of 189 bases (including initiation and termination codons) and a 3' non-coding region of about 360 bases (307 bases of 3' untranslated sequence followed by a poly(A) tail with average length of 50 bases), indicated that the 5' non-coding region was approximately 50 bases long.

2. Primer-extension on follicle RNA

As further confirmation of HGT-F mRNA length, the synthetic 14-mer oligonucleotide mix was employed as primer in a cDNA-extension experiment. The 14-mer mix was used in preference to the 25-mer as a primer, since the latter generally gave a smear (the cause of which was unknown) with a few indistinct extended cDNA bands (data not shown). The 14-mer mix was annealed to both total poly(A)$^+$ RNA and HGT-enriched RNA (see Chapter 4) with the extension reaction carried out in the presence of [α -^{32}P] dCTP to label the cDNA products. The resultant autoradiogram, following electrophoresis of the cDNA products on a denaturing polyacrylamide gel, is shown in Figure 5.8. Complex patterns of transcripts were obtained which was not surprising considering the large number of potential primers in the 14-mer oligonucleotide mix. However, two features of these patterns indicated that specific priming had occurred in both instances. Firstly, both RNA samples gave essentially

FIGURE 5.7

HGT-F mRNA SIZE ESTIMATION BY NORTHERN BLOT ANALYSIS

Total poly(A)$^+$ RNA (7.5µg) isolated from sheep wool follicles was disaggregated by glyoxal treatment, followed by electrophoresis on a 1% agarose gel according to the methods of Thomas (1980). Sheep rRNA (28S and 18S) markers were co-electrophoresed in parallel and visualized by staining with ethidium bromide.

Following transfer to nitrocellulose, the blot was hybridized (2.2.I.2) to the nick-translated (2.2.H.3) PstI insert of pSHGT-1. A single band of estimated size, 0.6 kb was detected on the resultant autoradiogram.


HGT-F
mRNA
28S
18S

FIGURE 5.8

POLYACRYLAMIDE GEL ELECTROPHORESIS OF EXTENDED cDNA PRODUCTS USING THE 14-MER OLIGONUCLEOTIDE MIX AS PRIMER ON WOOL FOLLICLE RNA

Extended cDNA products were synthesized using the 14-mer oligonucleotide mix as primer on both total poly(A)$^+$ RNA (2.2.C.2) and 'HGT-enriched' mRNA (4.2.B) templates, in reactions composed and carried out as described in 2.2.H.6. Samples were electrophoresed on a denaturing 6% polyacrylamide gel (2.2.L.2) and the cDNA products visualized by autoradiography. HpaII-digested pBR322 DNA, which was ^{32}P-labelled by the end-labelling procedure (2.2.H.5), was included on the gel as a molecular weight marker. Fragment sizes are given in bp.

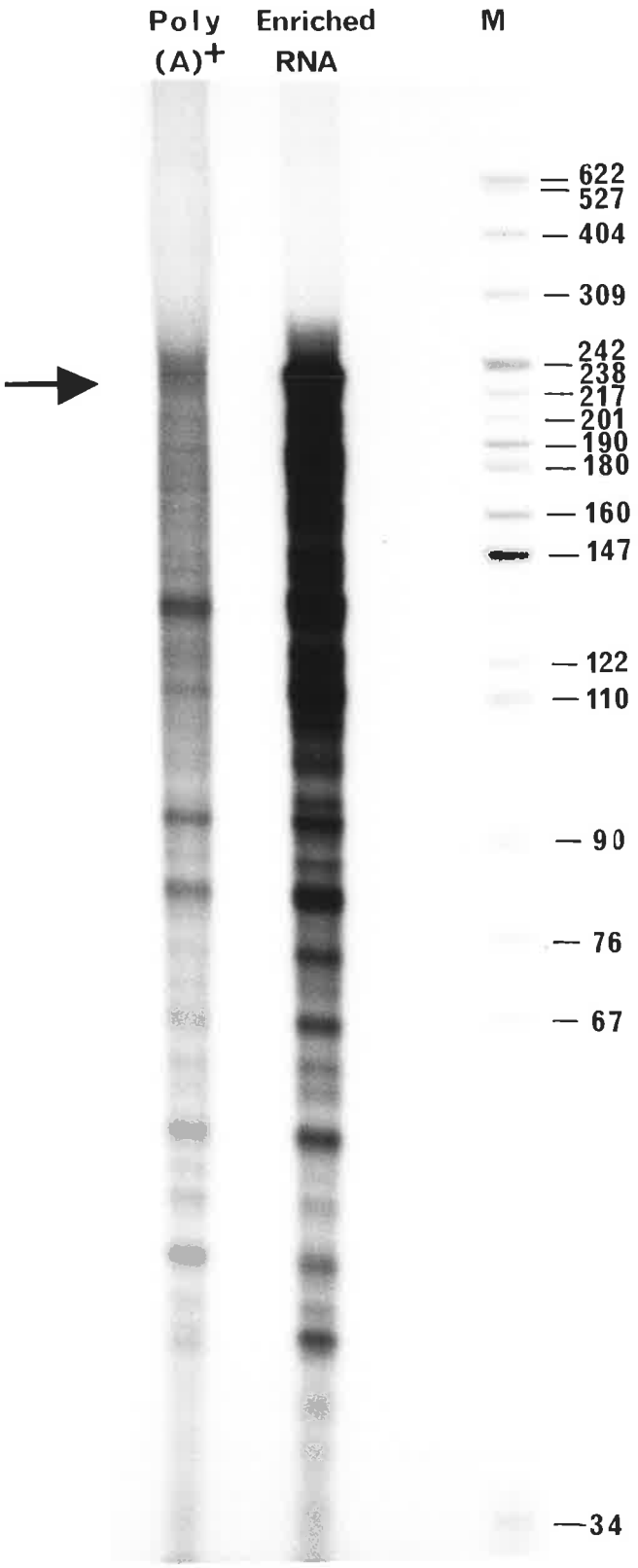
TRACKS :

Total wool follicle poly(A)$^+$ RNA-cDNA extension products.

'HGT-enriched' mRNA-cDNA extension products.

M. HpaII digest of pBR322 DNA.

The arrow indicates the position of the longest cDNA transcript obtained with both samples of RNA (discussed in 5.2.F.2).


Poly (A)+
Enriched RNA
M
622
527
404
309
242
238
217
201
190
180
160
147
122
110
90
76
67
34

identical patterns suggesting that the synthetic primer showed selectivity in hybridization and that the majority of products seen on the gel were most likely inter-related and possibly represented premature termination products due to secondary structure within the RNA. Secondly, the specific termination of transcripts in both tracks occured at a position (approximately 260 bases from the priming site) which was consistent with the 5' non-coding region being 45 to 50 bases long, as predicted by the Northern blot data (see part 1 above).

5.3 DISCUSSION

This chapter described the use of synthetic oligonucleotide probes specific for the HGT component F protein of wool, in detecting cloned HGT-F sequences by colony hybridization and by Southern blot hybridization to purified plasmid DNA.

Three plasmid clones, pSHGT-1, 3 and 4 were sequenced in their entirety and contained overlapping sequences which were identical and therefore derived from reverse transcription of the same mRNA. The amino acid sequence deduced from the combined DNA sequence obtained from these clones, corresponded to the published protein sequence for component F (Dopheide, 1973), except for three minor discrepancies which are indicated by the asterisks in Figure 5.6B. Firstly, at amino acid position 45 the DNA sequence codes for asparagine rather than aspartic acid; it is possible that the protein sequence is in error since amide assignments are difficult and the HGT proteins are susceptible to deamidation under the conditions used for their extraction and isolation (Frenkel, 1977). Secondly, at amino acid positions 47-48 there is a Ser-Gly pair, which is inverted (Gly-Ser) in the published protein sequence. All three amino acid discrepancies may reflect genetic polymorphism in the HGT-F component of wool arising from the different breeds of sheep used for the protein sequence work (Merino) and for the cDNA work (Merino-Dorset Horn cross).

However, as will be discussed in the next chapter, the sequence of the HGT-F gene isolated from a Merino genomic library also showed these same discrepancies with the protein sequence and therefore it is most likely that these amino acid differences are due to protein sequence errors.

The insert in pSHGT-1 encoded amino acid 29 through to the poly(A) tail of the HGT-F mRNA. pSHGT-3 prematurely terminated at nucleotide 220 but extended further in the 5' direction to include amino acid position 14. pSHGT-4 contained the smallest insert, beginning at amino acid position 16 and ending at nucleotide position 162. The 3' non-coding region of the HGT-F mRNA was 307 bases long, excluding the termination codon, TGA and poly(A) tail, with the putative polyadenylation signal, AATAAA (Proudfoot and Brownlee, 1976) at a distance of 21 bases from the poly(A) tail. This hexanucleotide sequence usually occurs 10-30 nucleotides upstream to the poly(A) tail of other eukaryotic mRNAs (Proudfoot and Brownlee, 1976). A function has been implicated for the AAUAAA sequence in poly(A) addition, either in formation of a recognition site, or processing of the RNA and recognition by the poly(A) polymerase. A number of exceptions to the AAUAAA homology exist and appear to function equally well. Some examples are AUUAAA in the chicken lysozyme mRNA (Jung et al., 1980), AAUAUA in rat pancreatic α-amylase mRNA (Tosi et al., 1981) and AAUUAA in anglerfish somatostatin II mRNA (Hobart et al., 1980). Apart from this putative polyadenylation signal there does not appear to be any other sequences common to most eukaryotic 3' non-coding regions (Baralle, 1983) and in fact the length of the 3' non-coding portion of eukaryotic mRNAs is highly variable. For example, the recently sequenced cDNA for the sodium channel protein of the electric organ of the eel (Noda et al., 1984) has a 1,386 nucleotide long 3' non-coding region, whereas that of mouse pancreatic α-amylase mRNA is only 30 nucleotides long (Hagenbuchle et al., 1980). Whether the length of the 3' non-coding region has a functional role in eukaryotic mRNAs had not been clearly

defined. Setzer et al., (1980) were unable to find any in vitro functional difference in four mouse polysomal dihydrofolate reductase mRNAs which differ only in the length of their 3' non-coding regions (Nunberg et al., 1980). And indeed removal of the 3' non-coding region did not effect the in vitro translation of rabbit β-globin mRNA (Kronenberg et al., 1979). However, in vitro studies (Gurdon et al., 1976; Huez et al., 1978) indicated that the poly(A) sequence appears to play some role in the stabilization of the mRNA in the cytoplasm, but is no longer thought to be associated with nuclear/cytoplasmic transport since RNA lacking poly(A) sequences can be transported at nearly normal efficiencies and enter into polyribosomes (Zeevi et al., 1982).

One puzzling feature of eukaryotic mRNA biogenesis concerns the exact location of the 3' terminus of the mRNA molecule. In each case, where it has been studied, transcription does not terminate at the poly(A) site but at some distance downstream (Nevins et al., 1980; Hofer and Darnell, 1981). More recently, Gil and Proudfoot (1984) suggested that a G+T-rich homology and a pentanucleotide sequence, CAYUG (where Y is a pyrimidine), downstream from the AATAAA sequence are required for rabbit β-globin mRNA 3' end formation. Both of these sequence elements have been found near the poly(A) site of a number of other genes (Taya et al., 1982; Berget, 1984) and the pentanucleotide sequence is complementary to regions of the RNA of the U4 small nuclear ribonucleoprotein, suggesting that U4 may mediate 3'-end processing. Therefore, the rule in poly(A)-containing mRNA formation may be that the poly(A) is a processing site and not a transcription termination site, but more cases need to be examined to establish the generality.

5' non-coding sequences were not present in the clones, pSHGT-1,3 and 4, and this can, in part, be explained by the nicking activity of S_1 nuclease used in the cloning protocol. The ability of single-stranded cDNA to form a hairpin structure at its 3' end was

exploited to prime second-strand DNA synthesis, and before inserting this double-stranded cDNA into the plasmid, pBR322, it was necessary to digest the single-stranded hairpin loop of the ds cDNA with S_1 nuclease. This invariably results in the loss of sequences corresponding to the 5' terminal region of the mRNA (Maniatis et al., 1976; Seeburg et al., 1977). The extent of 5' sequences lacking in these clones was greater than would be expected by the nibbling activity of the S_1 nuclease and since 3' terminal sequences were also missing in pSHGT-3 and pSHGT-4, the most likely explanation for these truncated cDNAs was that all three inserts resulted from an inefficient first-strand cDNA synthesis and clones pSHGT-3 and -4 were products of inefficient first-strand synthesis followed by partial second-strand synthesis. Although the wool follicle cDNA library was rescreened with the insert from pSHGT-1 (see below), the additional clones detected were not examined for insert length and therefore may have contained cloned sequences which extended into the 5' non-coding region of the HGT-F mRNA.

When the oligonucleotide probe sequence, used to isolate these clones from the cDNA library, was aligned with the corresponding region of the cDNA sequence, five mismatches were evident and are shown in Figure 5.2. Two of these, the terminal G:G (position 107) and the G:T mismatch (position 116) presumably contribute little to the destabilization during hybridization. However, the remaining G:A mismatches, more central in the probe sequence, may have decreased the effective probe length to 14 bases. The effect of these mismatches on signal strength and specificity as a function of hybridization and washing conditions were not rigourously tested.

To estimate the relative abundance of cloned HGT-F sequences within the wool follicle cDNA library, the insert from pSHGT-1 was resected, labelled with ^{32}P and used as a probe. Twenty colonies gave strong signals (data not shown) including the four which were detected by

the 25-mer probe and characterized by sequence analysis as described above. Since the signals from these colonies were strong, even after additional washes at 0.1xSSC/0.1% SDS (high stringency), the twenty clones were assumed to contain HGT-F cDNA sequences. These additional clones were not examined in any further detail and pSHGT-1 was used in all other transfer hybridization and screening experiments.

Based on (1) the weight of HGT proteins extracted as a proportion of the initial starting wool weight for the breed of sheep used in this work and (2) the contribution of component F to the total tyrosine-rich fraction of Merino wool estimated by Dopheide (1973), the proportion of clones detected by pSHGT-1 was almost twice that of the estimated abundance (about 0.45%) for the HGT-F component in the wool fibre. Unless the total follicle mRNA population was faithfully represented by the cDNA clones in the library, the proportion of presumptive HGT-F clones detected in this rescreening experiment, cannot be taken as a true indication of the abundance of component F in the fibre. If, as noted in section 4.3, a bias had occurred in the cloning of small molecular weight cDNA species, then the number of HGT clones in the library is likely be higher than the estimated 0.45%, with a correspondingly lower figure for the higher molecular weight cDNA species (in particular the low-sulphur sequences), as observed in 4.2.A.

In addition, the protein data needs to be considered with some caution, since only limited information is available for the proteins of the HGT family which have been difficult to study due to their insolubility below about pH 10. The proposed heterogeneity and relative abundance of different HGT components is as yet unresolved both for the Merino and the other less well studied breeds of sheep.

Although 5' non-coding sequences were not present in the isolated HGT keratin cDNA clones, the size of the HGT-F mRNA was estimated by Northern blot analysis to be about 600 bases long. Rather than

screening for longer cDNA clones or constructing new clones which would contain the 5' non-coding sequence absent from pSHGT-F, it was decided to use the characterized pSHGT-1 cDNA clone as a probe to a sheep genomic library and isolate the corresponding HGT gene. In this way, the complete primary structure of the HGT-F mRNA molecule could be determined and the gene sequence examined for the presence of possible consensus sequence elements involved in the expression of this gene in the wool follicle.

CHAPTER 6

HGT-F KERATIN GENE STRUCTURE

CHAPTER 6

6.1 INTRODUCTION

This chapter describes the isolation of the wool HGT-F gene from a sheep genomic library using the HGT-F cDNA clone (pSHGT-1) as a probe. The DNA sequence of the HGT-F gene and its flanking regions is presented and is compared with the relevant regions of other eukaryotic genes, in particular the high-sulphur wool keratin genes. The basis for this type of comparative study is the belief that gene sequences contain information, in addition to that required to code for a protein or functional RNA, which is concerned with the control of expression of a gene. Controlling sequences need not necessarily be conferred to non-coding regions, for example, selection for a particular codon and therefore tRNA population may play a significant role at the level of translation of the mRNA sequence into protein. Comparative analysis of sequences from genes which are not related and yet are expressed specifically, perhaps co-ordinately, in the one cell type may be useful in determining control regions of the gene which facilitate selective expression. Common sequences in the genes would therefore be candidates for control regions presumably indicating convergent evolution of essential sequences.

6.2 RESULTS

6.2.A HGT-F Constitutes a Unique Gene in the Sheep Genome

When nick-translated pSHGT-1 was hybridized to a sheep genomic DNA blot containing four different restriction digests of sheep DNA, a simple banding pattern was observed (Figure 6.1). The initial washing conditions, were low enough to allow detection of sequence-related DNA species - however no additional bands were visible which meant that other

FIGURE 6.1

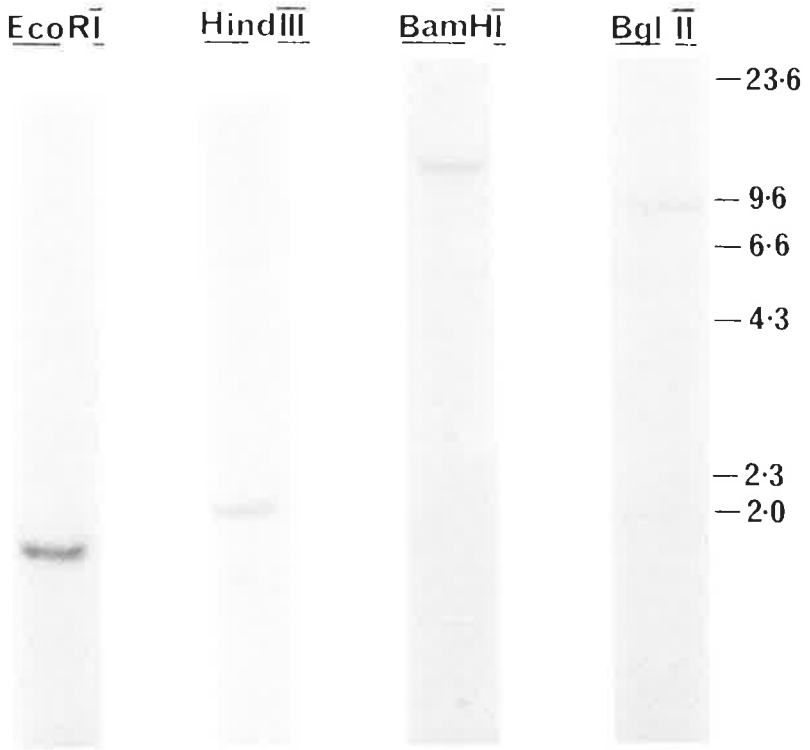
SHEEP GENOMIC DNA SOUTHERN ANALYSIS USING pSHGT-1 AS PROBE

Sheep liver DNA (10µg/track), prepared as described in 2.2.S, was digested with EcoRI, HindIII, BamHI and BglII, electrophoresed on a 1% agarose gel (2.2.M.1) and transferred to nitrocellulose. The filter was hybridized (2.2.I.2) to the nick-translated insert of pSHGT-1.

HindIII-digested DNA molecular weight markers were co-electrophoresed on the same gel and the fragment sizes in kb, are indicated on the right.

Single bands were detected in each track on the autoradiogram and the estimated band sizes were as follows:

HindIII	2 kb
EcoRI	1.7 kb
BamHI	14 kb
BglII	9.6 kb


EcoRI
HindIII
BamHI
Bgl II
23·6
9·6
6·6
4·3
2·3
2·0

HGT-F-like sequences were not present. The relevance of this Southern blot result to the complexity of the HGT keratin gene family will be discussed in Chapter 7.

6.2.B Isolation of the HGT-F Gene

1. Screening of the sheep genomic library

Approximately 4.5×10^5 pfu (of a λCh28 sheep genomic library kindly donated by Drs. P. Roche, P. Aldred and R. Crawford), that together contained approximately three genome equivalents, were lysed and bound to nitrocellulose filters by the method of Benton and Davis (1977), see section 2.2.T. The DNA on each of the duplicate filter sets was hybridized with the HGT-F cDNA probe (pSHGT-1) followed by low stringency washes at 2xSSC/0.1% SDS at 65°C. A single positive phage was identified (Figure 6.2A) and this recombinant phage, designated λsheep high-glycine/tyrosine-component F or λSHGT-F, was further plaque-purified through two more rounds of screening at low density (Figure 6.2B).

The isolation of only one genomic clone for HGT-F from such a large number of plaques was further evidence for the singular nature of this gene in the sheep genome, confirming the Southern blot result above (see 6.2.A).

2. Localization of the HGT keratin-coding sequence within λSHGT-F

The liquid culture method of Kao et al., (1982), as described in section 2.2.U, was used to prepare DNA from λSHGT-F. Digestion of the genomic clone with BamHI (which was compatible with the cloning site Sau3A used in construction of the genomic library) gave an insert fragment length of about 14 kb (Figure 6.3A). Digestion of λSHGT-F with EcoRI and HindIII generated a number of smaller fragments (Figure 6.3A). To try and locate the coding region of the gene to one of these smaller fragments,

FIGURE 6.2

DETECTION OF λ RECOMBINANTS CONTAINING SEQUENCES COMPLEMENTARY TO pSHGT-1

Approximately 4.5×10^5 pfu were screened in duplicate by the method of Benton and Davis (2.2.T), using the nick-translated insert of pSHGT-1 as probe. Part A shows duplicate filters (each containing approximately 2×10^4 pfu) containing the single recombinant phage detected by this probe. The final washing conditions were 2xSSC/0.1% SDS, 65°C.

Second and third round duplicate screenings at low density (100-200 pfu per filter) are shown in part B. The final washing conditions were 0.1xSSC/0.1% SDS, 65°C.

A.

B.

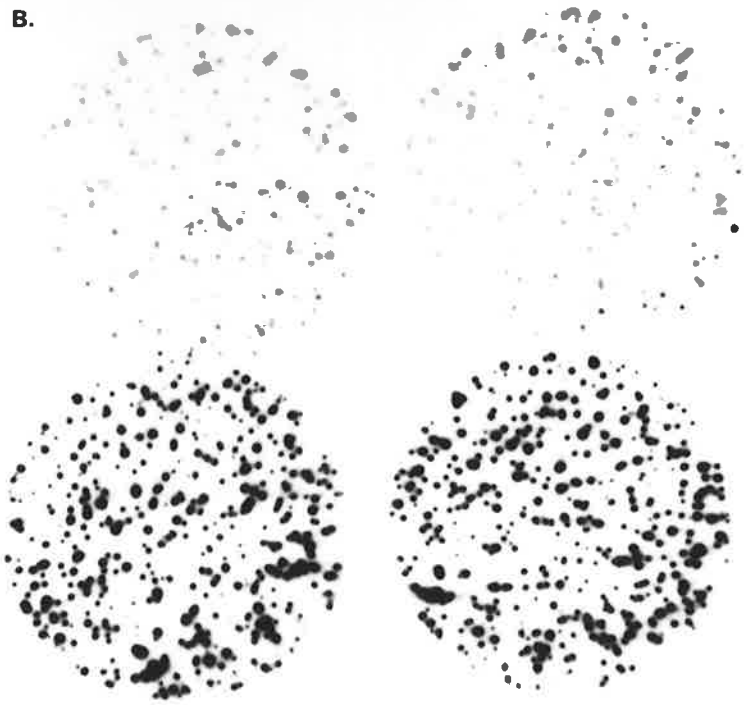


FIGURE 6.3

LOCALIZATION OF THE HGT-F GENE IN THE GENOMIC RECOMBINANT, λSHGT-F

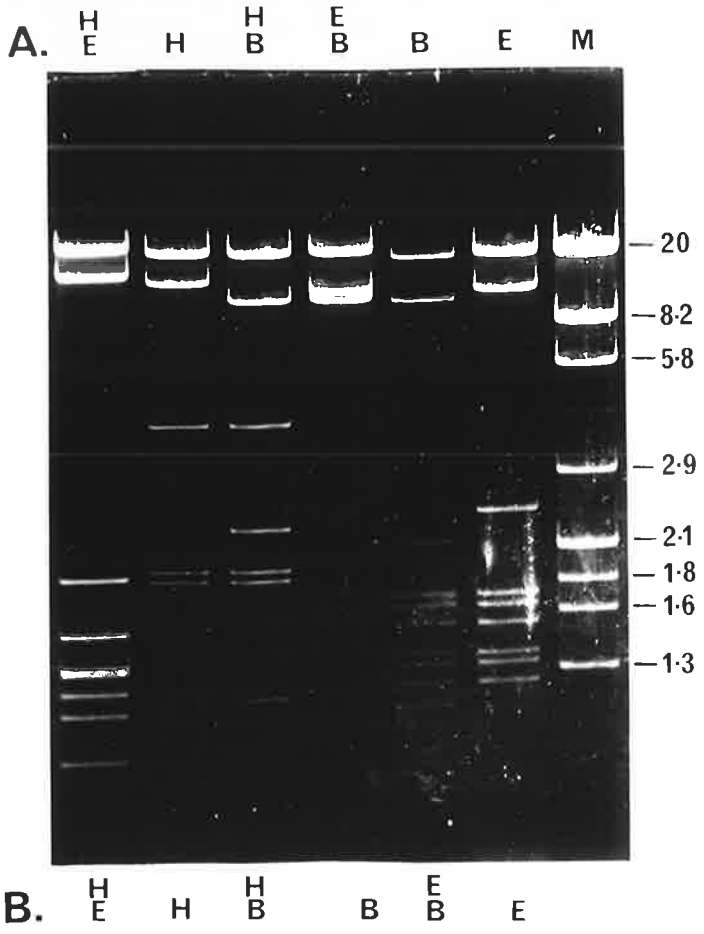
Phage DNA prepared (2.2.U) from the recombinant phage λSHGT-F (isolated as described in Figure 6.2), was digested with various combinations of the restriction endonucleases EcoRI, BamHI and HindIII. The digests were electrophoresed on a 1% agarose gel, alongside a molecular weight marker track (HindIII-digest of a characterized λ clone). The DNA was denatured in situ, transferred to nitrocellulose and hybridized to the nick-translated insert of pSHGT-1.

PART A : shows the ethidium bromide-stained gel of the various λSHGT-F digests and the molecular weight marker track.

PART B : shows the corresponding autoradiogram following hybridization to pSHGT-1. The sizes of the hybridizing fragments are indicated on the right and are in kb.

TRACKS :

H/E.	HindIII/EcoRI double digest
H.	HindIII digest
H/B.	HindIII/BamHI double digest
B.	BamHI digest
E/B.	EcoRI/BamHI double digest
E.	EcoRI digest
M.	Molecular weight marker


A.
H E
H
H B
E B
B
E
M
20
8·2
5·8
2·9
2·1
1·8
1·6
1·3
B.
H E
H
H B
B
E B
E
14
2·0
1·8

the gel was blotted and hybridized to nick-translated pSHGT-1 and the resulting autoradiogram is presented in Figure 6.3B.

In all tracks, single hybridizing bands were evident, which in the case of the BamHI, EcoRI and HindIII digests, were equivalent in size to the bands which hybridized in the sheep genomic blot (Figure 6.1). Given the small size of the HGT-F mRNA (0.6 kb as estimated in 5.2.F) the gene was most likely contained entirely within each of the small fragments. The 2 kb HindIII fragment was chosen for a more detailed analysis in preference to the other small hybridizing bands due to the relative ease with which it could be isolated free of the adjacent non-hybridizing fragments.

When a similar DNA blot of λSHGT-F was hybridized to total wool follicle ^{32}P-labelled cDNA, the pattern observed (data not shown) was essentially identical to that seen in Figure 6.3B. This result indicated that apart from the HGT-F gene, there was no other wool follicle keratin gene present on this genomic clone.

3. Subcloning of the 2 kb HindIII fragment from λSHGT-F

To facilitate the study of this 2 kb HindIII fragment, it was subcloned into pBR322 so that large amounts of DNA could be prepared relatively easily. Following transformation of E. coli (see section 2.2.V), recombinant colonies were screened by colony hybridization and plasmid DNA Southern blot analysis using pSHGT-1 as probe. This data is not shown but indicated that the pBR322 subclone, designated as pλSHGT-F, contained the HindIII fragment isolated from the λCh28 recombinant phage, λSHGT-F, and carried the HGT-F gene sequence.

6.2.C Sequencing Strategy and Analysis of pλSHGT-F

It was decided to sequence the entire insert within pλSHGT-F since it was expected to contain the complete HGT-F gene sequence as well as flanking DNA regions which could contain features important for the

expression of this gene. Rather than carrying out detailed restriction mapping of the insert in pλSHGT-F, use was made of the information provided by the cDNA clone, pSHGT-1, to choose relevant restriction enzymes which would generate cleavages within the mRNA coding region of the gene from which sequencing could be commenced. The DNA sequence was further extended by sequencing in from the HindIII ends of the insert and also from an EcoRI site about 200-300 bp in from one HindIII terminus, as implied from the Southern blot result in Figure 6.3 (see EcoRI/HindIII track). The cloning events for each specific restriction fragment into the appropriate M13 phage vector are detailed in the legend to Figure 6.4, with the sequencing strategy shown in part A of that figure. Where possible fragments were sequenced in both directions, or more than once in the same orientation, and restriction sites used for cloning were read through by sequencing begun at other restriction sites.

A diagrammatic representation of the sequence obtained from pλSHGT-F, showing the features relevant to the discussion below, is presented in Figure 6.4B, while the complete nucleotide sequence is presented in Figure 6.5.

Numbered from 1 to 2028 from the 5' end, the insert in pλSHGT-F (Figure 6.5) contained 837 bp of 5' flanking DNA upstream to the conserved 'CAAT' box (see discussion below) and 548 bp of 3' flanking DNA downstream from the presumed polyadenylation site. The HGT-F gene sequence, including putative 5' regulatory signals, is enclosed within a box and spans nucleotides 838 to 1480. A detailed description of the gene sequence is given below.

6.3 DISCUSSION

The Keratin-coding Region

The keratin protein-coding region in pλSHGT-F encoded the 61 amino

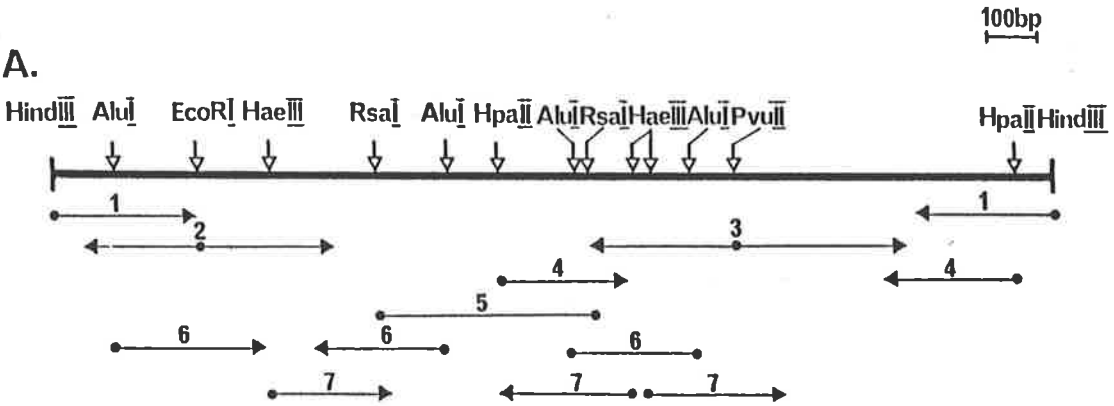
FIGURE 6.4

SEQUENCING STRATEGY AND A STRUCTURAL OVERVIEW OF THE DNA SEQUENCE WITHIN THE SUBCLONE, pλSHGT-F

PART A shows the sequencing strategy for the HindIII fragment of λSHGT-F which hybridized to the HGT cDNA clone, pSHGT-1 (see Figure 6.3). Prior to sequencing, this ≃ 2 kb fragment was subcloned into pBR322 and the subclone subsequently designated, pλSHGT-F. Extent and direction of sequencing is indicated by the arrows, and where possible DNA sequence was obtained in both directions. The numbers in parentheses indicate the approach used for sequencing each fragment as described below:

(1) Sequencing initiated at HindIII sites after subcloning into M13mp93 HindIII vector
(2) Sequencing initiated at EcoRI sites after subcloning into M13mp93 EcoRI/HindIII vector
(3) Sequencing initiated at PvuII sites after subcloning into M13mp93 PvuII/HindIII vector
(4) Sequencing initiated at HpaII sites after subcloning into M13mp93 AccI vector
(5) Sequencing initiated at RsaI sites after subcloning into M13mp93 SmaI vector
(6) Sequencing initiated at AluI sites after subcloning into M13mp93 SmaI vector
(7) Sequencing initiated at HaeIII sites after subcloning into M13mp93 SmaI vector

PART B is a diagrammatic representation of the structural features of the HGT-F gene and its flanking regions as derived from the nucleotide sequence in pλSHGT-F (presented in full in Figure 6.5). A detailed description of the structural features are given in the text.


A.
100bp
HindIII AluI EcoRI HaeIII RsaI AluI HpaII AluI RsaI HaeIII AluI PvuII HpaII HindIII
1
2
3
4
5
6
7

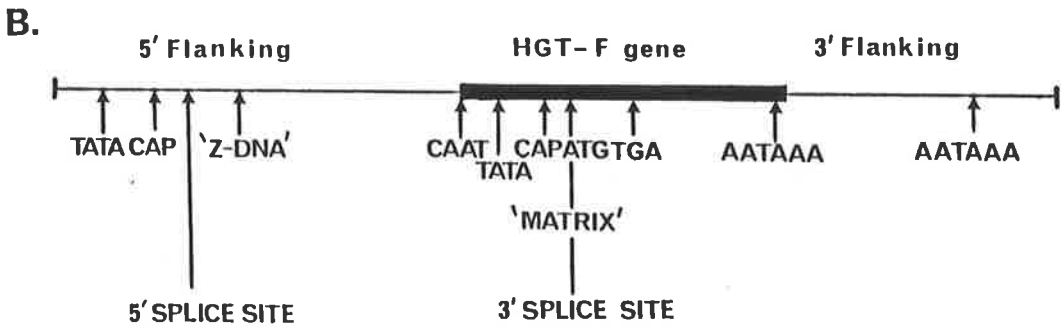

B.
5' Flanking
HGT-F gene
3' Flanking
TATA CAP
'Z-DNA'
CAAT
TATA
CAP
ATG
TGA
AATAAA
AATAAA
'MATRIX'
5' SPLICE SITE
3' SPLICE SITE

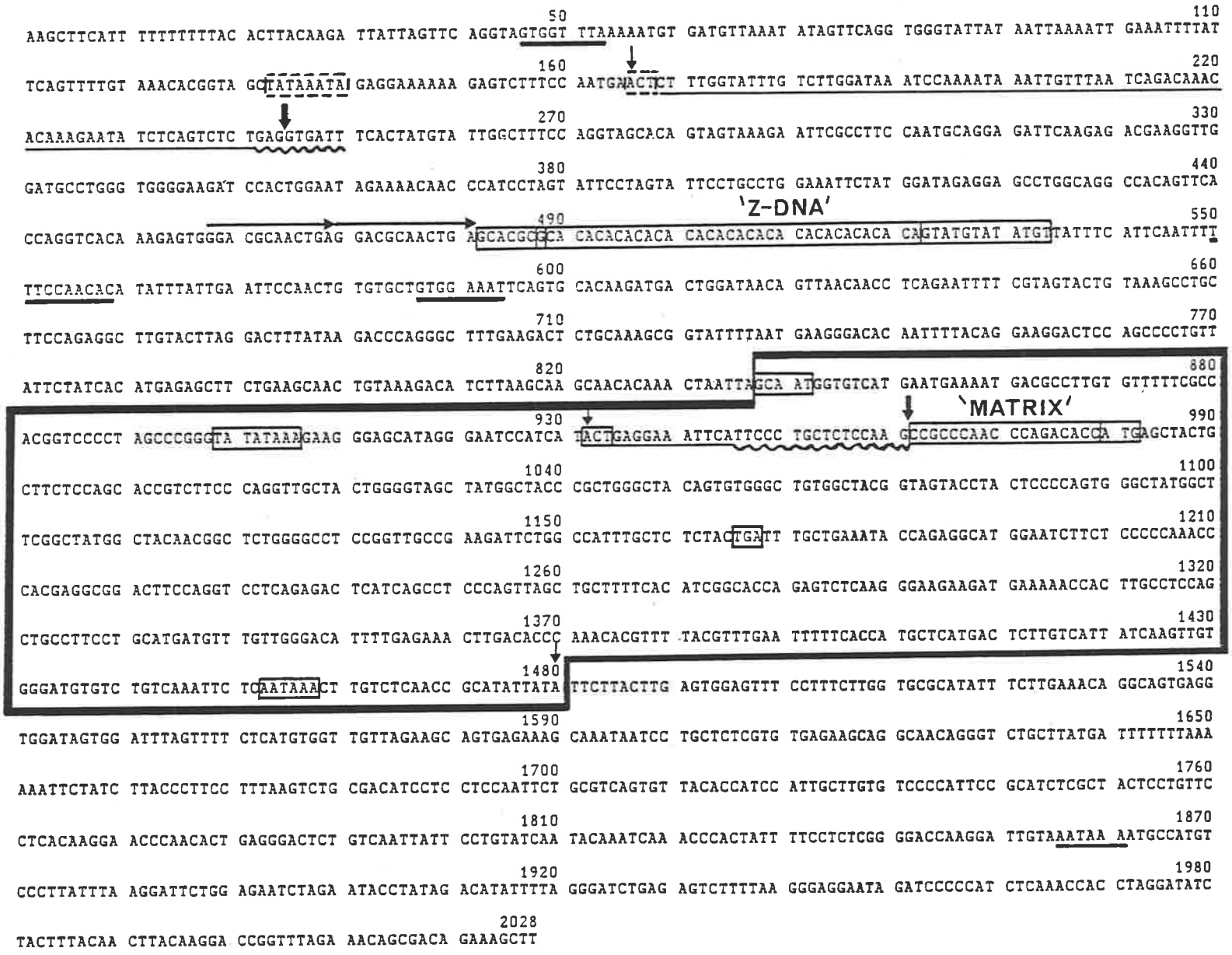
FIGURE 6.5

THE NUCLEOTIDE SEQUENCE OF THE HGT-F GENE ANT ITS FLANKING REGIONS

The nucleotide sequence of the mRNA strand of the HGT-F gene and its 5' and 3' flanking regions is presented in the 5'→3' direction. The DNA sequence which encodes the HGT-F gene is enclosed within the large box. The consensus 'CAAT', 'TATA', AATAAA sequences are boxed, as are the initiation (ATG), termination (TGA) codons, the possible cap site sequence (ACT) and the conserved 'matrix-gene-specific' element. The 5' and 3' residues presumed to be the sites for capping and poly(A) tail addition respectively, are indicated by a small vertical arrow. The wavy line immediately upstream to the 'matrix' element corresponds to the 3' splice site, with its 5' counterpart found in the 5' flanking region, also shown by a wavy line. The likely positions of the splice junctions are indicated by the large vertical arrows.

In the 5' flanking region of the HGT-F gene, upstream to the 5' splice site mentioned above, a cap site (hatched box and arrow) and a 'TATA' sequence (hatched box) are present. A putative 5' non-coding region is shown by a thin line extending between the cap and 5' splice site. Direct comparison of this sequence with the corresponding region in the HGT-F gene is given in Figure 6.8. Other interesting features include three viral 'enhancer-like' sequences (solid bars) and two 12 bp direct repeats (horizontal arrows) which directly precede a sequence capable of Z-DNA formation (boxed).

In the 3' flanking region of the HGT-F gene a second polyadenylation signal, AATAAA is underlined by a solid bar.


'Z-DNA'
'MATRIX'

amino acid long HGT-F protein, and provided the nucleotide sequence for the remainder of the NH_2-terminal region of component F not present in the cDNA clone, pSHGT-1, used to screen for the genomic clone. The gene sequence was identical to that of pSHGT-1 and contained the same amino acid discrepancies observed between the published protein sequence for HGT-F and the cDNA clone (see section 5.3). This suggested that since two different sources were used in determining the nucleotide sequences (Merino for the genomic DNA, Merino-Dorset Horn cross for the cDNA sequence), then the most likely explanation for the discrepancies was that of protein sequence error, rather than genetic polymorphism between HGT-F protein components of the wool from the two different breeds of sheep.

3' Non-coding Sequences

The 3' non-coding region extended for about 300 bp after the termination codon, TGA (position 1166), through the polyadenylation signal AATAAA at position 1453, which was 21 bp upstream to the presumed polyadenylation site, shown by an arrow at position 1480 (Figure 6.5). This site was identified by location of poly(A) residues at the 3' end of the mRNA sequence obtained from pSHGT-1 (see previous chapter, section 5.3). It is uncertain whether the A residue at position 1480 is transcribed from the genomic DNA and forms the first nucleotide of the poly(A) tail, but this is unlikely to be critical for correct 3'-end formation since exact sequences at the polyadenylation site are not important (Gil and Proudfoot, 1984). A second polyadenylation signal with the same hexanucleotide sequence was observed almost 400 bp downstream from the first site. Multiple polyadenylation signals and processing sites have been reported elsewhere (Moore and Sharp, 1984; Chin et al., 1984) but the Northern blot data in section 5.2.F and presence of a poly(A) tail downstream from the AATAAA sequence at position 1453 is consistent with this being the functional polyadenylation signal utilized

in the follicle cells. Whether the second site is used in another tissue where HGT proteins are found, for example hoof and horn, is not known.

Comparision of the 3' non-coding region of the HGT-F gene with the corresponding regions of the sequenced high-sulphur (HS) wool keratin genes showed no homology with the four conserved sequence elements present in these other matrix-protein-coding genes (Powell et al., 1983).

3' Flanking Sequences

Comparison of the 3' flanking DNA region with the HS keratin genes and with a small number of other eukaryotic genes showed no significant sequence homology or any other distinctive structural features apart from the putative second polyadenylation site mentioned in the above section.

5' Non-coding Sequences

The 5' non-coding sequence of the HGT-F mRNA was determined from the genomic DNA sequence since the corresponding region was not present in the cDNA clones analyzed in Chapter 5. It had a length of 48 bases which was consistent with the estimate made from the Northern blot and primer-extension experiments (section 5.2.F). The cap site at position 932 was identified by, using the length determined for the mRNA 5' non-coding region, the limited sequence homology with other capping sites and its position relevant to the 'TATA' sequence (see section below) further upstream. As pointed out by Sadler et al., (1983), the cap site generally begins with an A and is followed by a C and one or more T's within the next five bases, so the 5'-ACT-3' sequence boxed and marked with an arrow in Figure 6.5 was assumed to be the HGT-F mRNA cap site, since it was situated 25 nucleotides downstream from the 'TATA' box, a characteristic distance found between the 'TATA' sequence and capping site of many eukaryotic genes (Corden et al., 1980).

The sequence surrounding the ATG triplet at position 980 agrees well with the favoured sequence that flanks functional initiation

codons in eukaryotic mRNAs (Kozak, 1984), that is, 5'-CC$^{A}_{G}$CCAUG(G)-3' (where (G) is only semiconserved), however the 5'-terminal C and the 3'-terminal G of the consensus sequence are replaced by A residues in the HGT-F sequence. Since only 10 out of the 211 eukaryotic mRNA sequences examined by Kozak (1984) conformed perfectly to the consensus sequence, it is unlikely that the mismatches found in the HGT-F sequence will prevent it from serving as a functional initiation codon.

Comparative sequence analysis of the 5' non-coding region of the HGT-F gene with six HS keratin genes revealed a highly conserved homology block immediately upstream to the initiation codon, ATG. The conserved element of 18 bp, as shown in Figure 6.6, was present in three B2 and two BIIIB HS keratin genes but surprisingly did not occur in the BIIIA gene sequence which represents a third sub-family of HS keratins (Powell et al., 1983; Powell and Rogers, 1985). This conserved sequence is not found in either the sheep globin (Kretschmer et al., 1981) or the sheep metallothionein gene (J. Mercer, personal communication) and therefore may be specific for sheep keratins. When the relevant 5' sequences of the low-sulphur keratin genes become available, it will be of interest to determine whether this same homology is conserved or if, as presumed from the present data, the 18 bp conserved element is specific for matrix-protein coding genes. Its proximity to the initiation codon may suggest a regulatory role in the translation of these mRNAs, as opposed to a role in the control of transcription of these genes, however, further experiments need to be carried out to elucidate the function of this highly conserved sequence.

5' Flanking Sequences

'TATA' and 'CAAT' sequences

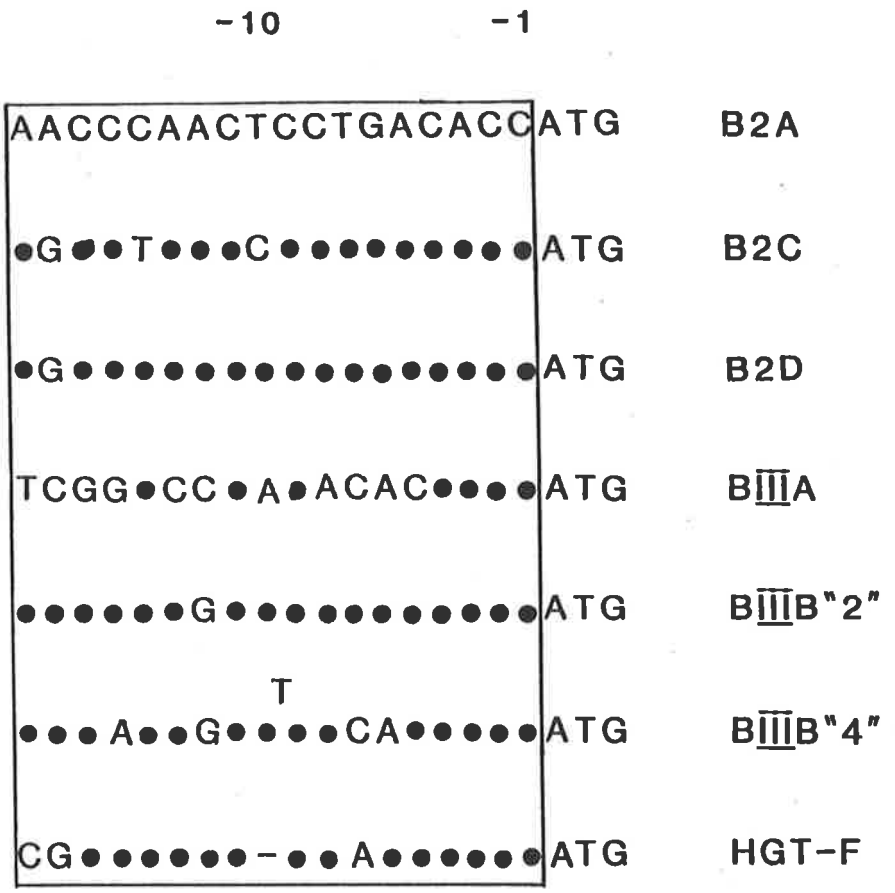
DNA sequence comparisons in the 5' regions of a number of genes transcribed by pol II (protein-coding genes) have identified two promoter

FIGURE 6.6

COMPARISON OF THE 5' NON-CODING REGIONS OF SIX SHEEP HIGH-SULPHUR AND ONE HIGH-GLYCINE/TYROSINE (HGT) KERATIN GENE

A comparison of the first 18 nucleotides of six high-sulphur and one HGT keratin gene is presented. Nomenclature for the high-sulphur genes is as given in Table 1.1, with "2" and "4" representing different members of the same protein sub-family. Nucleotides are numbered backwards from the initiation codon, ATG, with a dot indicating nucleotide homology, a dash indicating a deletion and nucleotides above the line representing insertions at that point.

Reprinted with permission from Powell and Rogers (1985).


-10
-1
AACCCAACTCCTGACACCATG
B2A
●G●●T●●●C●●●●●●●●ATG
B2C
●G●●●●●●●●●●●●●●ATG
B2D
TCGG●CC●A●ACAC●●●●ATG
BIIIA
●●●●●●G●●●●●●●●●●ATG
BIIIB"2"
T
●●●A●●G●●●●CA●●●●●ATG
BIIIB"4"
CG●●●●●●-●●A●●●●●●ATG
HGT-F

elements (for review see Breathnach and Chambon, 1981). The 'TATA' box (Goldberg, 1979), with the sequence 5'-TATA$^{T}_{A}$AAG-3' occurs 25-30 bp upstream to the transcription initiation or cap site. Although the 'TATA' box appears to be related in sequence to the Pribnow box of prokaryotic promoters (Pribnow, 1975), it clearly differs in that deletion of the TATA sequence does not abolish in vivo and in vitro transcription (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Grosveld et al., 1982a,b). These studies showed that in its absence, initiation occurs at several sites, so the function of 'TATA' sequence appears to be in the accurate positioning of the start of transcription.

The second conserved element, the 'CAAT' box (Benoist et al., 1980; Efstratiadis et al., 1980) with the consensus sequence, 5'-GG$^{T}_{C}$CAATCT-3', occurs 70-80 bp upstream from the start site of transcription and acts a modulator of gene expression as evidenced by the decrease in the efficiency of transcription following constructed mutations in and around the 'CAAT' box of the rabbit β-globin gene (Jones et al., 1982) and sea urchin histone gene (Grosschedl and Birnstiel, 1980). The 'TATA' and 'CAAT' sequences of the HGT-F gene are found 25 bp and 89 bp respectively, upstream to the mRNA start (Figure 6.5) and therefore correlate well with the location of these sequences in most of the eukaryotic genes studied thus far (Corden et al., 1980; Efstratiadis et al., 1980).

In addition to these two well-documented 5' regulatory signals, a number of other interesting DNA structures were observed for which functional roles have been postulated in other gene systems but whether they are utilized in the same manner for the expression of the HGT-F gene is not known.

Distal regulatory sequences

Sequences essential for transcription have been found even further

upstream to the 'TATA homology' associated with structural genes of a broad range of eukaryotic organisms including :

yeast	-	cyc1 (Guarante and Mason, 1983)
Drosophila	-	hsp70 (Pelham, 1982)
silkworm	-	fibroin (Tsuda and Suzuki, 1981)
sea urchin	-	histone H2B (Grosschedl and Birnstiel, 1980)
mammals	-	rabbit β-globin (Dierks et al., 1981)
		human α-globin (Mellon et al., 1981)
		mouse metallothionein (Brinster et al., 1982)
animal viruses	-	SV40 T-antigen (Fromm and Berg, 1982)
		HSV tK (McKnight et al., 1984)

These so called "distal signals" (McKnight et al., 1984) localized between 50 and 400 base pairs upstream to their respective mRNA cap sites, have been shown for most of the cases cited above, to play positive regulatory roles in the transcription initiation process. Although these signals require proximity to the transcription start sites it is unlikely that they represent the primary RNA polymerase binding site since various in vitro transcription studies have shown that accurate transcription can occur on truncated structural genes that retain a 'TATA' box but lack distal signals (see review by Breathnach and Chambon, 1981).

Some of the more extensively studied distal signals include the four sequences located 10-270 base pairs up from the 'TATA' box of the human metallothionein gene (Karin et al., 1984) and which consist of two metal ion-responsive elements, one hormone-reponsive element and one element containing alternating purine/pyrimidine residues - the latter element being required for the basal expression of the gene. The two distal signals required for efficient transcription of the tK gene of the Herpes simplex virus (McKnight et al., 1984) are located 47-105 base pairs upstream to the 'TATA' box. The first signal contains the hexanucleotide 5'-CCGCCC-3', and the second signal the inverted complement of this sequence, 5'-GGGCG-3'. From these studies with the tK gene, McKnight and co-workers suggested that a positive effect on RNA polymerase II

initiation could occur as a consequence of the action of a sequence-specific DNA binding protein that recognizes the two individual halves of the inverted repeat. Although the hexanucleotide sequences have yet to be shown to represent transcription recognition sites in the tK gene, these workers speculate that the expression level of the gene is likely to vary according to position, orientation and reiteration of the signals. Recently, Dynan and Tjian (1983), have identified a transcription factor, termed SP1, that binds specifically to a site upstream of the SV40 promoter region which contains six tandem copies of the 5'-CCGCCC-3' hexanucleotide and is known to be a critical component of the SV40 "early" promotor (Everett et al., 1983). These, hexanucleotide sequences are also present upstream to the metal ion-responsive elements in the human metallothionein gene (Karin et al., 1984),in the 5' presumptive promoter regions of the chicken histone H5 and chicken liver ALA-synthetase genes (L. Coles and I. Borthwick, personal communications) and the "housekeeping" HMG CoA reductase gene (Reynolds et al., 1984) which lacks a characteristic'TATA' and'CAAT'box. The function of the hexanucleotide sequences in these genes has not been established. Interestingly, the hexanucleotide, 5'-CCGCCC-3', was found in the 5' non-coding region of the HGT-F gene and formed the 5' terminus of the 18 bp conserved sequence discussed earlier (Figure 6.5). The same hexanucleotide sequence has also been found downstream to the 'TATA' box of the chicken liver ALA-synthetase gene (I. Borthwick, personal communication) which, unlike the HGT-F gene, contains multiple repeats of this sequence upstream to the 'TATA' box as well. Therefore the finding of the 5'-CCGCCC-3' sequence in the HGT-F gene is probably fortuitous and is unlikely to play a similar role to that described for the SV40 "early" promoter (Everett et al., 1983).

Enhancers

The positional dependance of distal signals to the transcription start site distinguishes them from "enhancers" (for review see Khoury and Gruss, 1983). These DNA segments, typified by the SV40 "21 bp repeats" are able to enhance transcription of viral genes and of heterologous cloned genes by acting in a cis fashion over large distances (up to 10,000 bp) irrespective of their position and orientation. The precise DNA sequence requirement for enhancer activity is unclear but a 'core' sequence of 5'-GTGG$^{AAA}_{TTT}$G-3' (Khoury and Gruss, 1983), is found in several viral enhancers and is essential for activity in SV40. Furthermore, sequences partially homologous to this element are found in the hormone-responsive element of mouse mammary tumour virus (Hynes et al., 1983) and variations of the sequence, 5'-TTTCCACACC-3' (found near the 'core' enhancer sequence of SV40), have been located within the polyoma virus enhancer (DeVilliers and Schaffner, 1981) and more recently, as part of the adenovirus internal transcription control region (Osborne et al., 1984). In this latter example, the DNA sequence required for transcription is located almost 400 nucleotides from the cap site, within the protein-coding region of the gene. The precedent of internal transcription control regions already exists for genes transcribed by RNA polymerase III (Bogenhagen et al., 1980) while more recent findings suggest the presence of transcriptional enhancers located well within the transcribed region of the immunoglobulin (Banerji et al., 1983) and globin genes (Charnay et al., 1984; Wright et al., 1984). Thus the existence of the cellular homologues (see also Conrad and Botchen, 1982) to the viral sequences seems to imply that enhancer elements may be essential for tissue-specific expression of a variety of eukaryotic genes.

Three short sequences highly homologous to known enhancers were found in the 5' flanking region of the HGT-F gene. The first of these, at position 550-560, conforms in 9 out of 11 bases with the consensus SV40 'core' enhancer element (Khoury and Gruss, 1983), while

located 26 bases further downstream is a sequence with 7 out of 8 matches with the mouse mammary tumour virus (MMTV) hormone-responsive element (Chandler et al., 1983). A similar MMTV-like sequence was also found much further upstream, at position 46-53. Whether these sequences have a function in the expression of the HGT-F gene remains to be elucidated.

Repetitive sequences

Examination of related genes at the gross level of gene family organization where gene-gene distances are a few to tens of kilobases apart, has shown that repetitive DNA sequences occur in intergenic regions as well as in introns (Wahli et al., 1981; Nowock and Sippel, 1982). It is now well established that repetitive DNA sequences, generally 200-300 bp in length and present in many copies, are ubiquitous components of most eukaryotic genomes (Stumph et al., 1981). A sheep-specific highly repetitive 110 bp repeat, first observed within the flanking regions of the HS keratin genes (B. Powell, personal communication), was also present in the HGT-F gene 5' flanking region stretching from nucleotide 336 to 446 (Figure 6.5). The function of repetitive sequences in eukaryotic genomes is unclear, although it has been postulated that they are involved in the regulation of gene expression at the level of transcription and/or RNA processing (Davidson and Britten, 1979).

Potential Z-DNA structures

A singular type of repetitive sequence is the simple sequence repeat (Tautz and Renz, 1984), most notably alternating purine/pyrimidine stretches, such as $d(CA)_n$ where n is usually less than fifty. The d(CA) dinucleotide repeat has attracted interest since it has potential to form Z-DNA or a non-B, left-handed DNA conformation (for review see Rich et al., 1984 and references therein). The ubiquitous occurrence of potential Z-DNA elements has led to many proposals for the biological roles of this type of structure such as : (1) a positive regulatory transcription signal

associated with the SV40 enhancer (Khoury and Gruss, 1983), and with the human metallothionein gene promoter (Karin et al., 1984) (2) a negative regulatory signal for RNA polymerase III transcription (Hipskind and Clarkson, 1983) (3) in recombination events leading to expansion of the globin gene family (Slightom et al., 1980) (4) in chromatin structure (Rich, 1983). Nordheim and Rich (1983) found that the viral control elements for replication and enhancer function were contained in a nucleosome-free region of the SV40 minichromosome and that DNase I hypersensitive sites (usually associated with transcriptionally active chromatin) within this region were determined in part by the presence of Z-DNA binding proteins. Many types of Z-DNA binding proteins are known to exist but their sequence specificity and activity has yet to be characterized. Methylation of cytosines in m^5CG sequences in vivo is associated with gene inactivation (Doerfler, 1983) and also stabilizes Z-DNA structures (Behe and Felsenfeld, 1981), but whether the two phenomena are related is not certain. In a variety of eukaryotic genomes, especially SV40, the greatest Z-DNA forming potential resides in the enhancer regions, supporting the possible involvement of Z-DNA structures in the transcriptional regulation of gene expression. Nordheim and Rich (1984) have postulated that a change in chromatin structure, such as the interconversion of Z- and B-forming DNA, may be mediated by sequence specific Z-DNA binding proteins which affect proximal or distal sites resulting in activation or inactivation of a gene.

In view of this speculative association of Z-DNA structures and enhancer activity, it was interesting to observe within the HGT-F flanking DNA region, a stretch of 53 nucleotides capable of forming Z-DNA with two enhancer-like sequences (discussed in the section above) located directly downstream (Figure 6.5). The very distinctive alternating purine/pyrimidine "ladder" structure of this sequence, shown in Figure 6.7, is composed of the following dinucleotide sequences, $(\text{purine/C})_3$

FIGURE 6.7

POTENTIAL Z-DNA SEQUENCE IN THE 5' FLANKING REGION OF THE HGT-F GENE

Autoradiogram of a DNA sequencing gel (2.2.R.7), showing the alternating purine/pyrimidine-rich region potentially involved in Z-DNA formation. The tracks are marked A, G, C, T for each base-specific sequencing reaction (2.2.R.6), with the DNA sequence read 5'→ 3' from the bottom to the top of the gel. The unusual 53 nucleotide purine/purimidine sequence shown below is preceded by two 12-base direct repeats (GGACGCAACTGA) indicated by the two arrows.

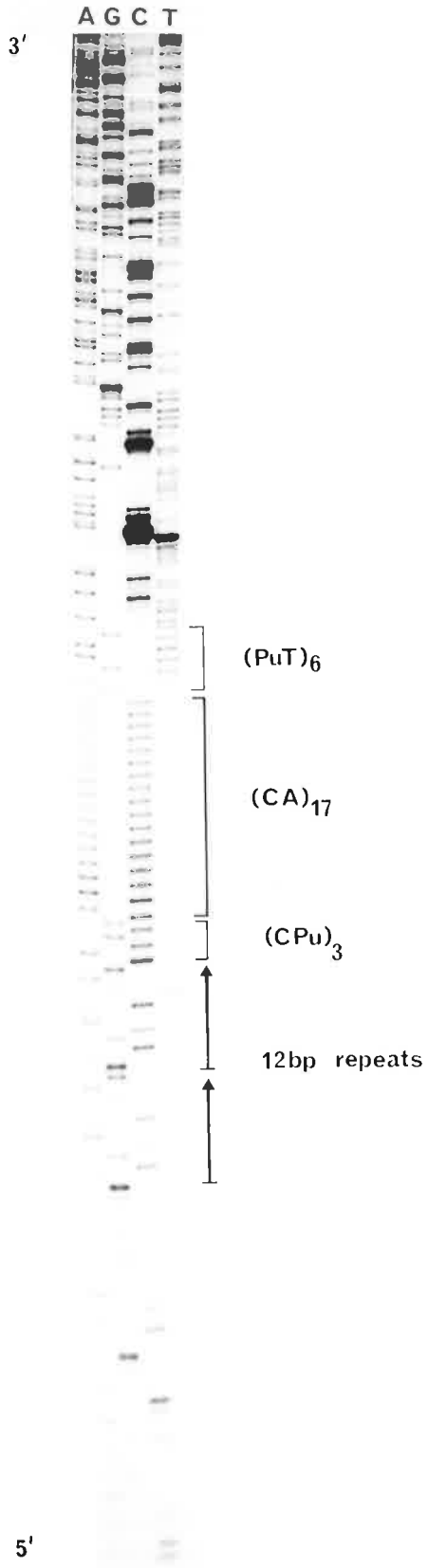
'Z-DNA' region:

5'-GCACGCGCACACACACACACACACACACACACACACACACAGTATGTATATGT-3'

or

$(PuC)_3G(CA)_{17}(PuT)_6$

where G is the single base not in alternation.


A G C T
3'
(PuT)6
(CA)17
(CPu)3
12bp repeats
5'

$(CA)_{17}$ and (purine/T)$_6$. Also shown in this figure are two 12 bp direct repeats directly upstream to the potential Z-DNA structure. Rogers (1983) has suggested that similar short direct repeats, usually found flanking alternating purine/pyrimidine structures, are involved with a mechanism of integration of these sequences within the genome.

A Putative Intron in the 5' Non-coding Region

The majority of eukaryotic genes, both cellular and viral are interuppted by introns (for examples see Berget et al., 1977; Tilghman et al., 1978) so that the transcription of these so-called split genes requires one or more processing (splicing) steps. No general pattern has emerged for the distribution or the structure of introns that might provide some clues to their function. The size and arrangement of introns varies enormously and examples are present in cellular, mitochondrial and chloroplast genes, so that the diversity in structure almost certainly represents diversity in function (for review see Gilbert, 1978 and references therein).

However, a common feature of introns has emerged following the examination of sequences over 130 splice junctions (Mount, 1982). This comparative study revealed a consensus sequence, 5'- AG/GT$^{A}_{G}$AGT-3', at the donor or 5' splice site and 5'-$(^{T}_{C})_{n}$N$^{C}_{T}$AG/G-3' at the acceptor or 3' splice site. Most notably the first two (GT) and last two (AG), underlined in the consensus sequences, are strictly conserved. Since variations of the consensus sequence occur they cannot explain the selection of splice sites, but they are most likely involved in the cleavage and ligation process. This was shown by the analysis of transcripts produced from defective globin genes (Treisman et al., 1982; Fukumaki et al., 1982; Felber et al., 1982) where alternate splice sites were utilized in favour of the normal sites.

Lerner et al., (1980) has proposed that the small nuclear RNA,

present as ribonucleoprotein particles are intimately involved in the splicing step, with the RNA sequences in the small ribonucleoprotein particles capable of base-pairing to the consensus junction sequences, imparting sensitivity to nuclease attack, carried out by the proteins associated with these particles. The actual mechanism by which splicing occurs, and in particular, how the splice sites are selected is not known and awaits the development of *in vitro* splicing systems for mRNAs to study the components of mRNA processing. Considerable success has been achieved in tRNA splicing systems in yeast (Abelson, 1979) yet the equivalent for mammalian mRNA lags far behind.

Examination of the nucleotide sequence of the HGT-F gene and its flanking regions revealed two sequences, separated by approximately 700 bp, which corresponded to the 3' and 5' junctions of an intron. The donor, or 5' site, was located at position 244 while the acceptor, or 3' site, was within the 5' non-coding region of the gene at position 961 (Figure 6.5). Although the splice junctions adhered to the GT/AG rule (Breathnach *et al.*, 1978) there were variations in some of the other positions relative to the consensus sequence. The low-sulphur keratin genes are known to contain introns (B. Powell, K. Ward personal communications) whereas the high-sulphur keratin genes do not (Powell *et al.*, 1983). If the HGT genes resemble the other matrix-protein coding genes in this respect (ie., high-sulphur genes) and do not contain introns, then it may be fortuitous that sequences similar to splice junctions are present in the gene sequence. This is supported by the Northern blot data (see section 5.2.F) which showed only one transcript for the HGT-F gene in the wool follicle cells, which corresponded in size to a mRNA with the functional start site of transcription at position 932.

However, it is interesting to consider the likely organization of the HGT-F gene if these splice sites are utilized. Firstly, if the

variant splice junctions reflect inefficient processing sites, then it is possible that other structural features of the gene, such as the Z-DNA region and enhancer-like sequences, may have a function in compensating for the poor splice sites. For example, a change in the DNA conformation may make these sites more accessible to the splicing enzymes. If the intron was then spliced out, a potential regulatory region is apparent upstream to the 5' splice junction and consists of a cap site (position 166) and a 'TATA' box (position 133) separated by a distance which is equivalent to that seen for the cap and 'TATA' sequences in the regular HGT-F gene (See Figure 6.8). No 'CAAT' box was apparent but a MMTV enhancer-like element (7 out of 8 matches), was found about 80 bases upstream to the 'TATA' box. This was especially noteworthy since enhancer-like elements have been found in the 5' promoter region of several genes which lack the 'TATA' and 'CAAT' homologies and are thought to play a role in regulation of transcription of these genes (Reynolds et al., 1984).

Figure 6.8 compares the 5' region of the regular HGT-F gene with the corresponding region of the putative spliced gene. The latter gene sequence has a 5' non-coding region which is almost twice that of the regular gene, with very little sequence homology apart from the 18 bp conserved matrix-element immediately, adjacent to the initiation codon. If splicing does occur, then the 3' splice site would reside at the 5' terminal nucleotide of this conserved sequence reinforcing the likely importance of this element for HGT gene expression. Introns solely within the 5' non-coding region of genes rarely occur, with the only known examples being that of chicken feather and scale keratin and chicken 'Fast-protein' genes (Molloy et al., 1982; S. Wilton, P. Morris personal communications). In these genes, the intron is wholly contained within the 5' non-coding region so that the regular upstream promoter sequences are utilized. The putative spliced HGT-F gene arrangement bears some

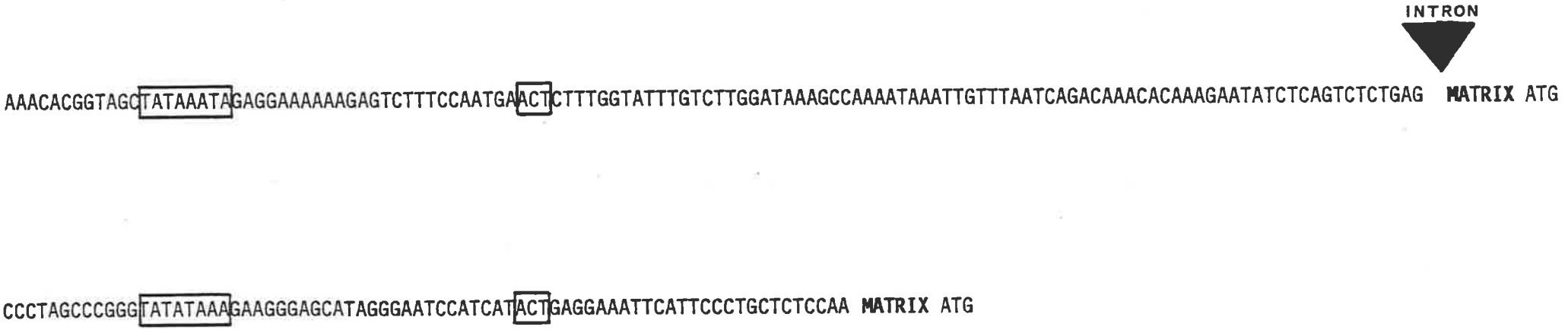
FIGURE 6.8

A COMPARISON OF SEQUENCES UPSTREAM TO THE INITIATION CODON OF THE HGT-F GENE AND THE PUTATIVE SPLICED GENE

A comparison of the 5' non-coding and flanking DNA sequence of the HGT-F gene (bottom line) and a corresponding region adjacent to the putative upstream 5' splice site (top line).

The 'TATA' sequences and cap sites are boxed. MATRIX refers to the gene-specific conserved element shown in Figure 6.6, which directly precedes the translation initiation codon, ATG.

If the 5' and 3' splice sites (indicated in Figure 6.5) are utilized, then an intron of about 700 bp, immediately upstream to the MATRIX sequence, would be spliced out. The 5' non-coding, cap and 'TATA' sequences of the putative spliced HGT-F gene are shown in the top line of this figure.


INTRON
AAACACGGTAGCTATAAATAGAGGAAAAAAGAGTCTTTCCAATGAACTCTTTGGTATTTGTCTTGGATAAAGCCAAAATAAATTGTTTAATCAGACAAACACAAAGAATATCTCAGTCTCTGAG MATRIX ATG
CCCTAGCCCGGGTATATAAAGAAGGGAGCATAGGGAATCCATCATACTGAGGAAATTCATTCCCTGCTCTCCAA MATRIX ATG

resemblance to the mouse α-amylase mRNA system (Young et al., 1981) where the expression of two mRNAs, present in vastly different amounts in the liver and salivary gland, requires the use of two different promoters for a single cellular gene. Differential splicing generates overlapping but non-identical mRNAs differing solely in their 5' non-coding regions. The HGT proteins are more abundant in the hoof and horn structures of sheep as compared with wool (Gillespie, 1972), so that the length disparity in the 5' non-coding regions may affect translational efficiences resulting in quantitative changes in the HGT mRNA levels within the different tissues. The same type of mechanism has been proposed for the variable α-amylase concentrations found in the mouse liver and salivary glands where the length of the 5' non-coding regions have been thought to be involved in tissue-specific processing, although the signals which initiate the tissue-specific splicing events have not been determined.

Obviously in vitro manipulation of the structural features, discussed in the sections above, are required to establish the importance of these sequences in HGT-F gene expression. Such studies are currently underway (G. Cam, J. Forrest, E. Kuczek, personal communication). A variety of factors, such as genetic, dietary, chemical and hormonal influences are known to affect HGT keratin synthesis (Frenkel et al., 1974; Reis, 1979; Gillespie and Marshall, 1980; Gillespie et al., 1983), so that a complex array of transcriptional and/or translational controls are likely to be involved in the differential expression of these keratins in the wool fibre.

CHAPTER 7

HETEROGENEITY OF THE HGT WOOL KERATIN GENES: CHARACTERIZATION OF A HGT-C2 cDNA CLONE

CHAPTER 7

7.1 INTRODUCTION

The existence of multiple HGT protein components in wool (Gillespie, 1972) would seem to suggest that these keratins are encoded by a multigene family. However, the protein data remains equivocal because many of the chromatographic properties of the HGT proteins are determined by interactions between the aromatic residues and the chromatographic support, so it is possible that the apparent heterogeneity arises from different conformational states of aromatic residues (Marshall et al., 1980).

Therefore, the isolation of individual gene members through the use of specific DNA probes seemed a feasible method by which the nature of the heterogeneity of these proteins could be resolved. This chapter describes the use of the synthetic probe approach (see 5.2.A) to isolate cDNA clones encoding sequences for the second major HGT protein species present in wool, component C2 (Marshall et al., 1980).

The high-sulphur keratin proteins contain sufficient amino acid sequence homology within each sub-family to enable the isolation of all members of a gene family using a single gene probe (Powell et al., 1983). The limited protein data available for the HGT proteins indicated that such sequence cross-hybridization would not occur between the genes encoding these protein components, even though the two HGT proteins for which amino acid sequence data is known (components C2 and F), are members of the same sub-family (Type I).

In Chapter 5, the HGT-F gene was characterized by DNA sequencing and Southern blot anaylsis, which revealed that it was unique in the sheep genome and failed to cross-hybridize with any other sequences even under low-stringency conditions, thereby precluding its use as a probe for other HGT genes. Therefore, the synthetic probe approach was

employed again to synthesize a HGT-C2 specific DNA probe. In this way, a HGT-C2 cDNA clone subsequently isolated from the wool follicle cDNA library could be used to detect related sequences in the sheep genome.

7.2 RESULTS

7.2.A Design of the HGT-C2 Specific Oligonucleotide Probe

HGT amino acid sequence data is limited to the two protein sequences shown in Figure 5.1. The overall direct sequence homology is very low and restricted to small blocks of up to 3 amino acids in length. The longest stretch of residues in common is Gly-Cys-Gly-Tyr-Gly, which occurs in different regions of the two proteins. Therefore it was decided to construct a probe which would specifically detect HGT-C2 protein-coding sequences.

The HGT-C2 specific oligonucleotide probe, as shown in Figure 7.1, was a mixture of all the possible coding permutations complementary to a small portion of the component C2 sequence close to the carboxyl terminus and utilized the unique tryptophan and other low ambiguity codons, such as phenylalanine and tyrosine. The probe was therefore a mixture of 128 different 17-base sequences and was characterized for its size and sequence (data not shown), essentially as carried out for the other oligonucleotide probes used in this thesis (see Figure 5.3).

7.2.B A HGT-C2 Keratin Clone Isolated from the Wool Follicle cDNA Library

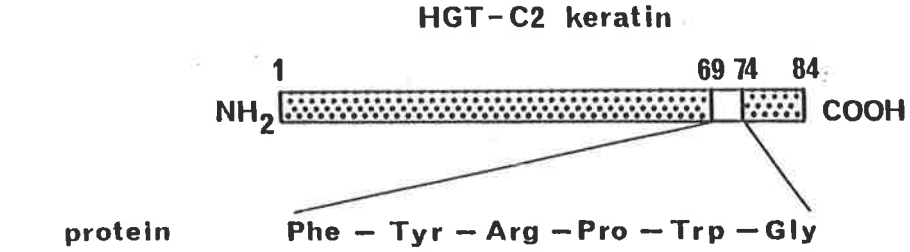
The wool follicle cDNA library was screened with the ^{32}P-labelled 17-mer mix using the hybridization conditions described in 2.2.I.1. Strong hybridization to all colonies was apparent which indicated that the 17-mer probe contained a sequence which was homologous to a pBR322 sequence. However, 12 colonies gave hybridization signals

FIGURE 7.1

DESIGN OF THE HGT-C2 SPECIFIC OLIGONUCLEOTIDE PROBE

The sequence of the oligonucleotide probe used to isolate HGT-C2 cDNA clones from the wool follicle cDNA library (7.2.B) is presented. The 17-mer mix, whose sequence was confirmed by the sequencing method described in 2.2.Q, was a mixture of 128 oligonucleotides. N = A, G, C or U

The complementary DNA sequence obtained from the cDNA clone is shown beneath the probe sequence.



possible codons	5′	UU(U/C)	UA(U/C)	C(A/G/U/C)(A/G)	CCN	UGG	GGN	3′
probe	3′	AA(A/G)	AT(A/G)	G(T/C)(A/G/C/T)	GG(A/G/C/T)	ACC	CC	5′
cDNA clone	5′	TTT (75)	TAT	AGG	CCA	TGG (85)	GG	3′

above background (data not shown) and since it was not feasible to examine all twelve clones in detail, only the clone showing the strongest hybridization signal was chosen for DNA sequence analysis.

DNA from this presumptive HGT-C2 cDNA clone was first characterized by Southern blot analysis. Digestion of the DNA with several restriction enzymes, followed by blotting and hybridization to the oligonucleotide probe showed hybridization to the insert fragment in the PstI and HpaII degests (Figure 7.2, tracks 2 and 4) but not to the HaeIII insert fragment (track 3). This result suggested the presence of a HaeIII restriction site in the region of the cDNA clone complementary to the probe, which effectively decreased the length of DNA hybridizing to the probe, so that the resultant hybrid was too short to remain stable under the washing conditions used. No hybridization was visible to the insert from pSHGT-1 (the HGT-F cDNA clone), although strong background hybridization to pBR322 was evident in all tracks (Figure 7.2). The degree of homology between the pBR322 sequence(s) and sequence(s) in the 17-mer mix was not determined, but was probably quite high as was apparent from the strong hybridization signals.

7.2.C Sequence Analysis of pSHGT-C2.1

PstI resection of the presumptive C2-cDNA clone, designated pSHGT-C2.1 (or plasmid sheep high-glycine/tyrosine - component C2. clone 1), revealed an insert size of about 230 bp. Sequencing was initiated at the HaeIII site predicted by Southern blot analysis (see section 7.2.B above) and extended in both directions to the PstI sites at both ends of the cloned DNA fragment. An AluI site, deduced from the DNA sequence, was then utilized to confirm the sequence across the HaeIII site.

The insert within pSHGT-C2.1 contained a cDNA sequence encoding approximately 50% of the HGT-C2 protein region and 66 bp of the 3' non-coding region (Figure 7.3). The conceptual amino acid sequence

FIGURE 7.2

SOUTHERN BLOT ANALYSIS OF THE 17-MER SELECTED cDNA CLONE

Plasmid DNA was purified from the 17-mer selected cDNA clone (see 7.2.B) and digested with PstI, HaeIII and HpaII. These digests along with a PstI digest of pSHGT-1 were electrophoresed on a 2% agarose gel (2.2.M.1), followed by transfer of the DNA to nitrocellulose. The blot was then hybridized to the ^{32}P-labelled 17-mer essentially as described in 2.2.I.1, except that the hybridization was performed at 30°C overnight in the presence of 10% dextran sulphate. The blot was washed at room temperature in 6xSSC.

PART A : shows the ethidium bromide-stained gel of the various plasmid DNA digests. Insert fragments of the putative HGT-C2 cDNA clone are indicated by a dot.

PART B : shows the corresponding autoradiogram following hybridization to the 17-mer probe. Asterisks indicate background hybridization of the mixed probe with sequences in pBR322.

TRACKS :

1. PstI digest of pSHGT-1
2. PstI digest
3. HaeIII digest
4. HpaII digest

} of 17-mer selected clone

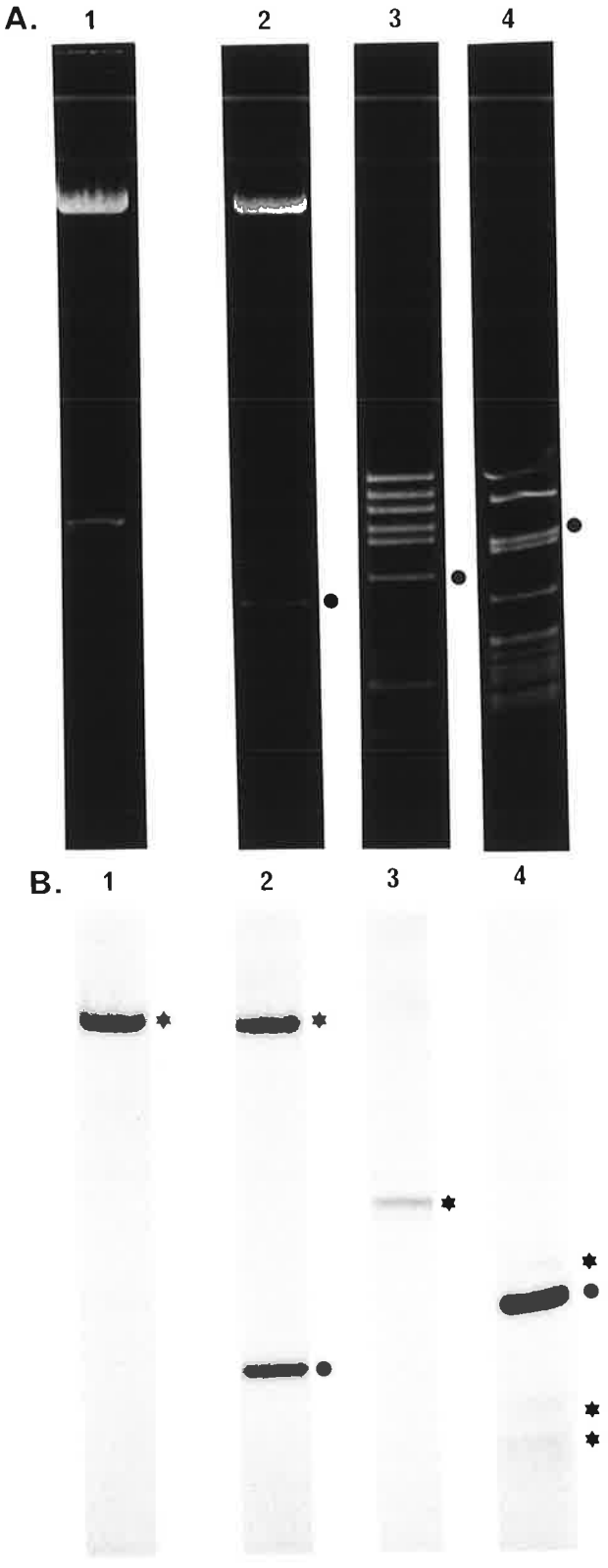
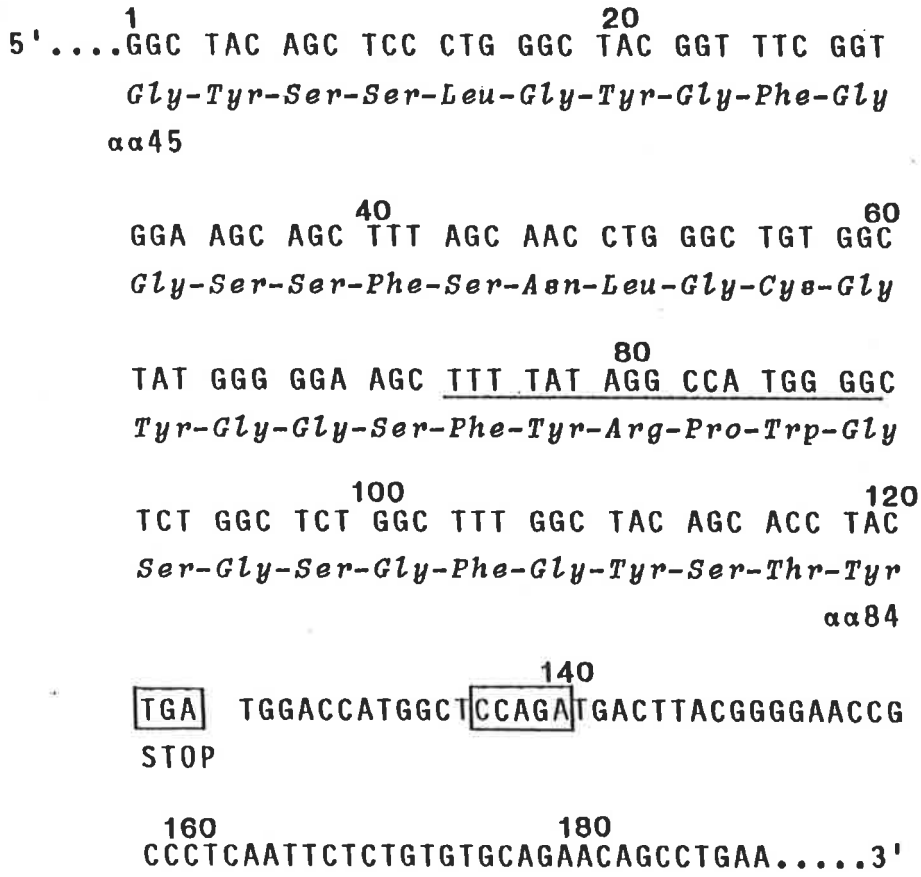

A.
1
2
3
4
B.
1
2
3
4

FIGURE 7.3

DNA SEQUENCE OF THE cDNA CLONE, pSHGT-C2.1

The DNA sequence of the insert from pSHGT-C2.1 is presented. Sequencing was initiated from the HaeIII (position 80) and AluI (position 70) sites following subcloning into SmaI/PstI-digested M13 mp19 vector (2.2.R). The sequence complementary to the 17-mer probe is underlined, while the termination codon (TGA) and pentanucleotide sequence, CCAGA, are boxed. This latter sequence element is also found in the same position in the HGT-F cDNA sequence (Figure 5.6B). The derived protein sequence is shown beneath the DNA sequence, with the numbers referring to the corresponding amino acid in the published sequence of HGT-component C2 (Marshall et al., 1980).


1
20
5'....GGC TAC AGC TCC CTG GGC TAC GGT TTC GGT
Gly-Tyr-Ser-Ser-Leu-Gly-Tyr-Gly-Phe-Gly
αα45
40
60
GGA AGC AGC TTT AGC AAC CTG GGC TGT GGC
Gly-Ser-Ser-Phe-Ser-Asn-Leu-Gly-Cys-Gly
80
TAT GGG GGA AGC TTT TAT AGG CCA TGG GGC
Tyr-Gly-Gly-Ser-Phe-Tyr-Arg-Pro-Trp-Gly
100
120
TCT GGC TCT GGC TTT GGC TAC AGC ACC TAC
Ser-Gly-Ser-Gly-Phe-Gly-Tyr-Ser-Thr-Tyr
αα84
140
TGA TGGACCATGGCTCCAGATGACTTACGGGGAACCG
STOP
160
180
CCCTCAATTCTCTGTGTGCAGAACAGCCTGAA.....3'

derived from the cDNA sequence was identical to the known protein sequence (Gillespie et al., 1980).

Comparison of the 3' non-coding regions of pSHGT-F (see Figure 5.6B) and pSHGT-C2.1 (Figure 7.3) showed no significant homology other than a pentanucleotide sequence, 5'-CCAGA-3', present in corresponding positions, 12 nucleotides downstream from both termination codons. Short conserved elements have been found amongst 3' non-coding regions of the high-sulphur keratin genes (Powell et al., 1983) and as for this pentanucleotide sequence, their function, if any, is unknown.

7.2.D Size Estimation of the HGT-C2 mRNA

Northern blot analysis of wool follicle mRNA showed the presence of a single broad band (Figure 7.4) when hybridized with the ^{32}P-labelled insert from pSHGT-C2.1. The HGT-C2 mRNA was therefore approximately 850±50 bases long, as judged by the co-electrophoresis of rRNA markers. Since the protein-coding region of the 84 amino acid HGT component C2 would account for 252 bases, the remainder of the mRNA constitutes the 5' and 3' non-coding sequences and as such must have a total length substantially longer than those seen for the HGT-F mRNA (50 and 307 bases respectively, see 5.2.3). The exact lengths of these regions will be determined once longer HGT-C2 cDNA clones and the corresponding gene are sequenced.

7.2.E Genomic Southern Analysis with the HGT-C2 cDNA Probe

Southern blot analysis using the ^{32}P-labelled insert from pSHGT-C2.1 as a probe to an EcoRI digest of sheep genomic DNA is shown in Figure 7.5. One major band (about 3.2 kb) as well as several other higher and lower molecular weight weakly hybridizing fragments were observed. This result suggested that in addition to the HGT-C2 gene (presumably located on the 3.2 kb EcoRI fragment), there were at least five other

FIGURE 7.4

HGT-C2 mRNA SIZE ESTIMATION BY NORTHERN BLOT ANALYSIS

Total wool follicle poly(A)$^+$ RNA (5µg) was disaggregated by glyoxal treatment, followed by electrophoresis on a 1% agarose gel (2.2.P). Sheep rRNA (28S and 18S) and B. amyloliquefaciens rRNA (23S and 16S) markers were co-electrophoresed in parallel and visualized by staining with ethidium bromide.

Following transfer to nitrocellulose, the blot was hybridized (2.2.I.2) to the nick-translated insert of pSHGT-C2.1. A single band of estimated size, 0.85 kb was detected on the resultant autoradiogram.

HGT-C2
mRNA

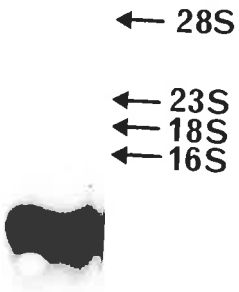


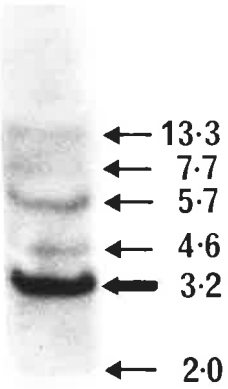
FIGURE 7.5

GENOMIC SOUTHERN BLOT ANALYSIS WITH pSHGT-C2.1

Sheep genomic DNA (10µg) was digested with EcoRI, electrophoresed on a 1% agarose gel (2.2.M.1), then transferred to nitrocellulose. The filter was then hybridized to the ^{32}P-labelled (2.2.H.7) insert of pSHGT-C2.1.

HindIII-digested DNA was co-electrophoresed on the same gel and used as a molecular weight marker.

The major hybridizing fragment, about 3.2 kb, is indicated by the thick arrow, while the weaker hybridizing fragments are indicated by the thin arrows and range in size from about 13.3 kb to 2 kb.


13·3
7·7
5·7
4·6
3·2
2·0

HGT-C2 related sequences in the sheep genome.

To determine whether the HGT-C2 gene was closely linked with the HGT-F gene, the ^{32}P-labelled insert of pSHGT-C2.1 was used to probe several digests of λSHGT-F. No bands were seen to hybridize (data not shown), implying that the HGT-C2 gene was not located on this genomic clone. This result supports the hybridization data obtained when a similar genomic blot was probed with total wool follicle ^{32}P-labelled cDNA (see 6.2.B.2), which indicated that other than the HGT-F gene, no other follicle keratin gene was present in the genomic clone, λSHGT-F.

7.3 DISCUSSION

This chapter described the isolation and characterization of a cDNA clone for the second major HGT Type I protein found in the wool fibre (Marshall et al., 1980). The derived amino acid sequence from this clone, pSHGT-C2.1, was in complete agreement with the protein sequence data, while the 3' non-coding sequence had negligible homology (apart from the pentanucleotide sequence, 5'-CCAGA-3') with pSHGT-1, the HGT-F cDNA clone (see Figure 5.6B). The negligible homology between the two cDNA clones extended into protein-coding region, where it was observed, that despite there being a common five amino acid sequence and smaller blocks of direct amino acid homology (see 7.2.A above), at the nucleotide level this homology was not retained. Differences in the third base position of various codons were apparent, and extensive homologies could not be found for any significant length of the two sequences. Therefore, the lack of cross-hybridization between HGT-F and C2 DNA sequences was not surprising. A more detailed comparison at the nucleotide level awaits the isolation of the HGT-C2 gene which is currently underway.

Since it has been shown that two high-sulphur keratin genes are closely linked in the sheep genome (Powell et al., 1983), several experiments were performed to determine whether other keratin (including

the HGT-C2) genes were located in the genomic clone, λSHGT-F, characterized in Chapter 6. The results (6.2.B.2 and 7.2.E) indicated that only the HGT-F gene was present in this clone, but the possibility exists that the HGT-C2 and other wool keratin genes reside on genomic segments adjacent to the 14 kb DNA fragment within λSHGT-F. This will be confirmed once overlapping genomic clones contiguous with λSHGT-F are isolated. At present, a sheep cosmid library is being screened (G. Cam, E. Kuczek) for the HGT-C2 gene. Since cosmid vectors (Collins and Hohn, 1978) are capable of accepting larger insert fragments than λbacteriophage vectors, isolation of a cosmid clone containing the HGT-C2 gene will not only provide sequence data for a comparative analysis with the HGT-F gene, but also allow an examination over a greater distance of the sheep genomic DNA for clues to the possible chromosomal arrangement of the HGT and other wool keratin genes.

Southern blot analysis of sheep genomic DNA (Figure 7.5) indicated that in addition to the HGT-C2 gene (presumed to reside on the strongly-hybridizing 3.2 kb EcoRI fragment), other sequence-related DNA fragments were also observed. In an attempt to determine the homology between the HGT-C2 gene sequence and the weaker-hybridizing fragments, the following questions need to be answered. Firstly, does the broad band on the Northern blot result (Figure 7.4) represent a single hybridizing RNA species or are there several mRNA species of similar size lengths which co-migrate? Secondly, do the other 17-mer-selected cDNA clones (see 7.2.B) encode other members of the HGT protein family which are related to component C2? This could be tested by using the 3' non-coding sequence of pSHGT-C2.1 as a probe to these other clones. Lack of hybridization would suggest that these clones encode C2-related proteins but which differ in their 3' non-coding sequences.

Such information will be crucial to clarifying the ambiguous heterogeneity of the HGT proteins. The early protein-chemical work

(Gillespie and Frenkel, 1974 and see Figure 7.6) showed that component C2 was a member of protein fractionation group C, which also contained other closely related proteins, components C3 and C4. However, subsequent amino acid sequence data showed components C2 and C3 to be identical (Marshall et al., 1980). This led to the proposal by Marshall and co-workers that the different electrophoretic mobilities of these components, as seen in Figure 7.6A, were likely to be due to small differences in the degree of amidation caused by the extraction and/or fractionation conditions. This was supported by Frenkel(1977), who prepared the HGT proteins at neutral pH conditions and found that each protein fractionation group contained a single protein species. It has also been reported that interactions occur between the tyrosine residues of the HGT proteins and various chromatographic supports (Gillespie and Frenkel, 1976; Marshall et al., 1980) so that the complex electrophoretic pattern, seen in Figure 7.6A, may result from an anomalous fractionation of a small number of species in various conformational states, rather than a pattern produced by a range of completely different molecular species. The two-fold range in tyrosine contents across this electrophoretic profile (see Figure 7.6B) could therefore reflect differing distributions of a small number of individual components in the various fractions.

In the absence of further protein and nucleic acid data the results discussed in this chapter can be interpreted as follows. Component F represents a unique member of the Type I HGT keratin gene family, whereas component C2 is part of a group of homologous sequences, whose size and sequence heterogeneity has yet to be established, which may constitute at least six related genes. It is possible that the weakly-hybridizing sequences (observed in Figure 7.5) are genes which encode the HGT species present in the other protein fractionation groups, other than C and F (see Figure 7.6A). If this is the case, then the component C2 sequence is more representative of the Type I HGT proteins

FIGURE 7.6

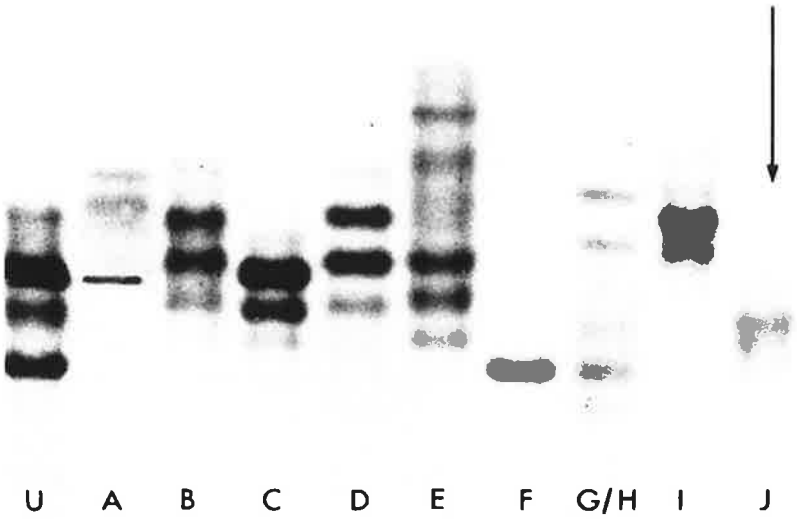
HETEROGENEITY OF THE HGT TYPE I WOOL PROTEIN FAMILY

The heterogeneity of the Type I HGT wool proteins, as implicated by the studies of Gillespie and Frenkel (1974), is shown in this figure.

PART A shows the electrophoretic pattern of the fractions (A →J) obtained from QAE-cellulose chromatography, after analysis on cellulose-acetate in 8M urea at pH 8.9. U is the pattern of unfractionated Type I HGT proteins; G/H = fractions G+H.

PART B shows the amino acid compositions (residues %) of a Type I HGT protein preparation and 10 purified fractions prepared from it. In identification of fractions, the prefixed letters (C, D, E, G+H, I) refer to QAE-cellulose groups and the numerals refer to components within a group separated on DEAE-cellulose. The range over which each amino acid varies is also indicated.

A.



B.

Amino acid	Whole HT-I	Fractions C2	C3	C4	D2	D3	D4	F	G+H2	G+H3	I4	Range
Lys	0·44	0·18	0·15	0·40	0·09	0·06	0·00	0·00	0·62	0·65	0·26	0·0–0·7
His	1·13	1·35	1·66	1·83	1·40	1·44	1·35	0·00	1·91	1·71	0·19	0·0–1·9
Arg	5·35	4·90	5·38	5·73	7·43	6·86	6·60	3·40	3·86	4·32	4·64	3·4–7·4
SCMCys	5·98	5·56	6·46	6·53	5·43	6·11	6·56	6·42	1·59	2·44	3·34	1·6–6·6
Asp	3·31	4·31	4·57	4·53	2·88	3·16	3·61	1·85	1·91	2·04	4·76	1·9–4·8
Thr	3·33	4·81	4·74	4·18	1·47	1·96	2·70	3·41	0·98	1·06	3·28	1·0–4·8
Ser	11·80	12·60	13·70	13·60	12·90	13·40	13·40	12·60	12·80	11·70	8·95	9·0–13·7
Glu	0·57	0·48	0·44	0·86	0·15	0·00	0·32	0·24	1·24	0·82	1·80	0·0–1·8
Pro	5·28	6·55	6·44	5·56	1·48	2·42	2·90	6·69	2·37	3·26	6·56	1·5–6·7
Gly	27·60	30·80	25·50	26·80	34·00	32·50	28·90	26·50	35·90	34·20	25·80	25·5–35·9
Ala	1·48	1·05	1·04	1·14	0·16	0·32	0·32	3·07	0·56	0·97	0·57	0·2–3·1
Val	2·09	1·64	1·57	1·66	0·16	0·32	0·64	4·70	0·36	0·48	3·22	0·2–4·7
Met	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
Ile	0·20	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	1·42	0·0–1·4
Leu	5·46	5·65	6·25	6·53	4·57	4·57	5·47	3·51	7·67	7·66	4·31	3·5–7·7
Tyr	15·00	10·10	11·10	10·40	13·80	13·60	14·30	18·00	19·80	19·70	20·20	10·1–20·2
Phe	10·30	10·00	11·00	10·70	13·90	13·40	12·80	9·63	8·29	8·88	9·46	8·3–13·9

than is component F. In this respect, the HGT-C2 cDNA clone can be used to isolate not only the HGT-C2 gene, but other HGT (and in this case, C2-sequence related) genes.

CHAPTER 8

HGT KERATIN GENES IN SHEEP AND OTHER MAMMALIAN GENOMES

CHAPTER 8

8.1 INTRODUCTION

Studies on mammalian and non-mammalian hard keratins (Gillespie, 1972; Gillespie and Frenkel, 1974; 1976) showed that proteins rich in the aromatic residues and glycine are responsible for the differences in tyrosine content observed between the various keratin structures (see Figure 1.2). Unlike amino acid compositions and molecular sizing, the charge differences observed between the numerous protein components indicated a great variation in the abundance of the HGT proteins in the individual keratin structures. In Figure 8.1, two-dimensional polyacrylamide gel analysis of a number of mammalian hairs shows the species-specific characteristic electrophoretic patterns of the three major protein families; the low-sulphur (LS), the high-sulphur (HS) and the high-glycine/tyrosine keratin (HGT) proteins. Further protein-chemical analysis of the homologous HGT proteins in mouse hair and sheep wool (Gillespie, 1983) has been carried out but amino acid sequence data is available only for the wool proteins, namely components C2(C3) and F, as discussed in this thesis.

The availability of sheep HGT-C2 and HGT-F DNA probes now allows an examination of other genomes for related HGT genes and may provide information concerning the number, similarities and possibly evolutionary relationships of these proteins much more rapidly than by comparative protein sequence analysis.

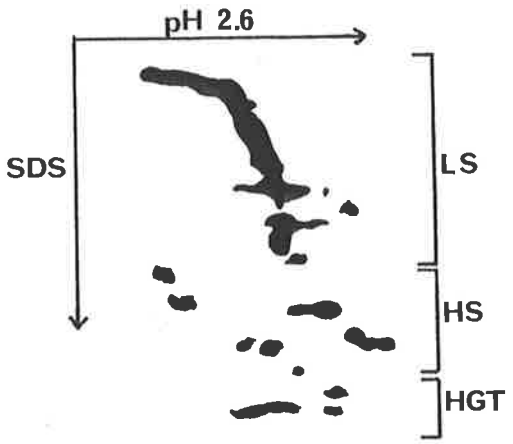
Genetic control of protein composition in wool is well established (Frenkel *et al.*, 1974), and of particular interest are the mutations causing dramatic changes in protein composition such as, the dominantly inherited 'Naked' trait of mouse which produces hair with less than half the normal content of the HGT proteins (Tennenhouse and Gold, 1976) and the Felting lustre-mutant Merino whose wool is virtually devoid

FIGURE 8.1

SDS-ACID POLYACRYLAMIDE GEL ELECTROPHORETIC PATTERN
COMPARISON OF VARIOUS MAMMALIAN HAIR SAMPLES

Fluorograms of two-dimensional polyacrylamide gel electrophoretic patterns of the proteins of various mammalian hairs. First dimension: electrophoresis in 8M urea at pH 2.6. Second dimension : electrophoresis in SDS. Fluorograms courtesy of R. Marshall.

The schematic diagram below shows the distribution of the three major protein families in this two-dimensional gel system: low-sulphur (LS), high-sulphur (HS), high-glycine/tyrosine (HGT).



ACID-SDS TWO-DIMENSIONAL ELECTROPHORESIS OF MAMMALIAN HAIRS

of HGT proteins (Gillespie and Darskus, 1971). In both cases, the whole range of HGT proteins are affected so the mutations are unlikely to be specific for a single protein species but must act concertedly over the whole protein family. Qualitative and quantitative differences have also been noted in the HS keratins of the 'Naked' mouse (Raphael et al., 1983) but they have not been examined in any great detail. However, other phenotypic lesions are associated with this mutation, which has led Raphael and co-workers to conclude that in the 'Naked' mouse, the N (Naked) locus is only indirectly involved in the synthesis of the structural proteins found in the mouse hair and that the decreased HGT content may be a secondary effect of a more general tissue malformation, such as the disturbance in the differentiation of the follicle cells. Similar studies examining the wool follicles of the Felting lustre-mutant Merino have not been carried out, but since other lesions, such as absence of crimping and complete absence of skinfolds, have been observed (Short, 1958), it is likely that a similar general tissue malformation as suggested for the 'Naked' mouse mutation, probably exists in the Felting lustre-mutant Merino.

This chapter describes a preliminary investigation of HGT-related sequences in the human and mouse genomes and in the DNA of phenotypically HGT-deficient sheep.

8.2 RESULTS

8.2.A Polyacrylamide Gel Electrophoretic Analysis of Human, Mouse and Sheep Hair Proteins

Human, mouse and sheep hair were the source of samples used in the experiments described in this chapter due to their relative availability compared to other mammalian hairs.

All hair samples were labelled with ^{14}C-iodoacetic acid by

S-carboxymethylation and analyzed following electrophoresis on a SDS-urea 15% polyacrylamide gel (see 2.2.E.3). The one-dimensional gel system was used in preference to the two-dimensional gel system of Marshall (1981), since it allowed more than one sample to be processed on the same gel to facilitate the direct comparison of each hair sample. Identification of proteins was based on their position in the gel, with the HGT proteins readily distinguished in this gel system ,by their characteristic location towards the bottom of the gel away from the slower-migrating LS and HS keratin proteins. Figure 8.2 shows the one-dimensional gel electrophoretic patterns of human and mouse hair, Merino, Merino-Dorset Horn cross, Lincoln and Felting lustre-mutant Merino wool samples. In order to make satisfactory comparisons between the proteins, several exposure times were used so that the resolution of specific proteins was optimal. As such, it was not possible to make quantitative comparisons between the groups of proteins (or individual components), so only relative amounts are compared.

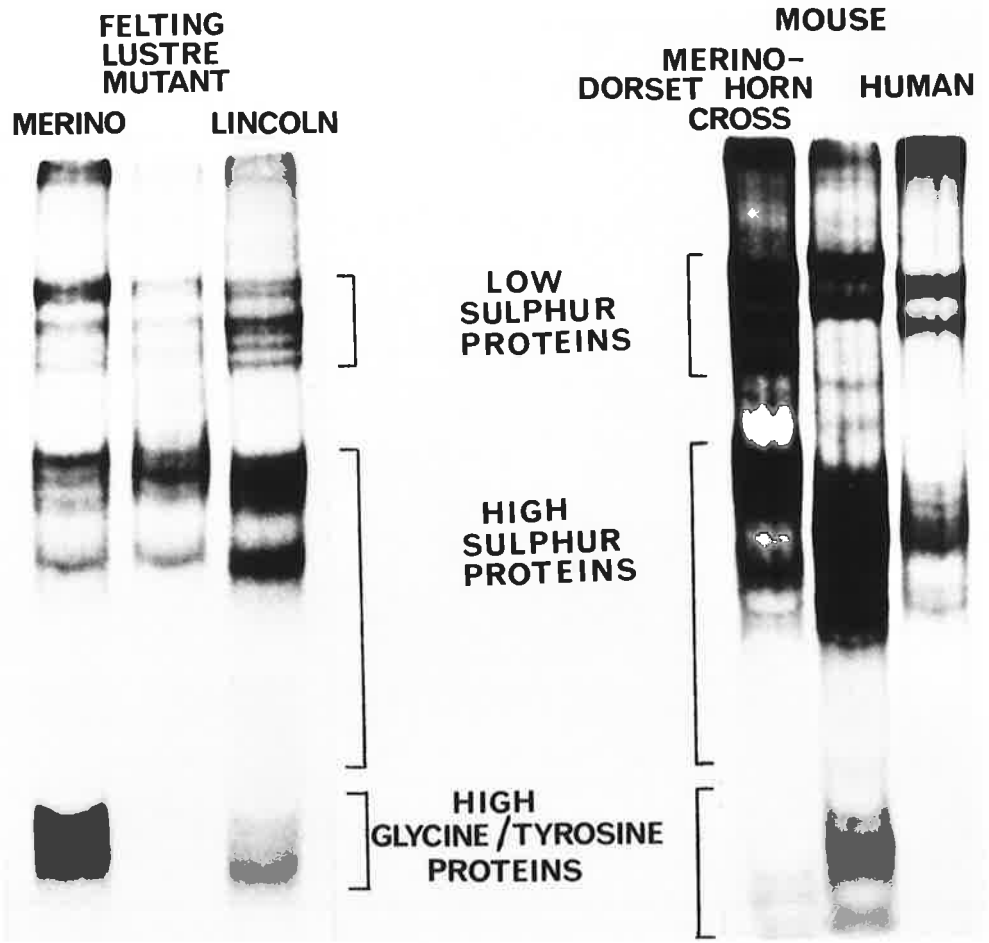
Intra and inter-species variations were evident for the LS and HS components, but only the differences observed for the HGT proteins will be discussed here. The relative abundance of the HGT proteins in these hair samples (Figure 8.2) was in agreement with previously reported studies (Gillespie and Frenkel, 1974; 1976; Gillespie and Marshall, 1980). Their observations were as follows: human hair (and nail) are devoid of HGT proteins, whereas in the mouse they constitute 18% of the total protein in the hair. For comparison, a wool sample from the breed of sheep used as the primary source of tissue in this thesis, Merino-Dorset Horn cross (HGT content about 3%), was co-electrophoresed on the same gel. The findings of Gillespie (1983) and co-workers indicated that fractionation of mouse HGT proteins, under the same conditions as those used for sheep wool, gave rise to electrophoretic patterns of similar heterogeneity to those seen for the wool proteins. Whether the

FIGURE 8.2

ONE-DIMENSIONAL SDS-UREA POLYACRYLAMIDE GEL ELECTROPHORESIS OF VARIOUS WOOL AND HAIR SAMPLES

^{14}C-labelled wool and hair proteins (2.2.E.1) were electrophoresed on SDS-urea 15% polyacrylamide gels (2.2.E.3). The distribution of the three major wool protein families are indicated for each fluorogram. Note: for the left-hand fluorogram, the HGT protein region of gel was exposed for a longer time relative to the regions of lower mobility.

Samples include Merino, Merino-Dorset Horn cross, Felting lustre-mutant Merino and Lincoln wool, as well as mouse and human hair.


FELTING
LUSTRE
MUTANT
MERINO
LINCOLN
MOUSE
MERINO-
DORSET HORN
CROSS
HUMAN
LOW
SULPHUR
PROTEINS
HIGH
SULPHUR
PROTEINS
HIGH
GLYCINE / TYROSINE
PROTEINS

co-electrophoretic similarities reflect amino acid sequence homology between the mouse hair and wool proteins is not known at present.

Wool containing low levels of HGT proteins have been reported (Gillespie and Darskus, 1971) and in Figure 8.2 samples from Merino wool (HGT content 3-12%), Lincoln wool (HGT content < 1%) and Felting lustre-mutant Merino wool (undetectable levels of HGT proteins) are compared. Even after long exposure of the autoradiogram (data not shown) labelled protein material in the Felting lustre-mutant Merino sample was not detected in the region of the gel where HGT proteins characteristically migrate. In Figure 8.3, the characteristic fleece appearance of these three breeds of sheep is shown and although differences between breeds are readily apparent, they are unlikely to be due solely to differences in the HGT protein content of the wool.

8.2.B Sequences Homologous to Sheep HGT Keratin mRNAs in Human, Mouse and Various Sheep Genomes

With the availability of HGT cDNA probes (isolated and characterized as described in Chapters 5 and 7) it was hoped to obtain answers at the molecular level to the following questions:

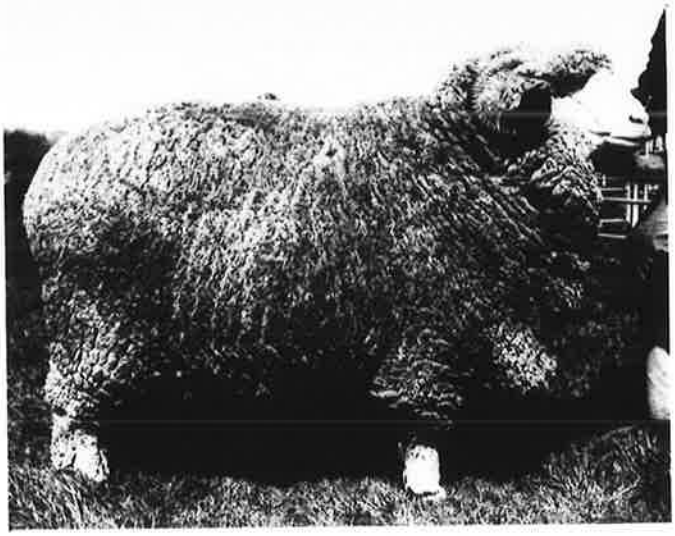
(1) is there inter-species homology between sheep and mouse HGT proteins
(2) is there intra-species homology between the Merino and Lincoln sheep breeds
(3) how extensive are any homologies
(4) what are the complexities of the HGT gene families
(5) can HGT genes be detected in the genomes of the 'Naked' mouse and Felting lustre-mutant Merino

Both the HGT-F and HGT-C2 cDNA clones (pSHGT-1 and pSHGT-C2.1) were used to probe genomic blots of human, several mouse and various sheep DNAs. The resultant autoradiograms, after washing at moderately low stringency (1xSSC/0.1% SDS at 65°C), are shown in Figure 8.4. The results

FIGURE 8.3

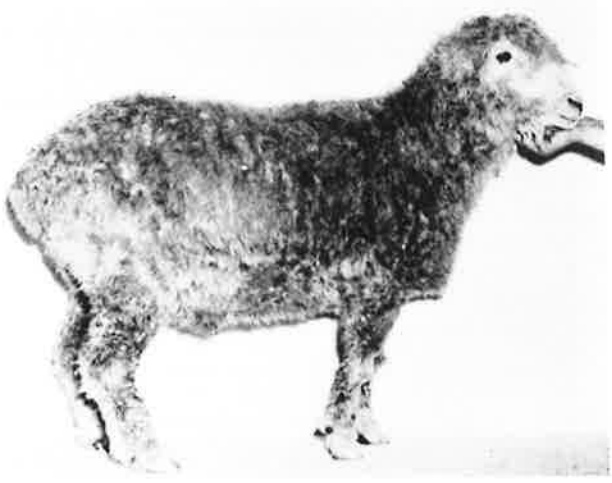
FLEECE CHARACTERISTICS OF MERINO, LINCOLN AND FELTING LUSTRE-MUTANT MERINO SHEEP

The photographs opposite show the characteristic fleece appearance of the three breeds of sheep discussed in the text. The main phenotypic difference between these three breeds, is the lack of crimp observed in the wool of the Felting lustre-mutant Merino.
Photographs courtesy of P. Reis and M. Fleet.



Merino

Lincoln



Felting Lustre
mutant Merino

FIGURE 8.4

DETECTION OF HGT-F AND HGT-C2 RELATED SEQUENCES IN SHEEP, HUMAN AND MOUSE GENOMES

Various genomic DNAs, prepared as described in 2.2.S, were digested with either HindIII (human DNA) or EcoRI (all other samples) and electrophoresed on 1% agarose gels. Following transfer of the DNA to nitrocellulose, the filters were hybridized to either ^{32}P-labelled (2.2.H.7) insert from pSHGT-1 or pSHGT-C2.1. A HindIII digest of λ DNA was co-electrophoresed on the agarose gels and used as a molecular weight marker. Fragment sizes are in kb.

Separate HindIII digests of human DNA were used for the hybridization to the C2 and F probes. However, for all other samples, the DNA blots were first hybridized the C2 probe, then reused for hybridization with the F probe.

The arrows, in the three sheep DNA samples, indicate fragments specifically hybridizing to pSHGT-1 i.e., present in the C2+F track but not the C2 track.

TRACKS:

Merino DNA

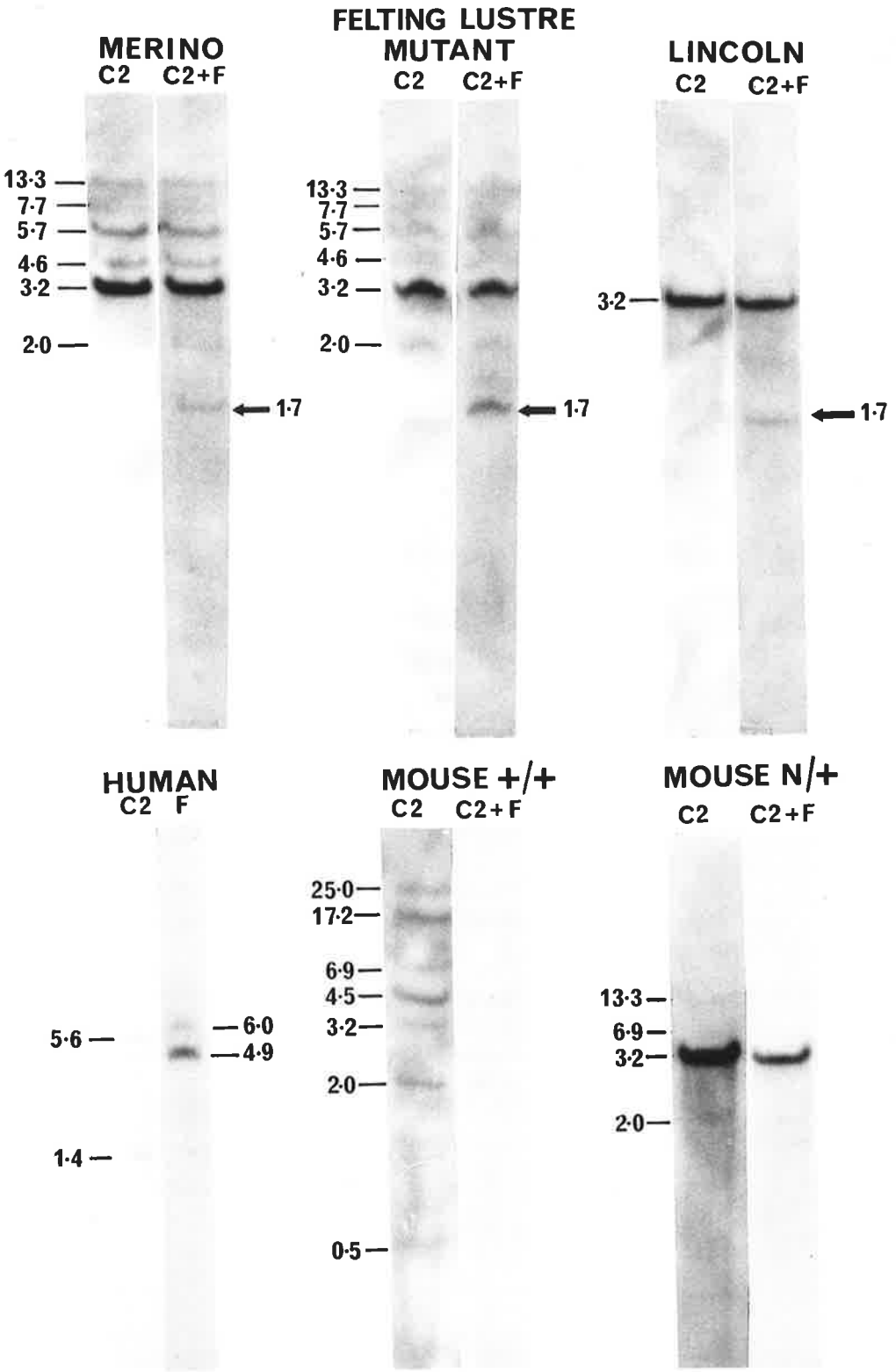
Felting lustre-mutant Merino DNA

Lincoln DNA

Human DNA

Mouse DNA - normal (+/+)

Mouse DNA - heterozygote (N/+) for the 'Naked' phenotype


MERINO
C2
C2+F
13·3
7·7
5·7
4·6
3·2
2·0
1·7
FELTING LUSTRE MUTANT
C2
C2+F
13·3
7·7
5·7
4·6
3·2
2·0
1·7
LINCOLN
C2
C2+F
3·2
1·7
HUMAN
C2
F
5·6
6·0
4·9
1·4
MOUSE +/+
C2
C2+F
25·0
17·2
6·9
4·5
3·2
2·0
0·5
MOUSE N/+
C2
C2+F
13·3
6·9
3·2
2·0

have been combined from several Southern blot experiments so that relative intensities of individual bands can not be directly compared.

The HGT-C2 probe detected multiple components, ranging in size from 0.5 kb to 25 kb, in most cases except Lincoln DNA which had one hybridizing fragment of 3.2 kb and human DNA which had two fragments 1.4 kb and 5.6 kb in size. On the other hand, the HGT-F cDNA probe detected two fragments, (4.9 kb and 6 kb) in human DNA, one fragment (1.7 kb) in each of the three sheep DNA samples, and no detectable hybridization with mouse DNA. These Southern blot results are discussed in more detail below.

8.3 DISCUSSION

DNA probes complementary to the only two HGT proteins sequenced to date, namely components C2 and F from Merino wool, were used to detect HGT-related sequences in the human, mouse and other sheep genomes. Hybridizing sequences were apparent in all cases with either the HGT-F, HGT-C2 or both probes (Figure 8.4), indicating the presence of HGT-related sequences in the various genomes. A similar but more extensive cross-species study, carried out by Fuchs et al., (1981) and Fuchs and Marchuk (1983), both at the protein and DNA level, has indicated homology between the human epidermal keratins and a wide variety of vertebrate species. The human epidermal and sheep low-sulphur wool keratins have similarities with each other, which places them within the cytokeratin protein class of the intermediate filaments (see 1.2.B.1).

Human sequences homologous to sheep HGT keratin mRNAs

The detection of HGT sequence-related DNA fragments in human DNA was a surprising result since HGT proteins have not been detected in either human hair or nail (Gillespie and Marshall, 1980). Glycine-rich regions are present in the human epidermal keratin proteins (Fuchs and

Hanukoglu, 1983), but comparison of DNA sequences from the HGT and human epidermal cDNA clones (Hanukoglu and Fuchs, 1982; 1983) suggested that there was insufficient sequence homology to account for the cross-hybridization observed in the genomic blots. Therefore the nature of these human HGT-related DNA fragments is not known and requires further investigation, such as their isolation and characterization by sequence analysis.

HGT-related sequences in the normal and 'Naked' mouse

The similarity of gel electrophoretic patterns for sheep and mouse HGT proteins (Gillespie, 1983) and the observation that low-sulphur and high-sulphur wool keratin genes cross-hybridize with mouse genomic sequences (B. Powell, personal communication), would tend to suggest that sequences homologous to sheep HGT keratins are likely to be found in the mouse genome. Southern blot analysis showed that HGT-C2 (but no HGT-F) homologous sequences were present, and were of a similar number to that found in the Merino. As mentioned below, this high copy number in both the Merino and mouse genomes, may suggest that the C2 sequence reflects a basic form of HGT protein structure conserved in a number of species.

When the blot data for the heterozygote mouse mutant ('Naked') was examined, it was found that in addition to the disappearance of three bands (17.2 kb, 6.9 kb and 0.5 kb), the 3.2 kb fragment was of greater intensity than the corresponding band in the normal mouse DNA. It is possible that the increase in intensity reflects a number of similar sized fragments detected by the probe. Whether this altered Southern blot pattern (as compared to the normal mouse results) can be directly correlated with decreased HGT protein content in the mutant mouse hair, cannot be determined until mouse HGT protein and/or gene sequence data becomes available.

It was not possible to obtain a DNA sample from a homozygote

'Naked' mouse, since these mice usually die in utero (Tennenhouse and Gold, 1976). It is interesting to speculate that perhaps the hybridization pattern for this mutant mouse DNA would show fewer hybridizing fragments than that seen in the heterozygote DNA blot, thereby implicating additional loss of HGT-related sequences from the mutant mouse genome.

HGT genes in the Lincoln and Felting lustre-mutant Merino

In the absence of HGT protein sequence information for Lincoln wool, it is not known whether these proteins are similar to the HGT proteins isolated and sequenced from Merino wool. However, the Southern blot data described in 8.2.C, showed single HGT-F and HGT-C2 related DNA fragments in the Lincoln genome, which corresponded in size to the HGT-F and C2 gene-containing fragments in Merino DNA. This result suggested that protein components highly homologous (and possibly identical) to components HGT-F and C2 were present in Lincoln wool. The lack of multiple C2-like sequences distinguishes this hybridization result from the mouse and other sheep DNA Southern blot results and may indicate genetic variations between the two breeds of sheep (Merino and Lincoln), resulting in the altered HGT protein content in the respective wool fibres.

Another surprising result was the finding that HGT-F and C2 sequences are present in the genome of the Felting lustre-mutant Merino which contains negligible levels of HGT proteins in its wool. In fact, the hybridization pattern of each probe to the two different Merino DNA samples was found to be identical. The detection of HGT sequences in the mutant Merino suggested that the lack of expression of HGT genes could not be attributed to the absence of these sequences, but that the negligible level of the corresponding proteins in this particular wool fibre was probably due to other defects occuring in the transcriptional or

translational processes. For example, since all the HGT proteins are affected in this mutation, the co-ordinate expression of the HGT keratin genes may be caused by a mutation in a common regulatory signal, so that HGT gene transcripts are not produced. Northern blot analysis could test this hypothesis by probing for the presence of HGT mRNAs in the follicles of Felting lustre-mutant Merino. Alternatively, the inability to efficiently translate HGT mRNAs (perhaps due to mRNA instability) may reflect a defect in the translational control of these proteins.

Finally, the finding that HGT-C2 related sequences were generally of a greater number and present in more species that HGT-F sequences, would seem to suggest that component C2 contains a basic amino acid sequence arrangement found in HGT proteins or analogous proteins in other keratin structures. This speculation is supported by the fact that the glycine-rich segment of scale keratin is closely related to component C2 but not component F. See 1.3.C and Figure 1.3.

The results presented in this chapter have raised a number of questions concerning gene complexity, evolutionary protein homology and possible control mechanisms in HGT gene expression. By using 3' non-coding, 5' non-coding probes isolated from the HGT clones and differing degrees of washing stringency, further investigation of the heterogeneity and sequence-relatedness of the HGT genes in the sheep and other mammalian genomes is now possible.

CHAPTER 9

SUMMARY AND CONCLUDING DISCUSSION

CHAPTER 9

SUMMARY AND CONCLUDING DISCUSSION

The work detailed in this thesis dealt with the preliminary investigation into the structure and organization of the HGT wool keratin genes and as such provided an extension to the concomitant structural studies on the other keratin gene families expressed in the wool fibre, namely the low-sulphur and high-sulphur keratin genes (Powell et al., 1983; K. Ward, B. Powell, M. Frenkel, personal communications). These studies have focussed on a number of questions, such as:

(1) the number of members in each gene family and how are they organized
(2) the relationship between linked keratin genes
(3) the sequence homology between the genes in the protein-coding and in the 5' and 3' non-coding regions
(4) sequence homology between the genes in the 5' and 3' flanking regions
(5) are the various gene families clustered.

As yet not all of the above, or other related questions such as, when during the growth of the fibre are the different gene families expressed and to what relative level, have been adequately answered for any one of the three keratin gene families and further gene isolation and sequence analysis is required.

The availability of specific keratin gene probes provides the means with which to study the molecular mechanisms of control exercised in the co-ordinated expression of the keratin genes during fibre growth and in the fluctuations observed in the wool keratin protein families in response to dietary, chemical and hormonal influences. These variabilities may reflect changes in gene transcription or mRNA translation, but nothing at present is known about the types of control

that might be involved with these changes.

In order to study the genomic organization and expression of the HGT keratin genes, the cloning of DNAs complementary to wool follicle mRNA was undertaken. As described in Chapter 3, a sheep wool follicle cDNA library was constructed by the standard procedure of G/C tailing into the plasmid vector, pBR322. The initial screening approach for the detection of HGT cDNA sequences in the cDNA library, involved the preparation of RNA enriched for HGT mRNAs (Chapter 4). Using a Merino-Dorset Horn cross sheep (with a wool HGT protein content of about 3%) as the source of starting RNA material, an RNA fraction was obtained which when examined by cell-free translation, encoded predominantly HGT protein products. Conversion of this RNA fraction into ^{32}P-labelled cDNA, followed by screening of the cDNA library with this cDNA probe, yielded a single strong positively-hybridizing colony. This clone, p11E6, was analyzed by DNA sequencing but could not be identified due to absence of protein-coding sequence. The nature of the probe or the clone p11E6, was not examined further so that an explanation for this screening result was not elucidated during the course of this work.

HGT cDNA clones were eventually isolated using synthetic oligonucleotide probes as described in Chapters 5 and 7. The cDNA clones were detected using synthetic DNA probes complementary to different regions of the two HGT wool proteins (component F and C2) whose amino acid sequences were known (Dopheide, 1977; Marshall *et al.*, 1980). Sequence analysis of the cloned DNA sequences and examination of the derived amino acid sequences showed that:

(1) the cDNA clones were not full-length as they did not contain 5' non-coding or complete protein-coding regions. However, the derived amino acid sequences matched well with the published protein sequences which established the identity of the clones.

(2) the derived amino acid sequence from the HGT-C2 DNA clone showed no

discrepancies with the published protein sequence, whereas that from the HGT-F cDNA clones had three discrepancies which were most likely due to protein sequence error. Firstly, at amino acid position 45, the DNA sequence coded for Asn rather than Asp and secondly, at amino acid positions 47-48, there was a Ser-Gly pair, which was inverted (Gly-Ser) in the published protein sequence.

(3) the 3' non-coding regions of the HGT-F and HGT-C2 cDNA sequences (although not full-length in the case of the HGT-C2 clone) showed no sequence homology except for a pentanucleotide sequence, 5'-CCAGA-3', located 12 nucleotides downstream from each termination codon. Whether this short conserved sequence has a role in the transcription or translation of the HGT sequences remains to be elucidated.

(4) as expected from the low homology observed between the two proteins and also between the 3' non-coding regions of the cDNA sequences, the HGT-F and HGT-C2 clones did not cross-hybridize, even though they are members of the same HGT sub-family.

(5) the two HGT cDNA clones did not contain any sequence homology with the clone, p11E6 isolated as described above.

In Chapters 6, 7 and 8, the HGT cDNA clones were used to detect the corresponding genes or related sequences in several mammalian, including various sheep genomes. Southern blot analysis indicated that sequences cross-hybridizing with the sheep HGT probes were present in the human genome, which was surprising since it has been shown by Gillespie and Marshall (1980), that human hair and nail do not contain HGT proteins. Significant homology was not apparent between the HGT proteins and the glycine-rich regions of the human epidermal keratin sequences (Hanukoglu and Fuchs, 1982; 1983), so that confirmation of the identity of these hybridizing fragments awaits their isolation and DNA sequence analysis.

Perhaps not so surprising was the Southern blot results which showed that mouse genomic sequences cross-hybridize with the sheep HGT

cDNAs. The precedent has been set for cross-species mouse/sheep hybridization in the detection of high-sulphur wool keratin-related sequences and also human epidermal keratin-related sequences in the mouse genome (B. Powell, personal communication; Fuchs and Marchuk (1983), respectively). Also Gillespie (1983) has reported the similarity between the gel electrophoretic patterns of the Type I HGT proteins extracted from mouse hair and wool. The cross-hybridization observed between the HGT probes and mouse DNA strongly suggests that sequence homology is likely to exist between the HGT proteins from these two species.

Since the physical nature of the HGT proteins has made their isolation and characterization difficult by physico-chemical techniques, the genomic Southern blot results with sheep DNA were of central importance to this thesis, as a means of determining the extent of HGT protein heterogeneity in wool. The genomic blot data revealed that HGT-F was an unique gene, while the HGT-C2 probe detected at least five weakly hybridizing species in addition to the HGT-C2 gene. Early protein fractionation data (Gillespie and Frenkel, 1974) indicated that component F was a single homogenous protein species, whereas component C2 had two other related species, one of which was sequenced and found to be identical to C2 (Gillespie et al., 1980). It is possible that genes for other members of the HGT protein family, for which protein sequence data is not available, reside on the multiple components hybridizing to the HGT-C2 probe. If this is the case, then the abundance of C2-like sequences may indicate that component C2 represents the most common structural form of Type I HGT proteins. This is supported by the finding that multiple HGT-C2 like sequences are present in the mouse and other sheep genomes, unlike component F which has at most two cross-hybridizing fragments in the human genome. Determination of the HGT gene family heterogeneity, both in size and sequence, is dependent on further protein or gene sequence information. In this respect, the HGT-C2 probe can be

used to isolate the C2-related sequences (observed in the Southern blot) from the sheep genome, and their coding capabilities established by DNA sequencing. However, amino acid composition data has indicated that the Type II proteins, which constitute the other HGT wool keratin sub-family, are likely to differ significantly from the Type I sequeneces. Therefore, Type II protein sequence data is required so that the synthetic probe approach (used in Chapters 5 and 7) can be utilized to isolate the corresponding Type II HGT genes. Once obtained, this information will allow further comparison of HGT protein sequence homologies.

The existence of genetic variants in various gene products, as seen by a phenotypic change in an individual, provides a means of examining the expression of the corresponding genes. The Felting lustre-mutant Merino is an example of mutant fleece-type which contains negligible levels of HGT proteins in its wool compared to a normal Merino. Use was made of the HGT cDNA probes to try and determine the genetic lesion affecting the expression of the HGT proteins in this breed of sheep. Southern blot analysis showed that both the HGT-F and C2 genes were present and the identical hybridization patterns obtained with Merino and mutant-Merino DNA, indicated that there was no gross rearrangement of the HGT genes. The mutation was therefore not due to the absence of the structural genes, but rather a defect in the expression of the genes either at the transcriptional or translational level was responsible.

The gene for HGT component F was isolated and sequenced (Chapter 6) in its entirety, together with about 1.5 kb of flanking DNA. The gene contained a number of interesting structural features such as:

(1) the lack of introns, also shown to be absent from the high-sulphur keratin genes which encode the other matrix proteins of wool (Powell et al., 1983)

(2) the conventional eukaryotic consensus signals; 'CAAT', 'TATA' and cap site sequences

(3) an 18 base sequence element directly preceding the initiation codon, ATG, and conserved in almost all other high-sulphur kerain genes. Whether this sequence is important for the expression of the genes encoding the wool matrix proteins is not known.

(4) a potential 'Z-DNA' forming region in the 5' flanking region in proximity to short viral-like enhancer sequences.

(5) a sheep-specific highly-repeated sequence in the 5' flanking region also found in the high-sulphur keratin genes (B. Powell, personal communication).

A cosmid library is currently being screened for the HGT-C2 gene, in order to analyse the gene at the gross level of organization with respect to possible linkage with other keratin genes, and also at nucleotide level to look for conserved sequences between HGT-F and C2, which may be required for the selective expression of the HGT keratin genes.

The ultimate determination of the role of specific DNA sequences in transcriptional regulation of the various keratin gene families, which leads to co-ordinated keratin protein synthesis during wool fibre formation, requires *in vitro* manipulation of gene sequences and testing the effect on the transcription of these genes both *in vitro* and *in vivo*. The fact that it has been shown that DNA can be inserted into a recipient genome (homologous or heterologous) in such a way that the inserted DNA is stably inherited and in some instances expressed as a gene product (Brinster *et al.*, 1982; Constantini and Lacey, 1983), clearly establishes a vast potential for the application of such technology to the wool industry. The wool keratin genes may be potentially involved in gene transfer technology aimed at animal improvement, such as altered wool composition. The isolation of the various wool keratin genes is therefore the major step towards investigating the expression of these economically important proteins at the molecular level, and subsequently through the

use of molecular biological techniques in the reorganization of conventional animal breeding techniques.

BIBLIOGRAPHY

Abelson, J. (1979) *Ann. Rev. Biochem.* 48, 1035-1069

Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci.* 69, 1408-1412

Banaszuk, A. M., Deugau, K. V., Sherwood, J., Michalak, M. and Glick, B.R. (1983) *Anal. Biochem.* 128, 281-286

Banerji, J., Olson, L. and Schaffner, W. (1983) *Cell* 33, 729-740

Baralle, F. E. (1983) *Int. Rev. Cytol.* 81, 71-106

Beaucage, S. L. and Caruthers, M. H. (1981) *Tetrahedron Letters* 22(20), 1859-1862

Behe, M. and Felsenfeld, G. (1981) *Proc. Natl. Acad. Sci.* 78, 1619-1623

Bendit, E. G. and Gillespie, J. M. (1978) *Biopolymers* 17, 2743-2745

Benoist, C and Chambon, P. (1981) *Nature* 290, 304-309

Benoist, C., O'Hare, K. O., Breathnach, R. and Chambon, P. (1980) *Nuc. Acids Res.* 8, 127-142

Benton, W. D. and Davis, R. W. (1977) *Science* 196, 180-182

Berget, S. M. (1984) *Nature* 309, 179-181

Berget, S., Moore, C. and Sharp, P. A. (1977) *Proc. Natl. Acad. Sci.* 74, 3171-3175

Birnboim H. C. and Doly, J. (1980) *Nuc. Acids Res.* 7, 1513-1523

Bogenhagen, D. F., Sakonju, S. and Brown, D.D. (1980) *Cell* 19, 717-728

Bolivar, F., Rodriguez, R. L., Greene, P. J., Bettach, M. C., Heymeker, H. L. and Boyer, H. W. (1977) *Gene* 2, 95-113

Bollum, F. J. (1968) *Methods in Enzymology* XIIB, 169-173

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) *Proc. Natl. Acad. Sci.* 75, 4853-4857

Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383

Brinster, R. L., Chen, H. Y., Warren, R., Sarthy, A. and Palmiter, R. (1982) *Nature* 296, 5852-5855

Brooker, J. D., May, B. K. and Elliott, W. H. (1980) *Eur. J. Biochem.* 106, 17-24

Casadaban, M. J. and Cohen, S. N. (1980) *J. Mol. Biol.* 138, 179-207

Chan, S. J., Noyes, B. E., Agarwal, K. L. and Steiner, D. F. (1979) *Proc. Natl. Acad. Sci.* 76, 5036-5040

Chandler, V. L., Maler, B. A. and Yamamoto, K. R. (1983) *Cell* 33, 489-499

Chapman, R. E. and Gemmell, R. T. (1971) *J. Ultrastruct. Res.* 36, 342-354

Charnay, p., Treisman, R., Mellon, P., Chao, M., Axel, R. and Maniatis, T. (1984) *Cell* 38, 251-263

Chin, D. J., Gil, G., Russell, D. W., Liscum, L., Luskey, K. L., Basu, S. K., Okayama, H., Berg, P., Goldstein, J. L. and Brown, M. S. (1984) *Nature* 308, 613-617

Choo, K. H., Gould, K. G., Rees, D. J. G. and Brownlee, G. G. (1982) *Nature* 299, 178-180

Colantuoni, U., Romano, V., Bensi, G., Santoro, C., Constanzo, F., Raugei, G. and Cortese, R. (1983) *Nuc. Acids Res.* 11, 7769-7776

Collins, R. and Hohn, B. (1978) *Proc. Natl. Acad. Sci.* 75, 4242-4246

Conrad, S. E. and Botchan, M. R. (1982) *Mol. Cell. Biol.* 2, 949-965

Constantini, F. and Lacey, R. (1981) *Nature* 294, 92-94

Corden, J., Wasylyk, B., Buchwalder, R., Sasson-Corsi, P., Kedinger, C. and Chambon, P. (1980) *Science* 209, 1406-1414

Crewther, W. G. (1976) *Proc. 5th Intl. Wool Text. Res. Conf., Aachen* (1975) 1, 1-101

Crewther, W. G., Dowling, L. M., Gough, K. M., Inglis, A. S. and Parry, D. A. D. (1982) *Proc. 12th Intl. Cong. of Biochem., Perth* Abst. POS 004-205

Crewther, W. G., Dowling, L. M., Steinert, P. M. and Parry, D. A. D. (1983) *Intl. J. Biol. Macromol.* 5, 267-274

Dagert, M. and Ehrlich, S. D. (1979) *Gene* 6, 23-28

Darnell, J. E., Wall, R. and Tushinski, R. J. (1971) *Proc. Natl. Acad. Sci.* 68, 1321-1325

Darskus, R. L. (1972) *J. Chromatogr.* 69, 341-348

Davidson, E. H. and Britten, R. J. (1979) *Science* 204, 1052-1059

Davis, R. W., Thomas, M., Cameron, J., St John, T. P., Scherer, S. and Padgett, R. A. (1979) *Methods in Enzymology* 68, 404-411

DeVilliers, J. and Schaffner, W. (1981) *Nuc. Acids Res.* 9, 6251-6265

Diaz-Ruis, J. R. and Kaper, J. M. (1978) *Prep. Biochem.* 8, 1-17

Dierks, P., Van Ooyen, A., Mantei, N. and Weissman, C. (1981) *Proc. Natl. Acad. Sci.* 78, 1411-1415

Doerfler, W. (1983) *Ann. Rev. Biochem.* 52, 93-124

Dopheide, T. A. A. (1973) *Eur. J. Biochem.* 34, 120-124

Dowling, L. M., Gough, K. H., Inglis, A. S. and Sparrow, L. G. (1979) *Aust. J. Biol. Sci.* 32, 437-442

Dowling, L. M., Parry, D. A. D. and Sparrow, L. G. (1983) *Bio. Sci. Rep.* 3, 73-78

Downes, A. M., Sharry, L. F. and Rogers, G. E. (1963) *Nature* 199, 1059-1061

Dynan, W. S. and Tjian, R. (1983) *Cell* 32, 669-680

Ebling, F. J. and Hale, P. A. (1983) in *Biochemistry and Physiology of the Skin* Ed. Goldsmith, L.A. Oxford University Press I, 522-525

Efstratiadis, A., Kafatos, F. C., Maxam, A. and Maniatis, T. (1976) *Cell* 7, 279-288

Efstratiadis, A., Posakony, J. N., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B., Weissman, S. M., Slightom, I. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulder, C. C. and Proudfoot N. J. (1980) *Cell* 21, 653-668

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

Everett, R. D., Baty, D. and Chambon, P. (1983) *Nuc. Acids Res.* 11, 2447-2464

Feinberg, A. and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6-13

Felber, B. K., Orbin, S. H. and Hamer, D. H. (1982) *Cell* 29, 895-902

Fraser, I. E. B. (1969a, b) *Aust. J. Biol. Sci.* 22, 213-229 and 231-238

Fraser, R. D. B., Gillespie, J. M. and MacRae, T. P. (1973) *Comp. Biochem. Physiol.* 44B, 943-947

Fraser, R. D. B. and MacRae, T. P. (1980) in *The Mechanical Properties of Biological Materials* Eds. Vincent, J. F. and Currey, J. D. Society for Experimental Biology, London 211-246

Fraser, R. D. B., MacRae, T. P. and Rogers, G. E. (1972) *Keratins Their Composition, Structure and Biosynthesis* Ed. Thomas C. C. Illinois

Fraser, R. D. B. and Rogers, G. E. (1955) *Aust. J. Biol. Sci.* 8, 288-299

Frenkel, M. J. (1977) *Proc. 10th Aust. Biochem.Soc. Meeting*, 21

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

Fromm, M. and Berg, P. (1982) *J. Mol. Appl. Genet.* 1, 457-481

Fuchs, E. V., Coppock, S. M., Green, H. and Cleveland, D. W. (1981) *Cell* 27, 75-84

Fuchs, E. and Hanukoglu, I. (1983) *Cell* 34, 322-334

Fuchs, E. and Marchuk, D. (1983) *Proc. Natl. Acad. Sci.* 80, 5857-5861

Fukumaki, Y., Ghosh, P. K., Benz, E. J., Reddy, V. B. and Lebowitz, P. (1982) *Cell* 28, 585-593

Geisler, N., Kaufmann, E. and Weber, K. (1982) *Cell* 30 277-286

Geisler, N. and Weber, K. (1982) *EMBO J.* 1, 1649-1656

Gil, A. and Proudfoot, N. J. (1984) *Nature* 312, 473-474

Gilbert, W. (1978) *Nature* 271, 501

Gillespie, J. M. (1968) *Proc. Aust. Biochem. Soc.* 1, 29

Gillespie, J. M. (1972) *Comp. Biochem. Physiol.* 41B, 723-734

Gillespie, J. M. (1983) in *Biochemistry and Physiology of the Skin* Ed. Goldsmith, L. A. Oxford University Press II, 475-510

Gillespie, J. M. and Darskus, R. L. (1971) *Aust. J. Biol. Sci.* 24, 1189-1197

Gillespie, J. M. and Frenkel, M. J. (1974) *Aust. J. Biol. Sci.* 27, 617-627

Gillespie, J. M. and Frenkel, M. J. (1975) *Comp. Biochem. Physiol.* 47B, 339-346

Gillespie, J. M. and Frenkel, M. J. (1976) *Proc. 5th Intl. wool Text. Res. Conf. Aachen.* (1975) 1, 265-276

Gillespie, J. M., Frenkel, M. J. and Reis, P. J. (1980) *Aust. J. Biol. Sci.* 33, 125-136

Gillespie, J. M. and Marshall, R. C. (1980) *Cosmetics and Toiletries* 95, 29-34

Gillespie, J. M. and Marshall, R. C. (1983) *J. Invest. Dermat.* 80, 195-202

Gillespie, J. M., Marshall, R. C., Moore, G. P. M., Panaretto, B. A. and Robertson, D. M. (1982) *J. Invest. Dermat.* 79, 197-200

Gillespie, J. M., Reis, P. J. and Schinckel, P. G. (1964) *Aust. J. Biol. Sci.* 548-560

Gillespie, J. M. and Reis, P. J. (1966) *Biochem. J.* 98, 669-677

Glisin, V., Crkvenjakov, R. and Byus, C. (1974) *Biochem.* 13, 2633-2637

Goddard, D. R. and Michaelis, L. (1934) *J. Biol. Chem.* 106, 605-614

Goldberg, M. L. (1979) Ph.D Thesis, Stanford University

Grosschedl, R. and Birnstiel, M. L. (1980) *Proc. Natl. Acad. Sci.* 77, 1432-1436 and 7102-7106

Grosveld, G. C., deBoer, E. and Shewmaker, C. K. (1982)a *Nature* 295, 120-126

Grosveld, G. C., Shewmaker, C. K., Jat, P. and Flavell, R. A. (1982)b *Cell* 25, 215-226

Grunstein, M. and Hogness, D. S. (1975) *Proc. Natl. Acad. Sci.* 72, 3961-3965

Guarante, L. and Mason, T. (1983) *Cell* 32, 1279-1286

Gurdon, J. B., deRoberts, E. M. and Darlington, G. (1976) *Nature* 260, 116-120

Hagenbüchle, O., Bovey, R. and Young, R. A. (1980) *Cell* 21, 179-187

Hanukoglu, I. and Fuchs, E. (1982) *Cell* 31, 243-252

Hanukoglu, I. and Fuchs, E. (1983) *Cell* 33, 915-924

Hipskin, R. A. and Clarkson, S. G. (1983) *Cell* 34, 881-890

Hobart, P., Crawford, R., Shen, L., Pictet, R. and Rutter, W. I. (1980) *Nature* 288, 137-141

Hofer, E. and Darnell, J. E. (1981) *Cell* 23, 585-593

Horio, M and Kondo, T. (1953) *Text. Res. J.* 23, 373-387

Huez, G., Marbaix, G., Gallwitz, D., Weinberg, E., Devos, R., Hubert, E. and Cleuter, Y. (1978) *Nature* 271, 572-573

Hynes, N., Van Ooyen, A. J. J., Kennedy, N., Herrlich, P., Ponta, H. and Groner, B. (1983) *Proc. Natl. Acad. Sci.* 80 , 3637-3641

Ish-Horowicz, D. and Burke, J. F. (1981) *Nuc. Acids Res.* 9, 2989-2998

Jones, R. W., Old, J. M., Trent, R. J., Clegg, J. B. and Weatherall, D. J. (1982) *British J. Haematol.* 50, 415-422

Jung, A., Sippel, A. E., Grez, M. and Schutz, G. (1980) *Proc. Natl. Acad. Sci.* 77, 5759-5763

Kaplin, I. J. and Whiteley, K. J. (1978) *Aust. J. Biol. Sci.* 31, 231-240

Kao, F., Martz, J. A., Law, M. L. and Davidson, J. N. (1982) *Proc. Natl. Acad. Sci.* 79, 865-869

Karin, M., Haslinger, A., Holtgreve, H., Richards, R., Krauter, P., Westphal, H. and Beato, M. (1984) *Nature* 308, 513-519

Kates, J. (1970) *Cold Spring Harbour Symp. Quant. Biol.* 35, 743-752

Kates, J. (1970) *Methods in Cell Biology* 7, 53-65

Khoury, G. and Gruss, P. (1983) *Cell* 33, 313-314

Knott, T. J., Robertson, M. E., Priestley, L. M., Urdea, M., Wallis, S. and Scott, J. (1984) *Nuc. Acids Res.* 12, 857-872

Kozak, M. (1984) *Nuc. Acids Res.* 12, 857-872

Kretschmer, P. J., Coon, H. C., Davis, A., Harrison, M. and Nienhuis, A. W. (1981) *J. Biol. Chem.* 256, 1975-1982

Kronenberg, H. M., Roberts, B. E. and Efstratiadis, A. (1979) *Nuc. Acids Res.* 6, 153-166

Kurtz, D. T. and Nicodemus, C. F. (1981) *Gene* 13, 145-152

Laemmli, U. K. (1970) *Nature* 227, 680-685

Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schütz, G. (19810 *Nuc. Acids Res.* 9, 2251-2266

Laskey, R. A. and Mills, A. D. (1975) *Eur. J. Biochem.* 56, 335-341

Lee, S. Y., Mendecki, J. and Brawerman, G. (1971) *Proc. Natl. Acad. Sci.* 68, 1331-1335

Legon, S., Glover, D. M., Hughes, J., Lowry, P. L., Rigby, P. W. J. and Watson, C. J. (1982) *Nuc. Acids Res.* 10, 7905-7919

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. *Nature* 283, 220-224

Ley, K. F. and Crewther, W. G. (1980) *6th Intl. Wool Text. Res. Conf.* Pretoria II, 13-28

Lindberg, U. and Persson, T. (1972) *Eur. J. Biochem.* 31, 246-254

Lindley, H. (1977) in *Chemistry of Natural Protein Fibres* Ed. Asquith, R.S. Plenum, New York 147-191

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular cloning: A laboratory manual* Cold Spring Harbour Press

Maniatis, T., Kee, S. G., Efstratiadis, A. and Kafatos, F. C. (1976) *Cell* 8, 163-182

Marchuk, D., McCrohon, S. and Fuchs, E. (1984) *Cell* 39, 491-498

Marshall, R. C. (1981) *Text. Res. J.* 51, 106-108

Marshall, R. C. (1983) *J. Invest. Dermat.* 80, 519-524

Marshall, R. C. and Gillespie, J. M. (1977) *Aust. J. Biol. Sci.* 30, 389-400

Marshall, R. C. and Gillespie, J. M. (1981) *Proc. 2nd Natl. Conf. on Wool Harvesting Res. and Develop.* Sydney Ed. Hudson, P.R.W. 117-121

Marshall, R. C. and Gillespie, J. M. (1982) *J. of the Forensic Sci. Soc.* 22, 377-385

Marshall, R. C., Gillespie, J. M., Inglis, A. S. and Frenkel, M. J. (1980) *Proc. 6th Intl. Wool Text. Res. Conf.* Pretoria II, 147-158

Maxam, A. M. and Gilbert, W. (1980) *Methods in Enzymology* 65, 499-560

McDonell, M. W., Simon, M. N. and Studier, F. W. (1976) *J. Mol. Biol.* 110, 119-145

McKnight, S. L., Kingsbury, R. C., Spencer, A. and Smith, M. (1984) *Cell* 37, 253-262

Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981) *Cell* 27, 279-288

Mercer, E. H. (1954) *Text. Res. J.* 24, 39-43

Mercer, E. H. (1961) *Keratin and Keratinization* Pergamon, Oxford

Mercer, E. H., Golden, R. L. and Jeffries, E. B. (1954) *Text. Res. J.* 24, 615-618

Messing, J. (1979) *Recombinant DNA Technical Bulletin* 2, 43-48

Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276

Molloy, P. L., Powell, B. C., Gregg, K., Barone, E. D. and Rogers, G. E. (1982) *Nuc. Acids Res.* 10, 6007-6021

Montagna, W. and Parakkal, P. F. (1974) in *The Structure and Function of Skin* 3rd Edition Academic Press, New York 172-258

Montgomery, D. l., Hall, B. D., Gillam, S. and Smith, M. (1978) *Cell* 14, 673-680

Moore, G. P. M., Panaretto, B. A. and Robertson, D. (1982) *Aust. J. Biol. Sci.* 35, 163-172

Moore, C. L. and Sharpe, P. A. (1984) *Cell* 36, 581-591

Mount, S. M. (1982) *Nuc. Acids Res.* 10, 459-472

Nevins, J. R., Blanchard, J. M. and Darnell, J. E. (1980) *J. Mol. Biol.* 144, 377-386

Nickoloff, J. A. and Hallick, R. B. (1982) *Nuc. Acids Res.* 10, 8191-8207

Noda, M., Shimizu, S., Tanabe, T., Takai, T., Kayano, T., Ikeda, T., Takahashi, H., Nakayama, H., Kanakoa, Y., Minamino, N., Kangawa, K., Matsuo, H., Raftery, M., Hirose, T., Inayama, S., Hayashida, H., Miyata, T., Numa, S. (1984) *Nature* 312, 121-127

Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* 26, 101-106

Nowock, J. and Sippel, A. E. (1982) *Cell* 30, 607-615

Noyes, B. E., Mevarech, M., Stein, R. and Agarwal, K. L. (1979) *Proc. Natl. Acad. Sci.* 76, 1770-1774

Nunberg, J. H., Kaufman, R. J., Chang, A. C. Y., Cohen, S. N. and Schimke, R. T. (1980) *Cell* 19, 355-364

Orwin, D. F. G. (1979) *Intl. Rev. Cytol.* 60, 331-374

Osborne, T. F., Avidson, D. N., Tyan, E. S., Dunsworth-Browne, M. and Berk, A. J. (1984) *Mol. Cell. Biol.* 4, 1293-1305

Palmiter, R. D. and Smith, L. T. (1973-1974) *Mol. Biol. Rep.* 129-134

Parnes, A. R., Velan, B., Felsenfeld, A., Ramanathan, L., Ferrini, U., Appella, E. and Seidman, J. G. (1981) *Proc. Natl. Acad. Sci.* 78, 2253-2257

Parry, D. A. D., Fraser, R. D. B. and MacRae, T. P. (1979) *Aust. J. Biol. Sci.* 26, 583-590

Peacock, S. L., McIver, C. M. and Monoban, J. J. (1981) *Biochem. Biophys. Acta.* 655, 243-250

Pelham, H. R. B. (1982) *Cell* 30, 517-528

Pelham, H. R. B. and Jackson, R. J. (1976) *Eur. J. Biochem.* 67, 247-256

Powell, B. C. and Rogers, G. E. (1985) in *Biology of the Integument* II Eds. Bereiter-Hahn, J. and Matoltsy, A. G. Springer-Verlag Berlin in press

Powell, B. C., Sleigh, M. J., Ward, K. A. and Rogers, G. E. (1983) *Nuc. Acids Res.* II, 5327-5346

Pribnow, D. (1975) *Proc. Natl. Acad. Sci.* 72, 784-788

Proudfoot, N. J. and Brownlee, G. G. (1976) *Nature* 263, 211-214

Quax, W., Egberts, W. V., Hendriks, W., Quax-Jenken, Y. and Bloemendal, H. (1983) *Cell* 35, 215-223

Rabbits, T. H. (1976) *Nature* 260, 221-225

Raphael, K. A., Marshall, R. C. and Pennycuik, P. R. (1984) *Genet. Res.* Cambridge 44, 29-38

Reis, P. J. (1979) in *Physiological and Environmental Limitations to Wool Growth* Eds. Black, J. L. and Reis, P. J. University of New England Publishing unit 223-242

Reyes, A. A., Johnson, M. J., Schold, M., Ito, H., Ike, Y., Morin, C., Itakura, K. and Wallace, R. B. (1981) *Immunogenetics* 14, 383-392

Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L. and Luskey, K. L. (1984) *Cell* 38, 275-285

Rich, A. (1983) *Cold Spring Harbour Symp. Quant. Biol.* 47, 1-12

Rich, A., Nordheim, A. and Wang, A. (1984) *Ann. Rev. Biochem.* 53, 791-846

Rigby, P. W. J., Deickmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251

Rimm, D. L., Horness, D., Kucera, J. and Blattner, F. R. (1980) *Gene* 12, 301-309

Roberts, B. E. and Paterson, B. M. (1973) *Proc. Natl. Acad. Sci.* 70, 2330-2334

Roman, R., Brooker, J. D., Seal, S. and Marcus, A. (1976) *Nature* 260, 359-360

Rogers, G. E. (1959)a *J. Ultrastruct. Res.* 2, 309-330

Rogers, G. E. (1959)b *Ann. NY. Acad. Sci.* 83, 378-399

Rogers, G. E. (1964) in *The Epidermis* Eds. Montagna, W. and Lobitz, W. C. Academic Press, New York 179

Rogers, J. (1983) *Nature* 305, 101-102

Rougeon, F. and Mach, B. (1976) *Proc. Natl. Acad. Sci.* 73, 3418-3422

Roychoudry R., Jay, E. and Wu, R. (1976) *Nuc. Acids Res.* 3, 863-877

Rudall, K. M. (1956) *Proc. Intl. Wool Text. Res. Conf.* Australia (1955) 176-185

Sadler, J. R., Waterman, M. S. and Smith, T. F. (1983) *Nuc. Acids Res.* 11, 2221-2230

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci.* 74, 5463-5467

Sargan, D. R., Gregory, S. P. and Nutterworth, P. H. W. (1982) *FEBS Letters* 147, 133-136

Seeburg, H. P., Shine, J., Martial, J. A., Baxter, J. D. and Goodman, H. M. (1977) *Nature* 270, 495-499

Setzer, D. R., McGeogan, M., Nunberg, J. H. and Schimke, R. T. (1980) *Cell* 22, 361-170

Short, B. F. (1958) *Nature* 181 1414-1415

Short, B. F., Wilson, P. A. and Schinckel, P. G. (1965) in *Biology of Skin and Hair Growth* Eds, Lyne, A. G. and Short, B. F. Angus and Robertson 409-426

Sims, R. T. (1967) *Nature* 213, 387-388

Singer-Sam, J., Simmer, R. L., Keith, D. H., Shively, L., Teplitze, M., Itakura, K., Gartler, S. M. and Riggs, A. D. (1983) *Proc. Natl. Acad. Sci.* 80, 802-806

Slightom, J. L., Blechl, A. E. and Smithies, O. (1980) *Cell* 21, 627-638

Smith, G. E. and Summers, M. D. (1980) *Anal. Biochem.* 109, 123-129

Sood, A. U., Pereira, D. and Weissman, S. M. (1981) *Proc. Natl. Acad. Sci.* 78, 616-620

Southern. E. (1975) *Methods in Enzymology* 68, 152-176

Sparrow, L. G. and Inglis, A. S. (1980) *6th Intl. Wool Text. Res. Conf.* Pretoria II, 237-246

Steinert, P. M., Idler, W. W. and Goldman, R. D. (1980) *Proc. Natl. Acad. Sci.* 77, 4534-4538

Steinert, P. M., Jones, J. C. R. and Goldman, R. D. (1984) *J. Cell Biol.* 99, 22s-27s

Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L. and Steven, A. C. (1983) *Nature* 302, 794-800

Stumph, W. E., Kristo, P., Tsai, M. J. and O'Malley, B. (1981) *Nuc. Acids Res.* 9, 5383-5397

Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. and Itakura, K. (1981) *Proc. Natl. Acad. Sci.* 78, 6613-6617

Sutcliffe,J. G. (1978) *Nuc. Acids Res.* 5, 2721-2728

Swank, R. T. and Munkres, K. D. (1971) *Anal. Biochem.* 39, 462-477

Swart, L. S. (1973) *Nature New Biol.* 243, 27-29

Swart, L. S., Joubert, F. J. and Parris, D. (1976) *5th Intl. Wool Text. Res. Conf.* (1975) Aachen II, 254-263

Szostak, J. M., Stiles, J. I., Tye, B. K., Chiu, P., Sherman, F. and Wu, R. (1979) *Methods in Enzymology* 68, 419-428

Tautz, D. and Renz, M. (1984) *Nuc. Acids Res.* 12, 4127-4138

Taya, Y., Devos, R., Tavernier, J., Cheroutre, H., Engler, G. and Fiers, W. (1982) *EMBO J.* 1, 953-958

Tennenhouse,H. S. and Gold, R. J. M. (1976) *Biochem. J.* 159, 149-160

Thomas, P. S. (1983) *Methods in Enzymology* 100B, 255-266

Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V. and Leder, P. (1978) *Proc. Natl. Acad. Sci.* 75, 725-729

Tosi, M., Young, R. A., Hagenbüchle, O. and Schibler, U. (1981) *Nuc. Acids Res.* 9, 2313-2323

Treisman, R., Proudfoot, N. J., Shander, M. and Maniatis, T. (1982) *Cell* 29, 903-911

Tsuda, M. and Suzuki, Y. (1981) *Cell* 27, 175-182

Tushinski, R. J., Sussman, P. M., Yu, L-Y. and Bancroft, F. C. (1977) *Proc. Natl. Acad. Sci.* 74, 2357-2361

Villa-Komaroff, L., Efstratiadis, S., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. and Gilbert, W. (1978) *Proc. Natl. Acad. Sci.* 75, 3727-3731

Vogt, V. M. (1973) *Eur. J. Biochem.* 33, 192-200

Wahl, G. M., Stern, M. and Stark, G. R. (1979) *Proc. Natl. Acad. Sci.* 76, 3683-3687

Wahli, W., Dawid, I. B., Ryffel, G. U. and Weber, R. (1981) *Science* 212, 298-304

Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, J., Kawashima, E. H. and Itakura, K. (1981) *Nuc. Acids Res.* 9, 879-894

Ward, K. A. and Kashmarik, S. E. (1980) *J. Invest. Dermat.* 75, 244-248

Weber, K. and Geisler, N. (1982) *EMBO J.* 1, 1155-1160

Wickens, M. P., Buell, G. N. and Schimke, R. T. (1978) *J. Biol. Chem.* 253, 2483-2495

Wilkinson, B. R. (1969) *Aust. J. Biol. Sci.* 23, 127-138

Woo, S. L. C., Lidsky, A. S., Fleming, G., Chandra, T. and Robson, K. J. H. (1983) *Nature* 306, 151-155

Wood, W. B. (1966) *J. Mol. Biol.* 16, 118

Woods, E. F. and Gruen, L. C. (1981) *Aust. J. Biol. Sci.* 34, 515-526

Wright, S., Rosenthal, A., Flavell, R. and Grosveld, F. (1984) *Cell* 38, 265-273

Young, R. A., Hagenbüchle, O. and Schibler, U. (1981) *Cell* 23, 451-458

Zahn, H. and Biela, M. (1968) *Eur. J. Biochem.* 5, 567-573

Zain, A., Sambrook, J., Roberts, R. J., Keller, W., Fried, M. and Dunn, A. R. (1979) *Cell* 16, 851-861

Zeevi, M., Nevins, J. R. and Darnell, J. E. (1982) *Mol. Cell. Biol.* 2, 517-525

APPENDIX

PAPERS PUBLISHED

Sheep Keratins: characterization of cDNA clones for the glycine + tyrosine-rich wool proteins using a synthetic probe. E. Kuczek and G. E. Rogers (1985) *Eur. J. Biochem.* 146, 89-93*

PRESENTED AT CONFERENCES

Comparison of genes coding for feather keratins in the chicken. P. L. Molloy, K. Gregg, E. Kuczek, J. Crowe, G. E. Rogers (1981) *Proc. of the Aust. Biochem. Soc.* 14, 85

* the above work was presented at:
6th Annual Conference - Organization and Expression of the Genome - Lorne, Victoria February, 1984 and
CSIRO research seminar - The Properties and Growth of Wool - Melbourne, Victoria November, 1984.

Kuczek, E. & Rogers, G. E. (1985). Sheep keratins: characterization of cDNA clones for the glycine + tyrosine-rich wool proteins using a synthetic probe. *European Journal of Biochemistry, 146*(1), 89-93.