STUDIES ON HAIR KERATIN GENES.

A thesis submitted to the

University of Adelaide

for the degree of

Doctor of Philosophy

by

Maurice Joseph Frenkel (M.App.Sci.)

Biochemistry Department,

University of Adelaide.

August, 1985.

TABLE OF CONTENTS

SECTION I.

THE ISOLATION AND CHARACTERIZATION OF GUINEA PIG AND SHEEP KERATIN mRNA.

CHAPTER 2. MATERIALS AND METHODS.

SECTION III.

APPENDICES.

SUMMARY

This thesis was undertaken as part of a major project directed towards the elucidation of the processes involved in the keratinization of epithelial tissues in general, and the follicle in particular.

The work is divided into two sections: (i) the isolation and characterization of guinea pig hair follicle mRNA and (ii) a study of two BIIIB high-sulphur keratin genes isolated from sheep genomic DNA libraries. Techniques used in the course of the work included mRNA isolation, translation in cell-free systems, cDNA cloning, genomic DNA library screening, gene mapping and DNA sequencing.

The following results were obtained:

1. Guinea pig hair follicle RNA was isolated, fractionated and translated in wheat-germ and rabbit reticulocyte lysate cell-free systems. Examination of translation products by single and two-dimensional electrophoresis indicated that proteins representative of the major keratin protein classes were being synthesized. The presence of a disproportionate amount of low molecular weight products suggested that the mRNA was partially degraded.

2. An RNA subfraction was used to prime the synthesis of dscDNA. The dscDNA was annealed to the plasmid vector pBR322 and the hybrid DNA was used to transform E.coli. Only a small number of cDNA clones was obtained. Three clones were sequenced but only one had an open reading-frame of significant length. On the basis of the derived amino acid sequence and composition it was unlikely that the clone coded for a component of a major keratin protein class. This result, together with the partially degraded nature of the follicle mRNA, made it impractical to attempt further characterization of the mRNA species from which the cDNA clone was derived.

3. A novel hair follicle harvesting method was developed utilizing a cold-cure acrylic resin to replace the commonly used wax sheet. This

method had the major advantage of enabling hair follicles to be harvested from live guinea pigs and led to the isolation of undegraded RNA.

While studies with guinea pig hair follicle mRNA were in progress, a number of sheep keratin cDNA clones were isolated by other workers. The availability of these clones and of a sheep genomic DNA library made further studies on the guinea pig system unwarranted.

4. The cDNA clone pSWK18, which codes for a BIIIB high-sulphur component was used to screen a Charon 4A sheep genomic DNA library. Seven identical clones (λSWK50) were isolated. Sequencing studies showed that the 13.2 kb insert contained a BIIIB pseudogene which was homologous to pSWK18. The pseudogene (ψ50BIIIB) had a number of features, including an intact TATA sequence, mutated initiation codon, in-phase stop codons and a 23 bp deletion near the 3' end of the coding region. The derived amino acid sequence indicated that the pseudogene originally coded for a BIIIB3-like protein. Hybridization studies suggested that no other keratin-coding genes were present in λSWK50 but that there were three regions containing repeated DNA sequences.

5. Screening of λ1059 and λgt10 sheep genomic DNA libraries failed to detect any pSWK18-positive clones. A number of positives were isolated from a Charon 28 library. Restriction mapping, subcloning and limited DNA sequencing of a single clone (λSWK61) indicated that the 9.6 kb insert contained the same BIIIB pseudogene as λSWK50.

6. A mixed synthetic oligonucleotide probe based on three cDNA sequences and corresponding to a 20 bp fragment deleted from the pseudogene failed to hybridize with any of the Charon 28-pSWK18-positive clones. The library was re-screened using pSWK18 and the 20-mer to probe duplicate filters. Only one clone hybridized with both probes. This clone, λSWK96, contained a 13.8 kb insert with a single BIIIB gene. The derived amino acid sequence showed that the gene product differed at only two positions from the protein BIIIB4.

7. Comparisons of the non-coding and flanking regions of the

λSWK50 and λSWK96 BIIIB genes revealed some extensive homologies. The sequences were homologous for 120 bp 5' to the initiation codon. In the 3' non-coding region there were two blocks of homology ; 50 bp immediately following the termination codon and 70 bp surrounding the polyadenylation addition sequence. The same 3' homologies were found in two cDNA clones coding for BIIIB3- and 4-like proteins but not in a BIIIB2 cDNA clone. These results support amino acid sequence data which indicate that BIIIB3 and 4 are more closely related to each other than to BIIIB2.

8. The 5' non-coding and flanking regions of λ96BIIIB and ψ50BIIIB shared a 70% homology over approximately 160 bp and both genes had limited homologies with the 5' non-coding sequences of B2, BIIIA and high-glycine-tyrosine genes. Interestingly, λ96BIIIB and B2A shared a 65% homology over a 100 bp region located in different 5' flanking positions, while a 17 bp sequence was conserved in the BIIIB gene and three other high-sulphur genes. These data, in addition to the protein sequence studies of other workers, suggest that the major high-sulphur keratin gene families have arisen from a single ancestral gene.

9. A pHC79 cosmid library was screened with pSWK18 and a single positive clone was isolated. Limited restriction mapping indicated that the insert (≃32 kb) contained only 1-2 BIIIB genes on a 5 kb HindIII fragment and that the genes were not the same as those isolated from the bacteriophage λ sheep genomic DNA libraries.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made in the text. I consent to the thesis being made available for photocopying and loan.

MAURICE FRENKEL

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following people for their help during the course of my work. I thank each of them for their valuable assistance.

Professor W. H. Elliot for permission to work in the Department of Biochemistry, University of Adelaide.

Professor G. E. Rogers for the opportunity to learn the techniques of molecular biology while investigating a fascinating aspect of hair complexity and for his enthusiasm, advice and support during the course of this work.

Rob Lock for his invaluable assistance during many of the guinea pig hair follicle harvestings and for introducing me to the intricacies of working with RNA.

Drs. Rob Saint, Barry Powell, Steve Wilton and Trevor Lockett as well as Joe Rothnagel, Phil Morris and Elizabeth Kuczek for their many helpful discussions and for making unforgettable my years in 'Keratin Korner'.

Dr. W. G. Crewther for permission to move from C.S.I.R.O. Division of Protein Chemistry, Melbourne to the University of Adelaide to undertake the work; for his support and encouragement.

Drs. J. M. Gillespie, A. A. Azad, T. C. Elleman and C. W. Ward of the Division of Protein Chemistry for their encouragement, enthusiasm and advice.

Drs. K. A. Ward, M. J. Sleigh, P. J. Kretschmer, U. Novak, R. Crawford, R. B. Saint and G. Cam for generously providing cDNA clones or genomic DNA libraries used during the course of this work.

Dr. J. Adams and members of his laboratory for their assistance and generosity in allowing me to use the facilities in the Walter and Eliza Hall Institute for Medical Research for the cloning of guinea pig hair follicle mRNA.

Evelyn Kirszbaum for the typing of this thesis, ably assisted by Louis.

Finally I would like to express my thanks to Lesley Crocker and Sue Barrett for excellent technical assistance and to Drs. R. J. Blagrove and B. C. Powell for valuable editorial comments.

ABBREVIATIONS

A_{600} - absorbance of light of wavelength 600nm (1cm path length)

bp - nucleotide base pairs

B2 (SCMKB2) - the B2 high-sulphur protein family

BIIIA (SCMKBIIIA) - the BIIIA high-sulphur protein family

BIIIB (SCMKBIIIB) - the BIIIB high-sulphur protein family

cDNA - single-stranded DNA complementary to RNA

cpm - counts per minute

ddNTP - 2',3'-dideoxynucleoside 5'-triphosphate

DMSO - dimethyl sulphoxide

DNase - deoxyribonuclease

dNTP - 2'-deoxynucleoside 5'-triphosphate

dscDNA - double-stranded DNA complementary to RNA

DTT - dithiothreitol

High-gly-tyr - high-glycine-tyrosine proteins

kb - kilobase pairs = 1,000 base pairs

mRNA - messenger ribonucleic acid

poly(A)$^+$ RNA - polyadenylated ribonucleic acid

RNase - ribonuclease

rRNA - ribosomal ribonucleic acid

SCMKA - S-carboxymethylated low-sulphur keratin proteins

SCMKB - S-carboxymethylated high-sulphur keratin proteins

SDS - sodium dodecyl sulphate

TCA - trichloroacetic acid

Tris - tris(hydroxymethyl)aminomethane

UV - ultraviolet

UVO - uridine-oxovanadium (IV) complex

Other abbreviations are as listed in Biochem. J. (1976) 153, 1-21.

CHAPTER 1.

GENERAL INTRODUCTION

CHAPTER 1

GENERAL INTRODUCTION

The keratins are a group of highly specialized proteins synthesized in certain epithelial cells of higher vertebrates. They form the bulk of the horny layer of the epidermis and epidermal appendages such as hair, nails, claws, scales and feathers. The economic importance of animal fibres has provided the driving force for studies of the chemical and molecular structures of keratins, whilst investigations of the biosynthesis of keratins have stemmed from interests in wool production and medical aspects of hair and skin growth. Although a considerable body of knowledge concerning the histology and protein chemistry of the hair follicle has been amassed (see for example Fraser et al., 1972; Crewther, 1976 for reviews), many basic questions about the molecular events leading to the formation of fully-formed keratin remain unanswered. Little is known about: (i) the size, number, or organization of genes coding for the enzymes and structural proteins produced in the follicle or associated with keratin biosynthesis and differentiation; (ii) the nature of the mechanisms which control transcription from these genes in the various layers of the tissue at different stages of their development; (iii) the number and size of follicle mRNA molecules; or (iv), the nature of the control mechanisms which operate at the level of mRNA translation.

1. Hair follicle morphology:

The mammalian keratin fibre consists of a column of dead epidermal cells which grows from a follicle. The follicle is an invagination into the dermis formed embryogenically by downward growth of the basal layer of the epidermis (Mercer, 1961). The morphologies and detailed histological descriptions of the follicle and fibre have been described on a number of occasions (see for example Mercer, 1961; Fraser et al., 1972; Montagna and Parakkal, 1974). For the purpose of this

thesis it is sufficient to briefly describe the location and composition of the five developmental zones (Fig.1.1). Information in this section has been taken from the three reviews referenced above.

(a) Undifferentiated zone:

Around the dermal papilla and at the base of the follicle the cells are spherical, undifferentiated and have a large nuclear volume in comparison to the total cell volume. This is a zone of rapid mitosis which forces the column of epidermal cells above it further upwards.

(b) Zone of differentiation:

In this region cells differentiate to form the characteristic tissue layers of the follicle. The cell cytoplasm contains large numbers of ribosomes, polysomes and mitochondria.

(c) Keratogenous zone:

Tissue layers in this region are fully differentiated. Cells show distinct morphologies and protein products, and contain protein filaments.

(d) Keratinization zone:

In this zone cells become dehydrated, nuclei and cytoplasmic organelles are broken down, and the cells begin to fill with their characteristic end products.

(e) Keratinized zone:

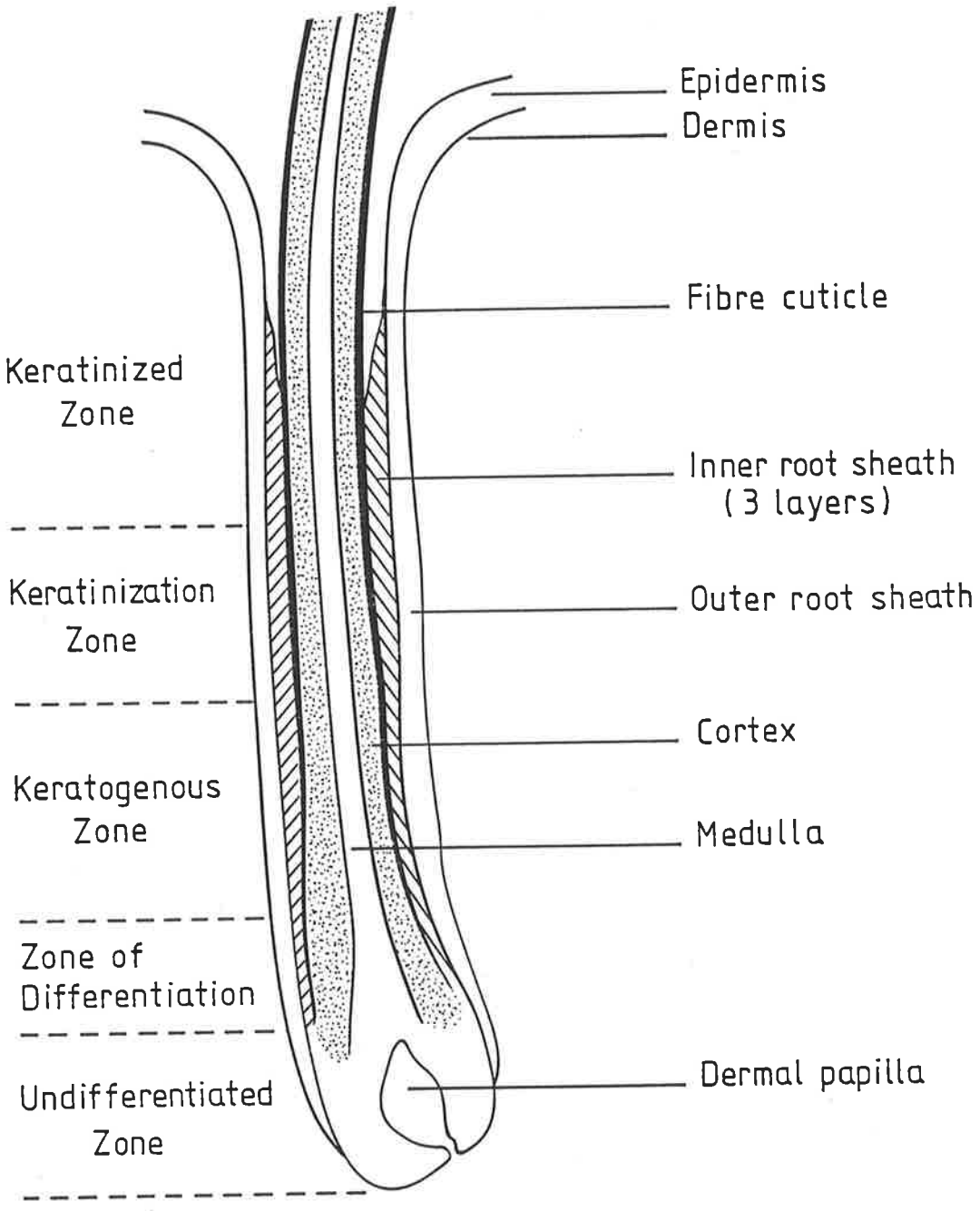
The cells in this region are dead and packed with protein. The remnants of their organelles have disappeared.

Above the level of the bulb where cell differentiation takes place, the follicle comprises a number of concentric sheaths of cells (Fig.1.1). Proceeding inwards, these are: (1) the outer root sheath; (2) the inner root sheath consisting of Henle's layer, Huxley's layer and the inner root sheath cuticle; and (3) the keratin fibre which projects beyond the epidermis and consists of a cuticle, cortex and often a medulla, although in some fibres this is absent.

FIGURE 1.1. DIAGRAM OF A LONGITUDINAL SECTION OF A MEDULLATED HAIR FOLLICLE.

Diagram of a longitudinal section of a medullated hair follicle showing the developmental zones and major histological regions. The diagram is schematic and not drawn to scale.

Adapted from Orwin (1976).


Epidermis
Dermis
Fibre cuticle
Keratinized Zone
Inner root sheath (3 layers)
Keratinization Zone
Outer root sheath
Cortex
Keratogenous Zone
Medulla
Zone of Differentiation
Dermal papilla
Undifferentiated Zone

2. Keratin proteins:

While most of the studies of keratin proteins have been carried out on wool, there are sufficient data available for other keratin fibres to indicate that the general characteristics found for wool apply to other keratins (see Crewther, 1976; and Gillespie, 1983 for comprehensive reviews of the structure and function of keratin proteins). Wool proteins have been studied most extensively as their S-carboxymethyl (SCM) derivatives and in this form they are readily fractionated into three major classes; the low-sulphur proteins - referred to collectively as SCMKA, the high-sulphur proteins - SCMKB, and the high-glycine-tyrosine proteins. The low-sulphur protein group contains relatively low levels of SCM-cysteine and high proportions of glutamic and aspartic acids, leucine, lysine and arginine. The high-sulphur proteins are rich in SCM-cysteine, proline, serine and threonine but contain little aspartic acid, lysine, alanine and isoleucine, whilst the high-glycine-tyrosine proteins contain high levels of glycine, serine and the aromatic amino acids tyrosine and phenylalanine.

(a) 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. There appear to be two distinct groups of proteins in SCMKA, termed Type I (originally referred to as component 8) and Type II (components 5 and 7), each of which contains 4 different proteins with molecular weights in the range 45,000-58,000 (Table 1.1; Crewther, 1976; Woods, 1979; Crewther *et al.*, 1980; L.G. Sparrow, personal communication). It should be noted that there is additional heterogeneity of the SCMKA proteins, particularly the component 7 family, as evidenced by a spread in isoelectric points in two-dimensional electrophoretograms (Gillespie, 1983). Amino acid compositions of a number of purified proteins indicate that within a component, constituent polypeptides are very closely related (Crewther, 1976). Partial amino

TABLE 1.1

SOME PROPERTIES OF THE MAJOR CLASSES OF KERATIN PROTEINS

A comparison of some properties of the three major keratin protein classes. The data is for wool and have been obtained mainly from Crewther (1976) and Woods (1979).

Protein class name	Protein families and approximate number of constituent polypeptides	Molecular weight	Approximate 1/2-cystine content as SCM-cysteine residues/100 residues
Low-sulphur (SCMKA)	Type I — Component 8 (4)	45,000-50,000	5
	Type II — Component 5 (1)	57,000	4
	Type II — Component 7 (3)	58,000	6
High-sulphur (SCMKB)	B2 } probably >50	19,000	23
	BIIIA } probably >50	16,000	25
	BIIIB } probably >50	11,000	17
High-glycine -tyrosine	Type I (>10)	6,000-9,000	6
	Type II (> 4)	6,000-9,000	10

acid sequence data has appeared for components 5 (Dowling et al., 1979), 7c (Sparrow and Inglis, 1980) and 8c-1 (Crewther et al., 1980, 1983). The complete sequence of component 8c-1 has been recently determined (L.M. Dowling, personal communication).

Amino acid sequence data for the low-sulphur keratins and a number of intermediate filament proteins have been compared (Geisler and Weber, 1982; Crewther et al., 1983; Dowling et al., 1983), and the striking homologies indicate that all of these proteins share common structural features.

(b) High-sulphur proteins:

The high-sulphur class of keratin proteins is located in the cuticle and in the matrix which surrounds the microfibrils, and constitutes up to 50% of the proteins of some hairs (Gillespie and Frenkel, 1974a). The proteins are heterogeneous in both size and charge (Table 1.1). When reduced and carboxymethylated, there are four main protein fractions with different molecular weights (SCMKB1-M_r 23,000-26,000; SCMKB2-19,000; SCMKBIIIA-16,000; and SCMKBIIIB-11,000), and each group comprises a number of closely related components (Haylett et al., 1971; Swart et al., 1976; Crewther, 1976). Joubert et al. (1968) used ion-exchange and gel filtration chromatography to separate the high-sulphur complex into 33 fractions, most of which showed multiple bands on disc gel electrophoresis (Joubert and Burns, 1967).

The full extent of heterogeneity of the high-sulphur proteins cannot be assessed by any single technique. Two-dimensional electrophoresis using charge separation at pH2.6 followed by size fractionation in the presence of sodium dodecyl sulphate (SDS) resolves at least 40 components (Gillespie and Marshall, 1981). However, there is an unresolved region in these electrophoretograms which contains the ultra-high-sulphur proteins (SCM-cysteine ≥30 residues/100 residues), a family of components with an almost continuous distribution of molecular sizes (Gillespie, 1983).

More wool high-sulphur proteins have been sequenced than any other type of keratin protein. Amino acid sequences have been determined for 3 proteins of the SCMKBIIIB family, 11 SCMKBIIIA proteins and 4 SCMKB2 components (Swart et al., 1976; Crewther, 1976). All of the sequenced proteins have a C-terminal SCM-cysteine residue and all the SCMKBIIIB and SCMKB2 components have a blocked N-terminal N-acetyl alanine. In contrast, all but two of the SCMKBIIIA proteins have threonine as the free N-terminal amino acid. Both the SCMKBIIIA and SCMKB2 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-(Arg or Gln)-Pro-X- where -X- can be serine, threonine or valine. Elleman et al. (1973) 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. More recently, Swart et al. (1976) compared the most likely codons for the amino acids of the SCMKBIIIA and SCMKB2 repeat units and found only two of the fifteen nucleotides were in conflict, further supporting the concept of a primordial single DNA coding sequence.

Amino acid sequence data for SCMKBIIIA and SCMKBIIIB proteins from mohair (Swart et al., 1976) reveals that there is a high degree of homology with the corresponding wool fractions. Peptide maps and amino acid compositions of the SCMKB2 fraction isolated from a variety of breeds of sheep, bovine hair, mohair , and sheep horn and hoof indicate that these proteins are homologous and have the same basic composition as the sequenced components of Lincoln wool (Gillespie et al., 1968).

(c) High-glycine-tyrosine proteins:

The high-glycine-tyrosine proteins are a group of small proteins (M_r<10,000), which form, with the high-sulphur proteins, the intermicrofibrillar matrix of the fibre. The proportion of these

aromatic-rich proteins in keratins ranges from almost zero in Lincoln wool up to 30-40% in echidna quill (Gillespie and Frenkel, 1974a). There are two major protein families (Table 1.1) which are readily distinguished by differences in their composition. The Type I components have relatively low levels of SCM-cysteine but are rich in phenylalanine, while the reverse is true for the Type II proteins. The occurrence and properties of these proteins have been recently reviewed (Gillespie and Frenkel, 1976; Marshall et al., 1980).

The high-glycine-tyrosine proteins of wool are heterogeneous (Table 1.1; Zahn and Biela, 1968; Brunner and Brunner, 1973; Gillespie and Frenkel, 1974b) although the extent of heterogeneity has not been fully determined. Ion-exchange chromatography of the Type I family separates the components into ten fractions which show little amino acid composition homology (Gillespie, 1972). Electrophoresis under alkaline conditions reveals that most of the fractions contain up to five components (Gillespie and Frenkel, 1974b), although at least some of this heterogeneity is an artifact of the extraction procedure used to isolate the proteins (Frenkel, 1977; Marshall et al., 1980).

Complete amino acid sequence data have been obtained for three Type I components from wool (Dopheide, 1973; Marshall et al., 1980; R.C. Marshall, personal communication). The two proteins sequenced by Marshall are members of the one sub-family and there is no difference in sequence 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). The limited amino acid sequence data available for one Type II component (T.A.A. Dopheide, personal communication) show no observable homology with the sequenced Type I proteins.

(d) Other proteins:

(i) Cuticle proteins:

Using standard alkaline-urea extraction techniques,

it is possible to solubilize up to 30% of the wool cuticle (Ley and Crewther, 1980). Although the extracted proteins are cystine-rich, in terms of amino acid composition and electrophoretic properties it is apparent that they are different to the high-sulphur proteins of the cortex (Ley and Crewther, 1980). The alkaline-urea insoluble residue of cuticle has an amino acid composition very similar to the soluble proteins. Ley and Crewther (1980) have suggested that γ-glutamyl-ε-lysyl residues and disulphide bonds resistant to reduction may play an important role in the cross-linking of cuticle proteins.

(ii) Citrulline-containing proteins:

The occurrence and properties of citrulline-containing proteins of the hair follicle have been reviewed recently (Rogers, 1983) and unreferenced material in this section has been taken from this source. These proteins are constituents of the inner root sheath and medulla, and are distinct from other keratin proteins in being extremely insoluble in the solvents used to extract keratin. However they are readily solubilized by treatment with proteolytic enzymes. Citrulline-containing proteins have a characteristically high level of glutamic acid, extremely low contents of half-cystine and proline, contain up to 15% of their residues as citrulline, and have a relatively high level of γ-glutamyl-ε-lysyl cross-links. It is very likely that the citrulline found in these proteins is formed by modification of arginine residues in arginine-rich precursors localized within trichohyalin granules of the inner root sheath and medulla. Studies on the precursor protein complex and the enzyme which converts the protein-bound arginine to citrulline are now in progress (J. Rothnagel and G.E. Rogers, personal communication).

3. Keratin Biosynthesis:

Mitosis occurs only in the basal layer of the follicle (Short et al., 1965; Fraser, 1965; Epstein and Maibach, 1969), and labelling studies with ^{3}H-thymidine indicate that DNA synthesis in the follicle is

also restricted to this layer (Epstein and Maibach, 1969; Downes et al., 1966a). When rat hair follicles are pulse-labelled with ^{3}H-cytidine or ^{3}H-uridine in vivo, radioactivity appears in the nuclei of cortical cells up to the lower regions of the keratogenous zone, although the basal cells are the most highly labelled (Sims, 1967). As cells move up the follicle into the upper part of the keratogenous zone, the radioactivity (and hence the RNA) is concentrated more in the cytoplasm than in the nucleus (Sims, 1967). Similar results have been found for guinea pig hair follicles (Fraser et al., 1972). Cells in the upper keratogenous zone lose their ability to incorporate ^{3}H-thymidine within their nuclei but radioactivity is retained in the cytoplasm (Fraser et al., 1972).

The appearance of keratin mRNA in the cytoplasm and the initiation of its translation do not begin until follicle cells have lost their ability to synthesize DNA and divide (Fraser et al., 1972). Messenger RNA synthesis occurs at an early stage in the keratogenous zone (Sims, 1967; Wilkinson, 1970b) and evidence suggests that the predominant mRNA species in wool and hair roots are stable (Wilkinson, 1970b; Fraser et al., 1972). Translation of mRNA persists into the keratinization zone; as the keratinocyte approaches maturity, synthesis of RNA and non-keratin proteins ceases and degradation of the nucleus ensues. Stabilization of the keratin begins, its synthesis declines and the keratinocyte finally dies filled with keratin protein (Fraser et al., 1972).

Mammalian hard keratin is a complex mixture of proteins and there is evidence to suggest that synthesis of the different protein families is not concurrent. Rudall (1956) reported that the sulphur content of horn keratin protein increased with the age of the keratinocyte, possibly reflecting the synthesis of low-sulphur proteins prior to that of high-sulphur proteins. This conclusion received some support from Rogers (1964) who obtained evidence from electron microscopy that microfibril synthesis appeared to precede deposition of the

intermicrofibrillar matrix. Studies measuring the level of ^{35}S specific activity of the low-sulphur and high-sulphur proteins at various levels in the follicle after administration of ^{35}S-cystine, supported a two-stage synthesis for these proteins (Downes et al., 1963), although later results were not in complete accord with this conclusion (Downes et al., 1966b).

Support for concurrent synthesis but different peak synthesis times for the two protein families comes from the work of Fraser (1969a,b). He isolated keratin proteins from active keratinocytes at different levels of the follicle and found that while the concentration of low-sulphur proteins increased steadily during maturation, the level of high-sulphur proteins increased sharply during the latter stages. The question of when the various protein classes are synthesized remains inadequately answered, particularly when one takes into account that there are no data available on the biosynthesis of the high-glycine-tyrosine proteins.

4. Factors affecting follicle growth and differentiation:

(a) Nutrition:

A great deal of research has been directed towards the dietary control of wool growth and the manipulation of its composition. When cysteine, methionine or proteins rich in these amino acids are infused into the abomasum of sheep, there is a substantial increase in wool growth rate and an increase of up to 45% in the cystine content of the wool (Gillespie and Reis, 1966). The change in wool composition is due to the increased synthesis of the ultra-high-sulphur proteins, an extremely heterogeneous family of polypeptides containing about 8.5% sulphur, which differ markedly from the high-sulphur keratin proteins (Gillespie, 1983).

The level of high-glycine-tyrosine proteins in wool can be substantially reduced by a number of dietary treatments, usually involving a nutritional imbalance. Examples include sheep on a diet of

wheat receiving an abomasal infusion of methionine, and sheep on a basal diet being 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 ultra-high-sulphur proteins which accompanies the reduction in high-glycine-tyrosine protein synthesis. It has been suggested that this increase results from greater availability of sulphur-containing amino acids as a consequence of the reduced wool growth rate caused by these treatments (Frenkel et al., 1975; Gillespie et al., 1981).

(b) Physiological factors:

Some variations in the composition of wool and hair appear to be unrelated to diet and suggest a physiological control over the synthetic activities of the follicle. The sulphur content of wool has been reported to change with the time of year and the physiological state of the sheep. It has been suggested (Doney and Evans, 1968) that these compositional changes result from variations in the wool growth rate which change the level of sulphur-containing amino acids available for ultra-high-sulphur protein synthesis. Gillespie (1983) has extended this by suggesting that the large number of physiological and environmental factors known to influence wool growth rate would all have the potential to change the protein composition of wool.

Extreme reductions in wool growth rate can be induced by administering chemical defleecing agents to sheep. Following administration of mimosine or cyclophosphamide, major changes in wool composition have been found to persist over a prolonged period of time. The high-glycine-tyrosine protein level decreased substantially but returned to pretreatment percentages after 10 weeks. On the other hand, the high-sulphur protein content increased initially, then decreased and remained below pretreatment levels after 12 weeks (Frenkel et al., 1975; Gillespie et al., 1981). In another study, the first wool or hair 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). It has been proposed that the changes in keratin protein proportions following defleecing may be characteristic of fibres synthesized by new or regenerating follicles (Gillespie et al., 1980).

(c) Species and breed:

Comparative studies of keratin composition suggest that the hard keratins are unique amongst homologous tissues in exhibiting a wide range of amino acid compositions (Gillespie, 1983). This variability in composition is a reflection not only of differing ratios of matrix to microfibrils but also of changes in the relative proportions of the two matrix protein classes. The levels of high-sulphur and high-glycine-tyrosine proteins vary over an extremely wide range when keratins from different species are compared (Gillespie and Frenkel, 1974a). Within a species it is also possible to observe variability, although over a narrower range (Frenkel et al., 1974). While the different keratin types produced by a single animal (e.g. horn, hoof, hair) contain identical groups of proteins, it has been shown that the relative proportions of the proteins can vary (Gillespie, 1972). This variability presumably reflects the different structure-function requirements of each keratin type.

Variation in the composition and proportion of the low-sulphur proteins is not as extreme as that found for the matrix components. Differences between and within species have been observed by Sparrow and Crewther (1972) who compared the low-sulphur protein electrophoretic patterns of Merino and Lincoln wool with mohair and porcupine quill.

(d) Hormones:

Adrenal, gonadal, thyroid and pituitary hormones all appear to effect hair and wool growth (various symposium papers in Lyne and Short, 1965; reviewed by 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 (Lyne and Short, 1965). Wool growth is also inhibited by the administration of mouse epidermal growth factor (Moore et al., 1982).

Studies on control factors acting within the follicle are still in the preliminary stages. Factors controlling growth and differentiation which have been identified, are as yet poorly characterized with respect to their modes of action at a molecular level.

5. Studies on the molecular biology of keratins:

(a) Chick feather:

The embryonic chick feather is a very useful system for investigating the molecular biology of keratins. This is due in part to the relative simplicity of feather keratin proteins, most components being of uniform size and highly homologous in composition (Fraser et al., 1972). However, the most important factor contributing to the success of molecular biology studies on the embryonic feather has been the ease with which undegraded keratin-coding mRNA may be isolated (Partington et al., 1973; Kemp et al., 1974). These early studies indicated that a 12S fraction isolated from 14 day chick embryo feather could be translated in a rabbit reticulocyte lysate cell-free system to give products which co-migrated with feather keratin proteins under various electrophoretic conditions, and which were immunoprecipitated by a specific keratin antiserum. Furthermore, the 12S RNA was polyadenylated and migrated as a single band of 2.5×10^5 daltons on denaturing gels. This result suggested that not only were the coding sequences of an uniform length (about 300 bases are required for the 10,000 dalton proteins), but there was also a conservation of the non-coding region length (500 bases, including a poly(A) tract of about 100 bases). The hybridization kinetics studies of Kemp (1975) indicated that there were 25-35 different feather keratin mRNA species and a total of

100-240 keratin genes in the chick genome. The keratin genes presumably contained a repetitive and unique sequence corresponding to the coding and non-coding regions respectively.

Powell et al. (1976) titrated feather mRNA species at different embryonic ages and showed that the amount of mRNA in feather cells increased about 700-fold between days 11 and 14. This increase was the direct result of an accumulation of keratin mRNA rather than the release of presynthesized mRNA from the nucleus, and indicated that major transcriptional activity was taking place at the onset of keratin protein synthesis.

The ability to isolate undegraded feather keratin mRNA has led to the construction of a cDNA library (Saint, 1979), nucleic acid sequence data for a number of cDNA clones (P. Morris, E. Kuczek, J. Crowe and G.E. Rogers, personal communication), and the isolation of a genomic clone containing five keratin genes (Molloy et al., 1982). The organization of the genes on this genomic fragment suggests that the clone is part of a longer cluster of tandemly spaced genes which has developed by a series of duplications. The genes studied by Molloy et al. (1982) are evenly spaced and are transcribed from the same DNA strand. Nucleic acid sequence data for one gene indicates that it codes for a protein of 97 amino acid residues which has a sequence typical of feather keratin. The gene has a single intron of 324 bp situated within the 5' non-coding region, 21 bp prior to the initiation codon, a feature which may be shared by the other genes in the clone (Molloy et al., 1982). Continued characterization of this gene cluster and studies on a chicken cosmid clone will provide further data concerning the organization of the feather keratin multigene family.

(b) Epidermal keratins:

The epidermal keratins are a family of at least 10 related proteins of 40,000-70,000 daltons that form cytoplasmic filaments in most vertebrate epithelial cells (for an extensive reference list, see Fuchs

et al., 1981). Working with cultured human epidermal cells, Fuchs and Green (1979) demonstrated that the four major keratin proteins were each translated from its own mRNA. Epidermal mRNAs can be grouped into two distinct classes on the basis of their nucleic acid sequence, each class encoded by about 10 genes which do not cross-hybridize with members of the other class. In addition, the two classes appear to be maintained as separate unlinked subfamilies throughout vertebrate evolution (Fuchs et al., 1981).

The epidermal keratins constitute one of five groups of cytoskeletal components that are referred to as intermediate filaments. The proteins of intermediate filaments share a number of common structural features including similar peptide domains and extensive regions of coiled-coil α-helix (Steinert et al., 1980; Geisler et al., 1982). The determination of the partial nucleic acid sequences of cDNA clones from each human epidermal keratin class (Hanukoglu and Fuchs, 1982; 1983), and the complete sequences of two mouse epidermal cDNA clones (Steinert et al., 1983; 1984) allowed direct comparisons to be made with the protein sequences of other intermediate filament types. These comparisons (Steinert et al., 1983; Hanukoglu and Fuchs, 1983) suggested that: (i) intermediate filaments contain an internal core of four helical domains separated by three regions of β-turns; (ii) the coiled-coil α-helical regions are structurally homologous and are of the same size; and (iii) the amino and carboxy termini portions are non-helical and variable in both size and amino acid sequence (although the epidermal sequences do contain at least one region extremely rich in glycine). The variability of the terminal, non-helical sequences probably accounts for the size heterogeneity among epidermal keratins. Sequence data for two low-sulphur wool proteins indicates that these microfibrillar components are also members of the intermediate filament class of structures (Geisler and Weber, 1982; Dowling et al., 1983).

Marchuk et al. (1984) have examined the organization of a human

epidermal keratin gene and have compared it with the hamster gene coding for the intermediate filament protein vimentin. Although the keratin and vimentin sequences are only 29% homologous at the amino acid level, and 42% homologous at the nucleotide level, there is a conservation of intron positions. Six of the introns show identical or nearly identical positioning in the two genes. However, neither the intron sizes nor their sequences have been conserved - while the keratin gene spans about 4.6 kb containing 3 kb of intron sequences, the vimentin gene comprises 10 kb of which 8 kb are intron s (Marchuk et al., 1984; Quax et al., 1983).

Interestingly, the highly conserved intron positions in the intermediate filament genes do not appear to mark any known structural boundaries of the proteins. Marchuk et al. (1984) suggest that the intron positions may define important functional domains which are common to the intermediate filament proteins.

(c) Hair and wool:

Initial studies on the molecular biology of the hair follicle centred about the use of cell-free protein synthesizing systems derived from follicle homogenates. The first such system was prepared by Rogers and Clarke (1965) and other early work included that of Freedberg (1970a) and Clarke and Rogers (1970). The systems used by these workers for incorporating labelled amino acids into proteins had low activities. Wilkinson (1970a,b) isolated polysomes from wool roots in good yield and concluded that the major proportion of follicle mRNA is probably stable. Steinert and Rogers (1971a,b; 1973a) showed that it was possible to prepare hair follicle cell-free systems with high amino acid incorporating activity by using very young guinea pigs as the source of follicle tissue. The use of young guinea pigs resulted in reduced contamination of the system with skin ribonuclease and allowed these workers to obtain polysome profiles similar to those of Wilkinson (1970a) for sheep wool follicles. From their studies on the synthesis of guinea

pig hair keratin proteins *in vitro*, Steinert and Rogers (1971b; 1973a), concluded that:

(i) low-sulphur and high-sulphur proteins identical to native proteins could be synthesized *de novo* in the cell-free system;

(ii) the mechanism of cell-free protein synthesis in the follicle homogenate was ribosomal-dependent, so discounting suggestions that the high-sulphur proteins might be synthesized by the addition to 'precursor' proteins of cysteine as free amino acid or short peptides;

(iii) *in vitro*, the low-sulphur proteins were synthesized at a greater rate than the high-sulphur components, implying that the mechanism of initiation of synthesis for the two protein classes was different, or that the rate of translation of high-sulphur mRNA was lower than for low-sulphur mRNA.

Gilmartin and Freedberg (1975) and Lock (1977) were unable to obtain the polysome profiles described by Steinert and Rogers (1971a), and this resulted in an extensive study of RNA extraction techniques in an attempt to isolate undegraded nucleic acids from guinea pig hair follicles (Lock 1977). Using the ribonuclease substrate analogue described by Gray (1974), prepared by complexing uridine with the oxovanadium IV ion, Lock (1977) was able to isolate follicle RNA. This RNA, and sucrose gradient fractions of it, stimulated the incorporation of ^{3}H-leucine into proteins in an S23 wheat germ cell-free translation system. Incorporation of radioactivity was low. The translation products were examined by measuring the level of ^{3}H-leucine in gel slices following SDS electrophoresis. While there appeared to be radioactivity incorporated into proteins with molecular weights in the range expected for keratins, the results were equivocal.

A study of protein synthesis in a wool root homogenate system stimulated by sheep follicle polysomes (Wilkinson, 1971) indicated that components which co-migrated with keratin proteins were being synthesized. Ward and Kasmarik (1980) extended this work by translating wool follicle mRNA in a reticulocyte lysate cell-free system, examining the products by gel electrophoresis. The results showed that the mRNA coded for proteins identical in molecular weight to low-sulphur and high-sulphur components. Fractionation of wool follicle polysomal RNA by sucrose gradient centrifugation enabled two peaks, of mean size 11S and 21S, to be collected. Electrophoresis of translation products indicated that the 21S peak coded for low-sulphur and larger high-sulphur proteins, while the 11S fraction coded for components with molecular weights of less than 20,000.

The isolation of sheep keratin-coding mRNA enabled K.A.Ward and M.J. Sleigh (personal communication) to construct a cDNA library and to isolate clones coding for proteins of the component 7 and 8 low-sulphur families, and the B2 and BIIIB high-sulphur families. Clones coding for SCMKBIIIA components or high-glycine-tyrosine proteins were not isolated from this cDNA library.

The cDNA clone pSWK20, which codes for the high-sulphur protein B2C, has been used as a probe to isolate homologous sequences from a sheep genomic DNA library (Powell et al., 1983). The two clones studied by Powell and his colleagues, λSWK1 and λSWK2, contain the genes for SCMKB2A and SCMKB2D (a previously unrecognized protein of this family) and SCMKB2C, respectively. The two λSWK1 genes are closely linked, being separated by only 1.9 kb, and are transcribed from the same DNA strand. Sequence data for the three genes (Powell et al., 1983), reveals extensive conservation of the 5' non-coding and coding sequences, but the 3' non-coding regions diverge both in length and sequence. There is no evidence for the presence of introns in the SCMKB2 genes.

6. Aims of the project:

(a) General aims:

The work presented in this thesis forms part of a major project directed to the elucidation of the processes involved in keratinization of epithelial tissues in general, and the follicle in particular.

The co-ordinate expression of keratin genes is an integral part of keratinization and its study requires the isolation and characterization of the individual genes. The development of recombinant DNA techniques enables genes to be purified and characterized. The use of these techniques in the study of keratin genes forms the basis of this thesis.

(b) Specific aims:

In order to study keratin genes at the molecular level, it is essential to have specific probes for each of the major keratin protein classes. The initial aim of the project was to extend the work of Lock (1977) on guinea pig hair follicle mRNA. The keratin mRNA was to be fractionated and translated in more efficient cell-free systems than used previously; the fractions, hopefully enriched in mRNA species coding for different protein classes, were to provide the starting material for the construction of cDNA libraries. Specific cDNA clones were to be sequenced and used to select keratin genes from genomic DNA libraries. The guinea pig was used as a convenient laboratory animal to obtain keratin-coding probes for future studies on the sheep wool system.

After the isolation of mRNA from guinea pig hair follicles had commenced, Dr. K.A. Ward, C.S.I.R.O. Division of Animal Production, Prospect, Australia, began parallel experiments using the wool follicle as a source of RNA. The two approaches were maintained until Dr. Ward and Dr. M.J. Sleigh constructed and partially characterized a sheep cDNA library. While the problems inherent in the guinea pig keratin mRNA system had been resolved, it was apparent that with the existence of a sheep cDNA library and a sheep genomic DNA library (Kretschmer et al.,

1980), continued studies on the guinea pig were no longer warranted.

Professor G.E. Rogers' group was provided with specific high-sulphur B2 and BIIIB probes from the sheep cDNA library and the use of the latter probe formed the basis for the second part of the project. The specific aims were to use the SCMKBIIIB cDNA clone to isolate genes for this protein family from a sheep genomic DNA library. The genes were to be mapped by restriction enzyme analysis and sequenced in order to obtain an understanding of their organization in the genome and their relationship to other keratin gene families.

SECTION I.

THE ISOLATION AND CHARACTERIZATION OF GUINEA PIG AND SHEEP KERATIN mRNA

CHAPTER 2.

MATERIALS AND METHODS

CHAPTER 2

MATERIALS AND METHODS

A. Materials.

1. Chemicals:

All chemicals used were of analytical reagent grade or of the highest available purity. Chemicals listed below were obtained from the sources indicated.

(a) General chemicals:

CsCl optical grade: Harshaw Chemical Co., Cleveland, Ohio.

Dextran sulphate, sodium salt: Pharmacia.

Diethylpyrocarbonate: Sigma.

Dimethylsulphoxide: B.D.H. Chemicals Ltd.

Ficoll 400: Pharmacia.

Formamide: Merck. Deionized with mixed-bed resin.

Glyoxal: B.D.H. Chemicals Ltd. Deionized with mixed-bed resin.

Phenol: B.D.H. Chemicals Ltd. Redistilled under N_2 and reduced pressure, stored at -20^0C under N_2 prior to use.

Polyethyleneglycol 6000: Merck.

Polyvinylpyrrolidone: May and Baker, West Footscray, Victoria.

POPOP (1,4-bis-(2,5-phenyloxazolyl)-benzene): Sigma.

PPO (2,5-diphenyloxazole): Sigma.

Sodium dodecyl sulphate: Sigma and B.D.H. Chemicals Ltd.

Sucrose - ultrapure: Schwarz-Mann.

Trizma base: Sigma.

Urea - ultrapure: Schwarz-Mann.

(b) Chemicals for electrophoresis:

Acrylamide: Merck, twice recrystallized from $CHCl_3$, and B.D.H. Chemicals Ltd. specially pure.

Agarose, Type I, low EEO: Sigma.

Ammonium persulphate: B.D.H. Chemicals Ltd.

Ethidium bromide: Sigma.

N,N' - methylenebisacrylamide: B.D.H. Chemicals Ltd., recrystallized from acetone.

N,N,N',N' - tetramethylethylenediamine (TEMED): B.D.H. Chemicals Ltd.

(c) Chemicals for RNA isolation:

Beeswax: Mr. F. Badenoch, Broadview, South Australia.

GuHCl - ultra pure: Schwarz-Mann.

Oligo(dT)-cellulose, Type 7: P-L Biochemicals Inc.

Resin, South American, grade-C: Dean's Art Wholesalers, Melbourne, Victoria.

Sarkosyl NL97 (N-lauryl-sarcosine): Ciba-Geigy Ltd.

Uridine, crystalline, A-grade: Calbiochem.

Vanadyl sulphate (Vanadyl (IV) oxide sulphate): Merck.

Vertex SC: Dentimex, Zeist, Holland.

(d) Chemicals for <u>in vitro</u> DNA synthesis and DNA sequencing:

Dimethyl sulphate: Aldrich Chemical Co. Inc.

dATP, dCTP, dGTP, dTTP: Sigma.

Hydrazine: Eastman Kodak Co.

Oligo$(dT)_{10}$: P-L Biochemicals Inc.

Piperidine: Sigma.

Sodium cacodylate: Fluka, Switzerland.

(e) Radiochemicals:

Unless otherwise specified, all radiochemicals were obtained from The Radiochemical Centre, Amersham.

Amino acid mixture: L-[4,5-^{3}H] Leucine, L-[4,5-^{3}H] Lysine, L-[2,4,6-^{3}H] Phenylalanine, L-[2,3,4,5-^{3}H] Proline and L-[2,3,5,6-^{3}H] Tyrosine, each with a specific activity within the range 60-110 Ci/mmol (2.2-4.1 TBq/mmol).

Adenosine 5'-[γ-^{32}P] triphosphate: >5000 Ci/mmol (>185 TBq/mmol): The Radiochemical Centre, Amersham and Dr. R.H. Symons, Biochemistry Department, University of Adelaide.

Deoxyadenosine 5'-[α-^{32}P] triphosphate: 2000-3000 Ci/mmol (74-110 TBq/mmol).

Deoxycytidine 5'-[α-^{32}P] triphosphate: 400 Ci/mmol (14.8 TBq/mmol): Dr. R.H. Symons, Biochemistry Department, University of Adelaide.

Deoxyguanosine 5'-[α-^{32}P] triphosphate: 400 Ci/mmol (14.8 TBq/mmol): Dr. R.H. Symons, Biochemistry Department, University of Adelaide.

Iodo [2-^{14}C] acetic acid: 40-60 mCi/mmol (1.5-2.2 GBq/mmol).

L-[4,5-^{3}H] Leucine: 40-60 Ci/mmol (1.5-2.2 TBq/mmol) and 130-190 Ci/mmol (4.8-7.0 TBq/mmol).

2. Proteins, nucleic acids and enzymes:

Albumin, bovine serum, Fraction V: Sigma.

Cytochrome C, horse heart, Type III: Sigma.

Creatine phosphokinase, rabbit muscle: Sigma.

DNA, E.coli, Type VIII: Sigma.

DNA, Salmon testes, sodium salt, Type III: Sigma.

Exonuclease III: New England Biolabs Inc.

Lysozyme: Sigma.

Nuclease S1, Aspergillus oryzae: Boehringer Mannheim.

Phosphocreatine, disodium salt, high grade: Sigma.

Polynucleotide kinase: Boehringer Mannheim.

Restriction endonucleases: New England Biolabs Inc; Bethesda Research Laboratories Inc; Boehringer Mannheim.

Ribonuclease A, bovine pancreas: Boehringer Mannheim.

RNA, 18S and 28S embryonic chick feather: S.D. Wilton, Biochemistry Department, University of Adelaide.

RNA dependent DNA polymerase (reverse transcriptase), avian myeloblastosis virus: Dr. J.W. Beard, Life Sciences Inc. Florida.

Terminal deoxynucleotidyl transferase, calf thymus: Ratliff Biochemicals, Los Alamos, New Mexico.

tRNA, E.coli: Boehringer Mannheim.

3. Sources of tissues:

Albino guinea pigs of both sexes were bred in the animal house of the Biochemistry Department, University of Adelaide and were also obtained from the Institute of Medical and Veterinary Sciences, Adelaide. Sheep from (Border Leicester X Merino) dams X Poll Dorset rams, were obtained from C.S.I.R.O. Division of Animal Health, Marybyrnong, Victoria.

4. Other materials:

Agar: Difco Laboratories, Detroit.

Amino acids: Mann Research Laboratories, New York.

Casamino acids: Difco Laboratories, Detroit.

GF/A, Glass fibre circles: Whatman Ltd. England.

Halberd wheat: Waite Agricultural Research Institute, South Australia.

Nitrocellulose membranes: Schleicher and Schuell, West Germany; Bio-Rad Laboratories, California.

Reticulocyte lysate, rabbit, nuclease-treated: The Radiochemical Centre, Amersham.

Tryptone: Difco Laboratories, Detroit.

Wheat germ extract translation system: Bethesda Research Laboratories Inc.

Yeast extract: Difco Laboratories, Detroit.

The bacterial strain ED8654: r_k^-, m_k^+, supE, supF, trpR used for transformation was obtained from Professor K. Murray (University of Edinburgh) by Dr. R.B. Saint.

B. Methods.

1. General methods:

(a) Buffers and preparation of solutions:

All solutions were prepared in glass-distilled water and either autoclaved or treated with diethylpyrocarbonate to eliminate any contamination with nuclease. Glassware, spatulas etc. were rendered nuclease-free by either autoclaving, incubation at 110°C overnight,

washing with 1M KOH followed by rinsing with sterile glass-distilled water, or by a combination of these procedures.

The buffers commonly used in this study were:

SSCE: 0.15M NaCl, 15mM $Na_3C_6H_5O_7$, 1mM EDTA, pH7.0.

SSPE: 0.18M NaCl, 10mM Na_2HPO_4, 8mM NaOH, 1mM EDTA, pH7.4.

TAE: 40mM Trizma base, 20mM CH_3.COONa, 1mM EDTA, pH8.2.

TBE: 90mM Trizma base, 90mM boric acid, 1.25mM EDTA, pH8.3.

TE: 10mM Tris-HCl pH7.5 (unless otherwise indicated), 1mM EDTA.

(b) Media:

Media used for this study are described in Chapter 4.

(c) Ethanol precipitation of RNA and DNA:

Unless otherwise stated, RNA and DNA were precipitated from aqueous solutions by the addition of 1/10 volume of 3M Sodium acetate pH6.0 and 2.5 volumes of ethanol, followed by storage at -20^0C for approximately 16h or -80^0C for 3h.

(d) Measurement of radioactivity:

The level of radioactivity of ^{32}P-containing materials was determined by measuring the Cerenkov emissions using the ^{3}H settings of a Packard tri-carb scintillation counter. The level of radioactivity of ^{3}H-containing materials was determined in either of two ways. Total counts were measured by spotting the material onto a GF/A filter, drying the filter and immersing it in liquid scintillation fluid (3.5g PPO, 0.35g POPOP per litre of toluene). Incorporation level of ^{3}H into TCA-insoluble material was determined by adding the sample to ice-cold 10% TCA (containing 20µg of carrier tRNA when studying DNA products), vortex mixing, filtering through GF/A filters, washing sequentially according to the requirements of the particular technique, drying and immersing the filter into toluene scintillation fluid for counting.

2. Preparation of hair and wool follicles:

(a) Isolation of guinea pig hair follicles using the wax-sheet method:

Albino guinea pigs less than 4 weeks old were killed by a blow to the head, their skins rapidly removed and the hair follicles exposed by the wax-sheet depilatory procedure of Ellis (1948) as modified by Clarke and Rogers (1970) and Lock (1977). All procedures were carried out at 4^{0}C. Sterile disposable gloves were worn; solutions and instruments were nuclease-free. Hair follicles were harvested using animal clippers (Model A-2, John Oster, Wisconsin) fitted with size 40 cutters, and placed into ice-cold buffer.

(b) Isolation of guinea pig hair follicles and sheep wool follicles using the self-curing acrylic resin Vertex SC:

Hair and wool follicles were isolated by a novel method based on the technique described by Wilkinson (1970a). Guinea pigs were anaesthetized by an intraperitoneal injection of Pentobarbitone (4mg/100gm body weight). Sheep were anaesthetized (with the assistance of Dr. P. Carter, C.S.I.R.O. Division of Animal Health, Parkville) by an intramuscular injection of Atropine Forte (1ml) followed by an intrajugular injection of Pentothal (1g). Anaesthesia was maintained by administration of Halothane:oxygen.

Follicles were harvested from both sides of guinea pigs and from an area of approximately 40x40cm from one side of a single sheep. The same technique was used for guinea pigs and sheep and differed only in the amounts of reagents and materials used. The description given here is that used for harvesting follicles from sheep. Wool was clipped away with Oster animal clippers using size 10 and then size 40 cutters, leaving a thin covering of wool on the animal. A thin layer of cold-cure acrylic resin (Vertex) was then applied by spatula to an area of about 7x20cm directed from the belly to the backbone. The resin consists of a powder (polymethylmethacrylate) and a liquid (methylmethacrylate), which for the purpose described here, were mixed in a 1:1 (w/v) ratio. After stirring the mixture for 2 min, it was applied gently to the side of the animal. A 45g:45ml mixture was sufficient for two 7x20cm strips. The

resin was quickly overlaid with a 7x30cm strip of glass-fibre tape (Ciba-Geigy Ltd), which had been dipped into resin to ensure that the tape was thoroughly wet. About 9 min after the resin ingredients had been mixed, polymerization, as evidenced by a rapid temperature increase, was well advanced. Gentle application of an ice pack eliminated the risk of skin burns and did not effect the polymerization. After about 12 min, polymerization was essentially complete and the follicles were harvested by pulling the glass-fibre tape away from the animal's skin and placing it immediately into liquid N_2.

Follicles were recovered from the resin backing by scraping with a single-edge razor blade into a beaker containing the RNA extraction buffer. This final stage of the harvesting procedure was aided by the follicles and attached fibres being extremely brittle at sub-zero temperatures.

Generally an area of about 600-700 sq.cm was harvested for sheep and 40-50 sq.cm for guinea pigs, yielding 2g and 0.2g of follicles respectively. This weight included wool and hair which was harvested with the follicles.

3. Isolation of RNA and mRNA from hair and wool follicles:

(a) The uridine-oxovanadium (UVO) method:

Guinea pig hair follicles harvested by the wax-sheet method as described above, were homogenized at 0^0C in a buffer modified from Lock (1977). The buffer contained 20mM Tris-HCl pH7.6, 150mM KCl, 5mM DTT, 10mM uridine and 1mM $VOSO_4$, and 30ml was used to extract the follicles from 12 guinea pigs. Because $VOSO_4$ is oxidized very rapidly above pH3.5 in the presence of oxygen [oxovanadium (IV) to (V)] (Lienhard et al., 1971), it was not added to the buffer until all the follicles had been harvested (up to 15 min for 12 guinea pigs). The buffer was thoroughly flushed with N_2 before use and kept securely sealed during isolation of the follicles.

Follicles were homogenized in a Potter-Elvehjem homogenizer

(AHT Co., Philadelphia) fitted with a loose teflon pestle driven by an electric drill at 100-150 rpm for about 30-40s, filtered through gauze to remove large particulate matter and then centrifuged at 17,000g(av) for 10 min. The clarified solution was then extracted with phenol (saturated with 1M Tris-HCl pH9.0, 10% SDS):chloroform (1:1). The aqueous phase was isolated following centrifugation and re-extracted with phenol: chloroform until the interphase was devoid of protein. The RNA was precipitated overnight at -20°C in the presence of sodium acetate and ethanol. RNA prepared in this manner had a detectable blue-green colour which could be removed by repeated ethanol precipitations.

(b) The GuHCl-CsCl method:

Follicles isolated using the acrylic resin-liquid N_2 method described above were extracted using a modification of the procedures described by Glisin et al. (1974), Kaplan et al. (1979) and Chirgwin et al. (1979) utilizing GuHCl in the place of GuHCNS.

Frozen follicles (2g) were scraped into 30ml of extraction buffer (7M GuHCl, 10mM Tris-HCl pH7.6, 5% v/v 2-mercaptoethanol, 1mM EDTA), which was being stirred vigorously in an ice bath to ensure adequate contact between the thawing tissue and denaturing buffer. Following a brief homogenization (3.5s maximum speed, Sorvall Omni-Mixer Model 17220), sarkosyl was added to 4% w/v and the homogenization repeated. The viscosity of the solution was reduced by 3x5s sonications using an MSE Ultra Sonicator and addition of buffer to give a tissue: buffer ratio of 5:1. Solid CsCl was added to the solution (0.5g/ml), the mixture layered over a cushion of 5.7M CsCl, and ultracentrifugation carried out in an SW40 rotor for 18-20h, 100,000g(av) at 20°C.

Following ultracentrifugation, the buffer and CsCl solutions were carefully removed to avoid any contamination of the RNA pellet. The pellet was briefly rinsed with 0.5ml of 0.3M sodium acetate-70% ethanol, suspended in 0.5ml of 0.2% w/v sarkosyl, and then precipitated in the presence of sodium acetate and ethanol. The RNA was sedimented by

centrifugation in an Eppendorf centrifuge (10 min at 4°C) and the pellet washed two times with 70% ethanol after disruption with a sterile applicator stick. The RNA was dried in vacuo, resuspended in low salt buffer (20mM Tris-HCl pH7.6, 2mM EDTA, 0.5% SDS) to give a concentration of about 2.5mg/ml, and disaggregated by heating at 65°C for 10 min and snap chilling in ice. Any insoluble material still present was removed by centrifugation for 5 min at room temperature. Solid NaCl was added to the supernatant to give a concentration of 3M and the solution was kept overnight at 0°C to precipitate the RNA away from DNA and low molecular weight RNA species (modified from Palmiter, 1974). Sedimented RNA was washed twice with 70% ethanol, dried in vacuo, resuspended in 10mM Tris-HCl pH7.6 and stored at -20°C.

4. Fractionation of RNA:

(a) Sucrose density gradient ultracentrifugation:

Guinea pig RNA prepared using the UVO method was resuspended in 20mM Tris-HCl pH9.0, 0.1% SDS, layered onto linear gradients of 10-40% (w/v) sucrose in TE containing 100mM NaCl, and subjected to ultracentrifugation in an SW40 rotor for 15h, 160,000g(av) at 4°C.

Fractions were collected by upward displacement using an Isco Density Gradient Fractionator and the absorbance monitored at 254nm. Fractions were collected and precipitated in the presence of sodium acetate and ethanol. RNA was resuspended in water or TE and stored at -20°C.

(b) Oligo(dT)-cellulose chromatography:

RNA isolated using the acrylic resin-liquid N_2 technique was fractionated by oligo(dT)-cellulose chromatography using a modification of the method described by Aviv and Leder (1972). Briefly, RNA in 10mM Tris-HCl pH7.6, 0.1% SDS was heated at 80°C for 5 min and chilled in ice, NaCl added to 0.5M and then layered onto a 2x0.5cm column of oligo(dT)-cellulose equilibrated with 10mM Tris-HCl pH7.6, 0.5M NaCl, 0.1% SDS. The column was washed with equilibration buffer until the A_{254}

(absorbance at 254nm, 1cm path length) was stable at zero. Poly(A)$^+$ RNA was eluted from the column using 10mM Tris-HCl pH7.6 and precipitated by the addition of sodium acetate and ethanol. Following drying *in vacuo*, RNA was resuspended in TE and stored at -20°C.

5. Electrophoresis of RNA:

Follicle RNA was examined by electrophoresis in 4.5% polyacrylamide gels in the presence of 98% formamide essentially as described by Pinder *et al.* (1974) other than the reservoir buffer was 20mM Trizma base, 10mM sodium acetate, pH8.1. Gels were stained with ethidium bromide (1µg/ml) and photographed under UV light.

Electrophoresis of RNA was also performed in 1.5% agarose gels using 10mM sodium phosphate pH6.8 as the gel and reservoir buffer. RNA was denatured in the presence of glyoxal and DMSO as described by Thomas (1980). Gels were stained with 0.02% toluidine blue and photographed under visible light.

6. Translation of RNA in cell-free systems:

(a) S30 wheat germ system:

The S30 wheat germ cell-free translation system was prepared exactly as described by Marcu and Dudock (1974) and was optimized for the amount of follicle RNA translated, and K^+ and Mg^{++} concentrations. Translation studies were carried out in 50µl assays essentially as reported by Marcu and Dudock (1974) except that the unlabelled amino acids were added to 20µM and [^{3}H]-leucine (1µCi) was used. Assays were incubated at 30°C for 1h and then stored at -20°C.

(b) Commercial wheat germ system:

The wheat germ translation system produced by Bethesda Research Laboratories was used as recommended. No additional K^+ or Mg^{++} were added to the assays and the concentrations of these in the final mix were 63mM and 2mM respectively. Incubation was at 25°C for 1h.

(c) Commercial, nuclease-treated, rabbit reticulocyte lysate system:

The rabbit reticulocyte lysate translation system (nuclease-treated) was used as specified by the manufacturer. The concentration of K^+ was 100mM and Mg^{++} was 1.6mM; incubation was at 30°C for 1h.

7. Measurement of [^{3}H]-leucine incorporation into translation products:

(a) Wheat germ systems:

In order to measure the level of incorporation of radio-labelled leucine into protein following translation in the S30 and commercial wheat germ systems, a modification of the methods of Bollum (1968) and Roberts and Paterson (1973) was used. Duplicate aliquots of 2µl were spotted onto GF/A filters and dried. Using 10ml of each solution per aliquot being assayed, the filters were washed successively (5 min on ice except where indicated) with 10% TCA-1% casamino acids (10 min), 5% TCA, 5% TCA (10 min, 90°C), 5% TCA, ethanol, ethanol:ether (3:1) and finally ether. Dried filters were placed into liquid scintillation fluid and counted for ^{3}H.

(b) Rabbit reticulocyte lysate system:

The manufacturer's instructions were followed exactly. Briefly, duplicate aliquots of 1µl were added to 1M NaOH, 5% H_2O_2 (0.5ml) and heated at 37°C for 10 min, the solution cooled and the proteins precipitated with 25% TCA-2% casamino acids (3ml). The proteins were collected on GF/A filters and washed thoroughly with 5% TCA, ethanol and finally ether. Dried filters were counted by liquid scintillation.

8. S-carboxymethylation of translation products:

Translation products from the S30 wheat germ system were S-carboxymethylated using the following procedure. To 46µl of translation mix was added 500µl 8M urea, 0.2M 2-mercaptoethanol, pH 10.8. Following reduction at 37°C for 2h, 300µl of iodoacetic acid: Trizma base: water (2:4:11) was added to the mixture and alkylation allowed to proceed at 20°C for 10 min. Excess iodoacetic acid was reacted with 30µl 2-mercaptoethanol and the mixture diluted with 910µl water. Proteins were

precipitated by the addition of TCA to 20% and incubation on ice for 1h. Following centrifugation in a bench centrifuge, the protein pellet was thoroughly washed in acetone:ether (3:1), dried in vacuo and resuspended in 15µl of gel loading buffer (20mM Tris-HCl pH8.0, 1% SDS, 10% glycerol, 0.1% bromophenol blue). Prior to electrophoresis the protein sample was heated at 100°C for 1 min.

Translation products obtained from the commercial cell-free systems were S-carboxymethylated using a simplified procedure. To 30µl of translation mix was added 24mg urea and 5µl of reduction buffer (0.5M Tris-HCl pH9.5, 0.5M DTT). After incubation at room temperature for 2h, the proteins were alkylated by the addition of 25µl of a solution containing Trizma base:iodoacetic acid:water (1.5:1:4). The reaction was terminated after 10 min at room temperature by the addition of 5µl 2 - mercaptoethanol. The samples were ready for direct loading onto alkaline-urea polyacrylamide gels (see Methods, 10b).

9. Preparation of S-carboxymethylated keratin proteins for use as electrophoresis standards:

Guinea pig hair keratin proteins were isolated, S-carboxymethylated and fractionated as described by Gillespie (1983). Extraction of hair and wool proteins and subsequent labelling with [^{14}C] -iodoacetic acid were achieved using the method of Marshall (1981).

10. Electrophoresis of translation products:

(a) SDS-polyacrylamide gel electrophoresis:

Polyacrylamide gel slabs (12.5%) were made as described by Laemmli (1970) but without the sample or stacking gels. Electrophoretic conditions found to give optimal resolution of the high-glycine-tyrosine proteins involved pre-electrophoresis at 20mA for about 1h during which the bromophenol blue tracking dye migrated about 1.5cm. Samples were then layered onto the gel and electrophoresis continued until the dye was within 4cm of the bottom of the gel. For protein staining, gels were treated with 0.2% Coomassie Brilliant Blue in acetic acid:methanol:water

(10:45:45) and destained in 10% acetic acid-10% methanol.

(b) Two-dimensional polyacrylamide gel electrophoresis:

Electrophoresis of proteins in 7.5% polyacrylamide alkaline-urea tube gels followed by separation in 10% polyacrylamide-SDS slab gels was carried out as described by Marshall (1981).

Fluorography of gels containing translation products or radiolabelled standard proteins was according to Bonner and Laskey (1974) and Laskey and Mills (1975).

11. In vitro synthesis of [^{32}P]-labelled DNA:

(a) Oligo(dT)-primed reverse transcription:

Oligo(dT)-primed reverse transcription of RNA was carried out in 25µl reaction mixtures containing up to 2.5µg RNA, 0.66mM each of dCTP, dGTP and dTTP, 5-60µM [^{32}P]-dATP (2500-3000 Ci/mmole), 8mM DTT, 10mM $MgCl_2$, 50mM Tris-HCl pH8.3, 10µg/ml oligo$(dT)_{10}$ and 100µg/ml actinomycin D. Reverse transcriptase (1µl, 13U/µl) was added and the solution incubated at 42°C for 15 min. The RNA template was removed by alkaline hydrolysis in 0.3M NaOH for at least 1h at 37°C. The solution was neutralized by the addition of Tris-HCl pH7.6 to 0.1M and HCl to 0.3M. Following extraction with phenol:chloroform (1:1) and chloroform alone, the aqueous phase was chromatographed through an 0.6x30cm Sephadex G-50 column in TE to remove unincorporated nucleotides.

(b) Random-primed reverse transcription:

Priming of 18S and 28S ribosomal RNA was achieved by the random hybridization of oligonucleotides of salmon sperm DNA, prepared as described by Taylor et al. (1976). Conditions for the synthesis of this cDNA were as described for the oligo(dT)-primed reaction, except that oligo$(dT)_{10}$ was replaced by a final concentration of 2mg/ml of oligonucleotides, and the synthesis was allowed to proceed for 1h at 37°C. The cDNA synthesized was isolated as described for oligo(dT)-primed synthesis.

12. Preparation of a cDNA library:

The method described here was developed by R.B. Saint (1979), and is a modification of the procedure reported by Rougeon and Mach (1976).

(a) Synthesis of double-stranded cDNA (dscDNA):

Synthesis of the first strand cDNA on the keratin RNA template was carried out using either oligo$(dT)_{10}$ or random primers as described above, other than actinomycin D was excluded from the mixture, the concentration of Tris-HCl pH8.3 was 10mM and a reaction mix of 100µl containing 10µg RNA was used. Non-radioactive dATP was added to 100µM while the amount of labelled dATP was sufficient to allow products of the synthesis to be detected.

Following synthesis of the first strand, the reaction mixture was heated at 100°C for 2 min, chilled on ice, and a further 0.25mM (final concentration) of nucleoside triphosphates and 10mM DTT were added. Reverse transcriptase (800U/ml) was added and the solution incubated at 37°C for 5h. Fresh reverse transcriptase was added after 2.5h of incubation.

(b) Nuclease S1 cleavage and size fractionation of dscDNA:

The reaction mix was diluted ten-fold in nuclease S1 buffer (30mM sodium acetate pH4.6, 300mM NaCl, 4.5mM $ZnCl_2$) containing 0.5 units of enzyme per µl and incubated at 37°C for 30 min.

Following nuclease S1 digestion, the dscDNA was extracted with an equal volume of a 1:1 mixture of phenol and chloroform, chromatographed on a Sephadex G-50 column to remove unincorporated nucleotides and concentrated by rotary evaporation. The dscDNA was then loaded onto a 5-20% (w/v) sucrose gradient in TE buffer plus 0.1M NaCl and centrifuged at 180,000g(av) for 16h at 4°C, or fractionated on a column (0.8x36cm) of agarose A-150 (Bio-Rad) in TE, 0.1M NaCl. Fractions were collected, assayed by agarose gel electrophoresis of aliquots, pooled and precipitated by the addition of magnesium acetate to 10mM and

3 volumes of ethanol. Following storage overnight at -20°C the dscDNA was pelleted by ultracentrifugation at 190,000g(av) for 2h at 2°C in an SW40 rotor, dried, and resuspended in TE buffer.

(c) Tailing and annealing of dscDNA to plasmid DNA and transformation of E.coli:

Homopolymeric nucleotide tails were added to the dscDNA using calf thymus terminal deoxynucleotidyl transferase using a modification of the method described by Deng and Wu (1981). Tritiated dCTP (500 pmoles) was dried down and resuspended in a solution of 140mM Na cacodylate, 30mM Tris-HCl pH7.6, 0.2mM DTT, 4mM $MgCl_2$ and 0.005-0.1 pmoles of dscDNA to give a final volume of 40µl. Terminal deoxynucleotidyl transferase (100 units) was added and the reaction at 4°C was followed by the conversion of the [^{3}H]-dCTP to a TCA-insoluble form. When an average of 10-20 nucleotides per end had been added the reaction was stopped by the addition of EDTA to 10mM. Unincorporated nucleotides were removed by Sephadex G-50 chromatography. Following precipitation, the dscDNA was resuspended in 50-100µl annealing buffer (see below).

Plasmid vector pBR322 DNA was cleaved with PstI and tailed with [^{3}H]-dGTP as described above. The vector DNA was annealed to an equimolar amount of tailed dscDNA in 200mM NaCl, 10mM Tris-HCl pH8.2 by heating for 10 min at 65°C, incubating for 1h at 45°C and finally allowing the solution to slowly cool to 4°C. Annealed DNA was stored at 4°C.

E.coli strain ED8654 was grown overnight at 37°C in Luria (L) broth, diluted 1/100 into fresh L broth and grown at 37°C to an A_{600} of 0.6-0.7. The cells were chilled on ice, pelleted by centrifugation and resuspended in 0.5 vol of ice-cold 100mM $CaCl_2$. The cells were pelleted immediately and resuspended in 0.05 vol of 100mM $CaCl_2$. The cells were kept on ice for at least 1h to make them competent, after which 0.2ml was added to 0.1ml of 0.1M Tris-HCl pH7.1 containing 1-5µl annealed DNA. The mixture was incubated on ice, with occasional mixing, for 30 min. The

cells were then heat-shocked at 42°C for 2 min, kept on ice for a further 30 min and warmed slowly to room temperature.

L broth (0.5ml) was added to the transformed cells and incubated at 37°C for 30 min. After mixing with 3ml of L broth containing 0.7% agar, the cells were plated on L + tetracycline plates (see Chapter 4) and incubated overnight at 37°C.

13. Detection of recombinants:

(a) In situ hybridization:

(i) Preparation of nitrocellulose filters:

Detection of recombinant plasmids by in situ colony hybridization was achieved by a modification of the procedure of Grunstein and Hogness (1975). Colonies from a transformation were transferred by toothpick to a master plate and to a circle of nitrocellulose which had been boiled in three changes of distilled H_2O, autoclaved for 10 min and placed onto an L-agar plate. The colonies were grown overnight at 37°C after which the nitrocellulose was transferred sequentially onto 3MM paper saturated with 0.5M NaOH (7 min), 1M Tris-HCl pH7.4 (2 min), 1M Tris-HCl pH7.4 (2 min) and finally 1.5M NaCl, 0.5M Tris-HCl pH7.4 (4 min). The lysed colonies were dried onto the nitrocellulose by suction through a scintered glass Buchner funnel. The DNA was then baked onto the filter at 80°C for 2h in vacuo.

(ii) cDNA and RNA probes:

The preparation of cDNA probes using oligo$(dT)_{10}$ or random-priming has been described above. End-labelling of alkali-fragmented RNA was performed as described by Sim et al. (1979).

(iii) Filter hybridization:

Nitrocellulose filters were pre-incubated overnight at 65°C in 2xSSCE containing 0.2% each of bovine serum albumin, ficoll and polyvinylpyrrolidone (modified from Denhardt, 1966). The filters were blotted between 3MM paper and dipped through the labelled probe in a minimum volume of 2xSSCE, 0.5% SDS, 100µg/ml of sonicated and denatured

salmon sperm or E.coli DNA, and hybridized at 65°C for 16-24h. Filters were washed to varying degrees of stringency, blotted dry and exposed to Fuji Rx medical X-ray film at -80°C using Ilford fast-tungstate intensification screens.

Alternatively, in situ hybridization was carried out at 42°C in the presence of 50% formamide and SSPE buffer as described by Maniatis et al. (1982).

(b) Miniscreen detection of recombinants:

Isolation of plasmids from small cultures for insert size estimation was carried out as described in Gough et al. (1980) except that the DNA solution was phenol extracted prior to ethanol precipitation. Plasmids were linearized by EcoRI digestion and then subjected to electrophoresis in agarose gels.

14. DNA sequence determination:

Recombinant plasmid DNA was prepared and then mapped with restriction enzymes as described in Chapter 4. The chemical cleavage method of Maxam and Gilbert (1980) was used for DNA sequence determination. Polyacrylamide gel electrophoresis was carried out in thin gels as described by Sanger and Coulson (1978).

15. Containment facilities:

The handling of viable organisms which contained recombinant DNA was carried out under C3, EK1, or P2, EK1 containment conditions as defined and approved by the Australian Academy of Science Committee on Recombinant DNA, and the safety committees of the Walter and Eliza Hall Institute of Medical Research, the University of Adelaide, and the C.S.I.R.O. Division of Protein Chemistry.

CHAPTER 3.

ISOLATION AND CHARACTERIZATION OF GUINEA PIG HAIR FOLLICLE AND SHEEP WOOL FOLLICLE mRNA

CHAPTER 3

ISOLATION AND CHARACTERIZATION OF GUINEA PIG HAIR FOLLICLE AND SHEEP WOOL FOLLICLE mRNA.

A. Introduction.

Early work on the molecular mechanisms of differentiation and protein synthesis in the hair follicle centred around the use of cell-free protein synthesizing systems derived from hair follicle homogenates. The first such system was prepared by Rogers and Clarke (1965) who used the wax-sheet depilatory method for the isolation of hair follicle tissue. The crude follicle homogenate incorporated labelled amino acids into protein but the incorporation time was limited. Other work with guinea pig hair follicle homogenates (Freedberg, 1970a; Clarke and Rogers, 1970) resulted in the isolation of polysomes containing abundant levels of monomers but low yields of heavy polysomes. Amino acid incorporating activity of the hair root systems was low and Freedberg (1970a) suggested that ribonuclease (RNase) could be a significant factor leading to degradation of ribosomes and the mRNA associated with them.

Wilkinson (1970a,b) developed a follicle harvesting method using the adhesive Araldite which resulted in the isolation of wool root polysomes and nucleic acids in good yield. A heavy polysome fraction was shown to synthesize proteins which co-migrated with low-sulphur components on alkaline starch gels (Wilkinson, 1971). There was no extension of these studies until Ward and Kasmarik (1980) reported the translation in a reticulocyte lysate cell-free system of mRNA isolated from wool follicle polysomes.

Steinert and Rogers (1971a) showed that the ribonuclease level in hair follicle tissue homogenates prepared from guinea pigs less than 1 month old was much lower than that from older animals. Using hair roots from young guinea pigs it was possible to prepare polysomes with profiles comparable to those produced from wool follicles by Wilkinson (1970a).

In addition, the polysome profile was unaffected by the temperature of the wax used for follicle isolation (in the range of 45^0-80^0C) indicating that there was no degradation of cellular components resulting from hot wax on the skin prior to exposure of the follicles. The polysomes isolated by Steinert and Rogers (1971a) were highly active in a guinea pig hair follicle cell-free homogenate and it was possible to demonstrate that low-sulphur and high-sulphur keratin proteins were synthesized (Steinert and Rogers, 1971b; 1973a).

Gilmartin and Freedberg (1975) were unable to obtain the polysome profiles reported by Steinert and Rogers (1971a) and suggested that the degradative effect of nucleases was limiting the yield of polysomes even in young guinea pigs. Attempts to isolate polysomes from new born guinea pig epidermis were also unsuccessful and the presence of a range of ribonucleases in the skin was proposed as the cause of polysomal breakdown during the preparative procedures (Freedberg and Gilmartin, 1977).

The first report of the successful isolation of hair follicle mRNA having activity in cell-free amino acid incorporating systems was that of Lock (1977). The preparation of active follicle mRNA fractions was due largely to the inclusion in the homogenization buffer of a ribonuclease substrate analogue prepared by the complexing of uridine with the oxovanadium-IV ion (VO^{++}) as described by Lienhard et al. (1971) and Gray (1974). The use of uridine-oxovanadium (UVO) as a ribonuclease inhibitor was the result of a comprehensive investigation of RNA extraction procedures applied to the preparation of mRNA from hair follicle tissue homogenates (Lock, 1977). This study included attempts to isolate polysomes using the method described by Steinert and Rogers (1971a), the addition of SDS and proteinase K to homogenization buffers and harvesting the follicles directly into buffer-saturated phenol. None of these methods, or variations of them, was as successful as the UVO system for the isolation of translationally active mRNA (Lock, 1977).

Total hair follicle RNA was fractionated by sucrose density gradient centrifugation and each fraction was shown to be active in stimulating amino acid incorporation in a heterologous (wheat germ S23) cell-free protein-synthesizing system (Lock, 1977). Studies using the S23 cell-free system indicated that wheat embryo ribosomes were not efficient in the translation of follicle mRNA. Incorporation of labelled amino acids into protein was low and it was difficult to identify which follicle proteins were being synthesized. Lock (1977) also observed that heated or formamide-denatured follicle RNA showed signs of degradation even when isolated in the presence of UVO.

The work of Lock (1977) enabled active mRNA to be isolated from guinea pig hair follicles and provided the foundation for investigating a number of questions, including: (i) is there a more suitable cell-free translation system for studying the follicle mRNA complex than the S23 wheat germ fraction used by Lock (1977)? (ii) can synthesis of the major hair keratin proteins be demonstrated? (iii) can the mRNA be fractionated so that fractions will code for different protein classes? and (iv), is the mRNA a suitable starting material for recombinant DNA studies?

While these questions were being investigated, Ward and Kasmarik (1980) reported the isolation of wool keratin mRNA. Using the Araldite adhesive method these workers prepared and separated follicle polysomes into two fractions. A 21S fraction coded for low-sulphur and larger high-sulphur proteins while an 11S fraction coded for proteins with molecular weights of less than 20,000. Ward and Kasmarik (1980) were able to purify keratin mRNA away from the ribosomal RNA by oligo(dT)-cellulose chromotography, a technique which had not been successful for guinea pig hair follicle mRNA (Lock, 1977).

After the work with guinea pig RNA described in this chapter had been completed Bertolino *et al*. (1982) described the isolation of mouse hair follicle mRNA from a hair root-enriched fraction prepared by

scraping the underside of frozen skin. Total follicle RNA was isolated using a modified guanidine hydrochloride extraction technique and translated in a reticulocyte lysate cell-free system. A large number of proteins was synthesized and immunoprecipitation or probing of gel transfers with keratin antiserum (raised against rat epidermal keratin) allowed visualization of at least 4 of the 5 keratin proteins in the low-sulphur protein region.

B. Results.

1. Isolation of guinea pig hair follicle RNA using the UVO method:

Guinea pig hair follicles were harvested at 4^0C using the wax-sheet method described in Chapter 2. From 2-12 guinea pigs were handled in single experiment and the wet weight of follicles, with their short lengths of attached hair, was approximately 0.5g per guinea pig. The follicles were disrupted by homogenization in the presence of UVO and total follicle RNA was prepared using the method described in Chapter 2. Yields of RNA generally ranged from 300-600µg of RNA per guinea pig although in a number of experiments up to 1mg of RNA per guinea pig was isolated.

2. Fractionation of follicle RNA by sucrose density gradient ultracentrifugation:

Follicle RNA was fractionated by ultracentrifugation through a sucrose density gradient as described in Chapter 2. A typical absorbance profile is shown in Fig.3.1. The 28S:18S ratio is about 2:1 which is expected for a preparation of undegraded RNA. Furthermore, the absorbance between the major RNA peaks is almost to the baseline, another indication that there is little degradation. Fractions were collected as indicated in Fig.3.1 and precipitated. Table 3.1 shows the average RNA yield per guinea pig obtained for each fraction.

Fraction 1, containing pooled 6-17S RNA, should contain mRNA coding for most if not all of the major keratin proteins (protein M_r <10,000-58,000; mRNA <300-1500 bases of coding sequence), therefore a

FIGURE 3.1. SUCROSE DENSITY GRADIENT PROFILE OF RNA ISOLATED FROM GUINEA PIG HAIR FOLLICLES IN THE PRESENCE OF UVO.

Ultracentrifugation was performed in sucrose density gradients as described in Chapter 2. Approximately 1mg total hair follicle RNA was loaded onto the gradient. Fractions 1-4 were collected as indicated by the bars.

Sedimentation coefficients (Svedberg units) are shown above the major peaks.

Vertical axis: Absorbance at 254nm.

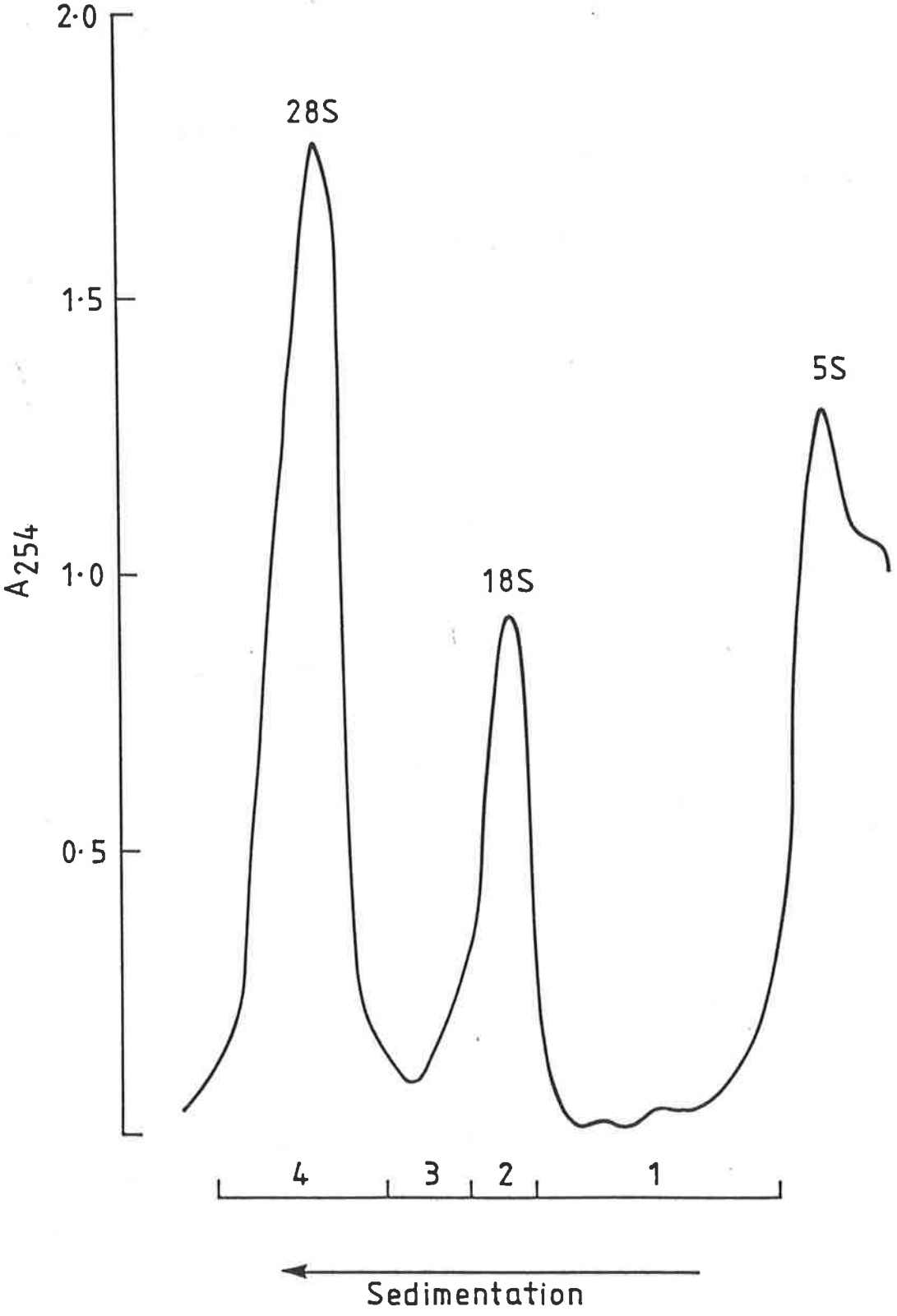

2·0
1·5
1·0
0·5
A254
28S
18S
5S
4
3
2
1
Sedimentation

TABLE 3.1

SEDIMENTATION COEFFICIENTS AND YIELDS
OF GUINEA PIG HAIR FOLLICLE RNA FRACTIONS AND SUBFRACTIONS
PREPARED BY SUCROSE DENSITY GRADIENT ULTRACENTRIFUGATION

Hair follicle RNA fractions were obtained by ultracentrifugation of total RNA (Fig.3.1), while subfractions A-D were recovered following recentrifugation of fraction 1 (Fig.3.2).

Yields (µg) represent average recoveries per guinea pig.

The approximate sedimentation coefficients (Svedberg units) of the various fractions are listed also.

Fraction	S value	Yield(µg)
1	6-17	60
2	17-19	140
3	19-21	30
4	21+	350
A	6-7.5	7
B	7.5-13	6
C	13-17	6
D	17-19	9

further fractionation of this RNA was attempted by repeating the sucrose density gradient ultracentrifugation. The absorbance profile obtained is shown in Fig.3.2. It can be seen that apart from the 5S and 18S RNA peaks, there are three other major peaks. While the relative proportions of the subfractions showed some variation between experiments, their position relative to the 5S and 18S peaks remained constant. Subfraction 1A contained pooled 6-7.5S RNA; 1B, 7.5-13S; 1C, 13-17S; and 1D, 17-19S RNA (Fig.3.2; Table 3.1). These subfractions were recovered in approximately equal yields (Table 3.1).

Attempts to use oligo(dT)-cellulose chromatography to fractionate RNA extracted in the presence of UVO were unsuccessful. A number of different commercial sources of oligo(dT)-cellulose were used but no fractionation was obtained. No significant fractionation was observed for feather keratin RNA which is known to contain polyadenylated mRNA. It would appear that the batches of oligo(dT)-cellulose used in these experiments were inefficient in binding poly(A)$^+$ RNA.

3. Electrophoresis under denaturing conditions of hair follicle RNA fractions and subfractions:

Total hair follicle RNA and fractions derived from it were examined by electrophoresis as described in Chapter 2. The electrophoretic patterns of undenatured total RNA and RNA denatured by glyoxal and DMSO treatments are shown in Fig.3.3. The undenatured follicle RNA has major bands corresponding to 18S and 28S ribosomal sequences and some minor bands between the 5S and 18S RNA (Fig.3.3a). The electrophoretic banding pattern of denatured follicle RNA (Fig.3.3b) shows an increase in the number of bands and in the relative proportion of RNA between 5S and 28S. There is a corresponding decrease in the amount of rRNA, particularly 28S rRNA.

Figure 3.4 shows a comparison of the electrophoretic patterns of hair follicle RNA fractions. Electrophoresis was carried out under denaturing conditions (98% formamide) and as expected from the

FIGURE 3.2. SUCROSE DENSITY GRADIENT PROFILE OF RNA ISOLATED FROM 6-17S (FRACTION 1) GUINEA PIG HAIR FOLLICLE RNA.

Ultracentrifugation was performed in a sucrose density gradient as described in Chapter 2. Approximately 0.2mg fraction 1 RNA (Fig.3.1) was loaded onto the gradient. Fractions A-D were collected as indicated by the bars.

Sedimentation coefficients (Svedberg units) are shown above the two major peaks.

Vertical axis: Absorbance at 254nm.

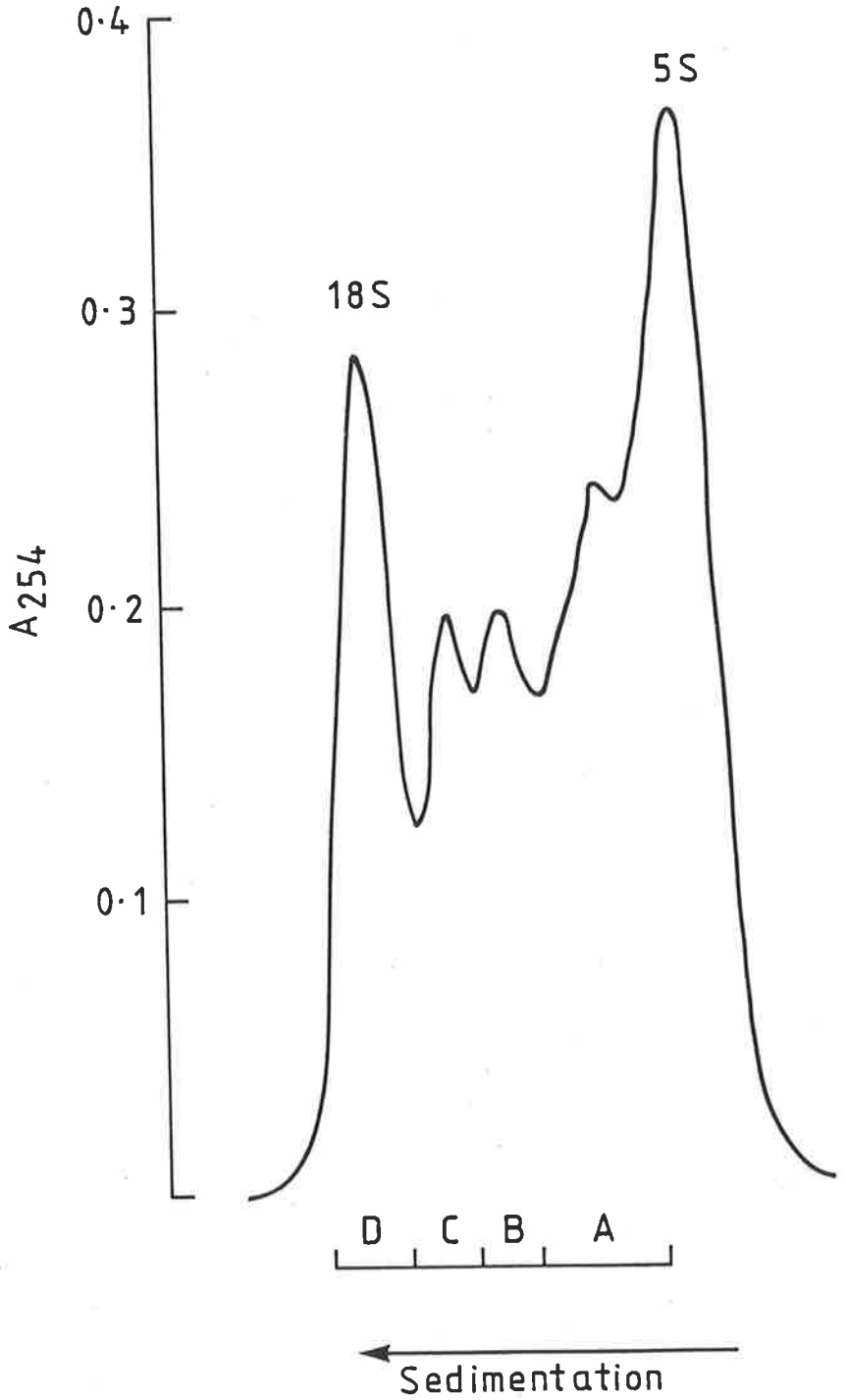

0·4
0·3
0·2
0·1
A254
18S
5S
D
C
B
A
Sedimentation

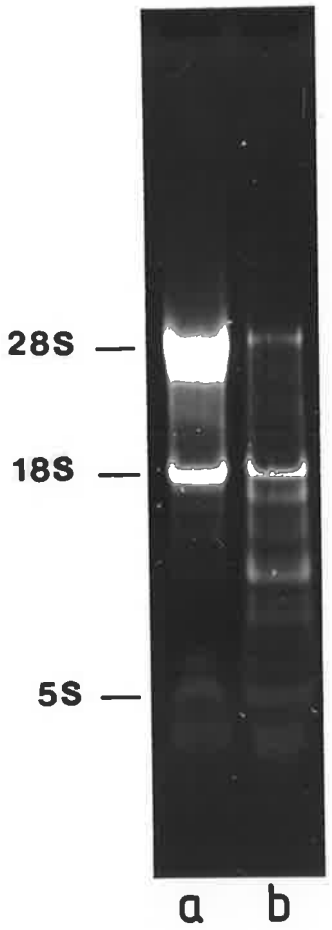
FIGURE 3.3. ELECTROPHORESIS OF TOTAL GUINEA PIG HAIR FOLLICLE RNA UNDER NON-DENATURING AND DENATURING CONDITIONS.

Total hair follicle RNA was resuspended in 10mM NaH_2PO_4 pH6.8 and electrophoresed with or without denaturation in the presence of glyoxal and DMSO (Chapter 2). Electrophoresis was in 1.5% agarose in 10mM NaH_2PO_4 buffer and the RNA was stained with toluidine blue.

(a) Non-denatured RNA

(b) Denatured RNA

28S, 18S and 5S: Relative mobility of RNA markers.


28S
18S
5S
a
b

electrophoretic pattern of denatured total RNA (Fig.3.3b), each fraction contains a relatively large number of components. Fraction 1 (Fig.3.4a) shows a spread of RNA species ranging from 5S up to 18S with a relatively major component of about 8S. Fraction 2 is derived from the 18S RNA peak (Fig.3.1). Under denaturing conditions it contains components down to about 7S and a number of minor components which co-migrate with bands in fraction 1 (Fig.3.4b). Fraction 3 is derived from RNA in the range of 19-21S (Fig.3.1) and while there is a trace of material above 18S, the majority of components lie between 6S and 18S. A number of the components co-migrate with RNA bands found in Fraction 2 (Fig.3.4c). Fraction 4 contains RNA from the 28S peak of follicle RNA (Fig.3.1). The electrophoretic pattern shows very little material at 28S but a large number of components down to about 6S in size. Many of these components have the same electrophoretic mobility as RNA bands in Fraction 3 (Fig.3.4d).

Fraction 1 RNA was further fractionated by sucrose density gradient ultracentrifugation to yield four subfractions, A-D (Fig.3.2). Figure 3.5 shows the electrophoretic patterns of these subfractions under denaturing conditions. Limited resolution of the subfraction peaks (Fig.3.2), results in cross-contamination of a number of components. However, some fractionation has been achieved. The major component occurring at about 8S in fraction 1 (Fig.3.4a) is resolved into three components with slightly different mobilities although the main region of fractionation is in the 12-16S range (Fig.3.5a-d). Subfraction D comprises RNA contained in the 18S peak (Fig.3.2) and this component is the major species present upon denaturation although there are minor bands down to about 8S in size (Fig.3.5d).

4. Optimization for translation of hair follicle RNA in the S30 wheat germ cell-free system.

The S30 wheat germ system of Marcu and Dudock (1974) was

FIGURE 3.4. ELECTROPHORESIS UNDER DENATURING CONDITIONS OF FRACTIONS DERIVED FROM TOTAL HAIR FOLLICLE RNA.

Fractions obtained from sucrose density gradient ultracentrifugation of follicle RNA (Fig.3.1) were denatured in formamide for 10 minutes at 60°C and electrophoresed in a 4.5% polyacrylamide gel containing 98% formamide (Chapter 2). The RNA was stained with ethidium bromide and photographed under UV light.

(a) Fraction 1 RNA

(b) Fraction 2 RNA

(c) Fraction 3 RNA

(d) Fraction 4 RNA

28S, 18S and 5S: Relative mobility of RNA markers.

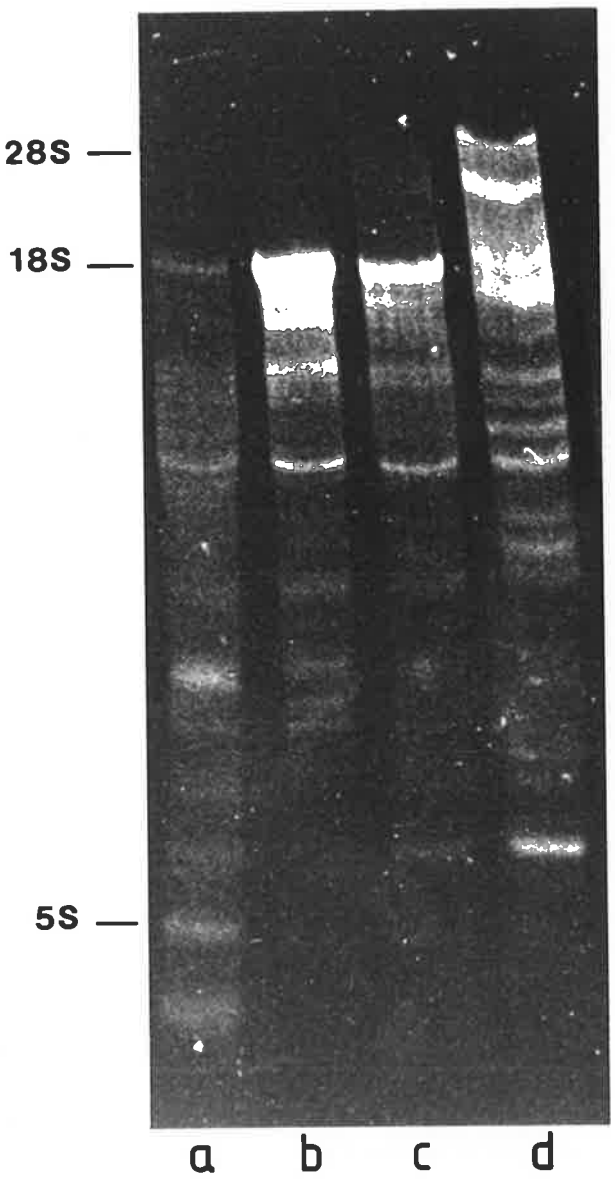

28S
18S
5S
a
b
c
d

FIGURE 3.5. ELECTROPHORESIS UNDER DENATURING CONDITIONS OF SUBFRACTIONS DERIVED FROM FRACTION 1 RNA.

Subfractions obtained from sucrose density gradient ultracentrifugation of fraction 1 RNA (Fig.3.2) were denatured in formamide for 10 min. at 60°C and electrophoresed in a 4.5% polyacrylamide gel containing 98% formamide (Chapter 2).

The RNA was stained with ethidium bromide and photographed under UV light.

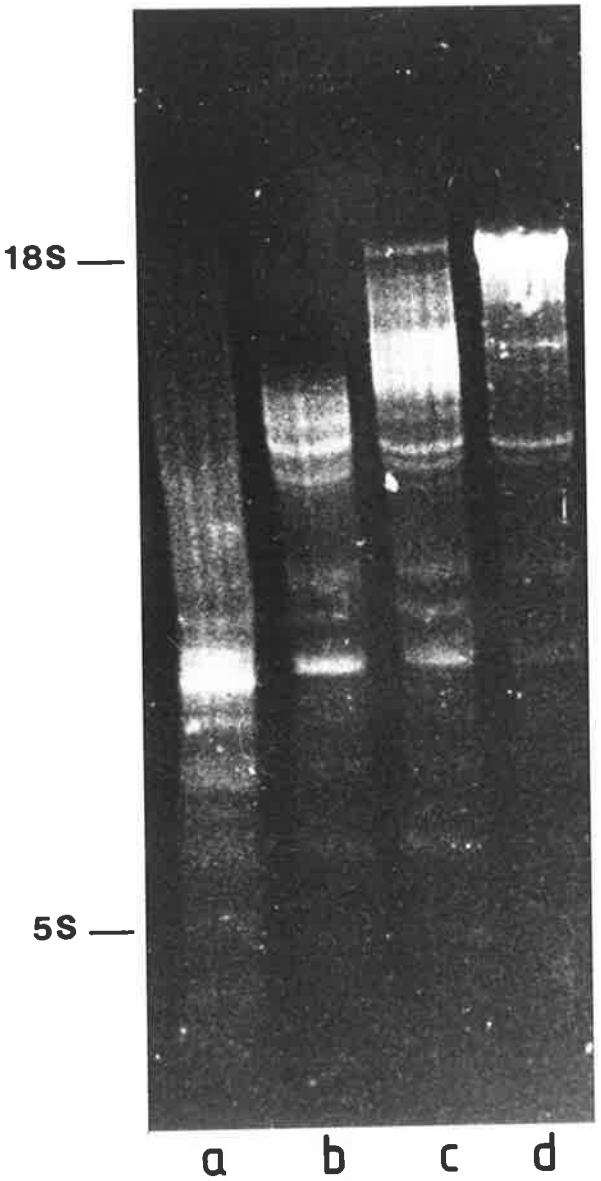
(a) Subfraction A RNA

(b) Subfraction B RNA

(c) Subfraction C RNA

(d) Subfraction D RNA

18S and 5S: Relative mobility of RNA markers


18S
5S
a
b
c
d

optimized for a number of parameters (concentration of Mg^{++} and K^{+}, the volume of S30 per 50µl mix and the amount of RNA per assay) using RNA fractions 1-4. The results obtained for fraction 1 are shown in Fig.3.6 and are typical of the data for fractions 2-4 except for differences in translation efficiency (see later results, this chapter). The optimal concentration of K^{+} was determined using the levels of Mg^{++} and S30 extract (2.5mM and 15µl respectively) reported by Marcu and Dudock (1974). It can be seen (Fig.3.6a) that there is a peak of activity in the range 50-60mM K^{+}. Using a constant concentration of 50mM K^{+}, the optimal level of Mg^{++} in the S30 system was approximately 2.5mM (Fig.3.6b) and for a number of wheat germ extract preparations was within the range 2.0-3.0mM. The volume of S30 extract per 50µl mix which led to maximum protein synthesis in the presence of 50mM K^{+} and 2.5mM Mg^{++} peaked at 15-20µl (Fig.3.6c) while the optimal amount of RNA per translation was 10µg per 50µl mix (Fig.3.6d). Protein synthesis increased almost linearly with RNA levels up to about 6µg, peaked at about 10µg, and then generally decreased with the further addition of RNA. One preparation of S30 extract was found to maintain optimal activity with 20µg RNA per 50µl translation mix (data not shown).

Subsequent translation studies of follicle RNA fractions used 50mM K^{+}, 2.5mM Mg^{++}, 15µl S30 extract and 10µg RNA. The optimal conditions determined for fraction 1 were assumed to apply to the subfractions derived from it although only 0.5-1µg of subfraction RNA was used per 50µl translation mix.

The translation efficiencies of the various hair follicle RNA fractions and subfractions are compared with globin mRNA in Table 3.2. The hair follicle RNAs show greatly reduced protein synthesis activity when compared with globin mRNA. Subfraction B, which is the most active of the follicle RNA fractions, has only about 12% of the globin activity in the S30 system.

Fractions 1-4 all exhibit some translation activity. For 10µg

FIGURE 3.6. OPTIMIZATION FOR TRANSLATION OF HAIR FOLLICLE FRACTION 1 RNA IN THE S30 WHEAT GERM SYSTEM.

Fraction 1 RNA, obtained by sucrose density gradient ultracentrifugation of hair follicle RNA (Fig.3.1) was used to prime cell-free translation in the S30 wheat germ system of Marcu and Dudock (1974).

Radioactivity incorporated into TCA-insoluble products was determined for 50µl mixes. Incubation was at 30°C for 1h in the presence of 1µCi [^{3}H]-leucine as described in Chapter 2.

●——● Fraction 1 RNA.

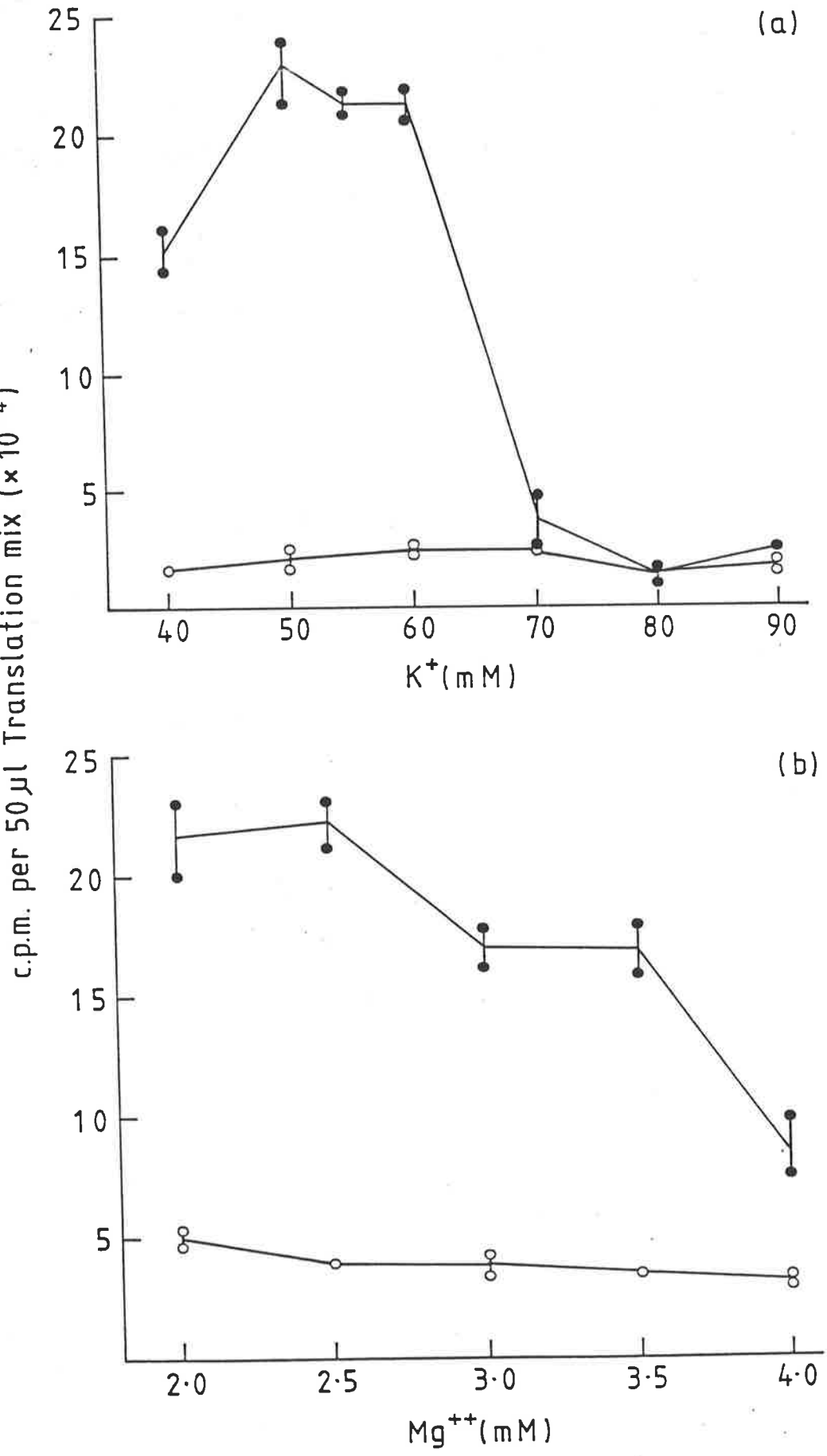
o——o Endogenous activity.

(a) Effect of K^+ concentration on S30 protein synthesis.
Mg^{++} 2.5mM; S30 15µl; RNA 10µg.

(b) Effect of Mg^{++} concentration on S30 protein synthesis.
K^+ 50mM; S30 15µl; RNA 10µg.

(c) Effect of S30 volume on protein synthesis.
K^+ 50mM; Mg^{++} 2.5mM; RNA 10µg.

(d) Effect of RNA content on S30 protein synthesis.
K^+ 50mM; Mg^{++} 2.5mM; S30 15µl.


(a)
25
20
15
10
5
40
50
60
70
80
90
K+(mM)
(b)
25
20
15
10
5
2·0
2·5
3·0
3·5
4·0
Mg++(mM)
c.p.m. per 50µl Translation mix (×10−4)

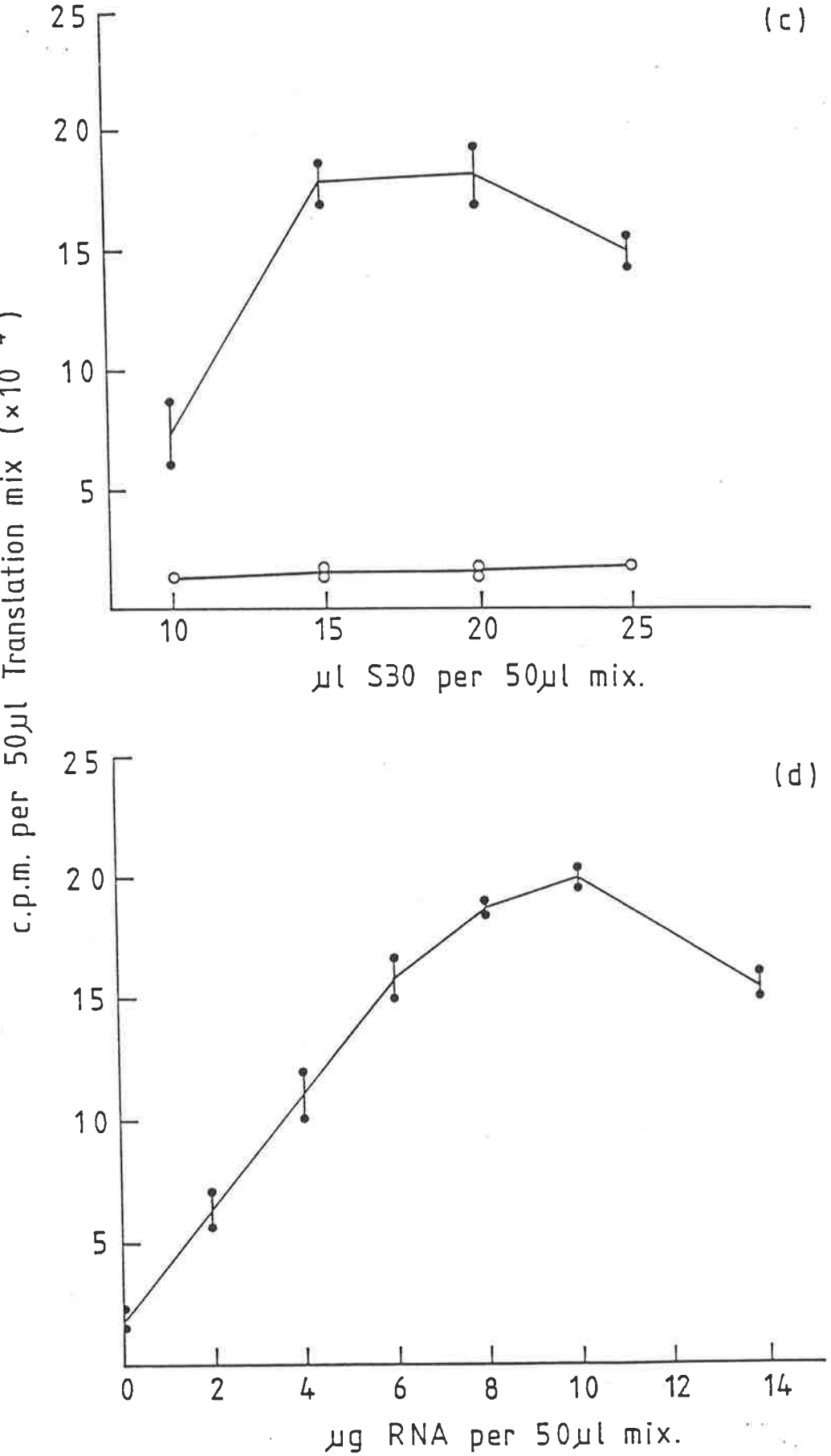

(c)
25
20
15
10
5
c.p.m. per 50µl Translation mix (×10⁻⁴)
10
15
20
25
µl S30 per 50µl mix.
(d)
25
20
15
10
5
0
2
4
6
8
10
12
14
µg RNA per 50µl mix.

TABLE 3.2

EFFICIENCY OF TRANSLATION OF HAIR FOLLICLE RNA FRACTIONS AND SUBFRACTIONS IN THE S30 WHEAT GERM SYSTEM

Hair follicle RNA, its fractions (10µg each), and subfractions (1µg), and globin mRNA (0.4µg) were translated in the S30 wheat germ system of Marcu and Dudock (1974).

The 50µl translation mix contained 2.5mM Mg^{++}, 15µl S30 extract, and 50mM K^{+} (90mM for globin).
Incubation was for 1h at 30°C in the presence of 1µCi [^{3}H]-leucine, and incorporated radioactivity was measured as TCA-insoluble products (Chapter 2).

The results are standardized to c.p.m.x10^{-3} per µg RNA.

RNA		c.p.m.x10^{-3} per µg RNA
Globin		1300
Total follicle RNA		6.5
Fraction	1	23.3
	2	4.1
	3	12.5
	4	3.0
Subfraction	A	93.7
	B	160.6
	C	106.4
	D	50.0

S30 Endogenous activity: 26x10^{3} c.p.m.

of RNA per translation mix the stimulation above background ranges from 1.1 for fraction 4 up to 9 for fraction 1. The poor stimulation by fractions 2 and 4 is to be expected since the RNA in these fractions is obtained from the 18S and 28S peaks respectively (Fig.3.1) - the limited activity indicates that there is mRNA present in these fractions. Fraction 3 shows considerable stimulation of the S30 system while fraction 1, which spans the RNA size range expected to code for most of the keratin proteins, is the most active fraction.

The subfractions A-C show 4-7 times the activity of fraction 1 when compared on a weight basis (Table 3.2). The refractionation of fraction 1 RNA reduces the level of 5S and 18S species and possibly other contaminating material and acts as an enrichment step. Subfraction D, which corresponds to the pooled 18S peak (Fig.3.2) has twice the activity of fraction 1. This translation activity is expected since the fractionation in sucrose gradients is only partial and there is cross contamination of subfraction D with RNA from C (Fig.3.5). It is also possible that subfraction D contains mRNA species in the 17-19S size range which contribute to the translation activity.

Translation of subfractions A-C in a commercial wheat germ system (63mM K^+, 2mM Mg^{++}, 1µg RNA) gave levels of incorporation and stimulation above background comparable to the S30 extract (data not shown).

The nuclease-treated rabbit reticulocyte lysate system was used also to study the translation activity of total hair follicle RNA and subfractions A-C (Table 3.3). As observed in the wheat germ system (Table 3.2), the translation efficiency of hair follicle RNA in the reticulocyte lysate is poor when compared with globin mRNA. The stimulation above background of the follicle RNAs is similar in the two translation systems although subfraction B is less active and C is more active in the reticulocyte lysate than in the S30 system. The manufacturers of the reticulocyte lysate claim that by diluting the

TABLE 3.3

EFFICIENCY OF TRANSLATION OF TOTAL HAIR FOLLICLE RNA AND SUBFRACTIONS A-C IN THE NUCLEASE-TREATED RABBIT RETICULOCYTE LYSATE SYSTEM

Hair follicle RNA (10µg), its subfractions (1µg) and globin mRNA (0.4µg) were translated in the nuclease-treated rabbit reticulocyte lysate system. The 11µl translation mix (100mM K^+, 1.6mM Mg^{++}) contained 8µl lysate unless otherwise indicated.

Incubation was for 1h at 30°C in the presence of 1µCi [^{3}H]-leucine and incorporated radioactivity was measured as TCA-insoluble products (Chapter 2).

The results are standardized to c.p.m.x10^{-3} per µg RNA.

RNA		c.p.m.x10^{-3} per µg RNA
Globin		2160
Total follicle RNA		7.4
Subfraction	A	98.2
	A(1)	70.0
	A(2)	32.7
	B	113.0
	C	173.9
Endogenous activity 18.5x10^3 c.p.m.		

A(1): 6.5µl lysate, 82mM K^+, 1.6mM Mg^{++}.

A(2): 5.0µl lysate, 63mM K^+, 1.6mM Mg^{++}.

lysate and adding Mg^{++} to its original level, the K^+ concentration can be lowered without adversely affecting the translation capabilities of the system. However, dilution of the lysate to give K^+ concentrations of 82mM and 63mM (compared with 100mM K^+ undiluted) led to decreased translation activity (Table 3.3).

The results of translation studies presented in Tables 3.2 and 3.3 show that all of the follicle RNA fractions and subfractions contain mRNA species which stimulate the incorporation of [^{3}H]-leucine into TCA-insoluble products. The nature of these translation products was examined by one- and two-dimensional electrophoresis.

5. Electrophoresis of hair follicle RNA translation products:

(a) SDS-polyacrylamide gel electrophoresis of hair keratin proteins:

Guinea pig hair keratin proteins were isolated, S-carboxymethylated and fractionated into the major protein classes as described by Gillespie (1983). When the total keratin proteins were subjected to electrophoresis in the Laemmli (1970) system, the high-glycine-tyrosine proteins migrated as a single band with the ion front (Fig.3.7a). If the proteins were loaded after the ion front had moved partially through the gel, the high-glycine-tyrosine proteins were resolved into a number of components (Fig.3.7c). The resolution obtained using this system is difficult to explain: when the ion front moves through the gel a continuous buffer system should be established behind it, yet when the proteins were electrophoresed under such continuous buffer conditions the resolution was greatly reduced (Fig.3.7b). The modified Laemmli (1970) system described in Chapter 2 was found to give good resolution of components of the major keratin protein classes, particularly the high-glycine-tyrosine groups (Fig.3.7d-g). The amount of protein loaded for the different fractions (legend to Fig.3.7) reflects the wide range of protein binding affinities for the Coomassie Brilliant Blue stain.

FIGURE 3.7. SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS OF SCM-GUINEA PIG HAIR KERATIN PROTEINS.

Guinea pig hair keratin proteins were extracted and fractionated as described by Gillespie (1983). Electrophoresis in a 12.5% polyacrylamide gel containing SDS was performed using a modified Laemmli (1970) system (Chapter 2) unless otherwise indicated. Proteins were stained with Coomassie Brilliant Blue.

(a) Total guinea pig hair keratin proteins (200µg loaded) - as in Laemmli (1970).

(b) Total guinea pig hair keratin proteins (200µg) - gel and reservoir buffers are both Tris-glycine (Laemmli, 1970).

(c) Total guinea pig hair keratin proteins (200µg).

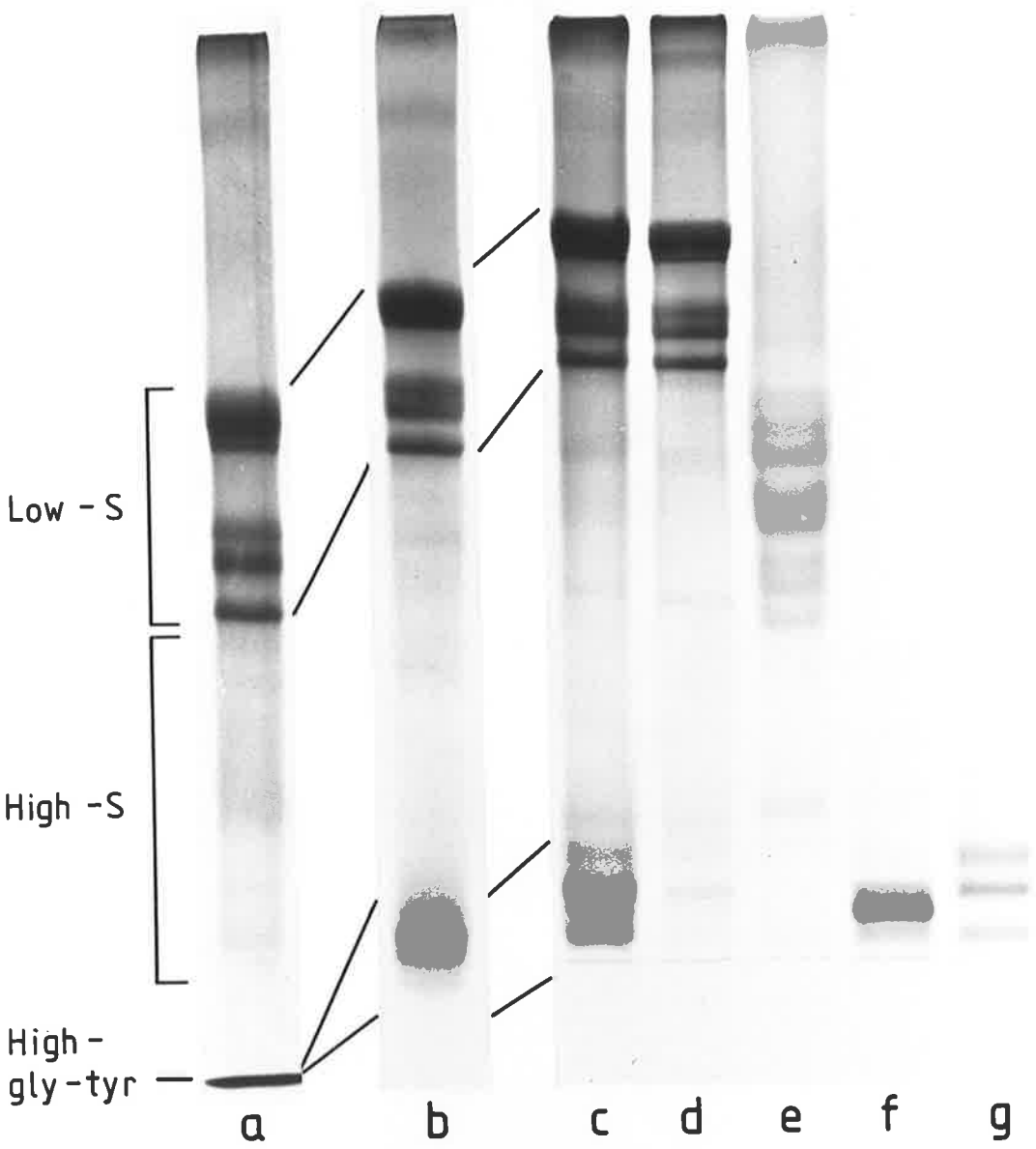
(d) Low-sulphur proteins (150µg).

(e) High-sulphur proteins (300µg).

(f) Type I high-glycine-tyrosine proteins (20µg).

(g) Type II high-glycine-tyrosine proteins (40µg).

Low-S: low-sulphur proteins; High-S: high-sulphur proteins; High-gly-tyr: high-glycine-tyrosine proteins.


Low - S
High -S
High -
gly - tyr
a
b
c
d
e
f
g

The major keratin protein classes fall into three separate molecular weight ranges (Chapter 1, Table 1.1) and this is reflected in their mobility in SDS-polyacrylamide gels (Fig.3.7). It should be noted that the high-sulphur and high-glycine-tyrosine proteins migrate anomolously in SDS-polyacrylamide gels (Marshall, 1983; Frenkel and Blagrove, 1975), so that the molecular weights of these components cannot be estimated from their relative electrophoretic mobilities.

(b) SDS-polyacrylamide gel electrophoresis of hair follicle RNA-S30 wheat germ translation products:

Translation products of total follicle RNA and fractions derived from it were S-carboxymethylated, electrophoresed in SDS and examined by fluorography, as described in Chapter 2. It can be seen (Fig.3.8) that a wide range of protein products have been synthesized in the S30 system. Total follicle RNA codes for proteins with sizes that span the molecular weight ranges of the three major keratin protein classes (Fig.3.8c). There is a low proportion of proteins co-migrating with the low-sulphur components and a number of bands, particularly in the high-sulphur protein region, which are present in the translation products but not apparent in the hair proteins themselves.

It is clear that there is cross-contamination of translation products between fractions 1-4 (Fig.3.8d-g). Components migrating in the lower part of the low-sulphur protein region are found in fraction 1 while fraction 4 contains products in the high-glycine-tyrosine protein region of the gel. However a degree of fractionation has occurred; fraction 1 has a greater proportion of translation products with relatively low molecular weights while fraction 3 contains greater amounts of bands co-migrating with the low-sulphur proteins. Fractions 2 and 4 contain translation products with similar molecular weight distributions to fractions 1 and 3 respectively (disregarding the fact that fractions 1 and 3 contained 4-6 times the amount of radioactivity loaded in fractions 2 and 4).

FIGURE 3.8. SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS OF S30 TRANSLATION PRODUCTS OF HAIR FOLLICLE RNA AND FRACTIONS DERIVED FROM IT.

Total follicle RNA and fractions derived from it (Fig.3.1) were translated in the S30 wheat germ system, S-carboxymethylated and electrophoresed in a 12.5% polyacrylamide gel containing SDS, as described in Chapter 2. Fluorography was according to Bonner and Laskey (1974) and Laskey and Mills (1975). Total hair keratin marker was radiolabelled with [^{14}C] as described in Chapter 2.

(a) Total hair keratin proteins.

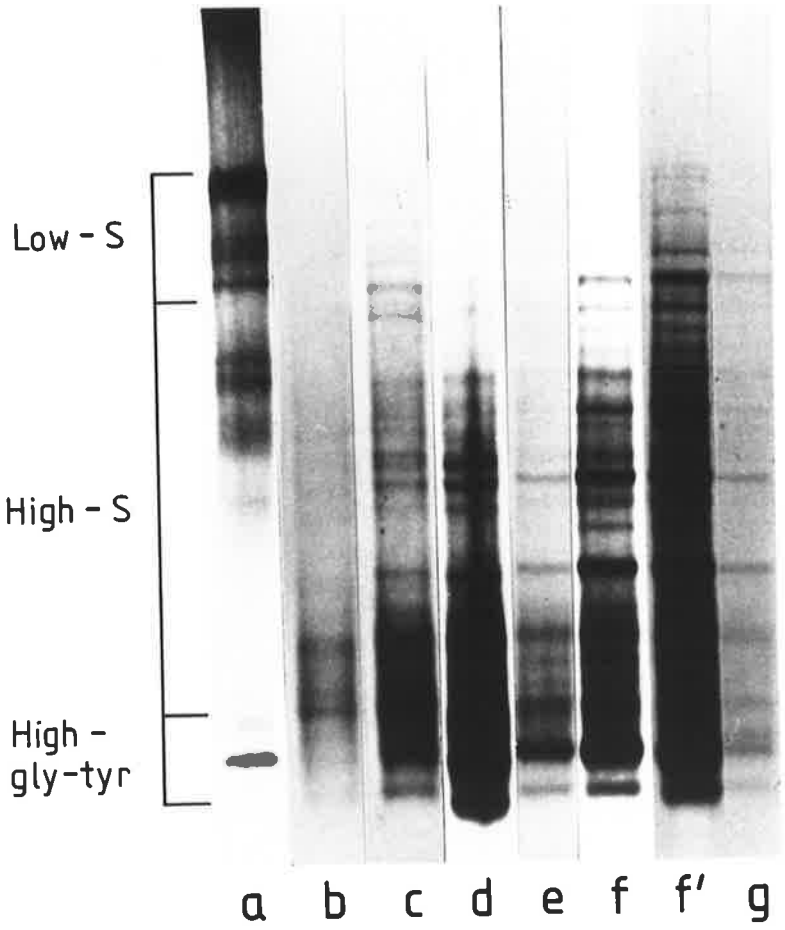
(b) Endogenous S30 activity.

(c) Total follicle RNA translation products.

(d-g) Translation products of RNA fractions 1-4.

f' is a longer exposure of f (fraction 3) to show the low-sulphur components.

Low-S: low-sulphur proteins; High-S: high-sulphur proteins; High-gly-tyr: high-glycine-tyrosine proteins.


Low - S
High - S
High - gly-tyr
a
b
c
d
e
f
f'
g

The electrophoretic separation of subfractions A-D translation products is shown in Fig.3.9. As seen with the products of fractions 1-4 (Fig.3.8), there is cross-contamination between subfractions. Some fractionation is occurring since the RNA subfraction of lowest S value (A) produces a predominance of components in the high-glycine-tyrosine protein region while subfractions with higher S values (C and D) have translation products with greater proportions of components in the high-sulphur and low-sulphur protein regions. The amount of protein in the low-sulphur region varied considerably in different experiments.

The hard keratins are an extremely complex group of proteins and no single dimension electrophoretic technique is capable of adequately resolving on one gel even the major components of the three protein classes (see Chapter 1). Two-dimensional electrophoresis, alkaline-urea followed by SDS-polyacrylamide, provides good resolution of many components and results in an excellent separation of the keratin protein classes (Marshall, 1981). In order to define more exactly which keratin proteins are synthesized in cell-free translation systems, the products from follicle RNA and a number of its fractions were examined by two-dimensional electrophoresis.

(c) Two-dimensional electrophoresis of SCM-guinea pig hair proteins:

Total hair keratin proteins were extracted, alkylated in the presence of [^{14}C]-iodoacetic acid and subjected to two-dimensional electrophoresis as described in Chapter 2. Figure 3.10 shows two exposures of a fluorograph of total hair keratin proteins. Components of the major protein classes are well-resolved and extreme over-exposure of the standard fluorograph differs only by the appearance of some additional minor spots. The positions of the protein classes have been assigned according to Marshall (1981).

The high-sulphur protein class appears to be the most abundant as a result of the labelling technique. While this class comprises only

FIGURE 3.9. SDS POLYACRYLAMIDE GEL ELECTROPHORESIS OF S30 TRANSLATION PRODUCTS OF HAIR FOLLICLE RNA SUBFRACTIONS

Hair follicle RNA subfractions (Fig.3.2) were translated in the S30 wheat germ system, S-carboxymethylated, and electrophoresed in a 12.5% polyacrylamide gel containing SDS as described in Chapter 2. Fluorography was according to Bonner and Laskey (1974) and Laskey and Mills (1975). Total hair keratin marker was radiolabelled with [^{14}C] as described in Chapter 2.

(a) Total hair keratin proteins.

(b) Endogenous S30 activity.

(c-f) Translation products of RNA subfractions A-D.

Low-S: low-sulphur proteins; High-S: high-sulphur proteins; High-gly-tyr: high-glycine-tyrosine proteins.

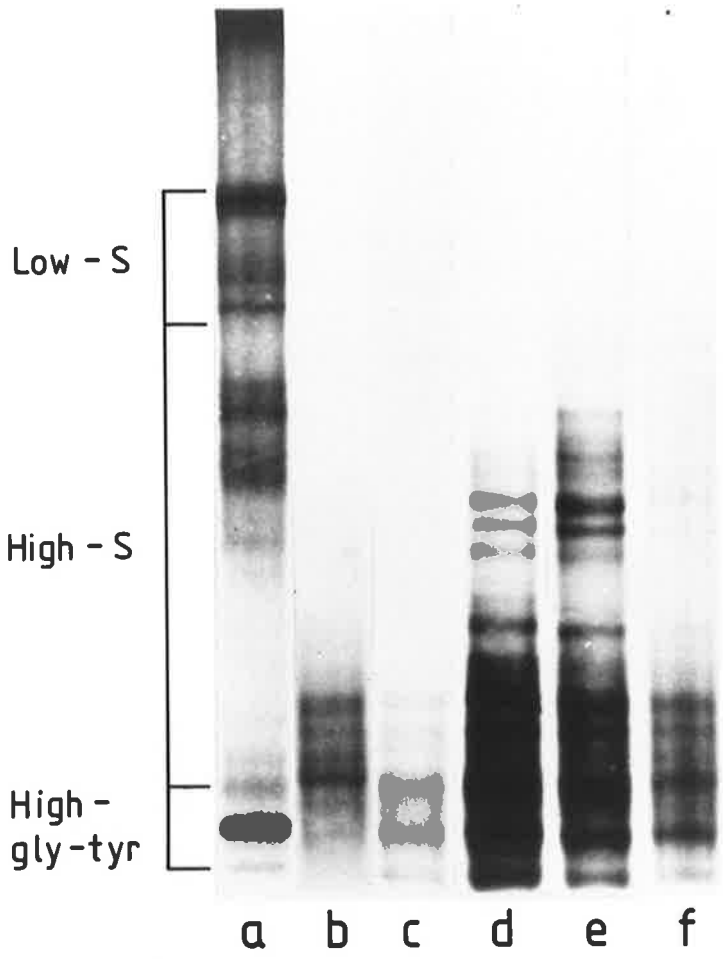

Low - S
High - S
High -
gly - tyr
a
b
c
d
e
f

FIGURE 3.10. TWO-DIMENSIONAL ELECTROPHORESIS OF TOTAL HAIR KERATIN PROTEINS

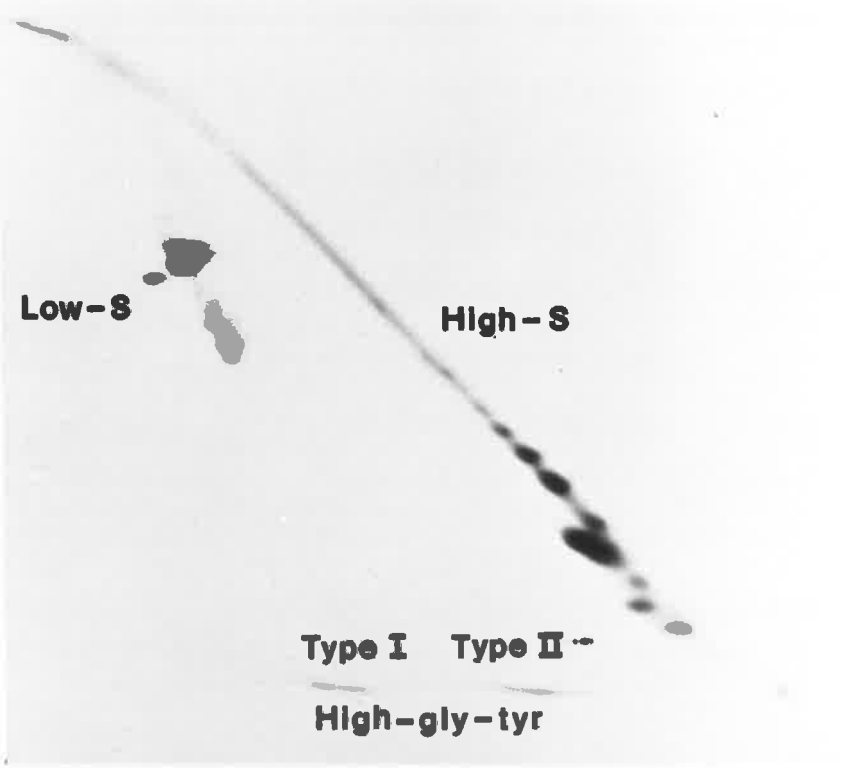
Guinea pig hair proteins were extracted, alkylated in the presence of [^{14}C]-iodoacetic acid and subjected to two-dimensional electrophoresis as described in Chapter 2. First dimension was an alkaline-urea 7.5% polyacrylamide gel, the second an SDS-containing 10% polyacrylamide gel. Fluorography was according to Bonner and Laskey (1974) and Laskey and Mills (1975). The positions of the major keratin protein families are based on the results of Marshall (1981).

(a) Fluorography for 7h.

(b) Fluorography for 30h.

Low-S: Low-sulphur proteins; High-S: High-sulphur proteins; High-gly-tyr: High-glycine-tyrosine proteins.

(a)



(b)

about 30% of the hair proteins (Steinert and Rogers, 1973b), the constituent polypeptides contain substantially more cystine than the other components and hence have a greater number of potential radiolabelling sites.

With regard to translation products, which were labelled with either leucine or a mixture of leucine, lysine, phenylalanine, tyrosine and proline, the low-sulphur components should be more heavily radiolabelled.

(d) Two-dimensional electrophoresis of hair follicle RNA - rabbit reticulocyte lysate translation products:

The endogenous products of the wheat germ cell-free translation system are of relatively low molecular weight (Fig.3.9b). In the two-dimensional electrophoresis system, these products migrated in the region occupied by the high-glycine-tyrosine proteins (data not shown). The endogenous products of the nuclease-treated rabbit reticulocyte lysate did not co-migrate with any of the major keratin proteins (Fig.3.11a) and was the translation system of choice for two-dimensional analysis. Figure 3.11(b-d) shows the fluorographs obtained following electrophoresis of follicle RNA subfraction translation products. A relatively large proportion of the incorporated radioactivity is located in low molecular weight products. Some spots can be seen in the Type II high-glycine-tyrosine region of Fig.3.11(d-e) although shorter exposure times revealed a large number of components across the lower portion of the gel (data not shown). A number of spots in the high-sulphur region co-migrate with components extracted from guinea pig hair (Fig.3.11b-e compared with Fig.3.10). While some subfraction C translation products did not include low-sulphur proteins (Fig.3.11d), others had small amounts of all the major low-sulphur components (Fig.3.11e). All translation products examined contained varying proportions of components which could not be found in the hair extract.

While it was apparent that the hair follicle RNA isolated in

FIGURE 3.11. TWO-DIMENSIONAL ELECTROPHORESIS OF RABBIT RETICULOCYTE LYSATE TRANSLATION PRODUCTS OF HAIR FOLLICLE RNA SUBFRACTIONS.

Hair follicle RNA subfractions (Fig.3.2) were translated in the rabbit reticulocyte lysate system in the presence of tritiated leucine, lysine, phenylalanine, tyrosine and proline. Following S-carboxymethylation, the translation products were subjected to two-dimensional electrophoresis as described in Chapter 2. First dimension was an alkaline-urea 7.5% polyacrylamide gel, the second an SDS-containing 10% polyacrylamide gel. Fluorography was according to Bonner and Laskey (1974) and Laskey and Mills (1975). The position of the major keratin protein families is based on the results of Marshall (1981).

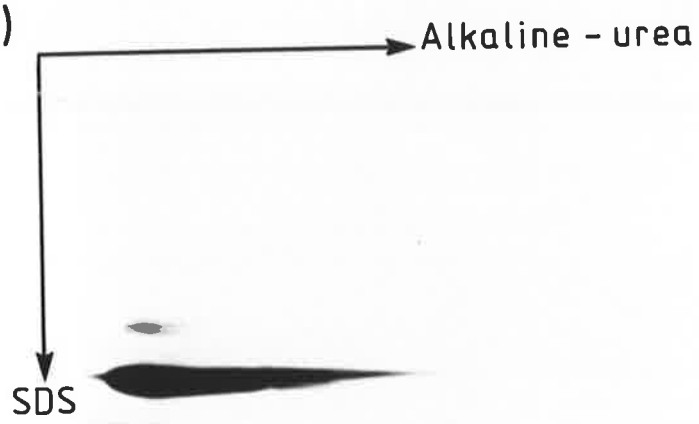
(a) Endogenous reticulocyte activity (exposure 2 weeks).

(b-d) Translation products of RNA subfractions A-C (6 weeks).

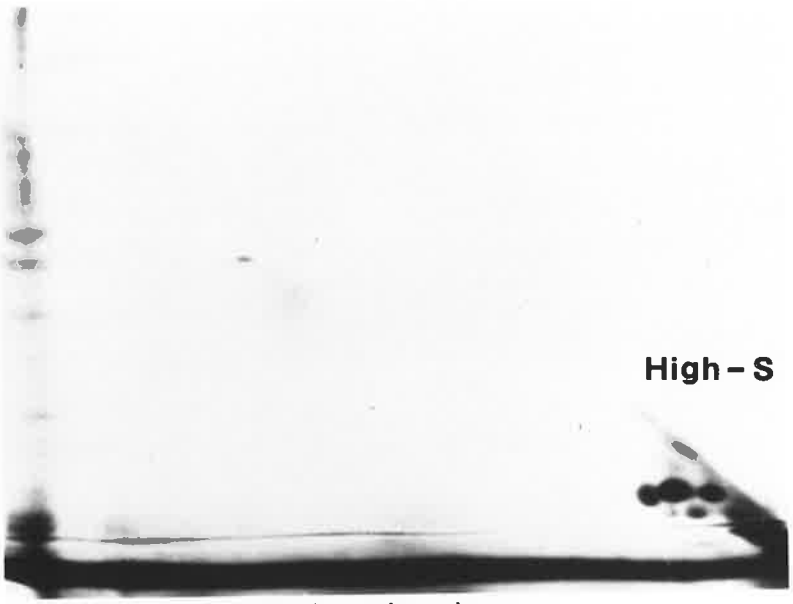
(e) Translation products of a different subfraction C (6 weeks).

Low-S: low-sulphur proteins; High-S: high-sulphur proteins; High-gly-tyr: high-glycine-tyrosine proteins.

(a)



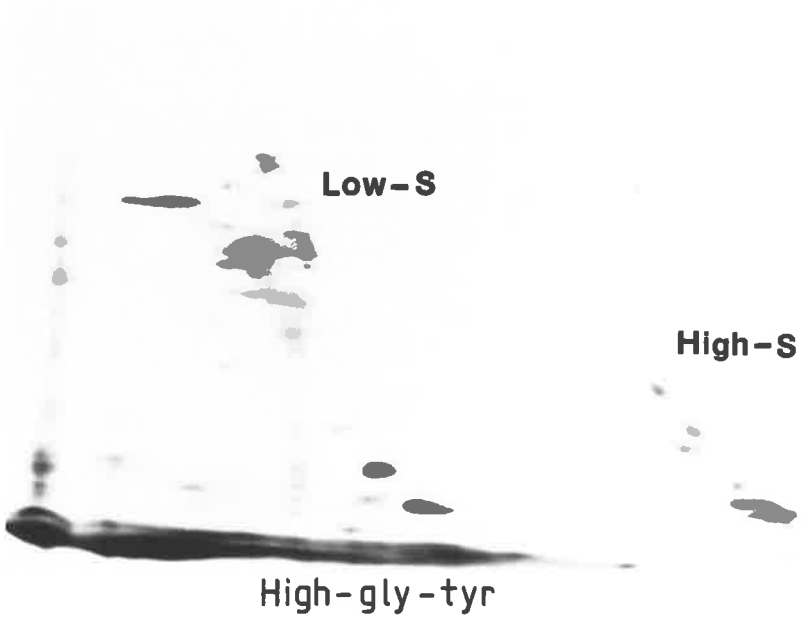
(b)



(c)

(d)

(e)



High-gly-tyr

the presence of UVO was partially degraded (Figs. 3.3, 3.8, 3.9, 3.11) the fluorographs of translation products indicated that full-length mRNAs coding for the major keratin proteins were present. Thus an attempt was made to construct a hair follicle cDNA library.

6. Construction and screening of a hair follicle cDNA library:

RNA subfraction C was chosen as the most suitable starting material for the construction of a follicle cDNA library. This subfraction was more likely to contain mRNAs coding for the low-sulphur proteins and at the same time was reduced in the proportion of small, presumably degraded, mRNA species (Fig.3.11e).

Double-stranded cDNA (dscDNA) was synthesized as described in Chapter 2, using the "loop-back" method. The first strand was prepared using either oligo(dT)- or random-priming. Random oligodeoxynucleotides were used as alternate primers for the synthesis of cDNA because of the low efficiency of first strand synthesis found with oligo(dT). The use of random primers meant that ribosomal cDNA clones would be produced. However the random nature of the cDNA products increased the probability that some coding sequences would be obtained from the mRNA species.

Using oligo(dT) as primer only 0.5-1% of the RNA was copied into cDNA. Following synthesis of the second strand, only 10% of the dscDNA was resistant to S1 nuclease digestion, so that approximately 0.1% of the RNA was copied into dscDNA. With random-priming, approximately 15% of the RNA was copied into dscDNA of which 20% was S1 nuclease-resistant. Therefore about 3% of the RNA was actually copied into dscDNA.

The synthesis of dscDNA was followed by removing aliquots at different stages and examining them under denaturing conditions in agarose gels. Figure 3.12 shows aliquots of cDNA, pre-S1 nuclease dscDNA and post-S1 nuclease dscDNA from an oligo(dT)-primed synthesis. The length of cDNA copies extended up to about 600 bases. Following second strand synthesis, the length was significantly increased while S1

FIGURE 3.12. ANALYSIS OF cDNA AND dscDNA BY ELECTROPHORESIS IN ALKALINE AGAROSE.

Electrophoresis of samples taken at various stages during the course of oligo(dT)-primed dscDNA synthesis.

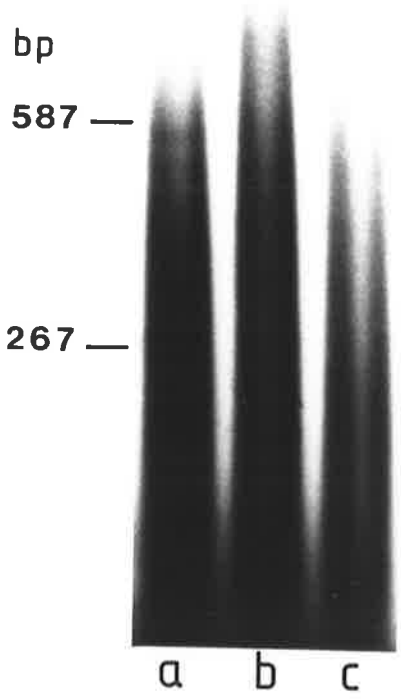
Samples were ethanol precipitated and resuspended in a loading buffer containing 50mM NaOH and 2mM EDTA. Electrophoresis was in a 2% agarose gel containing 30mM NaOH and 2mM EDTA and was followed by autoradiography.

(a) cDNA.

(b) dscDNA prior to S1 nuclease digestion.

(c) dscDNA after S1 nuclease digestion.

The sizes (bp) of some HaeIII-digested pBR322 fragments are indicated.


bp
587
267
a
b
c

nuclease digestion produced dscDNA of approximately the same size as the cDNA.

The dscDNA was fractionated on 5-20% sucrose gradients or by chromatography over agarose A-150 to isolate molecules greater than about 300 bp in size. Homopolymeric nucleotide tails of dCTP were added to the size-selected dscDNA and annealed to dG-tailed pBR322 as described in Chapter 2. Transformation of E.coli strain ED8654 yielded 32 recombinants from oligo(dT)-primed dscDNA and 19 from random-primed dscDNA which were not derived from rRNA.

A small amount of DNA was prepared from each recombinant, restricted with EcoRI and examined by electrophoresis in agarose gels using the miniscreen technique described in Chapter 2. Figure 3.13 shows the miniscreen result for a number of the recombinants. While the cDNA clones in this figure range in size from 0.3-1.2 kb, the majority of recombinants were in the range 0.2-0.5 kb. The size range was found to be similar for the oligo(dT)-and random-primed cloning experiments.

As described above, the efficiency of copying follicle RNA into cDNA using oligo(dT)-priming was found to be very low. This made colony hybridization screening of the cDNA clones difficult. Long exposure times resulted in only weak signals. Figure 3.14 shows the result of colony hybridization using cDNA to fraction 1 RNA as the probe; of the 51 colonies, only 8 show a definite positive hybridization.

Probes with a higher specific activity were prepared by kinasing fragmented RNA (Chapter 2). In situ colony hybridizations using [γ-^{32}P]-labelled subfractions A and C (Fig.3.15) produced much stronger signals. In addition, the different probes gave an overlapping set of positive clones. Hybridizations with probes derived from the same RNA subfraction from different experiments did not give consistent results (data not shown). However a number of clones were consistently positive when probed with cDNA and subfraction C RNA - pGPHF2, 32 and 40 (pGPHF = plasmid Guinea Pig Hair Follicle). Clones 2 and 32 were the result of

FIGURE 3.13. MINISCREEN ANALYSIS OF RECOMBINANT PLASMIDS.

Miniscreen amounts of DNA were prepared as described in Chapter 2, digested with EcoRI and electrophoresed in 1% agarose. The gel was stained with ethidium bromide and photographed under UV light.

(a) pBR322.

(b) pBR322 containing an insert of 1.5 kb.

(c) pBR322 containing an insert of 0.25 kb.

Clones 2, 32 and 40 refer to three recombinants used for DNA sequencing studies (section 6, this chapter).

a. b. c. 2. 32. 40. 49. 50. 51.

FIGURE 3.14. **IN SITU** COLONY HYBRIDIZATION OF HAIR FOLLICLE cDNA CLONES.

Nitrocellulose filters containing colonies of hair follicle clones were prepared and hybridized with [^{32}P]-labelled fraction 1 cDNA as described in Chapter 2. Filters were washed to a stringency of 1xSSCE/0.1%SDS at 65^{0}C and autoradiographed for 3 weeks in the presence of intensification screens.

(a) Autoradiograph of filter.

(b) Grid showing nomenclature of clones which gave positive hybridization.

(a)

(b)

1	2			5					
								29	
		32							
	34						40		
	44								

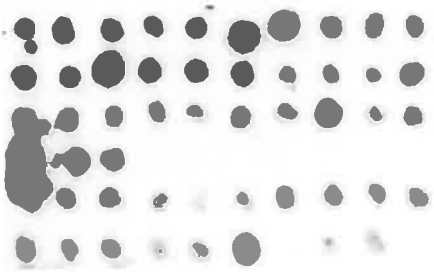
FIGURE 3.15. **IN SITU** COLONY HYBRIDIZATION OF HAIR FOLLICLE cDNA CLONES.

Nitrocellulose filters containing colonies of hair follicle clones were prepared and hybridized with [γ-^{32}P]-labelled RNA subfractions as described in Chapter 2. Filters were washed to a stringency of 1xSSCE/0.1% SDS at 65^{0}C and autoradiographed overnight in the presence of intensification screens.

(a) Hybridization using subfraction A RNA as probe. Autoradiograph and grid showing nomenclature of positive clones.

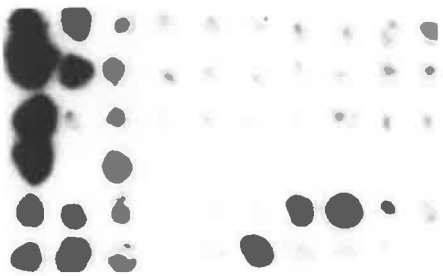
(b) Hybridization using subfraction C RNA as probe. Autoradiograph and grid showing nomenclature of positive clones.

(a)



	2			5	6	7			
		13	14	15	16				
					26		28		
31	32								
		35							
					48				

(b)



1	2	3							10
11	12	13							
21		23							
31		32							
33	34	35				39	40	41	
43	44	45			48				

oligo(dT)-primed dscDNA synthesis and contained inserts of about 240 bp and 1100 bp respectively, while pGPHF 40 was obtained from the random-primed library and had an insert approximately 240 bp long (Fig.3.13).

7. DNA sequencing studies on pGPHF2, 32 and 40:

Clones pGPHF2, 32 and 40 were mapped with a number of restriction enzymes and the chemical cleavage technique of Maxam and Gilbert (1980) was used to obtain sequence data (Fig.3.16). The sequence of pGPHF2 was obtained with the assistance of Dr. B. C. Powell. Clones pGPHF2 and 40 were completely sequenced and over 600 bp of pGPHF32 were determined (although data was not obtained for both strands). Analysis of the pGPHF40 and 32 data indicated that there were no significant stretches of coding sequence in either clone (data not shown). One of the reading frames of pGPHF2 gives an uninterrupted coding sequence (Fig.3.17). No other reading frame of this clone had significantly long regions of coding sequence.

Comparison of the amino acid composition of the deduced sequence shown in Fig.3.17 with those of the hair and hair follicle proteins reported by Steinert and Rogers (1973b) suggests that pGPHF2 does not code for a major keratin protein.

Further sequencing or other studies on the cDNA library were not justified. While the partial degradation of guinea pig hair follicle RNA had been assumed to be unavoidable, one aspect of the RNA isolation procedure remained open to further investigation - the follicle harvesting technique.

8. The isolation of intact hair follicle RNA:

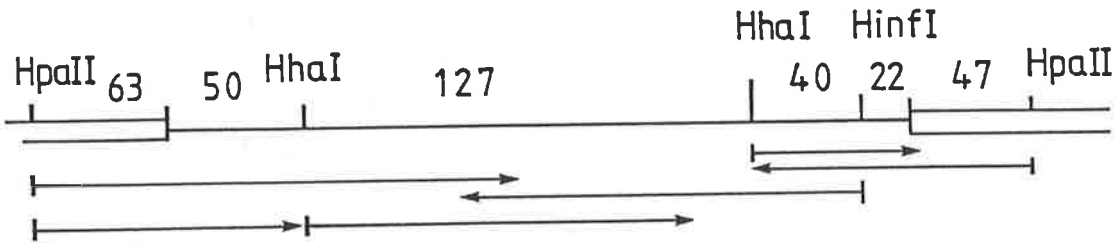
(a) Harvesting of follicles from live guinea pigs:

The Araldite technique of Wilkinson (1970a) which was used successfully for the isolation of wool follicles from sheep was unsuitable for guinea pigs because of the long curing time (G. E. Rogers, personal communication). A substitute for Araldite was required which was rapid-curing, non-toxic and strong enough to enable follicles to be

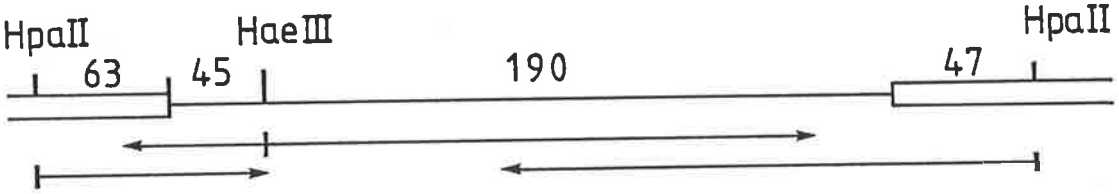
FIGURE 3.16. RESTRICTION ENZYME SITES AND DNA SEQUENCING STRATEGY FOR CLONES pGPHF2, 32 AND 40.

Restriction cleavage sites within the inserts of recombinant plasmids pGPHF2, 32 and 40 are shown. The single line represents the inserted DNA while the double line represents parental pBR322 DNA. Fragment sizes are expressed in base pairs. The arrows indicate the direction and extent of DNA sequence data obtained from each restriction fragment.

pGPHF 2



pGPHF 40



pGPHF 32

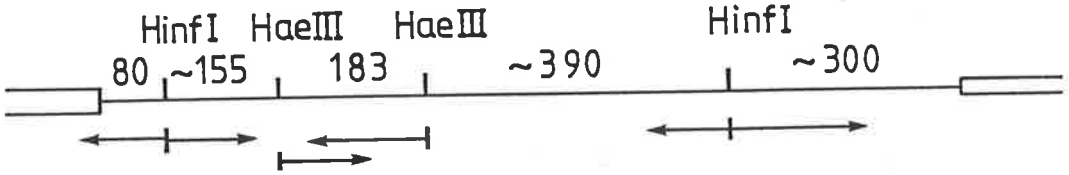


FIGURE 3.17. NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCE OF CLONE pGPHF2.

The nucleotide sequence of pGPHF2 was determined by the method of Maxam and Gilbert (1980) using the strategy outlined in Fig.3.16. The deduced amino acid sequence is shown below the nucleotide data. The dCTP and dGTP homopolymeric tails have not been included in the sequence.

```
5'                    15                        30                        45
  TTA.GAA.TCA.CTA.GTA.GGA.GTG.AGA.AAA.ACA.GCA.CGG.GGT.GGT.GCG
                                                10
  LEU GLU SER LEU VAL GLY VAL ARG LYS THR ALA ARG GLY GLY ALA

                      60                        75                        90
  CGT.TGT.ATG.GGT.GTG.TGT.TAC.ATT.ACA.TTC.ACT.AGT.AGG.AGT.GAG
                      20                                                  30
  ARG CYS MET GLY VAL CYS TYR ILE THR PHE THR SER ARG SER GLU

                      105                       120                       135
  AAA.AAC.ATC.ACG.GGG.AGG.TGC.GCA.TTT.GAT.GGG.TGT.GTG.TAC.ATA
                                                40
  LYS ASN ILE THR GLY ARG CYS ALA PHE ASP GLY CYS VAL TYR ILE

                      150                       165                       180
  CAT.CAT.GTG.GAG.TGG.AAA.ACA.TCA.CGG.GGT.GGT.GCG.CAT.TGT.ATG
                      50                                                  60
  HIS HIS VAL GLU TRP LYS THR SER ARG GLY GLY ALA HIS CYS MET

                                  201
  GGT.GTG.TGT.TAC.ATT.GGT.TTN.
                                  67
  GLY VAL CYS TYR ILE GLY PHE/LEU
```

harvested. I am indebted to L. N. Jones, C.S.I.R.O. Division of Protein Chemistry for suggesting that the acrylic resin Vertex SC could be a suitable adhesive.

Following a series of experiments investigating different liquid:powder ratios of Vertex SC, amount of resin applied, setting time allowed and handling following harvesting, the conditions described in Chapter 2 were found to be suitable for the isolation of hair follicles from live guinea pigs.

The resin consists of a liquid and powder which are mixed immediately prior to use. A ratio of 1:1 was found to give an adhesive suitable in its setting and handling properties. Following a thin application of resin, using glass-fibre tape as a backing, the follicles could be harvested about 12 min after the ingredients had been mixed. The hair follicles, attached by the resin to the glass-fibre tape were immediately placed into liquid N_2 and then stored at -80^0C until required. About 0.2g of follicles could be harvested from a single guinea pig. The guinea pigs subjected to this follicle-harvesting protocol recovered with no apparent ill-effects.

(b) Preparation of poly(A)$^+$ RNA:

Follicles isolated by the acrylic resin-liquid N_2 technique were extracted in GuHCl-2-mercaptoethanol as described in Chapter 2 and total RNA was isolated following ultracentrifugation through a CsCl cushion. Contaminating DNA was removed by high-salt precipitation of the RNA. The yield of total RNA was approximately 600µg per guinea pig. This compares with about 500µg per guinea pig using the UVO method described at the beginning of this chapter.

Poly (A)$^+$ RNA was prepared by chromatography over oligo(dT)-cellulose as described in Chapter 2. A single passage through the column was found to remove about 92% of the RNA, leaving 8% poly(A)$^+$ RNA although electrophoresis of this fraction indicated that some ribosomal species were still present (data not shown).

(c) Electrophoresis under denaturing conditions of RNA isolated from wax-sheet and Vertex SC-harvested follicles:

Total follicle RNA extracted from wax-sheet or Vertex SC-harvested follicles was denatured in the presence of glyoxal and DMSO and compared by electrophoresis as described in Chapter 2. It can be seen (Fig.3.18a) that RNA prepared by the UVO method from wax-sheet-harvested follicles has very little staining material at the position of 28S RNA. In addition, there are many components smaller than 18S present in this RNA preparation which is indicative of degradation. The RNA prepared from Vertex SC-harvested follicles (Fig.3.18b) has very little low molecular weight material and there is strong staining of 28S RNA.

(d) Synthesis of cDNA:

Subfraction C RNA, prepared from wax-sheet-harvested follicles (Fig.3.2), and poly(A)$^+$ RNA from Vertex SC follicles were used as templates for oligo(dT)-primed cDNA synthesis (Chapter 2). Only about 1% of the wax-sheet follicle RNA was copied into cDNA while about 7-10% of the Vertex SC follicle poly(A)$^+$ RNA was converted to cDNA. Following hydrolysis of the RNA with alkali, the cDNA was subjected to electrophoresis in an agarose gel (Fig.3.19). The poly(A)$^+$-derived cDNA had a much greater proportion of higher molecular weight species than subfraction C cDNA. While subfraction C cDNA was significantly smaller than 18S RNA, the cDNA synthesized from poly(A)$^+$ RNA contained material up to 18S (2000 bases).

9. The isolation of poly(A)$^+$ RNA from wool follicles harvested using the Vertex SC technique:

At the time that the Vertex SC follicle harvesting technique was being used to isolate hair follicles from guinea pigs, sheep were made available for the keratin gene project. Since the guinea pig was being used only as a convenient laboratory model and essentially all keratin protein data were obtained from wool, it was decided that the guinea pig studies should be discontinued in favour of an investigation

FIGURE 3.18 ELECTROPHORESIS OF TOTAL GUINEA PIG HAIR FOLLICLE RNA PREPARED FROM WAX SHEET OR VERTEX SC-HARVESTED FOLLICLES.

Total RNA was prepared using the UVO method for wax sheet-harvested follicles and the GuHCl-CsCl method for Vertex SC-harvested follicles. The RNA was resuspended in 10mM NaH_2PO_4 pH6.8 and electrophoresed following denaturation in the presence of glyoxal and DMSO (Chapter 2). Electrophoresis was carried out in a 1.5% agarose gel in 10mM NaH_2PO_4 buffer and the RNA was stained with toluidine blue.

(a) Wax sheet (UVO) RNA.

(b) Vertex SC (GuHCl/CsCl) RNA.

28S and 18S : Position of ribosomal RNA markers.

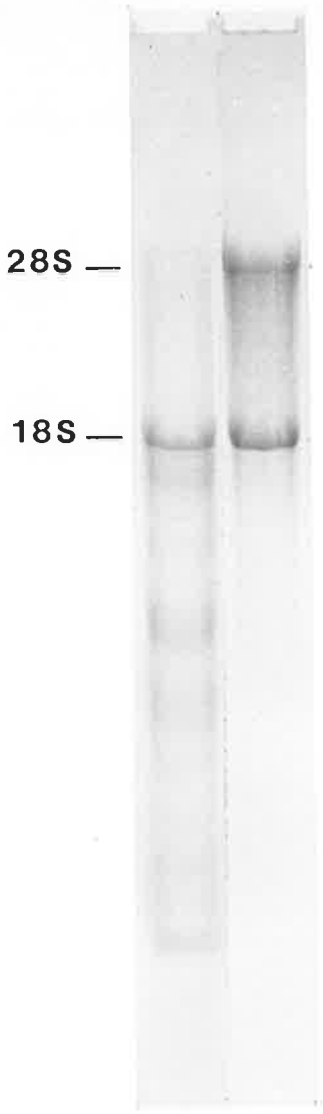

28S
18S
a
b

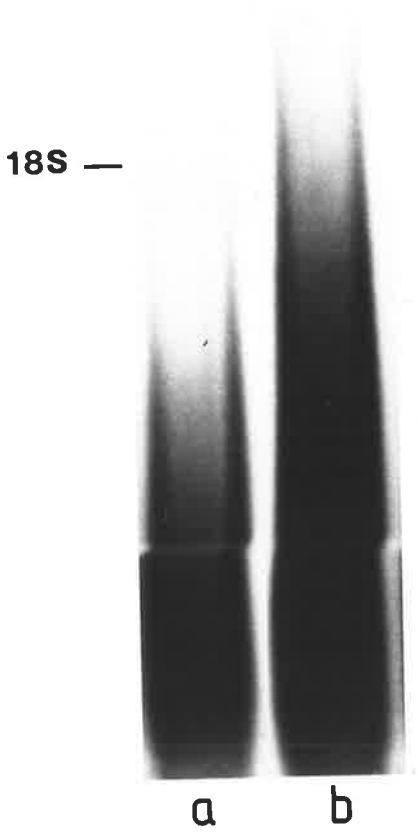
FIGURE 3.19. ELECTROPHORESIS OF cDNA SYNTHESIZED FROM WAX SHEET AND VERTEX SC FOLLICLE RNA.

cDNA was prepared using oligo(dT)-priming of subfraction C RNA purified from wax sheet-harvested follicles and poly(A)$^+$ RNA prepared from Vertex SC-harvested follicles (Chapter 2). The RNA was hydrolyzed with alkali and the cDNA electrophoresed in a 1.5% agarose gel in TAE buffer. The gel was then dried and autoradiographed.

(a) Subfraction C cDNA.

(b) Poly(A)$^+$ cDNA.

18S indicates the position of an 18S rRNA marker.


18S
a
b

of sheep follicle RNA.

Initial experiments indicated that wool follicles could be readily prepared using the Vertex SC technique. Individual sheep were used a number of times without showing any ill-effects. The RNA isolated from wool follicles was found to be undegraded and the poly(A)$^+$ fraction acted as a suitable template for the synthesis of cDNA greater than 2000 bases in length (data not shown).

Wool follicle RNA has been used successfully to construct a cDNA library which contains clones representative of the major keratin protein familes (Kuczek and Rogers, 1985; E. Kuczek, personal communication). In addition, the RNA has been used in Northern blotting experiments (Kuczek and Rogers, 1985) and for genomic Southern blotting studies (Chapters 5 and 6).

C. Discussion.

The aim of the work described in this chapter was to isolate and characterize guinea pig hair follicle RNA. The RNA was to be fractionated and used to construct cDNA libraries containing clones for the different major keratin protein classes.

The presence of RNases in the epidermis has made routine preparation of keratin-coding RNA very difficult. Methods for inactivation of RNase that are adequate for other cell systems are ineffective with epidermal tissues. Although Steinert and Rogers (1971a) were able to prepare large polysomes from the hair follicles of very young guinea pigs, Gilmartin and Freedberg (1975) and Lock (1977) were unable to overcome the degradative effects of RNases. Lock (1977) investigated a wide range of parameters that might improve the quality of polysomes and RNA isolated from guinea pig hair follicles and found that the most effective means of reducing RNase activity was inclusion in the extraction buffer of the UVO complex (see also Rogers et al., 1981). However, RNA isolated in the presence of UVO was still partially degraded.

In the current study the conditions used to prepare hair follicle RNA were essentially identical to those described by Lock (1977). The concentration of uridine and $VOSO_4$ were 10mM and 1mM respectively. Gray (1974) found that 2mM guanosine and 0.2mM $VOSO_4$ were sufficient to completely inhibit RNase activity during the isolation of French bean leaf RNA. More recently, Berger and Berkenmeier (1979) reported that levels of 10mM for individual ribonucleosides in combination with 10mM $VOSO_4$ were necessary for the preparation of intact lymphocyte cytoplasmic RNA. At 5mM $VOSO_4$ approximately 90% protection of the RNA was achieved, while at 1mM $VOSO_4$ only 65% protection was observed. It is unlikely that the limited degradation of follicle RNA was due to insufficient $VOSO_4$ since an increase to 5mM failed to improve the quality of the RNA isolated (data not shown).

The guanidine thiocyanate-CsCl technique of Chirgwin et al. (1979) which was used to isolate intact RNA from the pancreas, a tissue rich in RNase, resulted in a low RNA yield and in a reduction of translation activity when it was used to prepare hair follicle RNA (data not shown).

The subsequent studies on the characterization of guinea pig hair follicle RNA were undertaken with the assumption that limited degradation of RNA was unavoidable. While this assumption was later shown to be incorrect, the difficulty in preparing hair follicle RNA from small laboratory animals is highlighted by the fact that only one publication, describing the preparation of intact mouse hair follicle RNA, has appeared in the literature (Bertolino et al., 1982).

Accepting the limitations of the guinea pig hair follicle RNA preparation, an effort was made to characterize the RNA and to use it as the template for the construction of a cDNA library.

Lock (1977) had fractionated total hair follicle RNA in sucrose density gradients and found that a 6-17S fraction was the most active in an S23 wheat germ translation system. This fraction (fraction 1 in

Fig.3.1) should contain mRNA species for the high-sulphur, high-glycine-tyrosine and most, if not all, of the low-sulphur proteins (Table 1.1). Recentrifugation of fraction 1 RNA resulted in the isolation of four fractions (Fig.3.2). Electrophoretic studies on follicle RNA and the fractions derived from it revealed that a limited fractionation had occurred in the sucrose gradients and that some RNA degradation had taken place. The presence of distinct bands rather than a smear suggested that the RNA breakdown was relatively limited (Figs.3.3-3.5).

Lock (1977) attempted to characterize hair follicle RNA by translating it in an S23 wheat germ cell-free system. He found that the level of radiolabelled amino acid incorporation was very low and varied markedly with different S23 preparations. Translation products were electrophoresed in SDS-polyacrylamide gels and analyzed by measuring the radioactivity of gel slices. Fraction 1 RNA gave translation products which were resolved into a number of discrete peaks (M_r 57,000; 20,000; 14-16,000; 10-14,000) and other material of molecular weight less than 10,000. Products from fractions 2, 3 and 4 were poorly defined. The low level of incorporation meant that individual gel slices often contained very few counts, making interpretation of the results difficult.

The use of the S30 wheat germ and reticulocyte lysate translation systems as described in this chapter has led to better incorporation of labelled amino acids into proteins, while fluorography of gels has provided an improved means of comparing translation products.

The results obtained by electrophoretic comparisons of translation products (Figs.3.8, 3.9, 3.11) led to several conclusions;

(i) the hair follicle RNA contained full-length components which coded for the major keratin proteins. This was most apparent for the high-sulphur proteins which lay along a characteristic diagonal upon two-dimensional electrophoresis. The concentration of high-glycine-tyrosine proteins was difficult to ascertain because of contaminating low molecular weight products. However, in single dimension

electrophoretograms there was evidence of translation products co-migrating with high-glycine-tyrosine proteins. The low-sulphur protein concentration was low and this is more likely a reflection of the RNA quality than poor translation efficiencies of these particular components.

(ii) the RNA fractionation achieved by sucrose density gradient ultracentrifugation was marginal. While there was a gradation of translation product size, it was slight. Refractionation of fraction 1 RNA did result in an enrichment of species coding for larger proteins while at the same time it removed low molecular weight components (Fig.3.11). It was observed that the follicle mRNA species were aggregated to some extent since components coding for proteins co-migrating with high-glycine-tyrosine and low molecular weight high-sulphur proteins were found throughout the sucrose gradient (Figs.3.8,3.11). While RNA disaggregation could be readily achieved, it resulted in non-reproducible sucrose density gradient profiles and made fraction collection less convenient.

(iii) a large proportion of the labelled amino acid was incorporated into low molecular weight products. Presumably this was due mainly to translation of truncated mRNA species although, as noted previously, there was a significant proportion of high-glycine-tyrosine proteins present which contributed to the level of low molecular weight material.

(iv) there were many translation products which were not present in the total hair protein extract. The two-dimensional electrophoretograms in particular showed the presence of components with mobilities differing from those of the keratin proteins. Steinert and Rogers (1973b) compared guinea pig hair and hair follicle proteins and concluded that apart from some differences in the relative proportions of components, the types and properties of the proteins from both tissues were identical. It would appear that the protein spots which did not co-

migrate with the keratins and which were not a component of the low molecular weight peptides discussed above in (iii) were most probably translation products derived from relatively large 5' fragments of mRNA, some of which were present in high concentrations. However it cannot be overlooked that some of these proteins may represent the (partial?) translation products of mRNAs coding for hair follicle enzymes and other non-keratin components which are necessarily present in the tissue. The relatively high concentration of such non-keratin components could be a consequence of a greater translation efficiency of these mRNAs in heterologous cell-free systems compared with keratin-coding mRNA.

Recombinant DNA experiments were undertaken on the basis that the hair follicle RNA contained a proportion of full-length keratin-coding mRNAs which theoretically could be cloned and characterized. Subfraction C was chosen for these investigations because it comprised mRNAs coding for the three keratin protein classes and because it had a lowered concentration of small, presumably degraded, species. In addition, the limited RNA fractionation in sucrose density gradients precluded any attempt to clone different classes of coding sequences separately.

A total of only 51 cDNA clones was obtained, the majority of which were in the size range 0.2-0.5 kb. In situ colony hybridization using cDNA and RNA probes resulted in only 16-40% of the clones giving positive hybridization (Figs.3.14, 3.15). The cDNA library was enriched therefore in sequences present in low abundance in the original RNA. This possibly reflects differences in the efficiency with which mRNA species were copied into dscDNA. It has been reported (Rougeon and Mach, 1975) that AMV reverse transcriptase synthesizes the second strand of different mRNA species with different efficiencies. Interestingly, only 35% of the clones of a comprehensive wool follicle cDNA library hybridized when probed with sequences specific for the low-sulphur and two high-sulphur protein families (E. Kuczek, personal communication).

The homology of sequences within a family should have ensured about 90% positive hybridization assuming a direct relationship between the concentrations of follicle proteins and mRNAs and assuming that the synthesis and cloning efficiencies of all the dscDNA species were the same. While there was an enrichment of minor sequences in the library, the reason for this is presently unknown. Transcription and translation studies using purified genes may provide an explanation of the phenomenon.

There are no sequence data available for guinea pig hair proteins. However the characteristic amino acid compositions of the major keratin classes (see Chapter 1) should allow the classification of a nucleic acid sequence provided that sufficient coding data were available. The complete DNA sequences of two small clones and the partial sequence of a large clone were determined. Only one recombinant, pGPHF2 (200 bp without the homopolymeric tails) gave an open reading frame of significant length (Fig.3.17). The amino acid composition of the deduced protein sequence did not show any significant similarity with the keratin protein compositions reported by Steinert and Rogers (1973b).

The lack of proline and the relatively high levels of methionine, histidine and lysine indicate that the protein is not from the high-sulphur group. The low level of tyrosine and the presence of methionine, histidine and lysine suggest that it is not a high-glycine-tyrosine component. A secondary structure prediction analysis of the amino acid sequence - kindly carried out by Mr. T. P. MacRae using the method of Garnier et al. (1978) - did not predict any α-helix. If the pGPHF2 sequence is part of a low-sulphur protein, then the prediction analysis indicates that it is not derived from one of the helical regions. Nor is the sequence likely to be part of a N- or C-terminal non-helical region, since the data available for wool low-sulphur proteins (L. M. Dowling and L. G. Sparrow, personal communication) shows these regions to be rich in glycine, serine, tyrosine, phenylalanine and

valine.

It is possible to argue that pGPHF2 (i) is only part of a keratin-coding sequence; (ii) is derived from genomic DNA which has contaminated the RNA preparation and has been cloned; (iii) is part of a sequence coding for a keratin protein of unusual amino acid composition; (iv) is derived from a non-keratin follicle mRNA; (v) or is non-coding sequence which fortuitously has an open reading frame in one orientation. The compositional data presented above would tend to rule out the first possibility, while the second is unlikely since the pGPHF2 sequence hybridized with hair follicle cDNA. The available data make it difficult to choose which of the other alternatives is likely to be correct.

While pGPHF40 was probably a portion of 3' non-coding region, it is extremely unlikely that the long non-coding sequence of pGPHF32 occurs in a single mRNA. Artefactual sequences produced during molecular cloning of cDNA have been reported (see for example Fields and Winter, 1981; Weaver et al., 1981). Inversions of sequence within the same cDNA molecule are the predominant artefact observed. In addition, Fields and Winter (1981) described a clone which contained sequences derived from two different mRNAs. These workers suggested that the clone resulted from an incomplete cDNA molecule hybridizing to another mRNA template at a region of complementarity. The nature of the artefact in pGPHF32 is unknown although from the sequence data (covering over 50% of the clone) there was no evidence for sequence inversions. Interestingly, an artefactual clone comprising two high-sulphur sequences has been isolated from a wool follicle cDNA library (T. C. Elleman, personal communication). Since the sequence data on pGPHF32 contained no coding region, further studies on this clone were abandoned. The difficulties encountered with determining the origin of the sequenced cDNA clones made further investigations of the library unwarranted.

The quality of a cDNA library is basically a reflection of the RNA used to construct it. While translation studies showed that full-

length mRNA was present in the follicle RNA preparation, the cloning results indicated that the proportion of degraded RNA was unacceptably high. Within the small number of recombinants isolated, an enrichment of minor mRNA species further increased the difficulties of obtaining a keratin-coding sequence from the library.

Lock (1977) had thoroughly investigated a variety of RNA isolation techniques and failed to prepare intact follicle RNA. The possibility existed that RNA degradation was occurring during the time taken to harvest the follicles from the dead animal's skin and place it into extraction buffer. An alternate harvesting technique was developed which enabled hair follicles to be prepared from live guinea pigs.

The use of Vertex SC as an adhesive to replace wax and Araldite for the harvesting of follicles from guinea pigs and sheep respectively has two major advantages. Firstly, in the case of the guinea pig, the use of Vertex SC enables hair follicles to be harvested from live animals, thus minimizing the time during which RNA can be degraded. The technique should be suitable for the preparation of hair follicles from any small mammal. Secondly, when used for wool follicle harvesting, the very rapid curing time of Vertex SC means that sheep have only to be handled once; the Araldite method requires application of the adhesive on the first day, overnight penning of the animal to minimize buckling of the resin on the glass-fibre backing, and follicle harvesting the next day.

Bertolino et al. (1982) have described the preparation of mouse hair follicles by scraping the underside of frozen skin. The scrapings will contain surrounding tissue in addition to the follicles and so the mRNA will be contaminated with non-follicle sequences. Electrophoresis of translation products obtained from 'scraped' follicle preparations resolved a large number of components (with molecular weights up to 60,000) only some of which co-migrated with standard hair proteins (Bertolino et al., 1982). The Vertex SC method minimizes the level of

non-follicle tissue harvested and would be presumably quite suitable for the preparation of mouse hair follicles.

The harvesting of hair follicles from live guinea pigs enabled intact RNA to be isolated and cDNA up to 2000 bases long to be synthesized. Resolution of the problems associated with the preparation of intact RNA from guinea pig hair follicles coincided with the development of facilities to use sheep instead of guinea pigs as the source of experimental material. The existence of wool protein sequence data, combined with the fact that the guinea pig had been used only as a convenient laboratory model for studying mammalian keratins, indicated that studies using the guinea pig should cease in favour of an investigation of wool follicle mRNA.

Poly(A)$^+$ RNA was isolated from Vertex SC-harvested wool follicles and was copied into cDNA long enough to contain full-length sequences for the major keratin protein classes. Kuczek and Rogers (1985) used Vertex SC-harvested wool follicle RNA to construct a cDNA library. The library was successfully screened for the presence of low-sulphur and high-sulphur protein-coding sequences (E. Kuczek, personal communication). In addition, clones coding for a high-glycine-tyrosine component which constitutes only about 0.3% of the wool proteins were isolated and characterized (Kuczek and Rogers. 1985).

Parallel studies on wool follicle mRNA prepared from Araldite-harvested tissue led to the isolation and characterization of cDNA clones coding for members of the low-sulphur and B2 and BIIIB high-sulphur protein families (K. A. Ward and M. J. Sleigh, personal communication). A clone coding for a BIIIB high-sulphur protein was provided as a probe for isolating genomic clones and this work is described in the following chapters.

SECTION II.

THE ISOLATION AND CHARACTERIZATION OF BIIIB HIGH-SULPHUR KERATIN GENES

CHAPTER 4.

MATERIALS AND METHODS

CHAPTER 4

MATERIALS AND METHODS

Many of the materials and a number of the methods used in the studies described in the following chapters have been included already in Chapter 2.

A. MATERIALS

1. Chemicals:

(a) Chemicals for DNA cloning and sequencing:

ddATP, ddCTP, ddGTP, ddTTP: P-L Biochemicals Inc.

IPTG (Isopropyl-β-D-thiogalactopyranoside): Sigma.

Sequencing primers, 27-mer: Bethesda Research Laboratories Inc.

17-mer: New England Biolabs Inc.

X-gal (5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside): Sigma.

(b) Other Chemicals:

Ampicillin: Sigma.

Chloramphenicol: Sigma.

L-arginine: B.D.H. Chemicals Ltd.

Tetracycline: Sigma.

2. Vectors for DNA subcloning and sequencing:

Plasmids pBR322 and pBR325, and phages M13mp8 and M13mp9 were obtained from stocks held in the laboratory.

3. Bacterial Strains:

The strains of E.coli used in the following chapters were obtained from Dr. B. Egan, Biochemistry Department, University of Adelaide; Dr. R. B. Saint, Walter and Eliza Hall Institute for Medical Research, Melbourne; and Dr. U. Novak, Veterinary School, University of Melbourne.

(a) For growth of plasmids and cosmids:

(i) ED8654: r_k^-,m_k^+,supE,supF,trpR.

(ii) MC1061: ara,D139,Δ(ara,leu),7697,Δ(lacx74),galU$^-$,galK$^-$, hsr$^-$,hsm$^+$,strA.

(iii) HB101: F$^-$,hsdS20(r_B^-,m_B^-),recA13,ara-14,proA2,lacY1,galK2, rpsL20(Sm^r),xyl-5,mtl-1,supE44,λ^-.

(b) For growth of bacteriophage M13:

JM101: Δ(lacpro),supE,thi,F'traD36,proAB,lacI^q,zΔM15.

(c) For growth of bacteriophage λ vectors:

(i) LE392: F$^-$,hsdR514(r_k^-,m_k^+),supE44,supF58,lacY1 or Δ(lacIZY)6,galK2,galT22,metB1,trpR55,λ^-

(ii) K802: hsdR$^-$,hsdM$^+$,gal$^-$,met$^-$,supE.

(iii) Q359: hsdR_k^-,hsdM_k^+,supE,ϕ,P2.

(iv) RY1073: BNN93,hflA150[chr::TN10].

4. Sheep genomic DNA libraries:

The Charon 4A, Charon 28 and λ1059 sheep genomic DNA libraries were prepared essentially as described by Maniatis et al. (1978).

(a) Charon 4A:

This genomic library was the generous gift of Dr. P. J. Kretschmer, Bethesda Research Laboratories Inc. The library was constructed using foetal Dorset Horn liver DNA which had been partially digested with EcoRI and fractionated to yield DNA fragments 15-20 kb in size (Kretschmer et al., 1980). Over 10^6 unique recombinants were isolated and maintained in 14 separate, amplified fractions.

(b) Charon 28:

The Charon 28 (stock 2221) genomic DNA library was kindly supplied by Drs. P. Roche, P. Aldred and R. Crawford, Howard Florey Institute of Experimental Physiology and Medicine, Melbourne. The library was constructed using Merino liver DNA which had been partially digested with Sau3A and fractionated to yield DNA fragments 12-20 kb in size. About 0.75×10^6 unique recombinants were isolated; the library was maintained as one amplified stock.

(c) λ1059:

The λ1059 genomic library was kindly supplied by Dr. U. Novak, Veterinary School, University of Melbourne and was constructed using Merino lymphocyte DNA partially digested with BamHI and fractionated to yield fragments 15-24 kb in size. Over 0.6×10^6 unique recombinants were isolated and maintained in 4 separate, amplified fractions.

(d) λgt10:

The λgt10 library was constructed by Dr. R. B. Saint, Walter and Eliza Hall Institute of Medical Research, Melbourne, essentially as described by Kemp *et al.* (1983), using Dorset Horn DNA digested to completion with EcoRI. About 1.5×10^6 unique recombinants were isolated; the library was not amplified.

(e) Cosmid pHC79:

The pHC79 cosmid library was constructed by Dr. G. Cam, Biochemistry Department, University of Adelaide using Dorset x Border Leicester liver DNA partially digested with Sau3A and fractionated to yield fragments 33-49 kb in size. These fragments were inserted into the BamHI site of pHC79 (Hohn and Collins, 1980) and were packaged *in vitro* using Amersham packaging mix. Approximately 10^5 colonies were maintained on ten nitrocellulose filters as described by Hanahan and Meselson (1983).

5. Enzymes:

Bacterial alkaline phosphatase: Worthington Diagnostic Systems Inc.

Deoxyribonuclease I (bovine pancreas): Boehringer Mannheim.

DNA Polymerase I (*E.coli*): Bethesda Research Laboratories Inc.

DNA Polymerase I (*E.coli*; Klenow fragment): Bethesda Research Laboratories Inc.

Lysozyme (hen egg white): Sigma.

Restriction endonucleases: Bethesda Research Laboratories Inc.; Boehringer Mannheim; New England Biolabs Inc.

Ribonuclease A (bovine pancreas): Boehringer Mannheim.

T4 DNA ligase: Bethesda Research Laboratories Inc.

T4 DNA polymerase: Bethesda Research Laboratories Inc.

6. Other materials:

DEAE-membrane(NA45): Schleicher and Schuell.

Disposable pipettes and petri dishes: Sterilin Products.

Low-melting-temperature agarose: Bio-Rad Laboratories.

B. METHODS.

The preparation of buffers and reagents, and measurement of radioactivity have been described in Chapter 2.

1. Media for growth of plasmid, cosmid and bacteriophage vectors:

(a) Plasmids and cosmids:

E.coli transformed with plasmid vectors or transduced by cosmids were propagated in L broth containing 10g tryptone, 10g NaCl and 5g yeast extract per litre adjusted to pH7.5 with NaOH. Tetracycline (15µg/ml), ampicillin (50µg/ml) and chloramphenicol (150µg/ml) were added if required.

(b) Bacteriophage M13:

E.coli harbouring the bacteriophage M13 vectors were propagated in YT broth containing 8g tryptone, 5g NaCl and 5g yeast extract per litre. Individual M13 plaques were grown in 2YT broth containing 16g tryptone 5g NaCl and 10g yeast extract per litre.

(c) Bacteriophage λ:

E.coli required for bacteriophage λ infection were propagated in a number of broths.

(i) Charon 4A, Charon 28 and λgt10:

L broth containing 0.2% glucose (0.2% maltose was included in the liquid medium to prepare plating bacteria - see methods later this chapter).

(ii) λ1059:

CY broth - 10g Casamino acids, 3g NaCl, 5g yeast extract, 2g KCl per litre, adjusted to pH7.0, and supplemented with 25mM

Tris-HCl pH7.4 and 10mM $MgCl_2$

(iii) Large-scale liquid lysate medium:

For the large-scale preparation of purified phage, a modified NZCYM broth was used containing 10g tryptone, 5g NaCl, 5g yeast extract, 1g Casamino acids and 10mM $MgSO_4$ per litre pH7.0.

(d) Media containing agar or agarose.

E.coli were also propagated on solid media containing the broth and 15g agar per litre. Top agarose was prepared by the addition of 7g agarose per litre of broth. For most purposes the solid media contained the same nutrients as the liquid. Exceptions were for the growth of Charon 4A and λ1059.

(i) Charon 4A plates:

These contained 10g tryptone, 5g NaCl, 5g yeast extract, 2g glucose, 10mM Tris-HCl(pH7.5), and 10mM $MgCl_2$ per litre.

(ii) λ1059 plates:

These were prepared by the addition of 10mM Tris-HCl(pH7.5) and 10mM $MgCl_2$ to L broth.

2. Preparation of plating bacteria:

A single bacterial colony was used to inoculate 20ml of L broth containing 0.2% glucose and 0.2% maltose. After overnight incubation at 37°C, the bacteria were pelleted by centrifugation at 4000g(av) for 10min at room temperature. The pellet was resuspended in 10ml of 10mM $MgSO_4$ and was stored at 4°C for up to 2 weeks.

3. Long-term preservation of bacteria and bacteriophage:

(a) Bacteria:

Bacterial stocks and strains harbouring plasmids or cosmids were stored at -20°C or -80°C following the addition of glycerol to 40%.

(b) Bacteriophage M13:

Single-strand M13 stocks and clones were kept at 4°C. Bacteria containing the bacteriophage were stored as described above.

(c) Bacteriophage λ:

The genomic DNA libraries and bacteriophage at different stages of purification were stored at 4°C in phage storage buffer (PSB; 10mM Tris-HCl pH7.5, 100mM NaCl, 10mM $MgSO_4$, 0.05% gelatin), over 0.3% chloroform.

4. Screening of sheep genomic DNA libraries:

(a) Preparation of pSWK18 plasmid DNA probe:

(i) Plasmid pSWK18 DNA:

The plasmid pSWK18 was kindly provided by Dr. K. A. Ward, C.S.I.R.O. Division of Animal Production, Sydney. This plasmid was isolated from a Dorset Horn sheep wool follicle cDNA library constructed by Drs. K. A. Ward and M. J. Sleigh using the "loop-back" and "tailing" methods to clone dscDNA into the PstI site of pBR322. DNA sequence analysis of this clone indicated that the cDNA insert coded for the C-terminal 72 amino acids of a protein closely related to the high-sulphur keratin protein BIIIB3 (K. A. Ward and M. J. Sleigh, personal communication). Plasmid pSWK18 also contained an entire 3' non-coding sequence (360 bp), including at least some of the poly(A) tail. This plasmid, or the insert isolated from it, was used to screen all of the genomic DNA libraries described in the following chapters.

(ii) Preparation of pSWK18 DNA:

Plasmid DNA was prepared from overnight cultures (amplified in the presence of chloramphenicol) using a modification of the method of Clewell and Helinski (1969) where Triton X-100 replaced Brij 58 and sodium deoxycholate in the lysis buffer. Closed circular plasmid DNA was purified by CsCl buoyant density gradient centrifugation (Radloff et al., 1967).

(iii) Isolation of pSWK18 insert DNA:

Plasmid pSWK18 DNA (70μg) was digested to completion with PstI (120 units) by incubation at 37°C for 4h. The enzyme was inactivated by heating the sample at 67°C for 10min.

Following the addition of bromophenol blue and glycerol, the DNA solution was subjected to electrophoresis in TAE buffer in a horizontal 1% low-melting-temperature agarose gel (11x14x0.3 cm). DNA was stained with ethidium bromide and the insert band (about 660 bp) was cut out of the gel (0.5x6x0.3 cm; 900µl volume of gel). The gel slice was heated at 70°C for 5min, 4ml of TE buffer added, and the incubation was repeated to ensure complete melting of the agarose. The diluted agarose solution was extracted twice with buffer-saturated phenol and the aqueous phase was concentrated to 400µl with butan-1-ol. The DNA was precipitated in the presence of sodium acetate and ethanol. The pelleted DNA was resuspended in TE buffer to give a concentration of approximately 250ng/µl.

(iv) Nick translation of DNA:

DNA was labelled to high specific activity by nick translation using a modification of the method described by Rigby et al. 1977 (Dr. O. Bernard, personal communication). Radiolabelled dATP (30-50µCi, [α-^{32}P]) was dried in vacuo and was resuspended in 25µl of buffer containing 50mM Tris-acetate pH7.5, 5mM $MgCl_2$, 50mM KCl, 10mM 2-mercaptoethanol and 0.1mM each of dCTP, dGTP and dTTP. DNA (200-400ng in 1µl) was added, followed by 50pg DNase I (diluted in 20mM Tris-acetate pH7.9, 100mM NaCl, 5mM magnesium acetate, 50% glycerol) and 10 units of DNA polymerase I. The mixture was incubated at 15°C for 3h and the reaction was stopped by the addition of EDTA to 20mM. Phenol extraction and removal of unincorporated nucleotides were carried out as described in Chapter 2. The specific activity of the DNA was in the range 1-2x10^{8}cpm/µg.

(b) Synthesis of a mixed 20-mer oligonucleotide:

A mixed 20-mer oligonucleotide, based on a portion of the coding sequence of three cDNA clones, was synthesized by Dr. D. Skingle, Biochemistry Dept., University of Adelaide. The oligonucleotide

5'-ACACTCAGTCTAGCTGTGAA
T T C CC G

was synthesized by the solid-phase phosphite method (Beaucage and Caruthers, 1981) and was purified by reversed phase, high performance liquid chromatography. The sequence was confirmed by Dr. S. D. Wilton.

(c) Benton and Davis screening of the libraries:

(i) Plating of bacteriophage λ libraries:

The titre of bacteriophage in each library was determined using the required plating bacteria (prepared as described earlier in this chapter) essentially as reported by Maniatis et al. (1982). For each 15cm plate used in the first screen of a library, approximately $2-3x10^4$ bacteriophage were adsorbed to 0.3ml of plating bacteria for 20 min at 37°C, mixed with 9ml of top agarose (at 46°C) and quickly poured onto bottom agar medium. For subsequent screens, 9cm petri dishes were used with 0.1ml of plating bacteria and 3ml of top agarose. Plates were incubated overnight at 37°C.

(ii) Plaque hybridization:

The methods of Benton and Davis (1977) and Woo (1979) were carried out as described in Maniatis et al. (1982). After the bacteriophage DNA had been baked onto nitrocellulose circles, the filters were probed with pSWK18 total plasmid or insert DNA using the method of Maniatis et al. (1978) or Maniatis et al. (1982). Exposure of the filters to x-ray film for 16-30h was usually sufficient to observe positive hybridization.

Plaque hybridization using the synthetic oligonucleotide was carried out using a combination of the methods described by Hanahan and Meselson (1983) and Wallace et al. (1980). Briefly, the mixed oligonucleotide was labelled with adenosine 5'-[γ-^{32}P] triphosphate in a 10µl reaction mix containing 50mM Tris-HCl pH9.0, 10mM $MgCl_2$, 10mM DTT, 500ng oligonucleotide and 4 units of polynucleotide kinase. Following incubation at 37°C for 60min, the reaction was terminated by the addition of EDTA to 50mM. The labelled oligonucleotides were isolated by passage through a G-25 Sephadex column.

Nitrocellulose filters were prehybridized at 67°C for 4h in the presence of 6xNET (1xNET is 15mM Tris-HCl pH7.5, 150mM NaCl, 1mM EDTA), 5xDenhardt's solution, 0.5% SDS and 100µg/ml denatured E.coli DNA. Hybridization was performed for 16h at 42°C in 6xNET, 5xDenhardt's solution, 0.5% SDS, 250µg/ml E.coli tRNA and labelled oligonucleotide. Filters were washed at room temperature for 30min with three changes of 6xNET, 0.5% SDS and then for 1-2min at 54-56°C in the same buffer. Exposure of the filters to x-ray film was at -80°C for up to 3 days, in the presence of an intensification screen.

(iii) Plaque purification:

The area surrounding a putative positive plaque was picked into 1ml of PSB containing 0.3% chloroform and the bacteriophage were allowed to elute overnight at 4°C. The titre of bacteriophage was determined by serial dilution and a volume sufficient to yield 100-300 plaques was plated onto 9cm petri dishes. Plaque hybridization and replating at lower density were continued until all the plaques on a plate were positive when probed with pSWK18 DNA.

(d) Preparation of bacteriophage DNA:

Bacteriophage DNA was prepared using a method based on that of Kao et al. (1982). The method was found to give good yields of high-quality DNA without requiring centrifugation in CsCl. Bacteriophage at a multiplicity of infection (m.o.i) of approximately 0.1 were incubated for 20 min at 37°C in the presence of PSB containing 2mM $CaCl_2$ with about 10^9 freshly grown bacterial cells. The mixture was then added to 50ml of modified NZCYM broth and was incubated at 37°C with vigorous shaking until lysis occurred - generally about 6h. Chloroform was added to 0.5-1%, incubation was continued for 15min, and the lysate stored overnight at 4°C. Following incubation for 1h on ice in the presence of DNase I and RNase A (both at 1µg/ml), cellular debris was removed by centrifugation at 6000g(av) for 15 min. The bacteriophage were pelleted by centrifugation at 30,000g(av) for 3h and gently resuspended overnight

at 4°C in 0.4ml of buffer (100mM Tris-HCl pH8.0, 300mM NaCl). The bacteriophage were then gently extracted three times each with buffer (20mM Tris-HCl pH8.0, 1M NaCl, 1mM EDTA)-saturated phenol, phenol:chloroform (1:1) and ether, and the DNA precipited by the addition of 0.8ml ethanol. Sedimented DNA was gently washed with 70% ethanol, dried *in vacuo*, and resuspended in 1mM EDTA pH8.

(e) Screening of the cosmid library:

The pHC79 cosmid library was screened with the pSWK18 insert probe using the hybridization conditions for bacteriophage plaques described by Maniatis *et al*. (1982). The filters were washed in 0.5xSSC/0.1% SDS at 65°C and were exposed to x-ray film in the presence of an intensification screen. Colonies in the vicinity of the putative positive clone were picked from the master filter, diluted in L broth containing ampicillin and were plated onto nitrocellulose filters at a density of about 200 colonies per plate. Replicate filters were made and rescreened. (I am indebted to Dr. G. Cam for preparing the replicate filters for the initial screenings). Cosmid DNA was prepared using the technique described for plasmid DNA.

5. Restriction enzyme mapping of DNA:

Restriction enzymes were used in accordance with the manufacturer's recommendations. The orientation of restriction enzyme sites was found by electrophoresis (in acrylamide or agarose gels) of the products of single and double digests while the location of fragments containing genes was determined by Southern transfer of the digested DNA and probing with pSWK18 (see later this chapter).

6. Electrophoresis of DNA:

(a) Agarose:

DNA fragments larger than about 1 kb were examined by electrophoresis in 1% agarose gels in TAE buffer. Conventional vertical gels (16x20cm) or submersed horizontal gels (7.5x5cm mini gel, or 14x11cm) of 0.3cm thickness were used.

(b) Polyacrylamide:

DNA fragments smaller than 1 kb were examined by electrophoresis in 6-8% polyacrylamide gels in TBE buffer. Gels were prepared from a polyacrylamide stock solution containing 39g acrylamide and 1g N,N'-methylenebisacrylamide per 100ml which had been de-ionized by stirring for 1h with 5g of Bio-Rad AG501-X8 mixed-bed resin. Gels (16x20cm) were 0.15cm thick for analytical purposes and 0.04cm thick when isolating DNA for subcloning experiments.

(c) Staining and photography:

Following electrophoresis in agarose or acrylamide, DNA fragments were stained with ethidium bromide (1µg/ml) for 10-15min. After a brief destaining in water, the gels were photographed under UV light using a trans-illuminator and Polaroid Type 665 positive/negative film.

7. Subcloning of DNA restriction fragments into plasmid and bacteriophage vectors:

(a) Vectors:

Plasmids pBR322 and pBR325 and bacteriophage M13mp8 and M13mp9 were used as vectors for subcloning DNA fragments produced by restriction enzyme digestion of cloned genomic DNA. Vector DNA (5µg) was digested to completion with the appropriate restriction enzyme(s) in a volume of 20µl. Removal of 5'-phosphate groups was carried out by adding 10µl of 0.1M Tris-HCl pH8.2 to the DNA solution followed by 15µl of 10mM Tris-HCl pH8.5, 10mM $MgCl_2$ containing 1µl of bacterial alkaline phosphatase, and incubating for 15min at 45°C and 45min at 65°C. (The enzyme solution had been pre-treated at 80°C for 10min to inactivate contaminating nucleases). Following dephosphorylation, the reaction-mix was extracted with phenol (2x) and ether (3x) and then precipitated with sodium acetate and ethanol. Sedimented DNA was resuspended in H_2O at 40ng/µl.

(b) Isolation of genomic DNA restriction fragments:

(i) Agarose gels:

Restriction fragments were isolated from agarose using DEAE-membranes. Ethidium bromide-stained gels were examined under UV light and a slot was cut to the anodic side of the DNA band required for subcloning. A piece of DEAE-membrane was inserted into this slot and electrophoresis was continued until the DNA had migrated onto the membrane. After a brief wash in TE buffer, the DNA was eluted from the membrane by incubation at 70°C for 1h in a buffer containing 1M NaCl and 10mM arginine (R. B. Saint, personal communication) and was precipitated by the addition of ethanol.

(ii) Low-melting-temperature agarose gels:

The method used to isolate DNA from low-melting-temperature agarose is described earlier in this chapter. Occasionally isolated DNA was further purified by passage through a 200µl column of DEAE-cellulose. The DNA was bound to the resin in TE and eluted in the same buffer containing 1.5M NaCl. The eluted DNA was diluted to give 0.75M NaCl and precipitated by the addition of ethanol.

(iii) Polyacrylamide gels:

Polyacrylamide gels (6-8%, 0.04cm thick) were prepared as described above and used for the isolation of DNA fragments smaller than about 1 kb. After ethidium bromide staining, the required DNA band was cut out and incubated overnight at 42°C in the ammonium acetate buffer described by Maxam and Gilbert (1980). The DNA was precipitated in 70% ethanol at -20°C for 16h. Following resuspension in TE buffer, the DNA was phenol extracted, precipitated by the addition of sodium acetate, magnesium acetate (10mM) and ethanol and finally resuspended in H_2O.

(c) End-filling reactions:

Restriction fragments containing terminal sequences unsuitable for direct ligation to vector DNA were made blunt-ended by reaction with either the Klenow fraction of DNA polymerase I or T4 DNA polymerase.

For DNA having 5'-extensions, the restriction enzyme mix was

made 50mM in NaCl, dNTPs were added to 1mM each and incubation at 37°C for 15min was carried out in the presence of 1 unit of Klenow enzyme. For 3'-extensions, T4 DNA polymerase was used as described by O'Farrell et al. (1980). Reactions were terminated by incubation at 65°C for 10min, the DNA was extracted with phenol, and the unincorporated nucleotides were removed by Sephadex G-50 chromatography.

Radiolabelling was achieved using the same reaction conditions except that dATP was replaced by [^{32}P]-dATP.

(d) Ligation of restriction fragments to vector DNA:

Vector DNA (40ng) was ligated to an equimolar amount of restriction fragment DNA by incubation for 16h at 14°C in the presence of 50mM Tris-HCl pH7.4, 10mM $MgCl_2$, 10mM DTT and 1mM rATP (neutralized) in a 10-20µl volume. For ligation of DNA with 3'- or 5'-extensions 0.1 units of T4 DNA ligase were used; for blunt-end ligation 1 unit of enzyme was used.

(e) Transformation of bacterial strains:

(i) Recombinant plasmid DNA:

The method described in Chapter 2 was used to transform E.coli with recombinant plasmid DNA.

(ii) Recombinant bacteriophage M13 DNA:

E.coli strain JM101 was grown overnight and was used to inoculate YT broth at 1:100 dilution. When the fresh bacterial inoculum had reached an A_{600} of 0.6 the broth was cooled and the cells were sedimented. The bacterial pellet was resuspended in 0.5 vol 50mM $CaCl_2$, incubated on ice for 40 min, sedimented, and finally suspended in 0.1 vol 50mM $CaCl_2$. Ligated DNA (1-5µl) was added to 0.3ml of competent cells and was incubated on ice for 40 min. The cells were heat-shocked at 45°C for 2min, mixed with 3ml YT-top agarose containing 50µl X-Gal (20mg/ml in dimethylformamide), 10µl IPTG (24mg/ml in H_2O) and 0.2ml JM101 overnight culture, and were poured onto a YT agar plate. Although plaques could be seen after 6h, incubation was continued at 37°C for 16h.

(f) Detection of recombinants:

(i) Plasmids:

Recombinant plasmids were screened using the colony hybridization technique of Grunstein and Hogness (1975) or the mini-screen method described in Chapter 2.

(ii) Bacteriophage M13:

When E.coli JM101 is transformed by M13 using the conditions described above, parental phage plaques are blue whilst recombinant plaques are clear. Three techniques were used to screen for particular sequences:

(1) the plaque hybridization method of Benton and Davis (1977) as described earlier in this chapter.

(2) a mini-screen method was used to assess the DNA insert size. Clear plaques were picked into 2ml of 2YT broth and were incubated for 6h at 37°C. The inoculated broths were centrifuged in Eppendorf tubes for 8min at 4°C and the supernatants were used for mini-screens and DNA sequencing (described later in this chapter). For DNA insert size estimation, 8µl of supernatant were incubated at 68°C for 10min in the presence of 1µl 5% SDS and 2µl bromophenol blue-glycerol, and were subjected to electrophoresis in 1% agarose mini-gels. M13 clones of known size were used as markers for comparison.

(3) an annealing technique was used to locate clones containing sequences in the opposite orientation. Clear plaques were picked as described above and 4µl each of the supernatants from two clones were incubated with 1µl 5% SDS for 1h at 68°C. Dye-glycerol was added and the DNA was examined on 1% agarose mini-gels. Annealed DNA has slower electrophoretic mobility than non-annealed, single-stranded bacteriophage DNA.

(g) Preparation of subcloned DNA:

(i) Plasmids:

Recombinant plasmid DNA was prepared as described

earlier in this chapter.

(ii) Bacteriophage M13:

Replicative form M13 DNA was prepared essentially as described for plasmid DNA. E.coli JM101 overnight culture was used to inoculate YT broth at 1:100 dilution. The inoculum was then infected with M13 bacteriophage at a m.o.i. of 0.1 and was incubated for 16h at 37°C. The preparation of single-stranded DNA suitable for sequencing purposes will be described in the next section of this chapter.

8. DNA sequencing:

The 'chain-termination' sequencing technique described by Sanger et al. (1980) was used.

(a) Preparation of single-stranded DNA:

Cloned DNA was prepared from 2ml cultures as described by Sanger et al. (1980) except that phenol extraction (15min at room temperature) was followed by two ether extractions, and the DNA was resuspended in only 20µl of 10mM Tris-HCl pH7.5, 0.1mM EDTA.

(b) Primer annealing, sequencing reactions and electrophoresis:

Primer DNA (17-mer or 27-mer, 2.5ng) was annealed to 5µl of cloned DNA in the presence of 1µl 100mM Tris-HCl pH7.4, 100mM $MgCl_2$, 500mM NaCl by incubating the mixture at 100°C for 3min, 65°C for 20min and cooling slowly to room temperature.

The general procedure outlined in the M13 sequencing handbook supplied by Amersham was used for the sequencing reactions. The $X^{\bullet}$ and ddNTP mixes used are shown in Table 4.1. The reactions were terminated by the addition of 6µl of formamide-dye solution and the DNA was denatured by incubation at 100°C for 5min. Only 1-2µl of each reaction mix was used for electrophoresis in 6% or 8% polyacrylamide (7M urea-TBE buffer) gels as described in the Amersham M13 handbook. Gels (38x19x0.02cm) were electrophoresed at about 1200v/25mA, fixed in 10% acetic acid and were dried onto 3MM paper under vacuum. Autoradiography

TABLE 4.1. dNTP(X^0) MIXES (µl) AND ddNTP CONCENTRATIONS FOR DNA SEQUENCING REACTIONS.

	G^0	A^0	T^0	C^0
0.5mM dGTP	1	15	15	15
0.5mM dTTP	15	15	1	15
0.5mM dCTP	15	15	15	1
H_2O	10	10	10	10
50mM Tris-HCl pH8.0, 1mM EDTA	5	5	5	5

ddNTP	STOCK (mM)	WORKING SOLUTION (Stock + H_2O µl)
ddGTP	4	20 + 100
ddATP	4	5 + 100
ddTTP	10	20 + 80
ddCTP	4	5 + 100

was at room temperature for 16-40h.

9. Southern transfer:

Transfer of restriction fragments from agarose gel to nitrocellulose membranes was carried out as described by Southern (1975) using the modification reported by Wahl et al. (1979) or the bi-directional method of Smith and Summers (1980).

Nitrocellulose filters were probed as described in Chapter 2. Following hybridization, the filters were washed to varying degrees of stringency, blotted dry, and exposed to Fuji Rx medical x-ray film at $-80^{0}C$ using Ilford fast-tungstate intensification screens.

10. Wool follicle cDNA probe:

Wool follicle poly(A)$^{+}$ mRNA was prepared using the GuHCl method and cDNA prepared essentially as described in Chapter 2. Specific activities of $1\text{-}5x10^{7}$ cpm/µg were obtained.

11. Screening for repeated DNA sequences:

The method of Shen and Maniatis (1980) was used to detect sequences present in greater than 50 copies per genome. Briefly, high-molecular-weight genomic DNA was radiolabelled to high specific activity by nick translation and hybridized to a Southern transfer as described in this chapter. Filters were washed at $68^{0}C$ with a series of solutions, finally washed with 50mM Tris-HCl pH7.5, 1mM EDTA, dried and autoradiographed.

CHAPTER 5.

THE CHARACTERIZATION OF A BIIIB HIGH-SULPHUR PSEUDOGENE ISOLATED FROM A CHARON 4A SHEEP GENOMIC DNA LIBRARY

CHAPTER 5

THE CHARACTERIZATION OF AN SCMKBIIIB PSEUDOGENE ISOLATED FROM A CHARON 4A SHEEP GENOMIC DNA LIBRARY

A. INTRODUCTION

The study of mRNAs using recombinant DNA techniques has resulted in a wealth of information about the structure and composition of these molecules as well as the protein products for which they code. Cloning of cDNA sequences takes advantage of the fact that relatively pure mRNAs for specific genes can be isolated from cells specialized to produce large amounts of specific proteins. However, this approach is limited to genes that produce an RNA. While the study of cDNA is useful, these molecules do not contain sequences which are important for the regulation of gene expression, nor can they be used to provide any more than a small insight into gene organization. The collection of such information requires the isolation of genomic DNA.

Until recently, it had not been possible to readily purify genes from complex eukaryotic genomes. Genomic DNA fragments were concentrated using column chromatography or preparative gel electrophoresis and cloned to yield a limited pool of recombinants from which a specific gene could be isolated. However, these methods were unsuitable for studying multigene families. The approaches of Blattner *et al.* (1978) and Maniatis *et al.* (1978) enabled genomic libraries to be constructed which could be screened for any gene or family of genes for which a probe was available. The genomic library approach was made possible by the utilization of three technical advances: (i) the rapid *in situ* plaque hybridization technique of Benton and Davis (1977); (ii) the construction of suitable bacteriophage λ cloning vectors (Blattner *et al.*, 1977); and (iii), the development of *in vitro* packaging systems which greatly increased the efficiency of introducing bacteriophage DNA into bacteria (Hohn and Murray, 1977; Sternberg *et al.*, 1977).

The availability of genomic DNA libraries has made it possible to study in detail the organization of gene families - their location relative to each other and to other unrelated genes; the structure of the genes and their flanking sequences; the presence or absence of intervening sequences; the role of possible regulatory sequences during gene expression; and the evolution of genes and gene families.

Such data can also be collected on a larger scale using genomic DNA cloned into cosmid vectors (Collins and Hohn, 1978). Cosmids - plasmids carrying the cohesive end sites of bacteriophage λ - have the advantages of both plasmid and phage vectors. They carry antibiotic resistance genes, can be amplified, and are able to be introduced into bacterial cells by _in vitro_ packaging. Most importantly, they are capable of accommodating about 45 kb of foreign DNA compared with less than 23 kb for bacteriophage λ vectors.

A genomic clone containing feather keratin genes has been isolated and characterized by Molloy _et al_. (1982). The clone, purified from a chicken genomic DNA library using feather keratin cDNA as a probe, contains five keratin genes tandemly spaced with a regular centre-to-centre separation of 3.3 kb. The genes are transcribed from the same DNA strand and the arrangement of the cluster suggests that they evolved through a number of tandem duplications. This suggestion is supported by a comparison of the sequences of one of these genes with a feather keratin cDNA clone. There are only seven amino acid substitutions in 77 residues, six of which are the result of single nucleotide changes. In addition, the 3' non-coding regions which are of similar length, show 60% homology including two blocks of about 70 nucleotides which are highly conserved. The 5' non-coding region of the sequenced gene is characterized by the presence of an intron of 324 bp, a feature which is shared by the other genes of the cluster (R. Presland, personal communication). The study of the feather keratin gene family has been extended by the recent isolation of a cosmid clone (Presland and Rogers,

1984).

The DNA sequences of several chicken scale keratin genes have been determined (S. D. Wilton, personal communication). A comparison of the feather and scale coding sequences indicates that there is about 75% homology (Gregg et al., 1984). A number of insertions and deletions must be introduced to maximize the homology but the most striking feature is the presence of four 39 bp repeats in the scale sequence not found in the feather gene. The repeat codes for a glycine-rich peptide located between two feather-like domains. The scale and feather sequence homologies suggest that the proteins share a common ancestor. Since the evolution of scales in animals preceded the appearance of feathers, it can be reasonably assumed that feather keratins evolved from scale keratins. At some time after the deletion of the repeat unit from the scale gene, the resulting feather sequence appears to have undergone gene duplication to produce the present family of closely related structures (Gregg et al., 1984; Molloy et al., 1982). Further analysis of the scale-to-feather evolution will be possible when the non-coding and flanking sequences of the scale genes are available.

Marchuk et al. (1984) have recently reported the characterization of an epidermal keratin gene. As noted in Chapter 1, the epidermal keratins, together with the low-sulphur keratin components are members of the intermediate filament group of proteins. Many of the features of DNA organization found in the epidermal gene, and shared with another intermediate filament gene, vimentin (Marchuk et al., 1984; Quax et al., 1983), might be expected to occur in low-sulphur genomic sequences. Both epidermal keratin and vimentin genes have introns and the positions of a number of these are highly conserved in the two sequences. However, the size and sequence of the introns is not conserved. A striking feature of the gene organization is that the intron positions do not correspond to any known structural boundaries of the proteins and this has led Marchuk et al. (1984) to suggest that the

insertion sequences may delineate functional rather than structural domains.

The general introduction to this thesis described in detail the heterogeneity of mammalian keratin proteins. In spite of this complexity, the proteins can be considered as members of a limited number of families or subfamilies within which constituent proteins have homologous sequences. Presumably, if a specific probe is available for a particular gene, then it should be possible to probe for most, if not all, of the genes of that protein group. Powell *et al.* (1983) used the cDNA clone pSWK20, which codes for the high-sulphur protein B2C, to isolate the sheep SCMKB2C (B2C) gene and two other closely related genes.

The three genes, located on two genomic fragments, show an extremely high degree of nucleotide sequence homology in their coding regions. In addition, there is a highly conserved 18 bp sequence immediately 5' to the initiation codon. In contrast, the 3' non-coding region of the three genes diverge both in length and in sequence although there are four sequences of 10-17 bp which are conserved in two closely linked genes (Powell *et al.*, 1983).

While the sequence homologies between the genes are highly suggestive of gene duplication, the disparate 3' non-coding regions indicate that divergence was rapid after the duplication event. Two of the genes studied are closely linked, being only 1.9 kb apart, but there is no regular tandem repeat arrangement of the members of this family, unlike that found with the feather keratin system. There is also no evidence that the B2 genes contain introns (Powell *et al.*, 1983).

The genomic fragment which contains the B2C gene (Powell *et al.*, 1983) has been found also to contain a gene coding for a BIIIA high-sulphur protein (B. C. Powell, personal communication). Furthermore, a series of sheep genomic clones has been isolated which contain genes coding for components from the two low-sulphur families and from a high-glycine-tyrosine protein subfamily (M. J. Sleigh, K. A. Ward, E. Kuczek,

B. C. Powell and G. E. Rogers, personal communications). These genes are now under intensive study.

Data concerning the proteins of the BIIIB high-sulphur group have been reviewed by Swart et al. (1976). The proteins (M_r 11,000) are characterized by an acetylated amino terminal alanine residue and contain 97 or 98 residues per mole. A comparison of the amino acid sequences of three BIIIB components reveals that they are homologous members of a protein family. Furthermore, Swart et al. (1976) suggested that the BIIIB proteins may have evolved from the same ancestral gene as the B2 and BIIIA high-sulphur families. The isolation and characterization of genes coding for the SCMKBIIIB (BIIIB) high-sulphur components would provide valuable data required for an overall understanding of keratin gene organization, regulation and evolution. A cDNA clone, pSWK18, which codes for a BIIIB3-like protein (K. A. Ward and M. J. Sleigh, personal communication; Swart et al., 1976), was made available for screening sheep genomic DNA libraries for BIIIB genes.

B. RESULTS

1. Isolation of genomic clone λSWK50:

The Charon 4A sheep genomic DNA library prepared by Kretschmer et al. (1980) was screened using as the probe pSWK18 - a cDNA clone coding for a member of the BIIIB high-sulphur protein family (see Chapter 4 for screening methods). Over 7×10^5 plaques from 9 library fractions were screened and using moderately stringent washing conditions, 7 positives (all from one fraction) were obtained. The pSWK18-positive plaques were purified through several screenings until more than 95% of the plaques hybridized with the probe (Fig.5.1a-c).

Phage DNA from the 7 clones was prepared as described in Chapter 4 and compared by restriction enzyme digestion using EcoRI, PstI, BglII and BamHI. The restriction patterns for each clone were identical (data not shown), and it was assumed that the recombinants contained the same genomic fragment. All of the following studies were carried out

FIGURE 5.1. SCREENING OF A CHARON 4A SHEEP GENOMIC DNA LIBRARY USING AS PROBE THE cDNA CLONE pSWK18.

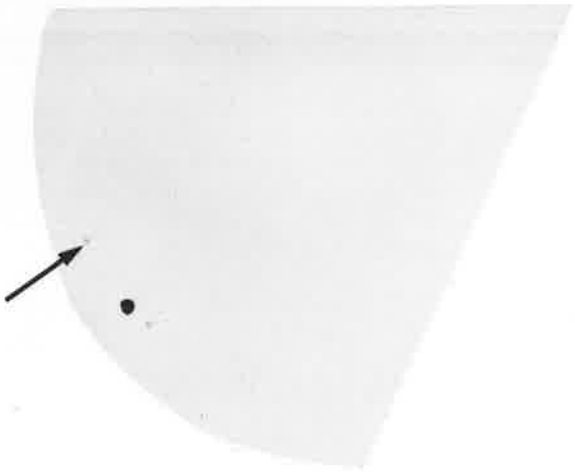
The sheep genomic DNA library of Kretschmer et al. (1980) was screened as described in Chapter 4. Autoradiography was for 1-3 days at -80°C using intensification screens. Hybridizations were carried out for 20h at 42°C (Chapter 2) and the filters were washed at 65°C.

(a) Portion of an autoradiograph of a nitrocellulose filter prepared from a 15cm plate containing approximately $2x10^4$ plaques. Following hybridization, the filter was washed in 1xSSC/0.1% SDS. The arrow points to the single positive plaque.

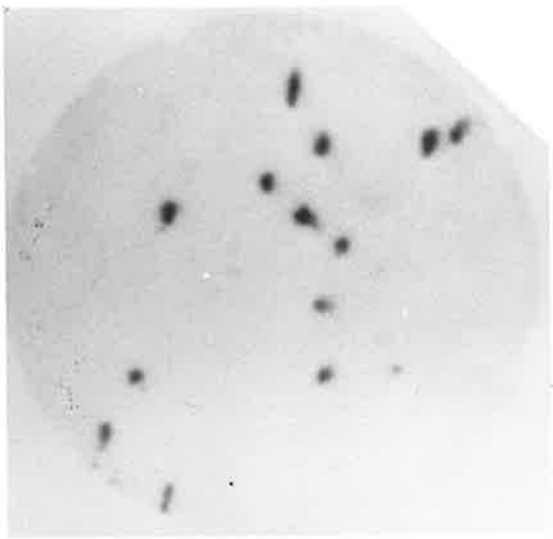
(b) Autoradiograph of a nitrocellulose filter prepared from a 9cm plate containing approximately 200 plaques. Following hybridization, the filter was washed in 0.2xSSC/0.1% SDS.

(c) Autoradiograph of a nitrocellulose filter prepared from a 9cm plate containing approximately 200 plaques. Hybridization was followed by washing in 0.1xSSC/0.1% SDS.

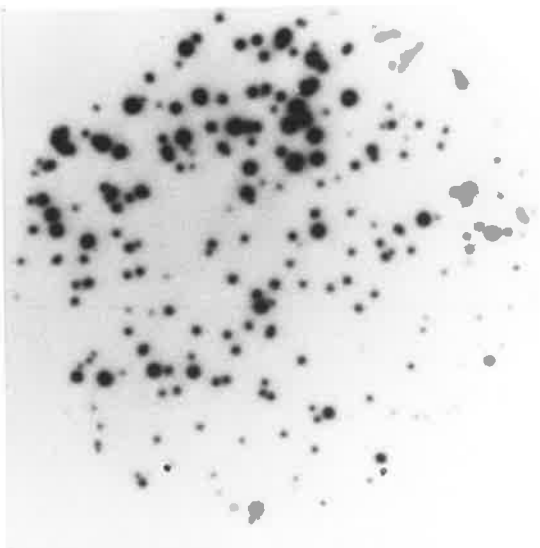
(a)



(b)



(c)



using DNA isolated from a single bacteriophage - λSWK50.

2. Restriction mapping of λSWK50:

The detailed restriction mapping of λSWK50 and subclones derived from it are presented in Appendix A. The restriction map of the λSWK50 insert is shown in Fig.5.2. The BIIIB gene was located on a 2.3 kb HindIII/BglII restriction fragment (Fig.5.2) and this was cloned into M13mp8 and M13mp9 to yield the subclones mp8λ50/2.3 and mp9λ50/2.3 respectively.

3. DNA sequence determination of the λSWK50 BIIIB gene (ψ50BIIIB):

(a) Sequencing strategy:

The strategy used to sequence the λSWK50 BIIIB gene (ψ50BIIIB) is shown in Fig.5.3. From the restriction map data it was possible to prepare a set of overlapping subclones in M13mp8 and M13mp9 which enabled the sequence of both DNA strands to be determined. (The extreme 5' and 3' sequences were determined in one direction only). The major part of the sequence was determined using flush-ended fragments (end-filled where appropriate) isolated from acrylamide gels and ligated into the HincII site of M13mp8, as described in Chapter 4.

The availability of M13 recombinant bacteriophage with the 2.3 kb HindIII/BglII fragment cloned in both orientations (mp8λ50/2.3 and mp9λ50/2.3) allowed rapid electrophoretic screening for selection of a subclone with a sequence in a particular orientation (Chapter 4). The method was applicable for determining the orientation of insert sequences as small as 50 bp (Fig.5.4). The single-stranded DNA of clones containing complementary sequences anneal under the conditions used and the resulting hybrid has a reduced electrophoretic mobility when compared with un-annealed DNA (Fig.5.4).

To obtain the DNA sequence overlapping the PstI site of ψ50BIIIB (Fig.5.3), the 70 bp AluI/PstI fragment was treated with exonuclease III and used as a sequencing primer on mp8λ50/2.3, as described by Hindley and Phear (1981).

FIGURE 5.2. RESTRICTION MAP OF λSWK50.

Restriction map of λSWK50 based on data presented in Appendix A. The location of the 2.3 kb HindIII/BglII fragment used for subcloning into M13 vectors is indicated by the black bar.

V_L and V_R represent the Charon 4A left and right arms respectively.

E: EcoRI sites.

P: PstI.

H: HindIII.

B: BamHI.

Bg: BglII.

K: KpnI.

λSWK50

1cm = 1kb

FIGURE 5.3. STRATEGY USED FOR DETERMINING THE NUCLEOTIDE SEQUENCE OF THE BIIB GENE.

Restriction map of mp8λ50/2.3 which shows only that portion which was sequenced. The restriction sites relevant to the sequencing strategy are shown (Av=AvaII, P=PstI, H=HinfI, A=AluI, Ha=HaeIII and Bg=BglII). Numbering (bp) is from the initiation codon site. The shaded bar on the map represents the location of the BIIIB coding region.

⟼ The line represents the extent of sequence determined from a particular restriction fragment.The vertical line and arrow point show respectively the 5' and 3' extremities of the sequence determined.

⟻⟼ Restriction fragments completely sequenced in both directions.

The 'chain termination' method (Sanger et al., 1980) using commercial primers was used to determine the DNA sequence of the restriction fragments shown below the map.

AluI/PstI* - this fragment was used as an internal sequencing primer, as described by Hindley and Phear (1981).

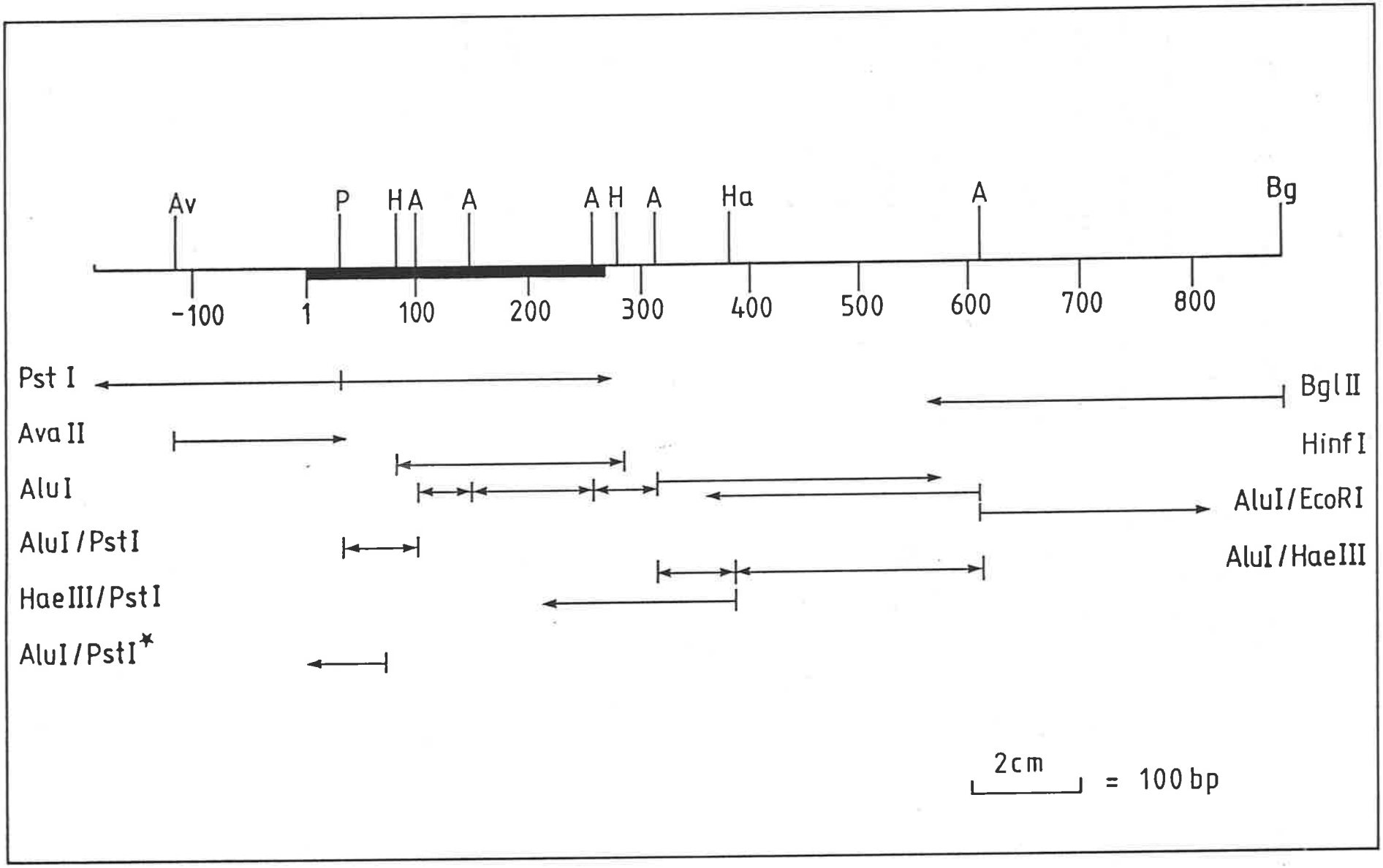

Av
P
H A
A
A H
A
Ha
A
Bg
-100
1
100
200
300
400
500
600
700
800
Pst I
Bgl II
Ava II
Hinf I
Alu I
AluI / EcoRI
AluI / PstI
AluI / HaeIII
HaeIII / PstI
AluI / PstI*
2cm
= 100 bp

FIGURE 5.4. RAPID GEL SCREENING OF M13 CLONES.

Clear M13 plaques were picked into YT broth and grown at 37°C for 6-16h. Aliquots (4µl) of supernatants were mixed with 1µl of 5% SDS and annealed at 65°C for 1h (Chapter 4). After the addition of dye, the mixture was electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light.

(a) mp8λ50/2.3.

(b) mp8λ50/2.3 + M13 clone containing a 50 bp insert (mp8/50).

(c) mp9λ50/2.3 + mp8/50.

a b c

(b) DNA sequence of ψ50BIIIB:

Figure 5.5 shows the complete nucleotide sequence of ψ50BIIIB from 190 bp 5' to the initiation codon site to 640 bp downstream from the termination codon. The restriction sites used in M13 subcloning experiments are shown. As noted in Appendix A (Section 3), the size of one AluI fragment estimated from electrophoretic mobilities was anomalously high. The "400" bp fragment lies between residues 503 and 794 and so is actually more than 100 bp smaller than estimated.

The gene has a TATA box positioned approximately 90 bp 5' to the initiation codon site, and a possible CAAT box located a further 60 bp upstream. While the TATA box is in excellent agreement with the consensus sequence 5'-TATATAT-3', the CAAT box has only a limited
A A

homology with the consensus sequence 5'-GGCCAATCT-3' (Breathnach and
T

Chambon, 1981). The cap site is usually an A residue located 26-34 bp downstream from the first nucleotide of the TATA box, is generally preceded by a C residue, and is in a region rich in pyrimidines (Cordon et al., 1980; Breathnach and Chambon, 1981). The boxed A residue in Figs.5.5 and 5.6 most closely fulfills these requirements and is the probable mRNA cap site.

The initiation and termination codon sites, based on cDNA clone sequence data are shown in Fig.5.5. The initiation codon has a C for T substitution, thus excluding initiation from this site. There are three poly(A)-addition signal sequences. Two of these overlap while the third is situated a further 180 bp downstream. Thirteen base pairs 3' to this latter poly(A)-addition sequence is a short stretch of A residues followed by a 20 bp sequence which is a limited direct repeat of a sequence surrounding the possible cap site (Fig.5.5).

The 3' flanking sequence of ψ50BIIIB is characterized by being rich in A and T residues. While the A/T percentage for the sequenced flanking region is 60%, this increases to 80% if only the first 140 bases

FIGURE 5.5. THE COMPLETE NUCLEOTIDE SEQUENCE OF ψ5OBIIIB.

The nucleotide sequence of the mRNA strand is shown in the 5' to 3' direction. The restriction sites used for subcloning and sequencing experiments are shown by a line under the sequence and the name of the relevant enzyme above the sequence.

===== indicates the consensus CAAT and TATA sequences.

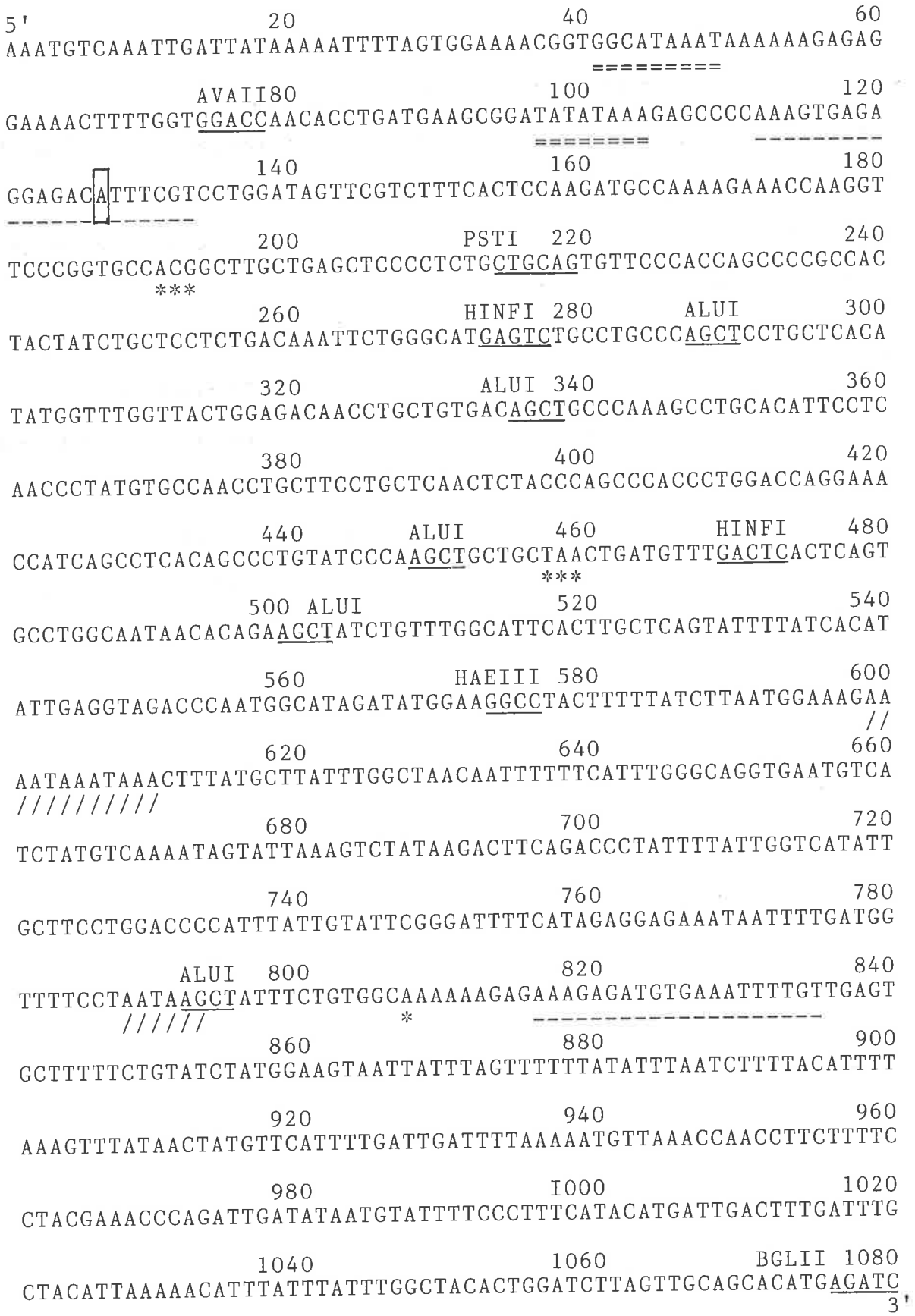
///// shows the poly(A)-addition signal sequences.

*** shows the location of the initiation and termination codon sites, based on cDNA sequence data.

----- indicates a 20 bp limited direct repeat.

* is the poly(A)-addition site of the cDNA clone pSWK18.

The possible cap site is boxed.


5'
20
40
60
AAATGTCAAATTGATTATAAAAATTTTAGTGGAAAACGGTGGCATAAATAAAAAAGAGAG
AVAII80
100
120
GAAAACTTTTGGTGGACCAACACCTGATGAAGCGGATATATAAAGAGCCCCAAAGTGAGA
140
160
180
GGAGACATTTCGTCCTGGATAGTTCGTCTTTCACTCCAAGATGCCAAAAGAAACCAAGGT
200
PSTI 220
240
TCCCGGTGCCACGGCTTGCTGAGCTCCCCTCTGCTGCAGTGTTCCCACCAGCCCCGCCAC
260
HINFI 280
ALUI
300
TACTATCTGCTCCTCTGACAAATTCTGGGCATGAGTCTGCCTGCCCAGCTCCTGCTCACA
320
ALUI 340
360
TATGGTTTGGTTACTGGAGACAACCTGCTGTGACAGCTGCCCAAAGCCTGCACATTCCTC
380
400
420
AACCCTATGTGCCAACCTGCTTCCTGCTCAACTCTACCCAGCCCACCCTGGACCAGGAAA
440
ALUI
460
HINFI
480
CCATCAGCCTCACAGCCCTGTATCCCAAGCTGCTGCTAACTGATGTTTGACTCACTCAGT
500 ALUI
520
540
GCCTGGCAATAACACAGAAGCTATCTGTTTGGCATTCACTTGCTCAGTATTTTATCACAT
560
HAEIII 580
600
ATTGAGGTAGACCCAATGGCATAGATATGGAAGGCCTACTTTTTATCTTAATGGAAAGAA
620
640
660
AATAAATAAACTTTATGCTTATTTGGCTAACAATTTTTTCATTTGGGCAGGTGAATGTCA
680
700
720
TCTATGTCAAAATAGTATTAAAGTCTATAAGACTTCAGACCCTATTTTATTGGTCATATT
740
760
780
GCTTCCTGGACCCCATTTATTGTATTCGGGATTTTCATAGAGGAGAAATAATTTTGATGG
ALUI 800
820
840
TTTTCCTAATAAGCTATTTCTGTGGCAAAAAAGAGAAAGAGATGTGAAATTTTGTTGAGT
860
880
900
GCTTTTTCTGTATCTATGGAAGTAATTATTTAGTTTTTTATATTTAATCTTTTACATTTT
920
940
960
AAAGTTTATAACTATGTTCATTTTGATTGATTTTAAAAATGTTAAACCAACCTTCTTTTC
980
I000
1020
CTACGAAACCCAGATTGATATAATGTATTTTCCCTTTCATACATGATTGACTTTGATTTG
1040
1060
BGLII 1080
CTACATTAAAAACATTTATTTATTTGGCTACACTGGATCTTAGTTGCAGCACATGAGATC
3'

of the flanking sequence is considered (Fig.5.5).

4. Comparison of the ψ5OBIIIB sequence with BIIIB cDNA sequences:

(a) 5' non-coding region:

Figure 5.6 shows the 5' flanking and non-coding sequence of ψ5OBIIIB. For comparison, the 5' non-coding sequence of the cDNA clone pSWK101 is shown in this figure - pSWK101 codes for a BIIIB2-like protein (Swart et al; 1976) and is the only BIIIB cDNA clone for which 5' non-coding data are available (K. A. Ward and M. J. Sleigh, personal communication).

While the homology between the 5' non-coding regions of ψ5OBIIIB and pSWK101 is not significant when the complete sequences are compared, there is a greater than 70% homology for 27 bp 5' to the initiation codon. As noted before, the initiation codon of ψ5OBIIIB has a base change in the second position.

(b) Coding region:

The nucleotide sequence of the ψ5OBIIIB coding region is shown in Fig.5.7. For reasons which will become more apparent when the protein sequence data is presented, the major comparison given here is with pSWK18. Since pSWK18 does not contain a complete coding sequence, the first 66 bases of ψ5OBIIIB have been compared with pSWK101. It is apparent that there is a strong homology between the gene and the cDNA sequences. If all the differences are considered, then there is a 75% homology between the gene and "hybrid" cDNA sequence. It should be noted that three of the differences between ψ5OBIIIB and pSWK101 result from base changes which, on the basis of protein sequence data, would not be expected for the pSWK18 sequence (if it were complete) - as noted previously, pSWK18 and pSWK101 code for BIIIB3-like and BIIIB2-like proteins respectively.

When only pSWK18 is compared with the gene sequence, the homology is 72%. If the 23 bp deletion near the 3' end of the coding region is disregarded, then the homology increases to 82%.

FIGURE 5.6. COMPARISON OF THE 5' FLANKING AND NON-CODING SEQUENCE OF ψ5OBIIIB WITH THE 5' NON-CODING REGION OF pSWK101.

The nucleotide sequence of the 5' flanking and non-coding region of ψ5OBIIIB is compared with the 5' non-coding region of pSWK101. Deletions (-) have been introduced to maximize the homology.

===== underlines the consensus CAAT and TATA sequences.

* marks the 5' terminal residue of pSWK101.

5O: ψ5OBIIIB.

101: pSWK101.

The possible cap site is boxed.

For pSWK101, only bases differing from ψ5OBIIIB are shown.

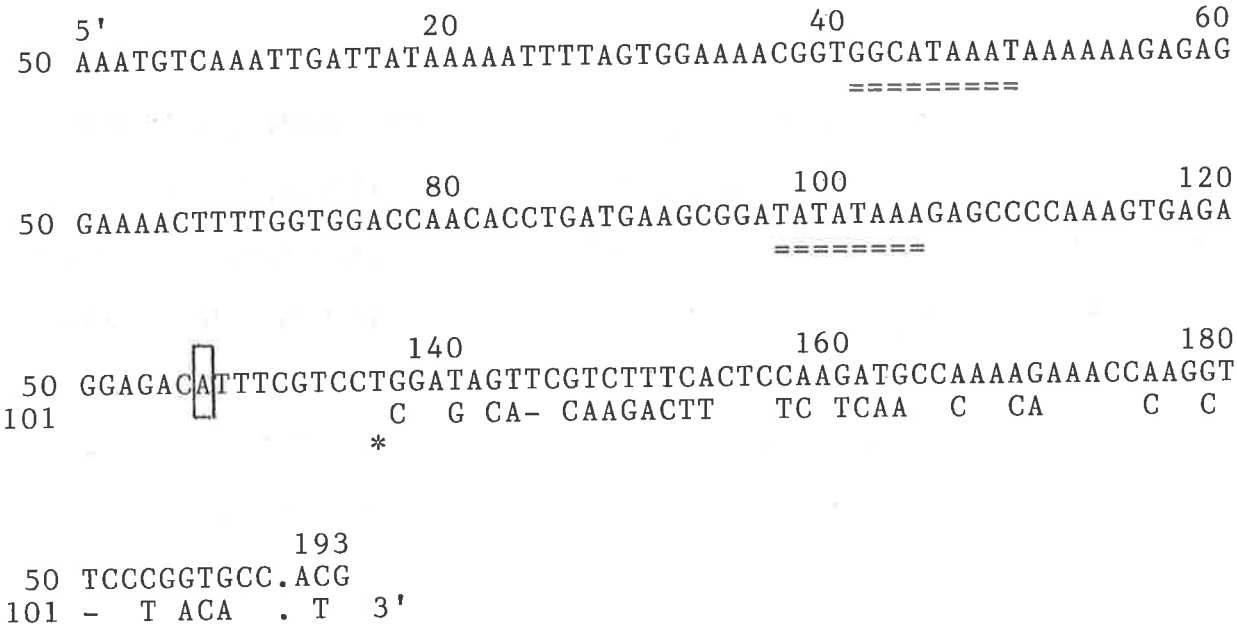

5'
20
40
60
50 AAATGTCAAATTGATTATAAAAATTTTAGTGGAAAACGGTGGCATAAATAAAAAAGAGAG
80
100
120
50 GAAAACTTTTGGTGGACCAACACCTGATGAAGCGGATATATAAAGAGCCCCAAAGTGAGA
140
160
180
50 GGAGACATTTCGTCCTGGATAGTTCGTCTTTCACTCCAAGATGCCAAAAGAAACCAAGGT
101 C G CA- CAAGACTT TC TCAA C CA C C
193
50 TCCCGGTGCC.ACG
101 - T ACA . T 3'

FIGURE 5.7. COMPARISON OF THE CODING REGION OF ψ50BIIIB WITH pSWK101 AND pSWK18.

The mRNA strand sequence of the ψ50BIIIB coding region is shown in the 5' to 3' direction. Comparison of the first 66 bases is with pSWK101, while the remainder of the sequence is compared with pSWK18.

For pSWK101 and pSWK18, only those bases differing from ψ50BIIIB are shown. Deletions (-) have been introduced to maximize the homology.

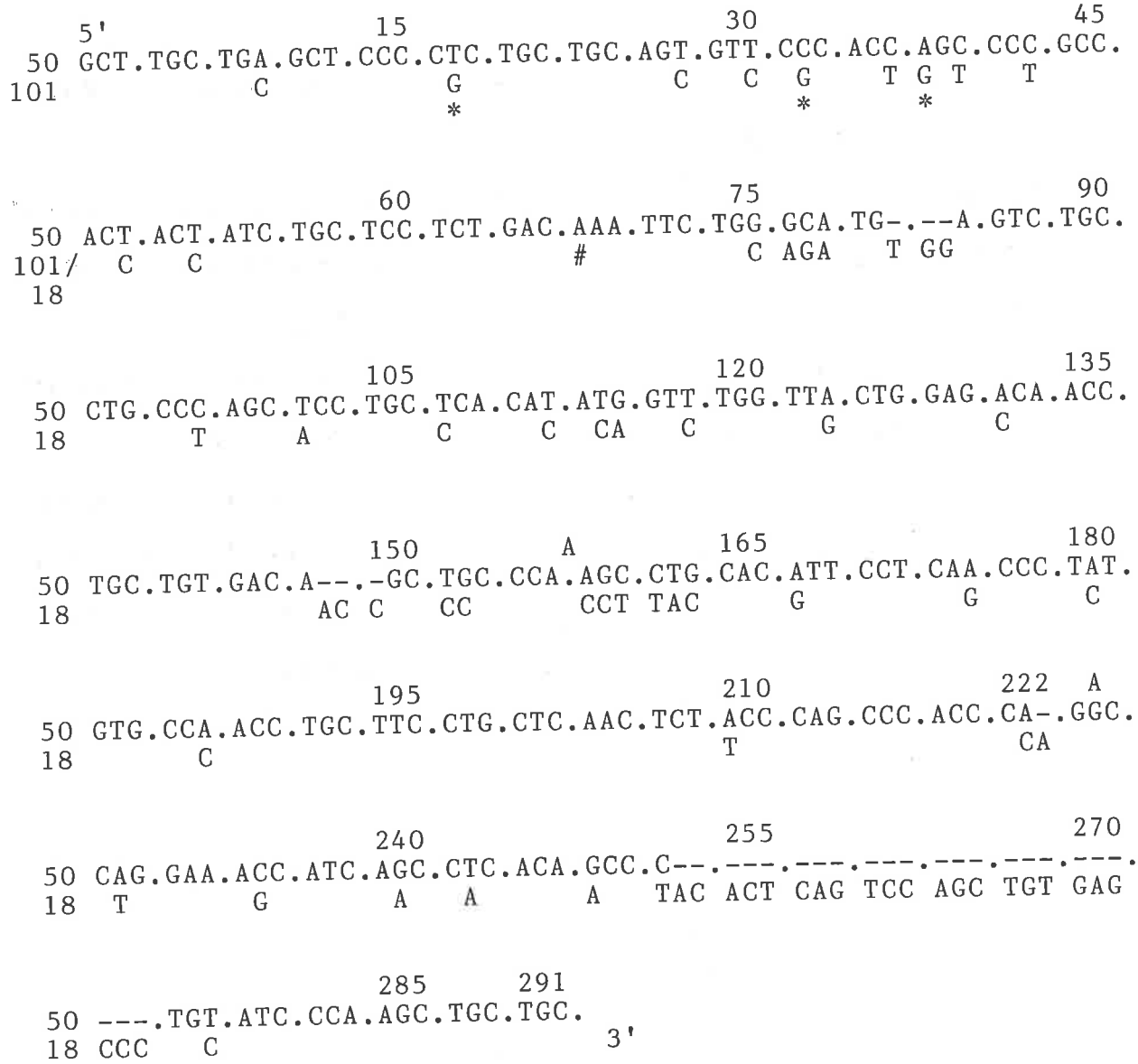
* denotes base differences leading to amino acid changes which would not be expected in the BIIIB3-like protein coded by pSWK18, but which are consistent with the BIIIB2-like protein coded by pSWK101.

\# The 5' terminal base of pSWK101.

50: ψ50BIIIB.

101: pSWK101.

18: pSWK18.


5'
15
30
45
50 GCT.TGC.TGA.GCT.CCC.CTC.TGC.TGC.AGT.GTT.CCC.ACC.AGC.CCC.GCC.
101 C G C C G T G T T
* * *
60
75
90
50 ACT.ACT.ATC.TGC.TCC.TCT.GAC.AAA.TTC.TGG.GCA.TG-.--A.GTC.TGC.
101/ C C # C AGA T GG
18
105
120
135
50 CTG.CCC.AGC.TCC.TGC.TCA.CAT.ATG.GTT.TGG.TTA.CTG.GAG.ACA.ACC.
18 T A C C CA C G C
150
A
165
180
50 TGC.TGT.GAC.A--.-GC.TGC.CCA.AGC.CTG.CAC.ATT.CCT.CAA.CCC.TAT.
18 AC C CC CCT TAC G G C
195
210
222 A
50 GTG.CCA.ACC.TGC.TTC.CTG.CTC.AAC.TCT.ACC.CAG.CCC.ACC.CA-.GGC.
18 C T CA
240
255
270
50 CAG.GAA.ACC.ATC.AGC.CTC.ACA.GCC.C--.---.---.---.---.---.---.
18 T G A A A TAC ACT CAG TCC AGC TGT GAG
285
291
50 ---.TGT.ATC.CCA.AGC.TGC.TGC.
18 CCC C 3'

It is apparent that with a mutated initiation codon and a 23 bp deletion in the coding region (Figs.5.7,5.8a), ψ50BIIIB is not a normal BIIIB gene. This is supported by other features of the coding sequence. There is an in-frame stop codon 6 bp from the 5' end (Fig.5.7), while a 3 bp deletion introduces another termination codon into the sequence between position 75 and 90 (Figs.5.7, 5.8b). There is another deletion immediately 5' to position 150, but the sequence remains in-frame until an A residue is inserted after position 156. The sequence continues out of frame until position 222 where there is a single base deletion, but the situation is then immediately reversed by an insertion 3 bp downstream. A further frame shift occurs as a result of the 23 bp deletion near the 3' end of the coding region (Fig.5.7).

Apart from the termination codons, frame shifts and deletions, there are also a number of base changes which lead to amino acid differences between the gene and pSWK18 protein sequences. The predicted protein sequence of ψ50BIIIB will be presented following the 3' non-coding and flanking data but it is apparent from the features of the coding region that ψ50BIIIB is a BIIIB pseudogene.

(c) 3' non-coding region:

Figure 5.9 compares the 3' non-coding sequences of ψ50BIIIB and pSWK18. The homology between the two sequences is 78%, comparable to that found for the coding region. The homology is strongest at the 5' and 3' ends of the non-coding region. Over the first 65 bp, the homology is 83% while over the 3' terminal 70 bp to the start of the pSWK18 poly(A) tail, the homology is 87%. Both the gene and the cDNA clone have a consensus AATAAA sequence approximately 20 bp from the poly(A)-addition site (although the sequence in ψ50BIIIB is AATAAG). The gene also has two AATAAA sequences which overlap approximately 150 bp downstream from the termination codon (Fig.5.9). Interestingly, the gene sequence has 6 A residues in a position that corresponds to the poly(A) tail of pSWK18 (Fig.5.9).

FIGURE 5.8. DNA SEQUENCING GELS OF TWO REGIONS FROM THE CODING SECTION OF ψ5OBIIIB.

(a) Gel analysis of a 195 bp HinfI fragment (Fig.5.7) cloned into the HincII site of M13mp8 and sequenced as described in Chapter 4.

* shows the site of a 23 base deletion.

(b) Gel analysis of a 69 bp PstI/AluI fragment (Fig.5.7) cloned into the HincII site of M13mp8 and sequenced as described in Chapter 4.

** shows the site of a 3 base deletion which results in the formation of an in-frame stop codon.

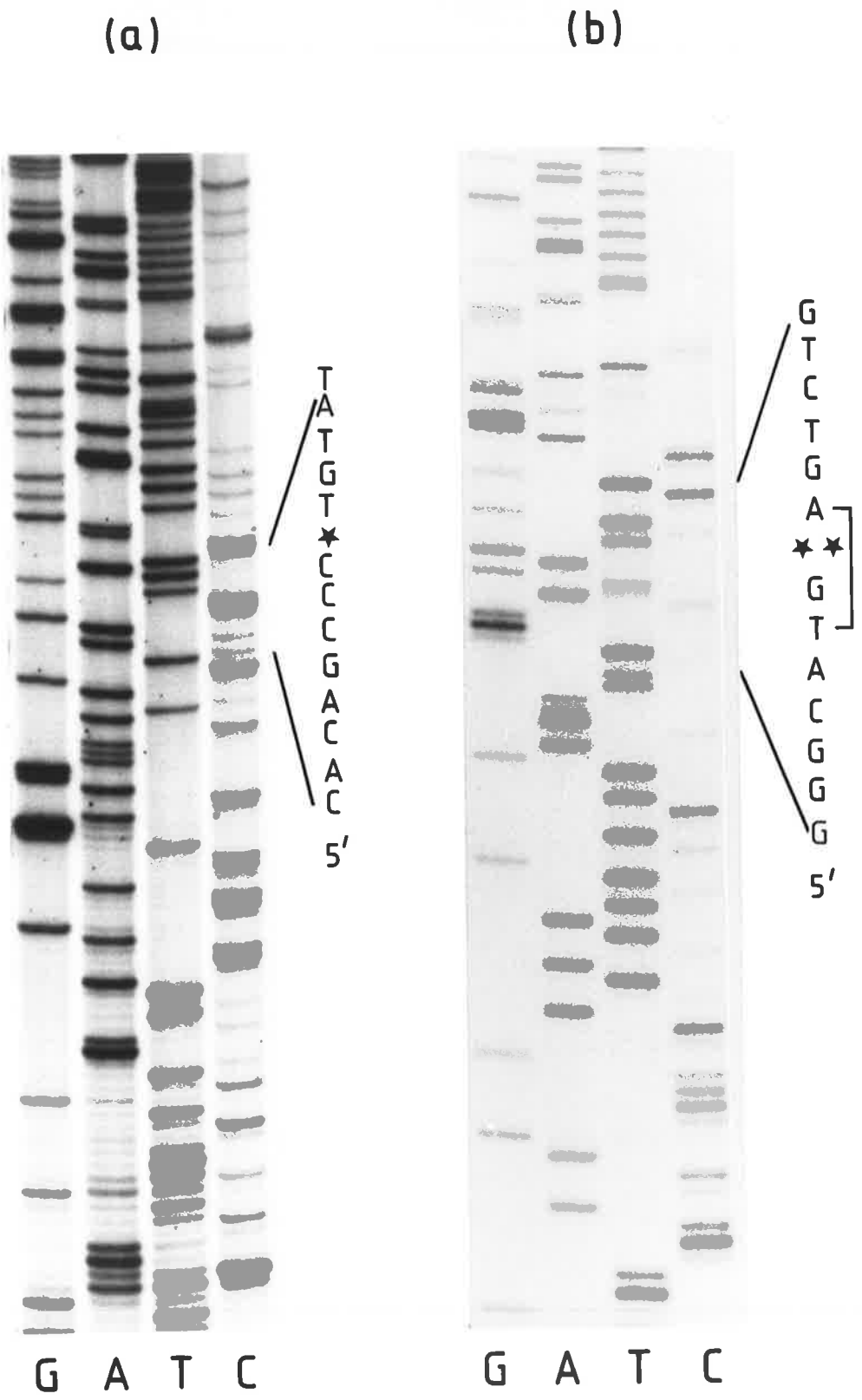

(a)
(b)
TATGT★CCCGACAC
5′
GTCTGA★★GTACGGG
5′
G A T C
G A T C

FIGURE 5.9. COMPARISON OF THE 3' NON-CODING REGIONS OF ψ5OBIIIB AND pSWK18.

The nucleotide sequence of the mRNA strand is shown for the 3' non-coding region of ψ5OBIIIB in the 5' to 3' direction. Deletions (-) have been introduced to maximize the homology.

The pSWK18 sequence is shown only where it differs from ψ5OBIIIB.

===== indicates the consensus AATAAA sequences.

5O: ψ5OBIIIB.

18: pSWK18.

```
   5'                  20                  40                  60
50 TAACTGATGTTTGACTCACTCAGTGCCTGGCAATAACACAGAAGCTATCTGTTTGGCATT
18  G  C  C GC                  AT                      A    AT

                    80                  100                 120
50 CACTTGCT-CAGTATTTTATCACATATTGAGGTAGACCCA-ATGGCATAGATATGGAAGG
18 T    T  T  A  G       G   G   T     T - G                  A-

                   140        ==========160                 180
50 CCTA-CTTTTTATCTTAATGGAAAGAAAATAAATAAACTTTATGCTTATTTGGCT----A
18     A    C-  -  C             GTT T  AA AC-             GAGT

                   200                 220                 240
50 ACAATTTT-TTCATTTGGGC------AGGTGAATGTCATCTATG-TCAAAATAGTA-TTA
18   CC   GG          ATAGATAG     ------       TA           G G

                   260                 280                 300
50 AAGTCTATAAGACTTCAGACCCTATTTTATTGGTCATATTGCTTCCTGGACCCCATTTAT
18   -          C      T       C       G       T   GT G   A C

                   320                 340         ======  360
50 TGTATTCGGG-ATTTTCATAGAGGAGAAATAA-TTTTGATGGTTTTCCTAATAAGCTA--
18       -  C              A      C                -     A  TTG

                376
50 TTTCTGTGGCAAAAAA
18           POLY A  3'
```

5. Amino acid sequence coded by ψ50BIIIB (ψBIIIB3A):

The amino acid sequence of the protein coded by ψ50BIIIB, disregarding the frame-shifts and deletions, is homologous to sequences of the BIIIB high-sulphur protein family (Fig.5.10). There is a 65% homology when ψBIIIB3A is compared with BIIIB3. If the deletions are disregarded, then the homology increases to 76%

Comparison of only the known BIIIB protein sequences (Fig.5.10) reveals that the differences between the proteins occur at particular amino acid residues, while the greater part of the sequences are completely conserved. A study of the ψBIIIB3A residues which lie in the variable regions indicates that 16 are identical to BIIIB3, 4 are the same as BIIIB2 and only the Ile at residue 57 is found in BIIIB4. Only 3 amino acids in the variable regions of ψBIIIB3A do not correspond to a residue found in the BIIIB family-Met at position 38, Leu at 55 and Thr at 79 (Fig.5.10).

There are a relatively large number of amino acid differences between ψBIIIB3A and other BIIIB proteins in the conserved regions. Many of these result from single base changes although a number are due to 2 or 3 base changes (Fig.5.7). Apart from the 8 amino acid deletion near the C-terminal end of ψBIIIB3A, the sequence differences between this protein and BIIIB3 appear to be randomly distributed between conserved and variable regions (Fig.5.10).

6. Probing λSWK50 for other keratin genes:

In order to determine whether any other keratin genes were located on λSWK50, restriction fragments of the phage DNA or pλSWK50 DNA were probed with either wool follicle cDNA or cDNA clones representing the low-sulphur and B2 high-sulphur sequences. Probing with cDNA failed to reproducibly detect any positive fragments other than those containing the BIIIB gene (data not shown). Similarly, screening with cDNA clones coding for low-sulphur and B2 high-sulphur proteins failed to detect any positive hybridization (data not shown). At the time that these

FIGURE 5.10. AMINO ACID SEQUENCE OF BIIIB3 COMPARED WITH BIIIB2, BIIIB4 AND ψBIIIB3A.

The amino acid sequence of the wool protein BIIIB3 is presented in full (Swart et al., 1976). The sequences of BIIIB2 and BIIIB4 (Swart et al., 1976) and ψBIIIB3A are shown only where they differ from BIIIB3. The sequence of ψBIIIB3A is presented disregarding frame-shifts, and deletions (---) have been introduced to maximize the homology.

\# indicates that the proteins sequenced by Swart et al. (1976) had an acetylated N-terminus.

*** denotes in-frame stop codons.

2,3 and 4 refer to BIIIB2, BIIIB3 and BIIIB4 respectively.

3A refers to ψBIIIB3A.

					5					10					15
2	#				PRO	ARG					ARG		GLY		
3	ALA	CYS	CYS	ALA	ARG	LEU	CYS	CYS	SER	VAL	PRO	THR	SER	PRO	ALA
4															
3A			***		PRO										

					20					25					30
2															
3	THR	THR	ILE	CYS	SER	SER	ASP	LYS	PHE	CYS	ARG	CYS	GLY	VAL	CYS
4															
3A										TRP	ALA	***	---		

					35					40					45
2								ASN	ILE	SER					
3	LEU	PRO	SER	THR	CYS	PRO	HIS	THR	VAL	TRP	LEU	LEU	GLN	PRO	THR
4											PHE				
3A				SER		SER		MET					GLU	THR	

					50					55					60
2			---			SER		VAL		CYS	VAL	TYR		ASP	THR
3	CYS	CYS	CYS	ASP	ASN	ARG	PRO	PRO	PRO	TYR	HIS	VAL	PRO	GLN	PRO
4										CYS		ILE			
3A			---		---	SER	CYS		SER	LEU		ILE			

					65					70					75
2	TYR											HIS			
3	SER	VAL	PRO	THR	CYS	PHE	LEU	LEU	ASN	SER	SER	GLN	PRO	THR	PRO
4															
3A	TYR										THR				LEU

					80					85					90
2			SER	GLY						PHE	ILE		PRO	GLY	
3	GLY	LEU	GLU	SER	ILE	ASN	LEU	THR	THR	TYR	THR	GLN	SER	SER	CYS
4													PRO		
3A	ASP	GLN		THR		SER			ALA	---	---	---	---	---	---

					95				99
2		ASN	VAL		GLU		ARG		---
3	GLU	PRO	---	CYS	ILE	PRO	SER	CYS	CYS
4									
3A	---	---							

experiments were undertaken, no high-glycine-tyrosine cDNA probes were available.

7. Probing λSWK50 for repetitive DNA sequences:

The presence of repetitive DNA in the genomic clone was investigated initially by probing PstI-digested λSWK50 with nick-translated sheep liver DNA and using washing conditions which allow detection of sequences present in greater than 50 copies per genome (Chapter 4). In the PstI digest (Fig.5.11), the 4.5, 4.2 and 2.0 kb fragments show positive hybridization. This result indicates that there are repetitive DNA sequences dispersed throughout most of λSWK50.

The B2 high-sulphur genes have been shown to have at least two repetitive DNA families associated with them - termed the XhoII and NcoI families (B. C. Powell, personal communication). Probing HindIII/EcoRI double-digests of λSWK50, the XhoII repetitive family sequence gave a weak positive hybridization to the 1.0 kb fragment which lies at the 5' end of the insert (Figs.5.2 and A.1c; i.e. figure 1c, Appendix A), while the NcoI probe resulted in moderate hybridization to the 2.5 kb HindIII fragment - data not shown. (I am indebted to Dr. B. C. Powell, University of Adelaide, for carrying out the XhoII and NcoI probing experiments).

8. Screening of other sheep genomic DNA libraries using the pSWK18 insert as the probe:

(a) λ1059 BamHI library:

The λ1059 sheep genomic DNA library consisted of 0.6×10^6 unique recombinants maintained in 4 amplified fractions. Approximately 0.2×10^6 plaques were screened for each fraction using nick-translated pSWK18 insert as the probe (Chapter 4). No positive recombinants were observed.

(b) λgt10 EcoRI library:

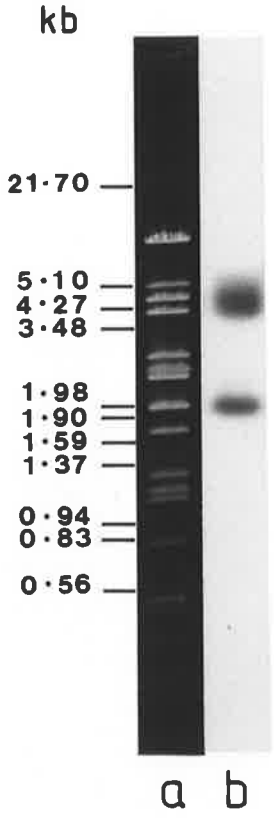
The λgt10 library consisted of 1.5×10^6 unique recombinants containing sheep genomic DNA which had been digested to completion with EcoRI (Chapter 4). The EcoRI fragment which carries the λ50BIIIB

FIGURE 5.11. HYBRIDIZATION OF λSWK50 REPETITIVE DNA WITH SHEEP GENOMIC DNA.

λSWK50 DNA was digested with PstI, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. Following Southern transfer, the digested DNA was hybridized with nick-translated sheep liver DNA for 40h at 65°C, as described in Chapter 2. The filter was washed as described by Shen and Maniatis (1980).

(a) Ethidium bromide stained gel. The DNA molecular weight marker was a HindIII/EcoRI digest of λ DNA.

(b) Autoradiography of (a) following hybridization analysis.


kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
a
b

pseudogene is 13.2 kb long and should not clone into λgt10. Southern analysis of EcoRI-digested genomic DNA using the pSWK18 insert as probe indicated that, apart from the 13.2 kb fragment, the major BIIIB gene-containing fragments were approximately 4-5 kb in size (Fig.5.12). Screening of 0.9×10^6 plaques from the unamplified λgt10 library failed to detect any pSWK18-positive recombinants.

(c) Charon 28 Sau3A library:

The Charon 28 genomic DNA library contained 0.75×10^6 unique recombinants maintained as a single amplified stock (Chapter 4). A screening of approximately 0.2×10^6 plaques using the pSWK18 insert as a probe and non-stringent washing conditions to optimize hybrid stability, resulted in the isolation of 20 possible positive recombinants. A second screening indicated that 6 of the isolates were positive. However, with third and subsequent screenings, the number of pSWK18-positive plaques decreased. A repeat of the screenings showed that the hybridization results were reproducible. By carefully repeating a series of screenings it was found that the titre of pSWK18-positive plaques was 10^3 lower than other plaques in the library. This had led to the loss of positive recombinants during the earlier screenings where there had been dilution with other bacteriophage. Another second screen was carried out which enabled individual positive plaques to be picked. Twelve recombinants were selected from this screening and one, λSWK61, was used in restriction mapping and DNA sequencing studies.

9. Restriction mapping and subcloning of λSWK61:

Detailed restriction enzyme mapping and cloning of λSWK61 are presented in Appendix B. The restriction map of a 7.2 kb BamHI fragment which contains the BIIIB gene is shown in Fig.5.13. The data obtained from the λSWK61 mapping studies (Appendix B) suggested that this genomic clone was homologous to a portion of λSWK50 (compare Figs.5.2 and 5.13; data presented in Appendices A and B).

FIGURE 5.12. SOUTHERN ANALYSIS OF EcoRI-DIGESTED SHEEP GENOMIC DNA USING pSWK18 AS THE PROBE.

High-molecular-weight sheep liver DNA (Dorset Horn) was digested to completion with EcoRI, electrophoresed in 1% agarose and transferred to nitrocellulose as described in Chapter 4. The filter was hybridized for 40h at 65°C with nick-translated pSWK18 insert and then washed in 0.5xSSPE/0.1% SDS at the same temperature (Chapter 2).The filterwas autoradiographed for 1 week at -80°C using intensification screens.
The DNA molecular weight marker was a HindIII/EcoRI digest of λ DNA.

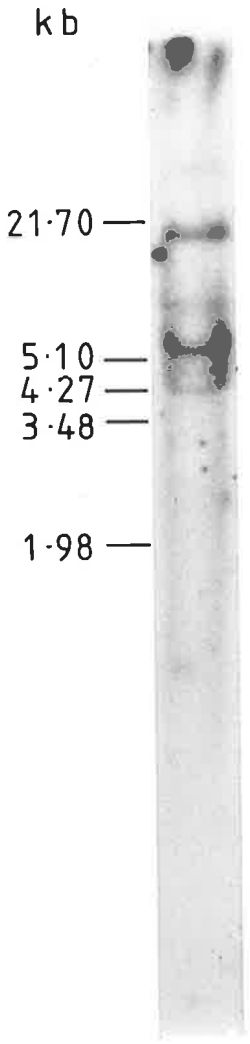

kb
21·70
5·10
4·27
3·48
1·98

FIGURE 5.13. RESTRICTION MAP OF A 7.2 kb BamHI FRAGMENT ISOLATED FROM λSWK61.

A 7.2 kb BamHI fragment, containing a BIIIB gene, was subcloned from λSWK61 into pBR322. The restriction map of the subclone (pλSWK61/7.2) is based on data presented in Appendix B. The orientation shown here is opposite to that of a homologous fragment found in λSWK50 (Fig.5.2). The arrow indicates the location of a BglII site present in λSWK50 but absent in λSWK61.

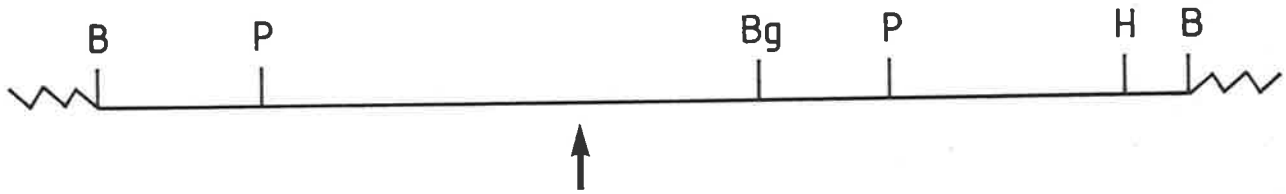
B: BamHI sites.

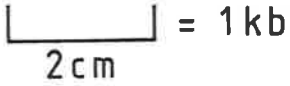
P: PstI.

Bg: BglII.

H: HindIII.

pλSWK61/7·2





10. DNA sequence analysis of a pλSWK61/7.2 PstI/BglII fragment:

The 0.9 kb PstI/BglII fragment of pλSWK61/7.2 (Fig.5.13) was cloned into PstI/BamHI-digested M13mp8 so that the insert would be sequenced inwards from the PstI end. If the assumption was correct that this fragment was homologous to the 0.9 kb PstI/BglII fragment of λSWK50, the DNA sequence would show an in-frame stop codon 53 bases from the PstI site. A single experiment indicated that the TGA codon was present (Fig.5.14), and that the surrounding sequence was identical to that found in the ψ50BIIIB pseudogene (Fig.5.7). Further analysis of this subclone or λSWK61 was considered to be unnecessary.

C. DISCUSSION

1. λSWK50 contains a BIIIB high-sulphur pseudogene:

The screening of a Charon 4A sheep genomic DNA library using as probe the BIIIB3-like high-sulphur cDNA clone pSWK18, resulted in the isolation of seven recombinants, each containing the same 13.2 kb insert. Restriction enzyme mapping and Southern probing with pSWK18 indicated that the recombinant chosen for study, λSWK50, contained a single BIIIB gene. The gene was isolated and its DNA sequence was determined (Fig.5.5). A comparison of the gene sequence with those of two BIIIB cDNA clones (Figs.5.6, 5.7) revealed that while there were extensive homologies between the sequences, the gene did not code for a normal high-sulphur protein. In addition to the mutated initiation codon, which would preclude translation, the gene contained two in-frame stop codons and a 23 bp deletion in the coding region. There were also a number of smaller deletions and two single base insertions which altered the reading frame. On the basis of these data, the gene - ψ50BIIIB - can be classified as a BIIIB high-sulphur pseudogene - the first to be reported for the keratin multigene complex.

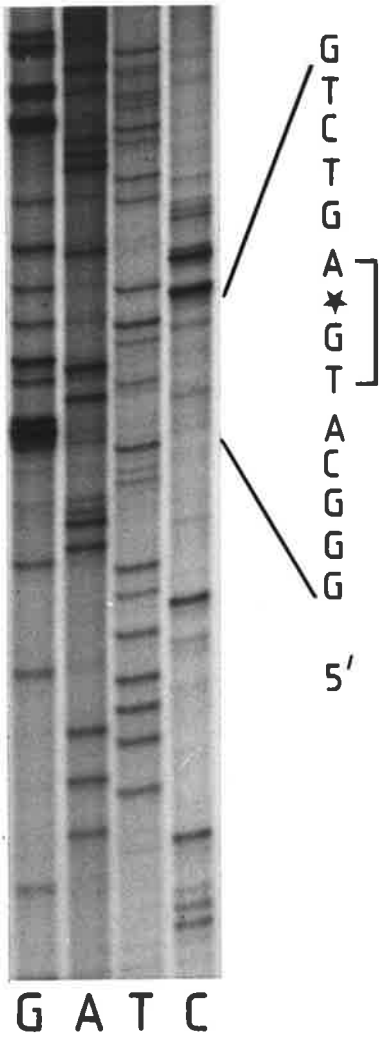
2. ψ50BIIIB originally coded for a BIIIB3-related protein:

Although ψ50BIIIB contains multiple genetic lesions which preclude the translation of any transcript to give a functional high-

FIGURE 5.14. DNA SEQUENCING GEL OF AN M13mp8 CLONE CONTAINING THE 0.9 kb pλSWK61/7.2 PstI/BglII FRAGMENT.

Gel analysis of the 0.9 kb pλSWK61/7.2 PstI/BglII fragment cloned into PstI/BamHI-cleaved M13mp8 and sequenced as described in Chapter 4.

* shows the site of the deletion which results in the formation of an in-frame stop codon.


G
T
C
T
G
A
★
G
T
A
C
G
G
G
5′
G A T C

sulphur protein, an analysis of the nucleotide sequence and the predicted protein sequence can provide valuable data concerning the possible nature of the original active gene and the evolution of the pseudogene.

The cDNA clone pSWK18 does not contain a complete coding region. However, the available data reveal that this clone codes for a protein which differs at only three residues from component BIIIB3, while the recombinant plasmid pSWK101 codes for a BIIIB2-like protein (K. A. Ward, personal communication, Swart et al., 1976). The coding region of ψ5OBIIIB has an 80% homology with pSWK18 and a 75% homology with pSWK101. This latter degree of homology is comparable to that found between the coding regions of the cDNA clones (data not shown).

Both ψ5OBIIIB and pSWK18 have 3' non-coding regions of approximately 350 bp while the equivalent sequence in pSWK101 is 270 bp. A comparison of the 3' non-coding regions indicates that ψ5OBIIIB has a homology of almost 80% with pSWK18, while neither the pseudogene nor pSWK18 show any significant similarity to pSWK101. The nucleotide sequence data suggest that ψ5OBIIIB is a BIIIB3-related pseudogene.

The protein sequence data provide further evidence that the pseudogene originally coded for a BIIIB3-like protein. A feature of the known BIIIB protein sequences is that 70% of the residues are conserved between the components (Fig.5.10). If the deletions are disregarded and only the 24 residues which are variable in the BIIIB protein family are considered, 16 amino acids in ψBIIIB3A (the protein coded for by ψ5OBIIIB) are identical to BIIIB3, 4 are found in BIIIB2, 1 occurs in BIIIB4 and 3 are unique to the pseudogene product (Fig.5.10). Furthermore, a comparison of ψBIIIB3A with BIIIB3 indicates that the ratio of amino acid mismatches in the conserved and variable regions is the same. These data suggest that ψBIIIB3A has mutated from a BIIIB3-like protein.

As described above, the mutations found in the residues of the variable region of ψBIIIB3A show a definite bias towards amino acids

found in other BIIIB family members. It is possible that the functional BIIIB3 gene underwent a duplication event which was followed by the accumulation in one of the active genes of some "acceptable" mutations - i.e. residues found in other BIIIB proteins. Subsequent changes in the coding sequence could have led to inactivation of the mutated gene, resulting in the BIIIB pseudogene. Such a hypothesis suggests that there are selection pressures acting to maintain the composition of the variable regions of the BIIIB gene family within certain limits.

3. Heterogeneity of the BIIIB gene family:

Southern analysis of Dorset Horn genomic DNA using pSWK18 as the probe (Fig.5.12) indicated that there were up to 12 genes present. Densitometric scanning of the autoradiograph indicated that the 13 and 5 kb bands were approximately four times as intense as the other fragments. If the four faint bands contain a single BIIIB gene then there are a total of 12 genes. If, however, the faint bands are derived from a portion of a gene, then there are only 6 BIIIB genes. The number of BIIIB genes determined using the pSWK18 probe should include the BIIIB3- andBIIIB4-related sequences since there appears to be a close relationship between the coding and 3' non-coding regions of these genes. (see Chapter 6). It is unlikely that the pSWK18 probe would form stable hybrids with any pSWK101 sequences under the washing conditions used for the analysis shown in Fig.5.12. There is no homology between the 3' non-coding regions and while there is a 75% homology of the coding sequences, the nucleotide differences are dispersed so that there are only relatively short stretches of uninterrupted homology (pSWK101 sequence - K. A. Ward, personal communication). If the pSWK101 sequence is representative of all BIIIB2-related genes, then the pSWK18-positive bands in Fig.5.12 reflect only the number of BIIIB3- and BIIIB4-like genes.

From the protein sequence data of Swart et al. (1976), the cDNA clones prepared by K. A. Ward (personal communication) and the genes

isolated from sheep genomic DNA libraries (see also, Chapter 6) there are at least 7 BIIIB3- or BIIIB4-related genes. Do these sequences represent different genes which are present in any individual sheep? The protein studies of Swart et al. (1969) indicated that the BIIIB components 2, 3 and 4 were homogeneous. The chromatographic fraction BIIIB1, which contained protein with a composition very similar but not identical to BIIIB2, and material which eluted at a greater salt concentration than that required for the isolation of BIIIB4 (Fig.3B in Swart et al., 1969) were not characterized and may contain minor proportions of different BIIIB components. Furthermore, the work of Powell et al. (1983) indicated that the same B2A and C high-sulphur sequences were present in Dorset Horn and Lincoln, suggesting that the spectrum of keratin proteins was the same in different breeds of sheep. However, Marshall et al. (1985) have reported that there are significant differences in the number of high-sulphur proteins when the wool from different sheep of the same breed are compared. This may, of course, be a reflection of differential expression of high-sulphur genes rather than an indication of their presence or absence in the genome. Since the cDNA clones, bacteriophage λ genomic clones and the cosmid clone described in this thesis have been isolated from different sheep covering three species, conclusions should not be made concerning the number of functional BIIIB3 and 4 genes in the sheep genome. However, there appears to be at least some conservation of BIIIB sequences between species since the same pseudogene was isolated from Dorset Horn and Merino genomic DNA libraries.

4. The origin of ψ5OBIIIB:

Two basic classes of pseudogenes have been described - referred to as processed and non-processed pseudogenes. Processed pseudogenes appear to have arisen via an RNA intermediate and are generally characterized by a number of features (reviewed by Vanin, 1984). Processed pseudogenes completely lack the intervening sequences found in their functional counterparts; the homology between the pseudogene and

the productive gene does not extend beyond the transcription initiation and termination points, although exceptions have been reported (Vanin, 1984); they have an oligo(A) tract immediately 3' to the transcription termination site; and they are flanked by direct repeats rangings from 9-14 bp in length.

On the basis of these criteria, it is unlikely that ψ5OBIIIB is a processed pseudogene. The data available for a single BIIIB gene (Chapter 6) and three B2 genes (Powell et al., 1983) suggests that high-sulphur genes do not contain introns. Thus, the absence of introns in processed pseudogenes is a criterion that cannot be applied in the case of ψ5OBIIIB.

There is a significant homology of the 5' flanking sequences of ψ5OBIIIB and the functional BIIIB gene described in Chapter 6. This homology extends for almost 60 bp 5' to the TATA box and suggests that ψ5OBIIIB is not derived from the genomic incorporation of an RNA intermediate.

The pseudogene ψ5OBIIIB has six A residues immediately 3' to the presumed transcription termination site (Fig.5.5). The short oligo(A) tract is followed by a 20 bp sequence which is a limited direct repeat of a 22 bp sequence located in the vicinity of the transcription initiation site (Fig.5.5). The degree of homology of the direct repeat is not as high as has been observed in some processed pseudogenes (Vanin, 1984) but is comparable to that reported by Karin and Richards (1982) for a metallothionein pseudogene. (The computer program of Larson, 1982, failed to detect any other direct or inverted repeats in ψ5OBIIIB).

While the presence of a short oligo(A) tract and a limited direct repeat could suggest that ψ5OBIIIB is a processed pseudogene, there are further data which do not support this conclusion. When a gene transcribed by RNA polymerase II is incorporated into the genome by means of a cDNA copy of the mRNA, it lacks transcriptional control sequences. The presence of the consensus TATA and CAAT sequences in ψ5OBIIIB

(Figs.5.5, 5.6) suggest that this pseudogene is not derived from an mRNA transcript. Furthermore, processed pseudogenes are generally derived from genes for which RNA polymerase II transcription occurs in the germ line. For example, of 10 studied β-tubulin genes, 5 are processed pseudogenes (Lee et al., 1983). It is highly unlikely that normal transcription of keratin genes would take place in germ line cells.

However, there are processed pseudogenes which are believed to have arisen by a mechanism which does not involve RNA polymerase II transcription (Vanin, 1984). For one such gene, the $\psi\alpha_3$ globin pseudogene, there is homology with the adult α_1 gene for at least 350 bp 5' to the transcription start site (Vanin, 1984). It has been proposed (Vanin, 1984) that this processed pseudogene has arisen by RNA polymerase III transcription of genes in germ line cells. RNA polymerase III could initiate transcription 5' to the normal transcription start site. If intron splicing and polyadenylation proceeded normally, the processed, aberrant mRNA could then be incorporated into the genomic DNA. While it cannot be ruled out that such a mechanism may have been involved in the formation of ψ50BIIIB, pseudogenes for genes which are not normally transcribed in the germ line usually arise by gene duplication followed by sequence divergence.

Non-processed pseudogenes are generally characterized by a number of features. They are closely linked to the functional gene (in an 8 gene portion of the goat β-globin locus, there are two non-processed pseudogenes - Townes et al., 1984); they retain the same DNA organization as their productive counterpart; and they arise relatively frequently in multigene families.

There are insufficient data available concerning the organization of the BIIIB genes to determine whether ψ50BIIIB is closely linked to other members of this gene family. The restriction map of λSWK50 (Fig.5.2) indicates that there are no other BIIIB genes within 4.7 kb upstream and 7.8 kb downstream of the pseudogene. Data reported in

Chapter 6 of this thesis show that a single, functional BIIIB4-related gene has 0.7 kb of genomic DNA 5' and 12.3 kb 3' to the gene, while a cosmid clone containing approximately 32 kb of genomic DNA contains only 1-2 BIIIB genes. If the BIIIB genes are clustered, it would appear that they are not very closely linked. A high degree of homology in the immediate 5' flanking region between ψ5OBIIIB and a BIIIB4-related gene (Chapter 6) suggests that the pseudogene arose as the result of a gene duplication event and subsequent mutation of one of the duplicated genes and not via an RNA intermediate.

Restriction enzyme mapping and limited DNA sequencing indicate that the same pseudogene is present in the Dorset Horn and Merino breeds. It would be of interest to determine whether other related breeds and species also have maintained this pseudogene. A long evolutionary history has been established for a number of pseudogenes (see for example Harris et al., 1984) although there are no apparent reasons for the conservation of such sequences other than possibly providing DNA for the generation of new genes by gene conversion (Martin et al., 1983).

5. Transcription control sequences in ψ5OBIIIB:

The consensus TATA, CAAT and AATAAA sequences are all conserved in ψ5OBIIIB (Fig.5.5). The 3' non-coding region is characterized by having overlapping polyadenylation signals 140 bp downstream from the termination codon and the sequence AATAAG located in a position homologous to the single AATAAA sequence of pSWK18 (Figs.5.5, 5.9). Higgs et al. (1983) have reported that AATAAG does not function as a normal polyadenylation signal, and while the presence of multiple polyadenylation signals is not unique (see for example Zehner and Paterson, 1983), some active genes contain additional AATAAA hexanucleotides that do not appear to function as signals for the addition of poly(A) tails (Perricaudet et al., 1980).

Powell et al. (1983) reported the presence of a conserved 18 bp sequence in three B2 genes and the BIIIB cDNA clone pSWK101. The

sequence, located immediately 5' to the initiation codon is also present in a high-glycine-tyrosine gene but is not as highly conserved in a BIIIA gene (E. Kuczek, personal communication; Powell and Rogers, 1985). When this region of the ψ50BIIIB sequence is compared with pSWK101 (Fig.5.6) or B2A, there is a greater than 60% homology. While the importance of this sequence is not understood, Powell et al. (1983) have suggested that it may play a role in the regulation of mRNA translation in the keratinocyte.

6. Repetitive DNA sequences in λSWK50:

Screening of λSWK50 with cDNA made to wool follicle poly(A)$^+$ RNA indicated that ψ50BIIIB is the only keratin-related gene in the genomic clone. However, the genomic insert does contain repetitive DNA sequences which are distributed over much of its length (Fig.5.11). Two restriction fragments located 5' to the pseudogene harbour repetitive elements which show homology with sequences (termed NcoI and XhoII elements) associated with the B2 genes (B. C. Powell, personal communication). The XhoII and NcoI sequences (each approximately 110 bp in length) share 25 bp of identity and in 3 out of 5 sequences investigated, both repetitive elements were joined at this region of homology (B. C. Powell, personal communication). Sequences homologous with NcoI and XhoII repetitive elements have been found associated with other genes and copy numbers of up to 10^5 per bovine genome have been reported (Watanabe et al., 1982; Schimenti and Duncan, 1984). The function of these repeats, as for other families of repetitive DNA sequences, remains unknown.

7. The search for a BIIIB gene other than ψ50BIIIB:

During the course of the work described in this chapter, four sheep genomic DNA libraries were screened using pSWK18 as the probe. The washing conditions used during the screenings were non-stringent (1-2xSSPE/0.1%SDS, 55^0-65^{0}C) and the probe should have hybridized to BIIIB3 and BIIIB4 sequences, and possibly some BIIIB2 sequences if the regions

of homology were longer than those found between pSWK18 and pSWK101.

Although Southern analysis of sheep genomic DNA using pSWK18 as probe indicated the presence of 6-12 genes (Fig.5.12), no BIIIB-related recombinants were detected in either the λ1059 or λgt10 libraries. The λgt10 library was constructed using Dorset Horn DNA which had been digested to completion with EcoRI. While Southern analysis indicated the presence of major pSWK18-positive bands in the 4-5 kb size range (Fig.5.12), these fragments were not detected during the library screening.

The Charon 4A genomic DNA library yielded 7 pSWK18-positive clones - all from the same library subfraction and each containing the same genomic fragment. The Charon 28 library contained at least 12 positive clones. One of these, λSWK61 was characterized sufficiently to determine that it comprised DNA homologous to a portion of λSWK50 and that it contained a gene with an in-frame stop codon in a position identical to ψ50BIIIB (Figs.5.13, 5.14, B.1 and B.2).

The Charon 28 recombinants which were pSWK18-positive after a second screen were probed using a mixed synthetic oligonucleotide based on sequences present in three cDNA clones but absent in ψ50BIIIB (Fig.5.7). None of the selected Charon 28 recombinants hybridized with the oligonucleotide probe (Chapter 6). This result indicates that the Charon 28 clones carry either the ψ50BIIIB pseudogene or that portion of the BIIIB gene which does not contain the sequence corresponding to the probe. If the genes have been cleaved by Sau3A (there are 3 potential sites in the coding region although only pSWK101 has a Sau3A site in this region) there should be recombinants which contain the sequence corresponding to the synthetic oligonucleotide. The inability to isolate more than one representative of the BIIIB gene family suggests that there is a selection against these sequences during cloning or propogation in bacteriophage λ vectors.

A number of recent reports have described the difficulties

associated with isolating particular genes from bacteriophage λ genomic DNA libraries. Respess et al. (1983) were unable to obtain any clones containing an intact 7.3 kb EcoRI fragment which harbours most of the human 28S rDNA gene - all the recombinants were found to have deletions in the insert and/or vector DNA. These workers suggested that there may have been a deleterious sequence in the gene-containing fragment which was transcribed from the lac promoter in the right arm of the vector (Charon 16A). Deletion of either the lac promoter or a portion of the insert could lead to stability of the recombinant sequence. The regeneration of parental bacteriophage from plaque-purified recombinants was also observed by Respess et al. (1983) and required the deletion of the entire insert - presumably by the mechanism of recombination.

The poor cloning efficiency of wheat DNA in conventional cloning vectors has been reported to result from recombination events involving regions of internal homology of the genomic DNA. Recombination during bacteriophage propagation in E.coli could change the size of the insert fragment and so exceed the requirements for viable bacteriophage (Murray et al., 1984). The extent of sequence repetition alone did not offer a complete explanation of the poor cloning efficiency, and Murray et al. (1984) suggested that distinct sequence elements dispersed throughout the wheat genome could interfere with bacteriophage propogation. A stable, representative wheat DNA library was established using Charon 32 grown on recA$^-$ hosts, although the library could be also propagated in DP50supF which is not recA$^-$ (Murray et al., 1984). These authors suggested that the stability of wheat DNA was affected by the recBC gene products.

The loss of immunoglobulin V_H genes during isolation and amplification of Charon 4A clones has been observed (Cohen and Givol, 1983), while the deletion of a 5 kb intergenic BamHI fragment from a γ-globin clone has been reported to occur frequently (Fritsch et al., 1980). In the latter case, recombination between homologous sequences

arranged in tandem was proposed as the mechanism by which the deletions occurred (Fritsch et al., 1980).

Deletions of 700-1000 bp have been observed between short direct repeats in lac 1-z fusion strains of E.coli (Albertini et al., 1982). The deletions occurred in both recA$^+$ and recA$^-$ strains although the frequency was greater in the recA$^+$ strains. Similarly, deletions of up to 4 kb between direct repeats as small as 7 bp were observed in a pBR322 recombinant (Jones et al., 1982).

The reason for the low cloning efficiency of BIIIB genes in bacteriophage vectors is unknown. The isolation of only a single clone from a cosmid library (Chapter 6) suggests that the difficulties are not confined to bacteriophage vectors. It is possible that one or more of the mechanisms described above - recombination between homologous sequences, deleterious transcribed sequences or deletions between short direct repeats - could be involved in the poor recovery of BIIIB genes from genomic DNA libraries. Whatever the reason, the mechanism does not operate on the genomic fragment which contains ψ5OBIIIB.

The difficulties encountered in plaque-purifying Charon 28 recombinants which contained pSWK18-positive sequences may provide some insight into the mechanism by which BIIIB genes are lost from genomic DNA libraries. During plaque-purification it was observed that the titre of pSWK18-positive plaques was 10^3 lower than background recombinants and that the DNA yield from preparative experiments was very low. It is possible that recombination or deletion events occurred during amplification and this led to a lowered number of viable recombinant bacteriophage. There was no evidence from plaque purification and restriction mapping of λSWK61 that partial deletions of DNA were occurring (Figs.B.1, B.2).

Although λSWK5O and λSWK61 both contained the pseudogene ψ5OBIIIB, the titre of bacteriophage was much greater in the Charon 4A clone than in the Charon 28 recombinant. This may reflect differences in

the stability of the same sequence in two vectors. On the other hand the lowered titre of λSWK61 may result from DNA sequences beyond those that are found in the isolated recombinant. It is possible that the original clone contained DNA which extended beyond the EcoRI site located upstream of ψ50BIIIB in λSWK50. If this portion of the DNA contained sequences deleterious to bacteriophage viability, then the lowered titre of λSWK61 could reflect the number of recombinants that have lost these sequences by a deletion event.

It is of interest that spontaneous deletions occurred during the growth of λSWK1 which contains two B2 genes, but not during the growth of λSWK2 which has a single B2 gene (B. C. Powell, personal communication). Since both genomic clones contain many copies of the NcoI and XhoII repeats, the major contributing factor leading to the deletions is probably the presence in λSWK1 of two highly homologous genes in close proximity to one another resulting in misalignment of sequences during DNA replication (B. C. Powell, personal communication).

Chapter 6 describes the isolation and characterization of a functional BIIIB gene from the Charon 28 library. It is of interest that the 5' end of the gene is located only 0.7 kb from the end of the insert and it is possible that any closely associated deleterious sequences have been separated from the gene, allowing it to be cloned. If there are sequences which are important in the viability of bacteriophage containing BIIIB genes, it is probable that they are closely associated with the genes. The use of the enzyme Sau3A to generate the fragments used in the construction of the Charon 28 library would be expected to separate the gene from the deleterious sequence if they were not closely associated. (The deleterious sequence suggested previously to be present in the original λSWK61 clone may have been closely associated with another BIIIB gene and the deletion of this sequence led to the loss of the gene as well. Since non-processed pseudogenes are often linked to their functional counterpart it is possible that ψ50BIIIB is downstream

from another BIIIB gene).

CHAPTER 6.

THE CHARACTERIZATION OF A FUNCTIONAL BIIIB HIGH-SULPHUR GENE ISOLATED FROM A CHARON 28 SHEEP GENOMIC DNA LIBRARY

CHAPTER 6

THE CHARACTERIZATION OF A FUNCTIONAL SCMKBIIIB GENE ISOLATED FROM A CHARON 28 SHEEP GENOMIC DNA LIBRARY

A. Introduction.

The work presented in Chapter 5 described the screening of four sheep genomic DNA libraries using pSWK18 as the probe. Two recombinants, λSWK50 and λSWK61, isolated from separate genomic DNA libraries were found to contain the same BIIIB pseudogene, ψ50BIIIB. The pseudogene was characterized and shown to have numerous mutations which would preclude its transcription into a functional BIIIB keratin-coding mRNA. A major feature of ψ50BIIIB was a 23 bp deletion situated near the 3' end of the coding region. A mixed oligonucleotide probe, based on the sequence deleted from the pseudogene, was synthesized and used to rescreen the Charon 28 sheep genomic DNA library.

B. Results.

1. Use of a mixed, synthetic 20-mer probe to isolate a BIIIB gene:

(a) Synthesis of the mixed oligonucleotide:

The pseudogene ψ50BIIIB has a 23 bp deletion near the 3' end of the coding region. The DNA sequences in this region are known for 3 cDNA clones - pSWK18, pSWK31 and pSWK101 (K. A. Ward, personal communication) - and these are shown in Fig.6.1. The mixed oligonucleotide, based on the sequences of residues 1-20 (Fig.6.1) was synthesized by Dr. D. Skingle, University of Adelaide, using the solid-phase phosphite method (Chapter 4) and is referred to as the 20-mer although it consisted of 64, 20 residue sequences.

(b) Hybridization and washing conditions for use of the 20-mer probe:

To determine suitable conditions for the use of the 20-mer probe in screening for BIIIB genes, a series of experiments was carried out using pSWK18 as the target DNA. The plasmid was spotted onto

FIGURE 6.1. SEQUENCES OF A 23bp OLIGONUCLEOTIDE PRESENT IN BIIIB cDNA CLONES BUT ABSENT IN ψ5OBIIIB.

A 23bp sequence which is present in BIIIB cDNA clones but which is absent in the pseudogene ψ5OBIIIB is shown. The sequence of the mRNA strand is shown in the 5' to 3' direction. The full sequence is that of pSWK18 and only those residues of pSWK31 and pSWK101 which differ from pSWK18 are shown. The synthetic oligonucleotides were based on the sequences between residues 1 and 20.

```
          5'                  10                  20          3'
pSWK18    AC.ACT.CAG.TCC.AGC.TGT.GAG.CCC.
pSWK31                        C
pSWK101   T   T               C T C               A AAT
```

nitrocellulose filters, and following prehybridization at 67°C and hybridization at 42°C (Chapter 4), the filters were washed at increasing temperatures to disrupt non-specific hybrids. It was found that a strong signal was still obtained following a 1-2 min wash at approximately 55°C, and under these conditions there was no detectable hybridization to control spots of pBR322, λSWK50 or λSWK61 (data not shown).

(c) Re-screening of pSWK18-positive, Charon 28 recombinants using the 20-mer as probe:

The 12 Charon 28 recombinants which were positive after a second screen using pSWK18 as probe (Chapter 5) were re-screened with kinased 20-mer. While the clones were still pSWK18-positive, none showed any significant hybridization with the 20-mer (data not shown).

(d) Re-screening of the Charon 28 genomic DNA library:

Approximately $2x10^5$ plaques from the Charon 28 genomic DNA library were screened using pSWK18 and the 20-mer as probes on duplicate filters. Only one plaque hybridized with both probes (Fig.6.2). The clone was plaque-purified using both pSWK18 and 20-mer probes at each screening. While the initial library screening had used a washing temperature of 52°C for the 20-mer probe, during the plaque purification, temperatures of 55-57°C were used without significantly reducing the hybridization (data not shown). The recombinant isolated from the genomic DNA library using the 20-mer probe in conjunction with the pSWK18 cDNA probe was referred to as λSWK96.

2. Restriction enzyme mapping of λSWK96:

The details of restriction mapping and subcloning of λSWK96 are presented in Appendix C. The restriction map of the 13.8 kb insert of λSWK96 is shown in Fig.6.3. The overlapping subclones pλ96/0.7 and pλ96/0.7A which contain the BIIIB gene and which are referred to in the following section are derived from HindIII and PstI fragments respectively, as shown in Fig.6.3.

FIGURE 6.2. SCREENING OF THE CHARON 28 SHEEP GENOMIC DNA LIBRARY USING AS PROBES pSWK18 AND THE 20-mer.

(a) Autoradiography of a nitrocellulose filter prepared from a 15cm plate containing approximately $2x10^4$ plaques from a Charon 28 sheep genomic DNA library. Following hybridization for 22h at 42°C using nick-translated pSWK18 insert DNA as probe, the filter was washed in 2xSSPE/0.1% SDS at 65°C (as described in Chapters 2 and 4). Exposure was at -80°C for 16h using intensification screens.

(b) Duplicate filter to (a). The filter was prehybridized at 67°C and then hybridized for 20h at 42°C using kinased 20-mer as probe. Following hybridization, the filter was washed 3 times, for 10min each in 6xSET/0.5% SDS at 20°C, and then for 75s at 52°C in the same buffer (Chapter 4). Exposure was at -80°C for 72h using intensification screens.

The arrows point to the location of the DNA spot which hybridized with both probes.

(a)

(b)

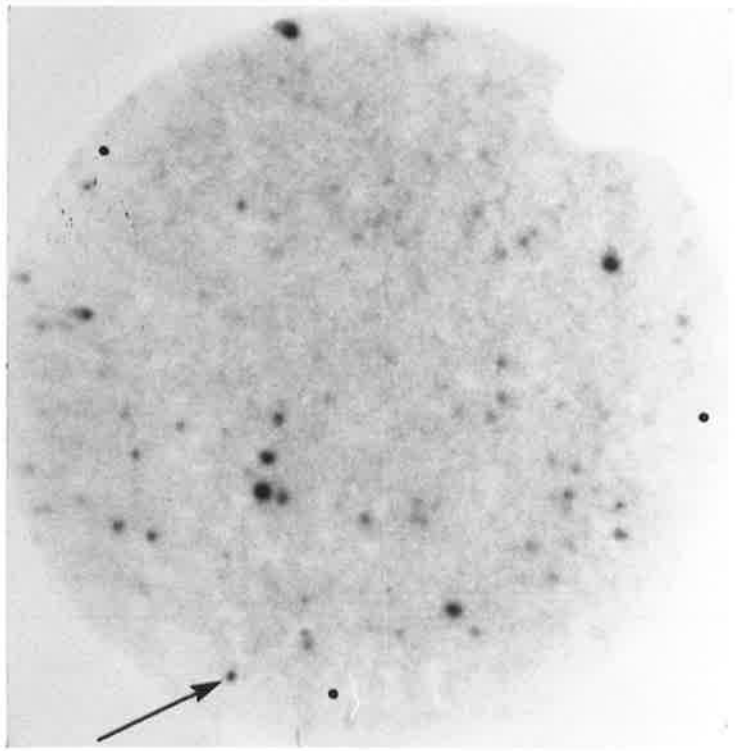


FIGURE 6.3. RESTRICTION MAP OF λSWK96.

Restriction map of λSWK96 based on data presented in Appendix C. The location and direction of transcription of the BIIIB gene is shown by the arrow above the map.

V_L and V_R represent the Charon 28 left and right arms respectively.

H: HindIII sites (the orientation of the sites shown by the sloped lines were not determined).

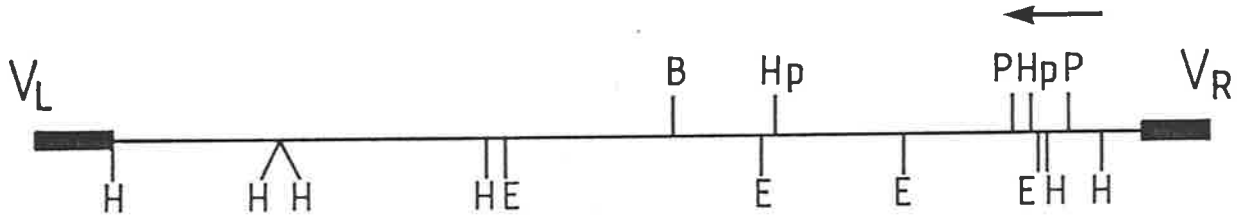
E: EcoRI.

B: BamHI.

Hp: HpaI.

P: PstI (only those sites used in obtaining the subclone pλ96/0.7A are shown).

λSWK96



= 1 kb

1cm

3. DNA sequence determination of the λSWK96 BIIIB gene (λ96BIIIB):

The strategy used to sequence the λSWK96 BIIIB gene (λ96BIIIB) is shown in Fig.6.4. Although the sequence was determined using two overlapping subclones - pλ96/0.7 and pλ96/0.7A - the data is combined in Fig.6.4. Restriction fragments were made flush-ended where required and cloned into the appropriate sites in M13mp8 and 9. There was sufficient sequence homology between the coding sequences and portions of the 3' non-coding sequences of λ96BIIIB and ψ50BIIIB to allow the orientation of M13 subclones to be determined by the annealing/electrophoresis technique described in Chapter 4. The subclones mp8λ50/2.3 and mp9λ50/2.3 (Fig.A.3c) were used for this analysis of the λ96BIIIB subclones.

The complete nucleotide sequence of λ96BIIIB, from 288 bp 5' to the initiation codon to 462 bp downstream from the termination codon, is shown in Fig.6.5. The restriction sites used to prepare DNA fragments for M13 subcloning and sequencing experiments are shown. The gene has a TATA box located approximately 90 bp 5' to the initiation codon site and a possible CAAT sequence a further 60 bp upstream - analogous to their positions in ψ50BIIIB (Fig.5.5). Corden et al. (1980) and Breathnach and Chambon (1981) surveyed the transcription start site for a number of genes and observed that the first nucleotide (an A) was located 26-34 bp downstream from the start of the TATA box and was surrounded by pyrimidines. Two potential transcription start sites for λ96BIIIB are shown in Fig.6.5.

The λ96BIIIB coding region of 291 bp is followed by approximately 360 bp of 3' non-coding sequence. The precise poly(A) addition site is unknown - pSWK31, a cDNA clone which is homologous with the λ96BIIIB gene did not have a poly(A) tail.

Interestingly, at the 3' end of the sequence there are CAAT and TATA boxes, approximately 60 bp apart, and a potential transcription start site 24 bp downstream from the beginning of the TATA box (Fig.6.5). If these data do indicate the presence of another gene in λSWK96, then it

FIGURE 6.4. STRATEGY USED FOR DETERMINING THE NUCLEOTIDE SEQUENCE OF λ96BIIIB.

Sequencing strategy used for the analysis of the λ96BIIIB gene. Only those restriction sites relevant to the sequencing strategy are shown (H: HindIII, D: DdeI, P: PstI, Pv: PvuII and E: EcoRI). Numbering (bp) is from the initiation codon. The shaded bar on the map represents the location of the BIIIB coding region. The arrow above the map shows the 5' to 3' direction of the mRNA strand.

⟼ The line represents the extent of sequence determined from a particular restriction fragment. The vertical line and arrow point show respectively the 5' and 3' extremities of the sequence determined.

⟻⟼ Restriction fragments completely sequenced in both directions.

The 'chain-termination' method of Sanger et al. (1980) using commercial primers was used for the sequence determination.

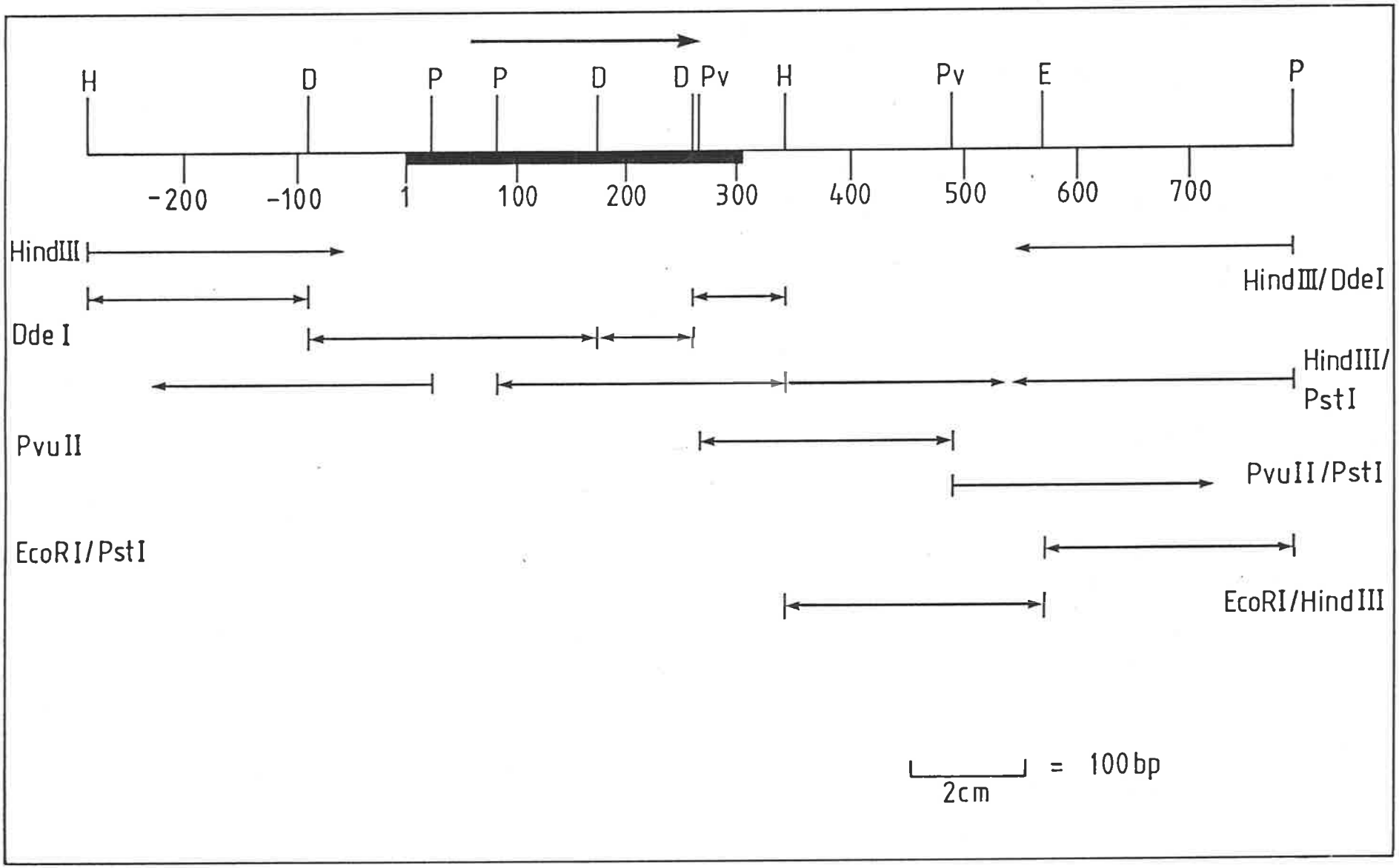

H
D
P
P
D
D
Pv
H
Pv
E
P
-200
-100
1
100
200
300
400
500
600
700
HindIII
HindIII/DdeI
Dde I
HindIII/
PstI
PvuII
PvuII/PstI
EcoRI/PstI
EcoRI/HindIII
= 100bp
2cm

FIGURE 6.5. THE COMPLETE NUCLEOTIDE SEQUENCE OF λ96BIIIB.

The nucleotide sequence of the mRNA strand is shown in the 5' to 3' direction. The restriction sites used for subcloning and sequencing experiments are shown by a line under the sequence and the name of the relevant enzyme above the sequence

----- indicates a possible CAAT box.

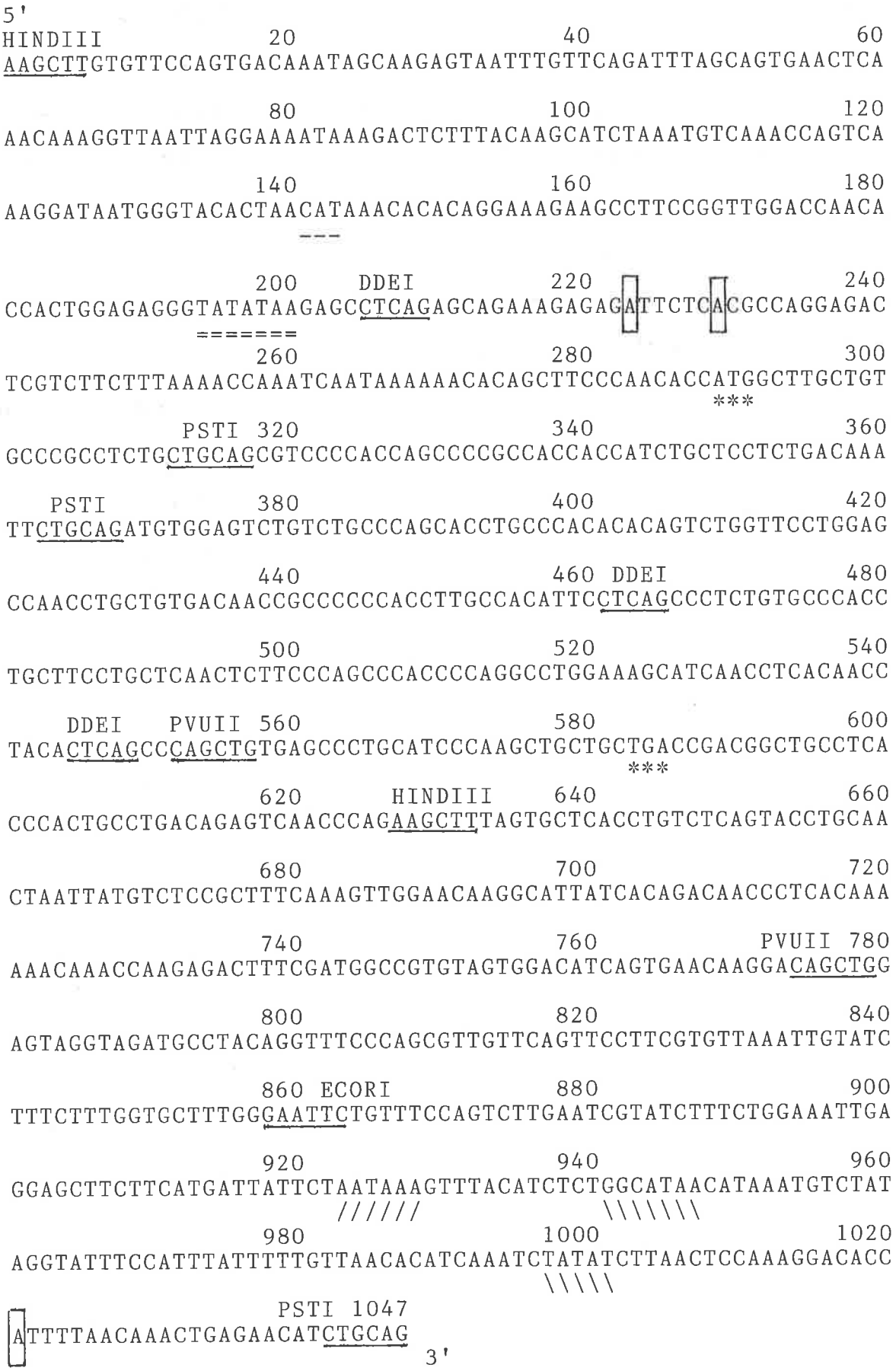
===== shows the consensus TATA sequence.

///// indicates the poly(A)-addition signal sequence.

*** shows the location of the initiation and termination codons.

\\\\\ indicates the consensus CAAT and TATA sequences for another gene downstream from the BIIIB gene.

The possible cap sites are boxed.


5'
HINDIII
20
40
60
AAGCTTGTGTTCCAGTGACAAATAGCAAGAGTAATTTGTTCAGATTTAGCAGTGAACTCA
80
100
120
AACAAAGGTTAATTAGGAAAATAAAGACTCTTTACAAGCATCTAAATGTCAAACCAGTCA
140
160
180
AAGGATAATGGGTACACTAACATAAACACACAGGAAAGAAGCCTTCCGGTTGGACCAACA
200
DDEI
220
240
CCACTGGAGAGGGTATATAAGAGCCTCAGAGCAGAAAGAGAGATTCTCACGCCAGGAGAC
260
280
300
TCGTCTTCTTTAAAACCAAATCAATAAAAAACACAGCTTCCCAACACCATGGCTTGCTGT
PSTI
320
340
360
GCCCGCCTCTGCTGCAGCGTCCCCACCAGCCCCGCCACCACCATCTGCTCCTCTGACAAA
PSTI
380
400
420
TTCTGCAGATGTGGAGTCTGTCTGCCCAGCACCTGCCCACACACAGTCTGGTTCCTGGAG
440
460
DDEI
480
CCAACCTGCTGTGACAACCGCCCCCCACCTTGCCACATTCCTCAGCCCTCTGTGCCCACC
500
520
540
TGCTTCCTGCTCAACTCTTCCCAGCCCACCCCAGGCCTGGAAAGCATCAACCTCACAACC
DDEI
PVUII
560
580
600
TACACTCAGCCCAGCTGTGAGCCCTGCATCCCAAGCTGCTGCTGACCGACGGCTGCCTCA
620
HINDIII
640
660
CCCACTGCCTGACAGAGTCAACCCAGAAGCTTTAGTGCTCACCTGTCTCAGTACCTGCAA
680
700
720
CTAATTATGTCTCCGCTTTCAAAGTTGGAACAAGGCATTATCACAGACAACCCTCACAAA
740
760
PVUII
780
AAACAAACCAAGAGACTTTCGATGGCCGTGTAGTGGACATCAGTGAACAAGGACAGCTGG
800
820
840
AGTAGGTAGATGCCTACAGGTTTCCCAGCGTTGTTCAGTTCCTTCGTGTTAAATTGTATC
860
ECORI
880
900
TTTCTTTGGTGCTTTGGGAATTCTGTTTCCAGTCTTGAATCGTATCTTTCTGGAAATTGA
920
940
960
GGAGCTTCTTCATGATTATTCTAATAAAGTTTACATCTCTGGCATAACATAAATGTCTAT
980
1000
1020
AGGTATTTCCATTTATTTTTGTTAACACATCAAATCTATATCTTAACTCCAAAGGACACC
PSTI
1047
ATTTTAACAAACTGAGAACATCTGCAG
3'

is located to the left (i.e.3') of the BIIIB sequence in the genomic fragment (Fig.6.3).

4. Comparison of λ96BIIIB with BIIIB cDNAs, ψ50BIIIB and other keratin gene sequences:

(a) 5' flanking sequence:

Figure 6.6a shows the flanking sequence of λ96BIIIB. A comparison of this sequence with that of the pseudogene indicates that there is a significant homology 3' to the CAAT box - this region is compared in Fig.6.6a.

A search for homology between the 5' flanking sequences of λ96BIIIB and other keratin genes has revealed two regions of sequence conservation (Powell and Rogers, 1985). The λ96BIIIB and B2A genes share approximately 65% homology over a 100 bp sequence located in different 5' flanking positions (Fig.6.6b), while there is a 17 bp sequence which is conserved to between 65 and 85% in four high-sulphur genes and which is located 160-195 bp 5' to the TATA box (Fig.6.6c).

(b) 5' non-coding sequence:

The sequence homologies in the 5' non-coding region between λ96BIIIB and 7 other sequences are shown in Fig.6.7a. With the exception of pSWK101, for which the complete 5' non-coding sequence is unknown, the sequences from the initiation codon back to the possible cap sites are shown (data are from Fig.5.6; Powell et al., 1983; E. Kuczek, B. C. Powell and K. A. Ward, personal communications).

If the sequences are compared 3' to residue 21, where all the genes are represented, there are a number of features which are apparent (Fig.6.7a). There are only 4 positions (apart from the initiation codon) at which a common base is found in each sequence (residues 36, 69, 77 and 78). There are however, 21 positions where only one of the sequences differs from the rest. As observed by Powell et al. (1983), there is a highly conserved sequence of 18 residues immediately 5' to the initiation codon. The λ96BIIIB sequence conforms to this pattern, although not to

FIGURE 6.6. SEQUENCE HOMOLOGIES IN THE 5' FLANKING REGION BETWEEN λ96BIIIB AND OTHER KERATIN GENES.

(a) The 5' flanking region of λ96BIIIB is shown in full. For comparison, the sequence of ψ50BIIIB 3' to the start of the CAAT sequence (indicated by *) is given, but only where it differs from λ96BIIIB. Deletions (-) have been introduced to maximize the homology.

96: λ96BIIIB.

50: ψ50BIIIB.

(b) The 5' 107 bp of λ96BIIIB are shown in full. The sequence of B2A (Powell et al., 1983) is shown only where it differs from λ96BIIIB. The B2A sequence covers the region 314-415 bp 5' to the initiation codon while the region 184-291 bp 5' to the initiation codon of λ96BIIIB is shown. Deletions (-) have been introduced to maiximize the homology. This figure is based on Fig.11a in Powell and Rogers (1985).

96: λ96BIIIB.

(c) The consensus sequence is shown in full and is derived from the four sequences below it. Only residues which differ from the consensus sequence are shown. The positions of the starting residues are numbered negatively with respect to the initiation codon. The approximate distances to the TATA box and initiation codon are also shown. This figure is based on Fig.11b in Powell and Rogers (1985).

96: λ96BIIIB.

(a)

```
   5'                  20                  40                  60
96 AAGCTTGTGTTCCAGTGACAAATAGCAAGAGTAATTTGTTCAGATTTAGCAGTGAACTCA

                       80                  100                 120
96 AACAAAGGTTAATTAGGAAAATAAAGACTCTTTACAAGCATCTAAATGTCAAACCAGTCA

                     140                 160                 180
96 AAGGATAATGGGTACACTAACATAAACACACAGGAAAGAAGC-CTTCCGGTTGGACCAAC
50                    *     T A A A  G  G AAA   TT   -

                     200                 220     229
96 ACCACTGGAGAG-GG-TATATAA-GAGCCTCAGAGC-AGAAAG-AGAGA
50    TGAT  -  C  A       A     C  A  TG   -- G   C
                                                   3'
```

(b)

```
   5'                  20                  40                  60
96 AAGCTTGTGTTCCAGTGACAAA-TAGCAAGAGTAATTTGTTCAGATTTAG-CAGTGAACT
B2A  GATCCT AA TG A    C CT   G   G -  --       -     A  ---

                       80                  100       111
96 CAAAC-AAAGGTTAATT-AGGAAAATAAAGACTCTTTACAAGCATCTAAAT
B2A  TGC T     C  TCAA    C T T T   --AAAG     CAGGGC A
                                                   3'
```

(c)

```
          A                   A
       5'-TTGTTGAGTATTAGCTG-3'

   96 -255      C  AT
  B2A -278    A     C  G   T  ]  160-195bp to ___ 75-90bp to
  B2C -246    G                    TATA box       ATG
BIIIA -254   T    TG  G
```

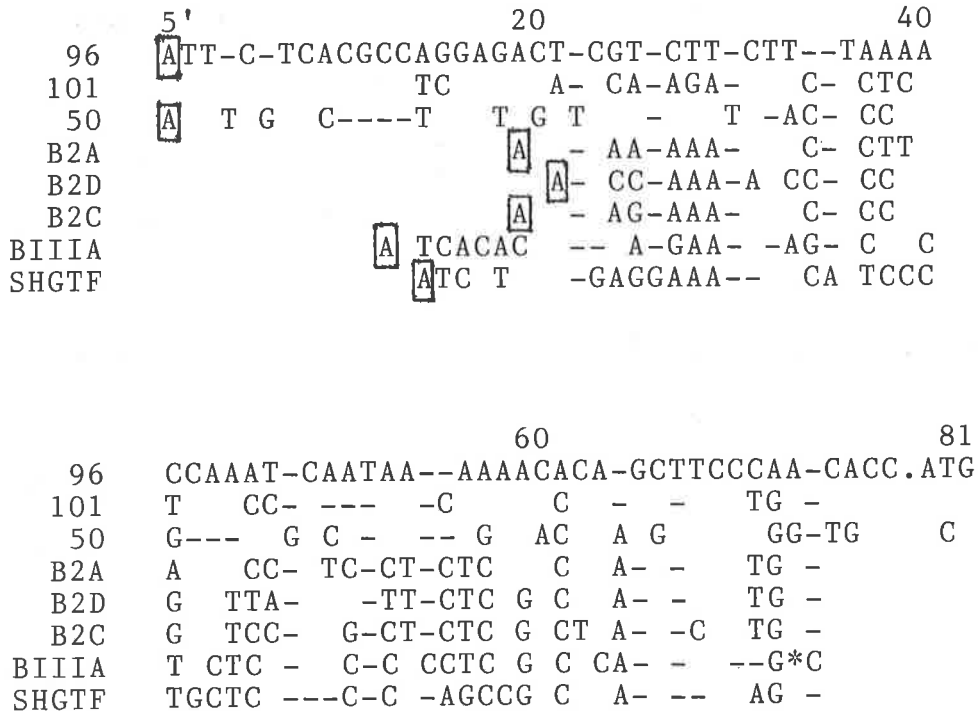
FIGURE 6.7. SEQUENCE HOMOLOGIES IN THE 5' NON-CODING REGION BETWEEN λ96BIIIB AND OTHER KERATIN GENES.

(a) The 5' non-coding region of λ96BIIIB is shown in full. Other sequences are shown only where they differ from λ96BIIIB. Deletions (-) have been introduced to maximize homology. In the BIIIA sequence, a 9 residue sequence, 5'-GCCCCACAC-3', has been deleted from the site indicated by *. The possible mRNA cap sites are boxed. The data for the B2, BIIIA and high-glycine-tyrosine genes were obtained from Powell et al. (1983), Powell and Rogers (1985) and from E. Kuczek and B.C. Powell, personal communications.

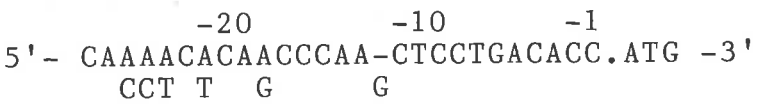
96: λ96BIIIB; 101: pSWK101; 50: ψ50BIIIB; SHGTF: High-glycine-tyrosine gene F.

(b) Consensus sequence derived from the eight sequences in (a). The sequence is numbered negatively with respect to the initiation codon. A single, possible deletion at -12 is indicated by a dash.

(a)


5'
20
40
96 ATT-C-TCACGCCAGGAGACT-CGT-CTT-CTT--TAAAA
101 TC A- CA-AGA- C- CTC
50 A T G C----T T G T - T -AC- CC
B2A A - AA-AAA- C- CTT
B2D A- CC-AAA-A CC- CC
B2C A - AG-AAA- C- CC
BIIIA A TCACAC -- A-GAA- -AG- C C
SHGTF ATC T -GAGGAAA-- CA TCCC
60
81
96 CCAAAT-CAATAA--AAAACACA-GCTTCCCAA-CACC.ATG
101 T CC- --- -C C - - TG -
50 G--- G C - -- G AC A G GG-TG C
B2A A CC- TC-CT-CTC C A- - TG -
B2D G TTA- -TT-CTC G C A- - TG -
B2C G TCC- G-CT-CTC G CT A- -C TG -
BIIIA T CTC - C-C CCTC G C CA- - --G*C
SHGTF TGCTC ---C-C -AGCCG C A- -- AG -

(b)


-20
-10
-1
5'- CAAAACACAACCCAA-CTCCTGACACC.ATG -3'
CCT T G G

the same degree as pSWK101 which differs at only one position from B2A (Fig.6.7a; Powell et al., 1983). The conserved nature of the region allows a consensus sequence to be defined which extends 27 bp 5' to the initiation codon (Fig.6.7b). For the BIIIA sequence, optimal homology is obtained by deletion of a 9 residue sequence, 5'-GCCCCACAC-3', which is almost a perfect direct repeat of the sequence immediately 5' to it (Fig.6.7b; B. C. Powell, personal communication).

When the 5' non-coding sequences of the keratin genes are compared with λ96BIIIB, the homologies range from 40-50% for the high-glycine-tyrosine (SHGTF), and BIIIA and B2 high-sulphur genes, up to 65% for the BIIIB sequences. If the deletions are disregarded, the homologies increase by approximately 10%.

(c) Coding region:

In Fig.6.8 the coding sequence of λ96BIIIB is compared with the available data for BIIIB3-like (pSWK18) and BIIIB4-like (pSWK31) cDNA clones. The comparison with pSWK18 indicates that the gene and cDNA sequences are greater than 97% homologous. Of the 6 nucleotide differences, two are third base changes which do not alter the amino acid. Another three differences change the amino acids to those found in the BIIIB4 protein but not in BIIIB3 (Fig.6.9), while the last, at residue 242, changes a His in pSWK18 - an unusual amino acid for this position - to Leu in λ96BIIIB, the residue normally found in BIIIB components (Fig.6.9). The nucleic acid sequence data available for pSWK31 is limited to the 3' non-coding region and 48 bp of coding sequence. There is only one nucleotide difference between λ96BIIIB and pSWK31 in the coding region and this does not lead to an amino acid change (Fig.6.8).

Table 6.1 compares the codon usage of three keratin genes - λ96BIIIB, B2A and SHGTF with pooled liver genes. Taking into account the unusual nature of the amino acid compositions of the keratin proteins, the general codon usage follows the trends found for the liver genes. Of

FIGURE 6.8. COMPARISON OF THE λ96BIIIB CODING REGION WITH pSWK18 AND pSWK31.

The mRNA strand sequence of the λ96BIIIB coding region is shown in the 5' to 3' direction. The sequence is compared with pSWK18 and pSWK31 which are shown only where they differ from λ96BIIIB.

* The 5' terminal bases of pSWK18 and pSWK31.

96: λ96BIIIB.

18: pSWK18.

31: pSWK31.

```
   5'                  15                  30                  45
96 GCT.TGC.TGT.GCC.CGC.CTC.TGC.TGC.AGC.GTC.CCC.ACC.AGC.CCC.GCC.

                      60                  75                  90
96 ACC.ACC.ATC.TGC.TCC.TCT.GAC.AAA.TTC.TGC.AGA.TGT.GGA.GTC.TGC.
18                                        *

                     105                 120                 135
96 CTG.CCC.AGC.ACC.TGC.CCA.CAC.ACA.GTC.TGG.TTC.CTG.GAG.CCA.ACC.
18       T                                   G

                     150                 165                 180
96 TGC.TGT.GAC.AAC.CGC.CCC.CCA.CCT.TGC.CAC.ATT.CCT.CAG.CCC.TCT.
18                                      A     G

                     195                 210                 225
96 GTG.CCC.ACC.TGC.TTC.CTG.CTC.AAC.TCT.TCC.CAG.CCC.ACC.CCA.GGC.
18

                     240                 255                 270
96 CTG.GAA.AGC.ATC.AAC.CTC.ACA.ACC.TAC.ACT.CAG.CCC.AGC.TGT.GAG.
18                       A                       T
31                          * T

                     285     291
96 CCC.TGC.ATC.CCA.AGC.TGC.TGC.
18                         .
31                         .
                             3'
```

TABLE 6.1

CODON USAGE IN KERATIN GENES COMPARED WITH LIVER GENES

The numbers represent the frequency of a particular codon.

A-BIIIB4A ;B-B2A (Powell et al., 1983) ;C-SHGTF(high-glycine -tyrosine gene F, E.Kuczek,personal communication) ;D-Pooled liver genes (Hastings and Emerson, 1983).

	A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
TTT Phe	0	1	1	17	TCT Ser	3	0	1	14	TAT Tyr	0	2	3	25	TGT Cys	4	10	1	32
TTC	3	1	5	27	TCC	2	9	2	24	TAC	1	2	8	26	TGC	13	29	3	41
TTA Leu	0	0	0	8	TCA	0	0	0	9	TAA	Termination				TGA	Termination			
TTG	0	0	0	12	TCG	0	0	0	2	TAG	Termination				TGG Trp	0	1	2	21
CTT Leu	0	0	0	15	CCT Pro	2	2	0	17	CAT His	0	0	0	10	CGT Arg	0	0	0	6
CTC	3	1	1	19	CCC	9	8	0	25	CAC	2	0	0	22	CGC	2	5	0	9
CTA	0	0	0	3	CCA	5	5	3	18	CAA Gln	0	1	0	15	CGA	0	0	1	6
CTG	4	2	1	41	CCG	0	3	1	7	CAG	3	15	0	38	CGG	0	0	0	10
ATT Ile	1	3	0	13	ACT Thr	1	1	0	20	AAT Asn	0	0	0	22	AGT Ser	0	1	2	7
ATC	3	6	0	22	ACC	8	16	2	31	AAC	3	0	1	35	AGC	6	14	3	18
ATA	0	0	0	9	ACA	2	1	0	24	AAA Lys	1	0	0	45	AGA Arg	1	0	1	22
ATG Met	0	0	0	26	ACG	0	0	0	9	AAG	0	0	0	66	AGG	0	1	0	13
GTT Val	0	1	0	12	GCT Ala	1	1	1	28	GAT Asp	0	0	0	31	GGT Gly	0	3	4	16
GTC	4	0	1	17	GCC	2	2	1	40	GAC	3	1	0	36	GGC	1	8	9	29
GTA	0	0	0	12	GCA	0	0	0	21	GAA Glu	1	0	0	45	GGA	1	3	0	26
GTG	0	4	2	39	GCG	0	2	0	4	GAG	2	5	0	51	GGG	0	1	1	8

interest is the 3:1 usage of the TGC:TGT cystine codons and the strong bias for some codons over others; for example, TCC and AGC (ser), CAG (Gln) and ACC (Thr).

(d) Amino acid sequence coded by λ96BIIIB (BIIIB4A):

In Fig.6.9, the amino acid sequence of the wool protein BIIIB4 is compared with other members of this keratin family and the protein coded by λ96BIIIB (referred to as BIIIB4A). The BIIIB4A component differs from BIIIB4 at only two positions - residue 43 is Glu and 48 is not present in BIIIB4A compared with Gln and Cys respectively in BIIIB4. It is of interest that the differences found in BIIIB4A are also observed in pSWK18 and ψBIIIB3A (Chapter 5). At all other residues, the protein coded by λ96BIIIB is identical to BIIIB4 and can be considered as a variant of this component.

(e) 3' non-coding region:

The 3' non-coding region of λ96BIIIB is approximately 370 bp in length and has a greater than 95% homology with pSWK31 (Fig.6.10). When the 3' non-coding region of λ96BIIIB is compared with other BIIIB sequences two stretches of homology can be found (Fig.6.11). The first region of homology between λ96BIIIB, pSWK31, pSWK18 and ψ50BIIIB occurs over 50 bp immediately 3' to the termination codon (Fig.6.11a). Compared with λ96BIIIB, the homology with pSWK31 is 95%, pSWK18 - 90% and ψ50BIIIB - 80% (disregarding deletions, which reduce the homology with the BIIIB3-related sequences by approximately 10%).

The second region of homology stretches for approximately 70 bp and is in the vicinity of the consensus AATAAA sequence (Fig.6.11b). Compared with λ96BIIIB, the homology with pSWK31 is greater than 90%. The homology with pSWK18 and ψ50BIIIB is approximately 75% when the deletions are disregarded. When the deletions are taken into account the homology is reduced by 10-15%.

Over the two regions of homology, the BIIIB4-like sequences are more closely related to each other than to the BIIIB3-like sequences.

FIGURE 6.9. AMINO ACID SEQUENCE OF BIIIB4 COMPARED WITH BIIIB2, BIIIB3 AND BIIIB4A.

The amino acid sequence of the wool protein BIIIB4 is presented in full (Swart et al., 1976). The sequences of BIIIB2 and 3 (Swart et al., 1976) and BIIIB4A (the protein coded by λ96BIIIB, Fig.6.8) are shown only where they differ from BIIIB4.

2,3 and 4 refer to BIIIB2, BIIIB3 and BIIIB4 respectively.

4A refers to the protein BIIIB4A.

# indicates that the proteins sequenced by Swart et al. (1976) had an acetylated N-terminus.

```
                          5                   10                  15
2     #                 PRO ARG                 ARG     GLY
3
4     ALA CYS CYS ALA ARG LEU CYS CYS SER VAL PRO THR SER PRO ALA
4A

                          20                  25                  30
2
3
4     THR THR ILE CYS SER SER ASP LYS PHE CYS ARG CYS GLY VAL CYS
4A

                          35                  40                  45
2                                       ASN ILE SER LEU
3                                               LEU
4     LEU PRO SER THR CYS PRO HIS THR VAL TRP PHE LEU GLN PRO THR
4A                                                  GLU

                          50                  55                  60
2             ---           SER     VAL         VAL TYR     ASP THR
3                                           TYR     VAL
4     CYS CYS CYS ASP ASN ARG PRO PRO PRO CYS HIS ILE PRO GLN PRO
4A            ---

                          65                  70                  75
2     TYR                                         HIS
3
4     SER VAL PRO THR CYS PHE LEU LEU ASN SER SER GLN PRO THR PRO
4A

                          80                  85                  90
2             SER GLY                     PHE ILE         GLY
3                                                   SER
4     GLY LEU GLU SER ILE ASN LEU THR THR TYR THR GLN PRO SER CYS
4A

                          95                  99
2        ASN VAL      GLU      ARG      ---
3
4     GLU PRO --- CYS ILE PRO SER CYS CYS
4A
```

FIGURE 6.10. COMPARISON OF THE 3' NON-CODING REGIONS OF λ96BIIIB AND pSWK31.

The nucleotide sequence of the mRNA strand is shown in the 5' to 3' direction. Deletions (-) have been introduced to maximize the homology. The pSWK31 sequence is shown only where it differs from λ96BIIIB.

===== indicates the consensus AATAAA sequence.

96: λ96BIIIB.

31: pSWK31.

```
   5'                  20                  40                  60
96 TGACCGACGGCTG-CCTCACCCACTGCCTGACAGAGTCAACCCAGAAGCTTTAGTGCTCA
31              -A          G

                       80                  100                 120
96 CCTGTCTCAGTACCTGCAACTAATTATGTCTCCGCTTTCAAAGTTGGAACAAGGCATTAT
31

                   140                 160                 180
96 CACAGACAACCCTCACAAAAAACAAACCAAGAGACTTTCGATGGCCGTGTAGTGGACATC
31                                                 C

                   200                 220                 240
96 AGTGAACAAGGACAGCTGGAGTAGGTAGATGCCTACAGGTTTCCCAGCGTTGTTCAGTTC
31                                                   A

                   260                 280                 300
96 CTTCGTGTTAAATTGTATCTTTCTTTGGTGCTTTGGGAA-TTCTGTTTCCAGTCTTGAAT
31 G   A                                     A   -

                   320                 340   ======        360
96 CGTATCTTTCTGGAAATTGAGGAGCTTCTTCATGATTATTCTAATAAAGTTTACATCTCT
31 G                 A     A                     -

       368
96 GGCATAAC
31   T    T
          3'
```

FIGURE 6.11. SEQUENCE HOMOLOGIES IN THE 3' NON-CODING REGION BETWEEN λ96BIIIB AND OTHER BIIIB SEQUENCES.

(a) The 3' non-coding sequence of λ96BIIIB immediately downstream from the termination codon is shown in full. Other sequences are shown only where they differ from λ96BIIIB. Deletions (-) have been introduced to maximize the homology.

(b) The 3' non-coding sequence of λ96BIIIB surrounding the AATAAA sequence is shownin full. Other sequences are shown only where they differ from λ96BIIIB. Deletions (-) have been introduced to maximize the homology.

===== indicates the consensus AATAAA sequence.

96: λ96BIIIB.

31: pSWK31.

18: pSWK18.

50: ψ50BIIIB.

(a)

```
   5'                  20                  40          52 3'
96 TGACCGACGGCTG-CCTCACCCACTGCCTGACAGAGTCAACCCAGAAGCTTT
31              -A          G
18               A -    T   G       T - - -   A         A
50  A  T  T TT   A -    T   G     G   - - -   A         A
```

(b)

```
   5'                320                 340          360
96 CGTATCTTTC-T---GGAAATTGAGGAGCTTCTTCATGATTATTC-TAATAAAGTTTACA
31 G                      A    A                    --
18 G C  -    A AGA       --A T- A   -  G   G  -   C-      C   GTT
50 G -  -    A AGA    G --A T- A-  -  G   G  -   C      GC A--TT
```

```
             373  3'
96 TCTCTGGCATAAC
31        T    T
18    G    POLYA
50    G    AAAAA
```

Also, the BIIIB3-like sequences show a greater homology to each other than to the BIIIB4-like sequences (Figs.6.11a and b). For the 3' non-coding sequence between the two regions of homology, there is no significant similarity when λ96BIIIB and pSWK31 are compared with pSWK18 and ψ50BIIIB (data not shown). Nor is there any homology between these sequences and pSWK101 (a BIIIB2-related cDNA clone) over the entire 3' non-coding region (data not shown).

5. Probing λSWK96 for other keratin genes:

In order to ascertain whether λSWK96 contained any other keratin genes, restriction fragments of the genomic clone were probed with cDNA made to wool follicle poly(A)$^+$ mRNA. Only those fragments which contained the BIIIB gene hybridized with the probe (data not shown). Probing with specific keratin-coding sequences was not undertaken.

6. Isolation of a cosmid clone containing BIIIB sequences:

A pHC79 sheep genomic DNA cosmid library, consisting of approximately 10^5 recombinants, was screened using pSWK18 insert as probe. A single positive colony was detected, purified through a number of colony screenings, and the DNA prepared. The clone, pcosBIIIB1, was digested with EcoRI, HindIII or BamHI and the fragments probed with pSWK18 and wool follicle cDNA (Fig.6.12). BamHI digestion results in very few fragments, while with EcoRI cleavage there are 11 fragments visible following electrophoresis (Fig.6.12a). The EcoRI fragments range in size from about 0.3 kb up to approximately 9.5 kb and give a total clone size of about 38 kb. This is in agreement with the size estimated from the HindIII fragments. The cosmid vector comprises 6 kb, indicating that the genomic insert in pcosBIIIB1 is approximately 32 kb in size.

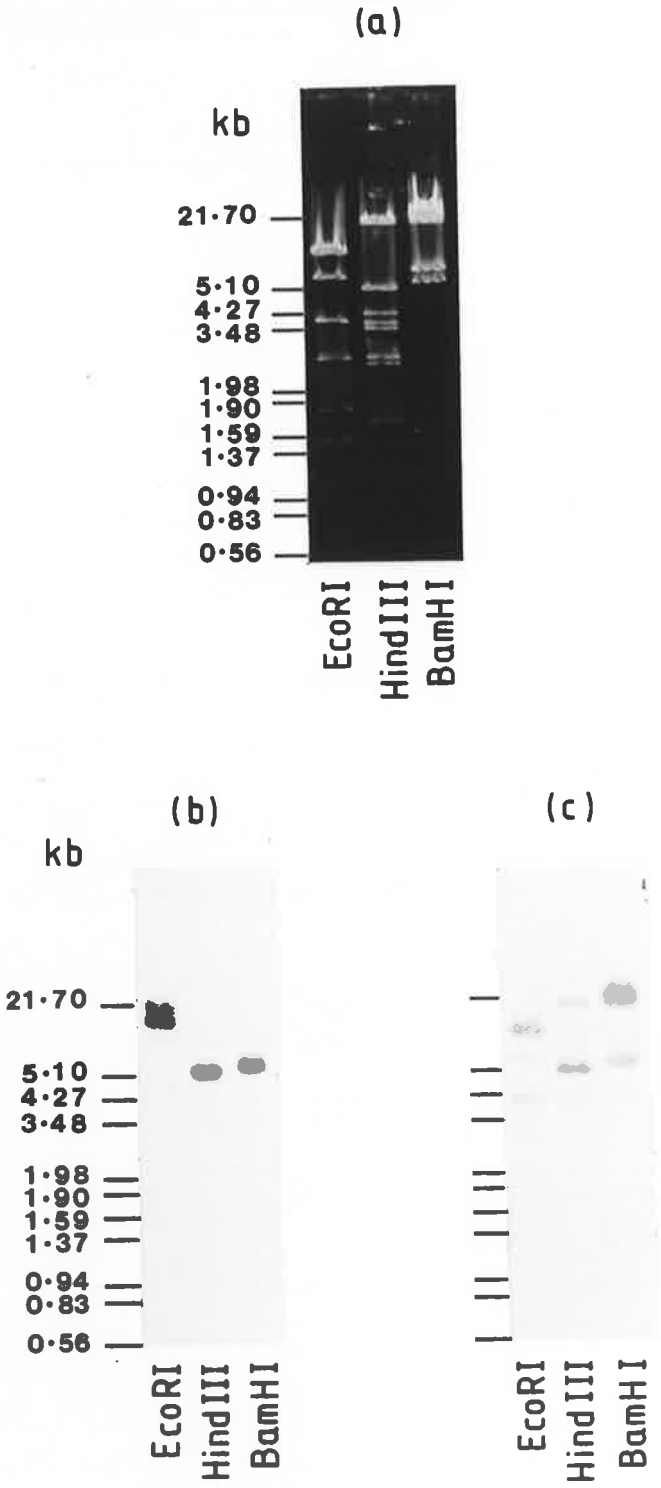
Probing pcosBIIIB1 with pSWK18 indicates that one of the 10 kb EcoRI bands, the 5 kb HindIII band and a 5.7 kb BamHI fragment contain BIIIB-homologous sequences (Fig.6.12b). By comparing the intensity of the hybridization signal obtained with the cosmid fragments with the

FIGURE 6.12. SOUTHERN ANALYSIS OF pcosBIIIB1 USING pSWK18 AND WOOL FOLLICLE cDNA AS PROBES.

(a) pcosBIIIB1 DNA was digested with EcoRI, HindIII or BamHI, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker is an EcoRI/HindIII digest of λ DNA.

(b) Autoradiograph following Southern transfer of the agarose gel shown in (a), hybridization with nick-translated pSWK18 for 18h at 42°C (Chapters 2 and 4), and washing in 0.5xSSPE/0.1% SDS at 65°C.

(c) As in (b) but using cDNA to wool follicle poly(A)$^+$ mRNA as probe and washing in 1xSSPE/0.1% SDS at 65°C.


(a)
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
EcoRI
HindIII
BamHI
(b)
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
EcoRI
HindIII
BamHI
(c)
EcoRI
HindIII
BamHI

signal obtained for λSWK50, it appeared that pcosBIIIB1 contains no more than 2 genes (data not shown). Probing pcosBIIIB1 with cDNA made to wool follicle poly(A)$^+$ mRNA results in hybridization with not only the BIIIB gene-containing bands but also other fragments (Fig.6.12c). EcoRI fragments of approximately 3.7 and 6.5 kb, HindIII bands of 3.6 and approximately 15 kb, and at least one of the high-molecular-weight BamHI fragments hybridize in addition to the pSWK18-positive bands.

C. Discussion.

1. Isolation of a BIIIB gene from a Charon 28 sheep genomic DNA library using a synthetic oligonucleotide probe:

Although the BIIIB keratin genes constitute members of a multigene family, it has been very difficult to isolate a functional gene from screens of a number of sheep genomic DNA libraries. A search of four libraries, using a BIIIB3-related cDNA clone as the probe resulted in the isolation of a pseudogene (see Chapter 5). A feature of the pseudogene was a 23 bp deletion near the 3' end of the coding region. Since this sequence must be present in the coding region of any functional BIIIB gene, a mixed oligonucleotide based on the sequences of 3 cDNA clones and corresponding to a 20 bp portion of the deleted region in the pseudogene, was synthesized and used as a probe (Fig.6.1).

The 20-mer was used initially to screen a number of Charon 28 pSWK18-positive clones but failed to hybridize with any of them. The Charon 28 library was then re-screened using the 20-mer and pSWK18 to probe duplicate filters - a single recombinant was found to hybridize with both probes (Fig.6.2).

The Charon 28 library contains the equivalent of 3-4 sheep genomes in 0.75×10^6 recombinants. Approximately 2 genome equivalents were screened and only a single clone was found to hybridize with the 20-mer and pSWK18 probes. This result supports the observation discussed in Chapter 5 that the sheep genomic DNA libraries examined have a very low representation of BIIIB gene sequences. Possible causes of the reduced

cloning efficiency of these sequences have been discussed in Chapter 5.

Based on the protein sequence data of Swart et al. (1976), the enzyme Sau3A, used to partially digest the sheep DNA in the construction of the Charon 28 library, has 3 potential cleavage sites in the BIIIB coding region. Two sites could be present in the genes coding for the sequenced proteins - Thr.Ile and Gly.Ile or Ser.Ile at residues 17-18 and 79-80 respectively (Fig.6.9), can lead to the sequence NNG.ATC. The limited codon usage data (Table 6.1) indicates that the most common Thr, Gly and Ser codons have C as the third base, thus reducing the possibility of Sau3A sites at these locations. The third site could occur in a variant combining a BIIIB2 residue with either a BIIIB3 or 4 amino acid, leading to Asp.Pro at residues 59-60 (Fig.6.9) and potentially giving rise to the nucleotide sequence GAT.CNN. From all of the BIIIB cDNA and gene data, covering over 4 kb of DNA sequence, only two Sau3A sites have been found. One site occurs in the codons for Gly.Ile in pSWK101 (K. A. Ward, personal communication) while the other is located in the 3' non-coding region of pSWK18. The low number of functional BIIIB genomic clones in the Charon 28, and indeed the λ1059 libraries, may result from a reduced number of Sau3A (and hence BamHI) sites within and surrounding the BIIIB genes, combined with a deleterious sequence closely associated with the gene. Interestingly the clone λSWK96, containing the functional gene, has only approximately 0.5 kb of DNA 5' to the concensus CAAT sequence and presumably has been separated from any sequence element which causes poor BIIIB gene cloning efficiencies.

It is also possible that there are Sau3A restriction sites, but they are not readily cleaved during the enzymatic digestion. Case (1982) has reported that Balbiani ring 1 DNA, digested to "completion" with MboI, has a "MboI-resistant" 30 kb fragment, while Balbiani ring 2 DNA is cut every 150-500 bp as would be expected from the random occurrence of the 4-base recognition sequence.

2. λSWK96 contains a functional BIIIB gene:

Restriction mapping and Southern analysis using pSWK18 as the probe indicated that the 13.8 kb insert of λSWK96 contains a single BIIIB gene (λ96BIIIB) located approximately 0.5 kb from the right arm of the Charon 28 vector (Figs.C.1-C.3). The gene was isolated and sequenced, and was found to contain an uninterrupted reading frame of 291 bp which codes for a BIIIB-related protein (Figs.6.5, 6.8, 6.9). Approximately 90 bp 5' to the initiation codon the gene contains a consensus TATA sequence while a possible CAAT box is situated a further 60 bp upstream. The 3' non-coding sequence extends for 340 bp from the termination codon to the single polyadenylation addition sequence (Figs.6.5, 6.10). The precise location of the poly(A)-addition site is unknown since the homologous cDNA sequence (pSWK31) does not include a poly(A) tail. Proudfoot and Brownlee (1976) have observed that the poly(A) site is situated between 11 and 30 bp downstream from the AATAAA sequence. Burget (1984) surveyed 61 vertebrate RNAs and found that the sequence CACTG was located either upstream or downstream from the site of poly(A) addition and polyadenylation occurred at the adenine nearest to the pentanucleotide. In λ96BIIIB, the sequence CTCTG occurs 9 bp downstream from the AATAAA hexanucleotide (Fig.6.10). The adenine which follows the CTCTG sequence is situated 16 bp 3' to the AATAAA and is in a position homologous to the poly(A)-addition site of pSWK18 (Fig.6.11b).

There is no indication that λ96BIIIB contains any introns - the same observation has been made for ψ50BIIIB (Chapter 5), the B2 genes (Powell et al., 1983), a BIIIA and a high-glycine-tyrosine gene (B. C. Powell and E. Kuczek, personal communications). The low-sulphur genes contain multiple intron sequences (B. C. Powell and K. A. Ward, personal communications).

The presence in λ96BIIIB of transcription control sequences found associated with other eukaryotic genes in addition to an uninterrupted coding sequence suggest that λ96BIIIB is a functional gene.

3. λ96BIIIB codes for a BIIIB4-related protein:

The predicted amino acid sequence coded by λ96BIIIB shows only two differences from the protein BIIIB4 and has been termed BIIIB4A (Fig.6.9). The differences - Glu at residue 43 instead of the Gln found in BIIIB4, and the lack of Cys at residue 48 in the gene product - are found also in ψBIIIB3A and the BIIIB3-related protein coded by pSWK18 (Chapter 5).

The peptide compositions and amino acid sequence data presented by Haylett et al. (1971) and Swart and Haylett (1971) indicate that the protein sequences reported by Swart et al. (1976) are correct. The lack of Cys at residue 48 is a feature of the BIIIB2 component but not BIIIB3 or 4 (Swart et al., 1976) and while the latter proteins are more closely related to each other than to BIIIB2, they do have some residues which are characteristic of the BIIIB2 component. BIIIB4 and 2, but not BIIIB3, have Cys at residue 55 and Pro at 88, while BIIIB3 shares with BIIIB2, but not 4, a Leu at residue 41 (Fig.6.9; Swart et al., 1976). The lack of Cys at residue 48 in ψBIIIB3A, BIIIB4A and the pSWK18 protein product may be examples of a similar type of heterogeneity.

A similar heterogeneity is found in the protein sequences available for the BIIIA high-sulphur keratin family (Swart et al., 1976). In the 11 wool protein sequences there are 13 sites at which amino acid variations have been observed. The differences - usually a choice of two residues at each site - include Glu for Gln substitutions and the presence or absence of Cys residues, and result in a complex mixture of homologous components displaying limited microheterogeneity. It would seem that comparable microheterogeneity exists in the BIIIB protein family, although the number of components is smaller than that described for the BIIIA family (see Discussion, Chapter 5).

4. Homologies between λ96BIIIB and other keratin genes:

(a) Homology with other BIIIB genes:

When the sequence of λ96BIIIB is compared with the available

BIIIB nucleotide sequence data, a number of homologies are apparent. Not surprisingly, there is a high degree of homology between the coding regions of the BIIIB sequences - greater than 95% between λ96BIIIB and pSWK18, and 80% between λ96BIIIB and pSWK101. If the third base changes which do not alter the amino acid sequence are disregarded, the homology with pSWK101 increases to almost 90%.

The 3' non-coding region of the BIIIB4-related sequences are highly homologous (Fig.6.10) and the same is true for the BIIIB3-like sequences (Fig.5.9). When the two groups of sequences are compared, there are two blocks of strong homology (Fig.6.11). The first extends for 50 bp 3' to the termination codon, while the second spans approximately 70 bp 5' to the poly(A)-addition site. There is no significant homology between the BIIIB3- and 4-related sequences in the region between the two homologous blocks, nor is there any homology with pSWK101 over the entire 3' non-coding sequence. Furthermore, while the BIIIB3- and 4-like sequences have 3' non-coding regions of 350-360 bp, the length in pSWK101 is 266 bp (K. A. Ward, personal communication).

There are significant homologies in the 5' flanking and non-coding regions of the BIIIB sequences (Figs.6.6a, 6.7). The λ96BIIIB and ψ50BIIIB genes are greater than 70% homologous from the CAAT sequence to the initiation codon, while a comparable similarity exists between λ96BIIIB and the available 5' non-coding sequence of pSWK101.

It is apparent from the protein sequences of Swart et al. (1976), from the gene sequences presented in this thesis and from the cDNA data of Ward (personal communication), that the BIIIB3 and 4 components are more closely related to each other than to the BIIIB2 components. It is possible that a single primordial BIIIB sequence gave rise, through gene duplication, to an ancestral BIIIB2 gene and BIIIB(3/4) gene. Further duplications and sequence divergence produced the multigene family as it presently exists. Presumably, the duplication which gave rise to the separate BIIIB3 and 4 genes is more recent than

that which resulted in the BIIIB2 genes.

In spite of the sequence diversity apparent in the 3' non-coding region of the BIIIB genes, there are constraints on divergence of the 5' non-coding and coding regions. It is possible that gene conversion events are taking place to maintain, within limits, sequence homogeneity of the coding region, as has been suggested for the B2 high-sulphur keratin gene family (Powell *et al.*, 1983) and the high-cysteine chorion gene family in the silkmoth *Bombyx mori* (Iatrou *et al.*, 1984).

(b) Homology with non-BIIIB keratin genes:

Powell *et al.* (1983) reported the conservation of an 18 bp sequence situated in the 5' non-coding region of three B2 genes and the BIIIB cDNA clone pSWK101. The availability of sequence data for BIIIB, BIIIA and high-glycine-tyrosine genes has enabled this conserved sequence to be re-examined (Powell and Rogers, 1985) and extended (Fig.6.7). Over a region of 50-60 bp 5' to the initiation codon there is a significant homology between λ96BIIIB and the B2, BIIIA and high-glycine-tyrosine genes (Fig.6.7a). If deletions are disregarded, the homology ranges from 50-60 %. While the B2 5' non-coding sequences are more related to each other than to λ96BIIIB, and while the BIIIA sequence more closely resembles the B2 genes than the BIIIB sequences, there is a significant homology between the 5' non-coding regions from the three different keratin gene families. The homology is sufficient to enable a consensus sequence to be proposed which covers 27 bp immediately 5' to the initiation codon (Fig.6.7b). While the sequence is not as highly conserved as the consensus TATA sequence, it could prove nevertheless to be of significance in relation to the control of translation of sheep keratin matrix mRNAs. (The extent of conservation of the sequence and its possible specificity in the matrix genes awaits future DNA sequencing studies).

There is some indirect evidence that the conserved sequence may have some function. Powell and Rogers (1985) noted that there is no

homology between the 5' non-coding regions of sheep keratins, globin or metallothionein suggesting that the conserved sequence may be keratin-specific. In addition, Kozak (1984) surveyed the 5' non-coding sequences of over 200 eukaryotic mRNAs and found that over one half of these had 3-4 nucleotides in common with a 5 bp sequence, 5'-CCACC-3', located immediately 5' to the initiation codon. There was no other apparent homology upstream from this sequence, thus further supporting a possible specific regulatory role for the conserved keratin sequence.

Powell et al. (1983) suggested that a portion of the conserved sequence may be involved in ribosome binding. While the heterogeneity in the size and nucleotide sequence of the 5' non-coding regions of mRNAs (all of which use the same ribosomes) indicates that the ATG initiation codon is the only universally recognizable signal sequence, the efficiency of ribosome binding was reduced by the removal of 5' non-coding sequences (Baralle and Brownlee, 1978). The conserved keratin sequence may have some site-specific translational control or, as suggested by Powell et al. (1983), may be involved in protein-mRNA interactions that accelerate keratin mRNA translation during the terminal differentiation stage of the keratinocyte.

While it is possible that the conserved sequence in the 5' non-coding region of keratin matrix genes may be present as a neutral sequence with no specific function, it is difficult to reconcile the homology observed in this region with the diversity found in the coding and 3' non-coding regions of the different keratin gene families.

The 5' flanking sequence of λ96BIIIB has one region which is homologous with a portion of the B2A gene and another region homologous with a short sequence found in the B2A, B2C and BIIIA genes (Figs.6.6b,c; Powell and Rogers, 1985). The homology which occurs between λ96BIIIB and B2A alone extends over a region of approximately 100 bp and begins respectively 184 and 314 bp 5' to the initiation codon. The full extent of the homology is unknown since it continues to the 5' limit of the

sequence data available for both genes (Fig.6.6b). The other conserved sequence is found in λ96BIIIB, B2A, B2C and a BIIIA gene, is only 17 bp in length and is situated between 160 and 190 bp 5' to the TATA box (Fig.6.6c).

Sequences upstream from the defined transcriptional control elements (TATA and CAAT sequences) have been implicated in gene regulation. In a study of insulin and chymotrypsin gene expression (Walker et al., 1983), it was observed that the deletion of sequences varying for the different genes from 170 to 300 bp 5' to the initiation codon resulted in dramatic reductions in cell-specific expression. The 5' border of the sequence corresponded approximately to a DNase I hypersensitivity site and it was proposed that the sequence could be involved in generating an open chromatin structure (Walker et al., 1983). A similar proposal was made by Breathnach and Chambon (1981) who also suggested that the upstream sequence element might be a binding site for proteins involved in positive control of transcription, or the sequence, brought by chromatin folding into proximity with the TATA box, could be involved in RNA polymerase II binding. Upstream control sequences, located between 50 and 400 bp 5' to the transcription initiation site appear to be required for the maintenance of transcription efficiency but tend not to effect the transcription start site (McKnight and Kingsbury, 1982). Any role in the control of high-sulphur keratin gene expression through the homologous sequences of the 5' flanking region remains to be demonstrated.

Although the overall codon usage of the BIIIB, B2 and high-glycine-tyrosine genes follows the general pattern found in vertebrate genes (Table 6.1; Hastings and Emerson, 1983), there are a number of interesting observations which can be made. The Cys codon TGC is used 3 times more frequently than TGT; the Thr ACC and Gln CAG codons are used almost exclusively, while the Ser codons TCC and AGC are present to a greater extent than might be expected from the data available for pooled

liver genes (Table 6.1). More coding sequence data are required in order to determine whether the unusual codon usage described above is a feature of each or individual keratin gene families. In a study of chorion genes, Jones and Kafatos (1982) observed that there is a selective codon usage which results in a third base composition extremely rich in pyrimidines. These workers suggested that codon usage in abundantly expressed genes is optimized through selection to reduce the risk of translation blocks resulting from the low prevalence of certain isoacceptor tRNAs and to maximize translational efficiency.

5. Was there a common ancestral gene which gave rise to all high-sulphur keratin genes?

Swart (1973) and Swart et al. (1976) provided evidence which suggested that the BIIIA, BIIIB and B2 wool high-sulphur protein families were derived from a single ancestral gene. Both the BIIIA and B2 families are characterized by the presence of a pentapeptide repeat - Cys-Cys-Gln-Pro-X and on the basis of nearest neighbour statistical analysis, Swart and Parris (1974) suggested that the BIIIB proteins were derived from the same sequence. Lindley and Elleman (1972) observed homology between residues of the C-terminal region of BIIIB and B2 proteins, while Swart (1973) extended the homology by aligning this region with residues 23-61 of a BIIIA component. The possible codon usage for a portion of the homologous region suggested an ancestral protein and a nucleotide sequence whereby amino acid differences between components could be explained by single-base mutations (Swart, 1973). The data were extended by Swart et al. (1976) who aligned 13 BIIIA and 3 B2 protein sequences to fit the pentapeptide repeating unit and then determined the most likely genetic code for the repeat in the two protein families. The results indicated that the codons differed at only two positions leading to an Arg and Thr in BIIIA substituting for Gln and Ser in the B2 family. The data led Swart et al. (1976) to suggest that the different gene families diverged from a single common region in periodic

DNA.

Elleman et al. (1973), using the same data as Swart (1973), proposed that the B2 and BIIIA protein families evolved by repeated partial duplications involving all or part of a decapeptide unit, and in addition, a subsequent larger duplication of 15 amino acid residues in the BIIIA family. Furthermore, Elleman et al. (1973) did not consider the proposed homology between the BIIIB and the B2 or BIIIA proteins to be significant.

The high-sulphur keratin gene data suggest that perhaps Elleman et al. (1973) and Swart et al. (1976) were each partially correct. In their study of the B2 genes, Powell et al. (1983) observed that a previously unrecognized component, called B2D, contained 6 tandemly repeated decapeptide units and that its gene was closely linked to the gene for B2A, a protein with 5 decapeptide units. The coding regions of the two genes showed a 93% nucleotide sequence homology. The data are highly suggestive that the B2 gene family has arisen either by duplication of a 30 bp repeat (equivalent to a decapeptide), or by deletion of this sequence from a gene containing multiple copies of the repeat unit. Jones and Kafatos (1982) have described the mechanisms whereby duplications or deletions of tandem repeats could have resulted in the diversification of the chorion gene families.

The nature of gene sequence heterogeneity in the BIIIA family and its relationship to the BIIIB and B2 genes must await complete nucleotide sequences of BIIIA components. However, the homologies in the 5' non-coding region and a 17 bp sequence in the flanking region of a BIIIA gene and members of the B2 and BIIIB families suggest some common ancestry. The 9 bp sequence in the 5' non-coding conserved region of a BIIIA gene which must be deleted to optimize the homology with the other keratin gene sequences is almost a perfect direct tandem repeat of the sequence immediately upstream from it (Fig.6.7a) and its presence can be explained using the mechanisms described by Jones and Kafatos (1982).

The homology observed at the C-terminal end between BIIIB and B2 proteins (Lindley and Elleman, 1972) is reflected in the nucleotide sequences. Disregarding deletions, there is a 75% homology between nucleotides 256-288 of λ96BIIIB (Fig.6.8) and 780-818 of B2A (Powell et al., 1983). The homologies observed in the 5' non-coding and flanking regions between λ96BIIIB and the B2 genes (Figs.6.6b,c; 6.7a) provide further evidence to support the concept of a common ancestral gene for the high-sulphur multigene complex.

Accumulation of sequence data, particularly for BIIIA and BIIIB genes, will allow the question of high-sulphur gene ancestry to be examined more comprehensively. More extensive DNA sequence data banks should also provide valuable insights into possible evolutionary relationships between the genes from different keratin families. It is important to note here that there are significant homologies in the 5' non-coding region between a high-glycine-tyrosine gene and some high-sulphur genes (Fig.6.7a).

6. Future work:

The high-sulphur gene studies have pointed to possible specific regulatory sequences in the 5' non-coding and flanking regions. The significance of these sequences should be investigated further in in vitro systems or in cultured keratinocytes.

While λSWK96 was shown by Southern analysis to contain no sequences homologous to wool follicle cDNA other than λ96BIIIB, the DNA sequence downstream from the BIIIB gene contained a possible CAAT box, a TATA box and an adenine residue in a location which could mark the transcription initiation site of another gene (Fig.6.9). The isolation and characterization of the 1.8 kb EcoRI fragment adjacent to the λ96BIIIB gene (Fig.6.5) would resolve whether there was a gene, and if so, its sequence would be of interest because it is so closely linked to the keratin gene.

Further studies on the cosmid clone, pcosBIIIB1, would also be

of value. The clone contains a 32 kb insert but carries only 1-2 BIIIB genes on a 5 kb HindIII fragment (Fig.6.12). The restriction fragments which contain the BIIIB gene(s) cannot be made to overlap the λSWK50 or λSWK96 restriction maps, indicating that the gene(s) in pcosBIIIB1 are different from those already characterized. Furthermore, pcosBIIIB1 is of interest because it contains other sequences homologous with wool follicle cDNA. Any other genes contained in the cosmid should be isolated, characterized and their relationship with the BIIIB genes determined.

SECTION III.

APPENDICES

APPENDIX A

RESTRICTION MAPPING AND SUBCLONING OF λSWK50

APPENDIX A

RESTRICTION MAPPING AND SUBCLONING OF λSWK50

1. Restriction mapping of λSWK50:

Bacteriophage λSWK50 DNA was digested with a number of restriction enzymes and subjected to electrophoresis in 1% agarose. The restriction sites and fragment sizes for the Charon 4A arms were obtained from the data of de Wet *et al.* (1980). Location of gene-containing fragments was made by hybridization of pSWK18 to Southern transfers of restriction fragments (Chapters 2 and 4).

Restriction enzyme digestion with EcoRI, SalI and SmaI did not cleave within the λSWK50 insert (data for SalI and SmaI not shown). Digestion with HindIII, BamHI, BglII and KpnI in the presence or absence of EcoRI, followed by Southern transfer and pSWK18 probing enabled a simple restriction map of λSWK50 to be determined (Fig.A.1a-c). These data indicate that the BIIIB gene lies between 3.5 and 5.8 kb from the left end of the insert.

To simplify further restriction mapping, the 13.2 kb insert of λSWK50 was subcloned into the EcoRI site of pBR325. Digestion of λSWK50 with EcoRI results in three fragments (Fig.A.1a). Only the internal 13.2 kb fragment has EcoRI cohesive ends and can be ligated into the plasmid. All of the recombinant plasmids examined had a 13.2 kb, EcoRI-resectable insert - the clone used for further study is referred to as pλSWK50.

The orientation of the insert in pλSWK50 was determined from a BamHI digest (Fig.A.2a). The presence of two components approximately 4.6 kb in size indicates that the orientation shown is correct. If the insert had been positioned in the opposite direction then BamHI fragments of 2.0 and 7.3 kb would have been observed. The orientation of the 1.0 and 1.6 kb BamHI fragments was obtained from a BamHI/KpnI double-digest (Fig.A.2a). KpnI cleaves pλSWK50 at one position, reducing the 1.6 kb BamHI fragment to 1.4 kb (the 0.2 kb fragment was not observed under the

FIGURE A.1. RESTRICTION ENZYME MAPPING OF λSWK50.

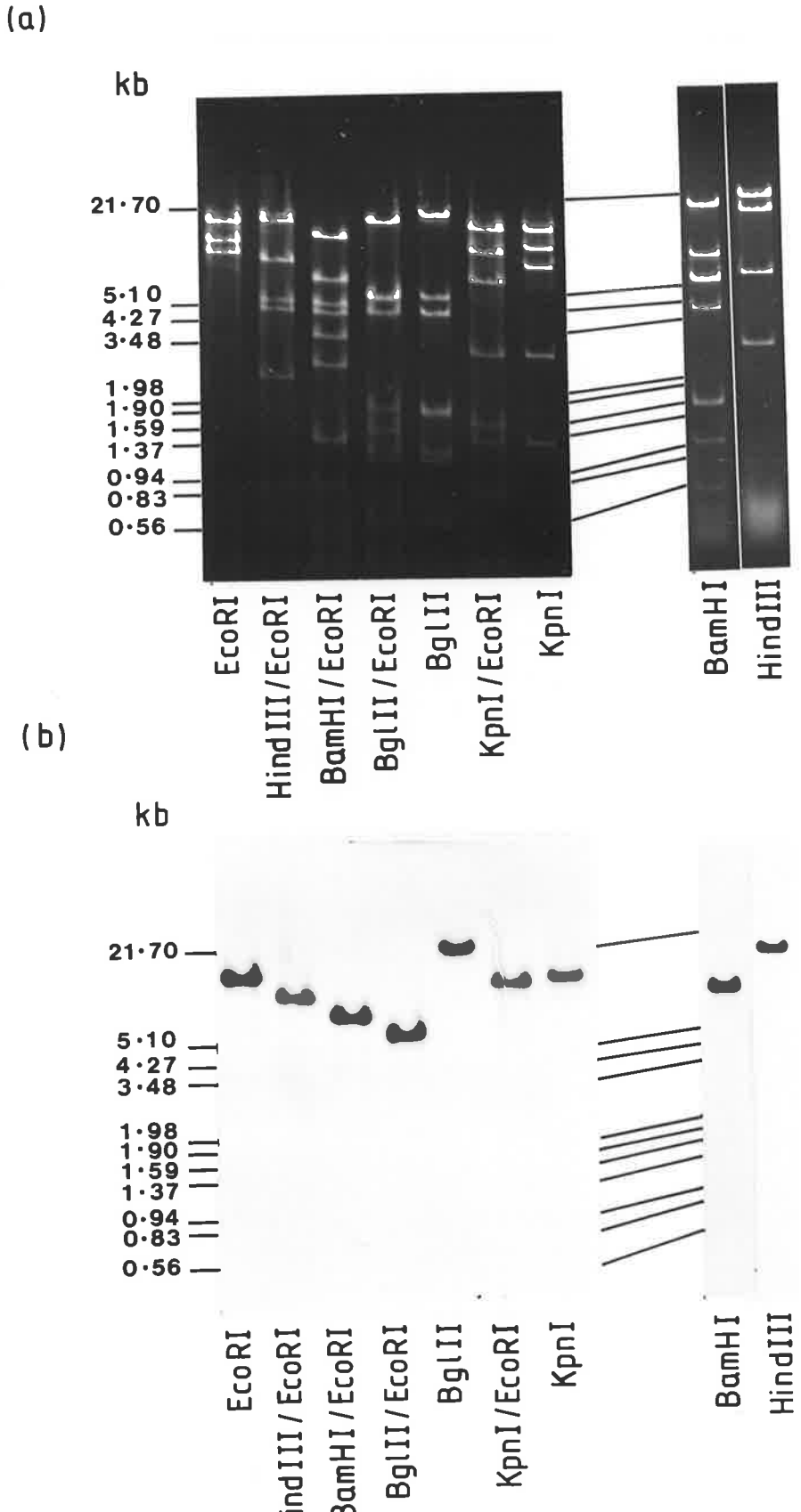
(a) λSWK50 DNA was digested with a number of restriction enzymes, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was a HindIII/EcoRI digest of λ DNA.

(b) Autoradiograph following Southern transfer of the agarose gels (a), hybridization with nick-translated pSWK18 for 24h at 67°C (Chapter 2), and washing in 0.1xSSPE/0.1% SDS at 55°C.

(c) Predicted restriction enzyme maps based on the results obtained in (a) and (b). Fragment sizes (kb) are shown. Restriction fragments for which orientations could not be determined are shown by sloped lines.

V_L and V_R represent the Charon 4A left and right arms respectively.

* indicates restriction fragments which hybridized with pSWK18 probe.


(a)
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
EcoRI
HindIII/EcoRI
BamHI/EcoRI
BglII/EcoRI
BglII
KpnI/EcoRI
KpnI
BamHI
HindIII
(b)
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
EcoRI
HindIII/EcoRI
BamHI/EcoRI
BglII/EcoRI
BglII
KpnI/EcoRI
KpnI
BamHI
HindIII

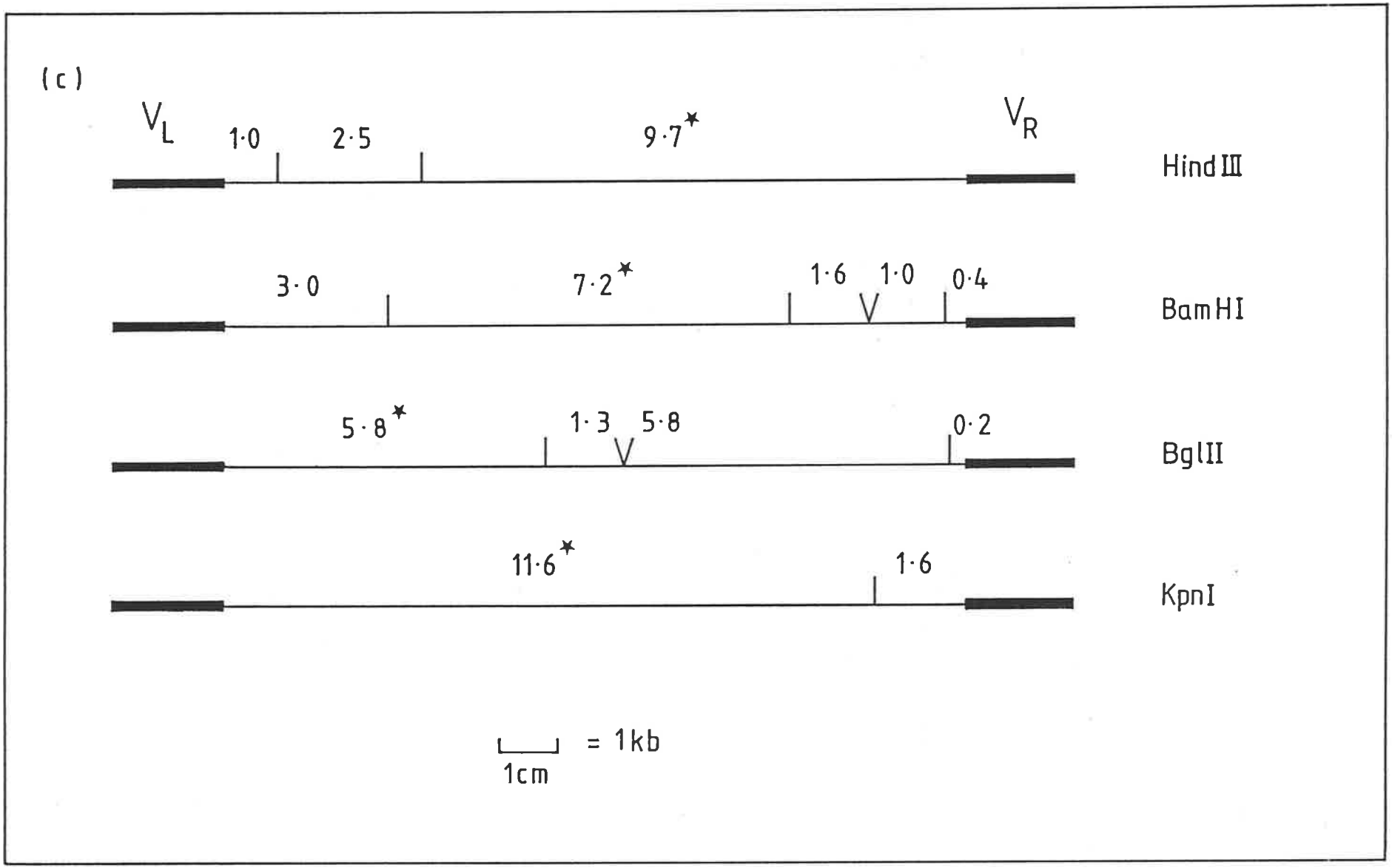

(c)
V_L
1·0
2·5
9·7*
V_R
Hind III
3·0
7·2*
1·6
1·0
0·4
BamHI
5·8*
1·3
5·8
0·2
BglII
11·6*
1·6
KpnI
= 1kb
1cm

FIGURE A.2. RESTRICTION ENZYME MAPPING OF pλSWK50.

pλSWK50 DNA was digested with the appropriate restriction enzyme(s), electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was a HindIII/EcoRI digest of λ DNA. The drawings show the location and sizes (kb) of the relevant restriction fragments. The striped region represents the pBR325 vector.

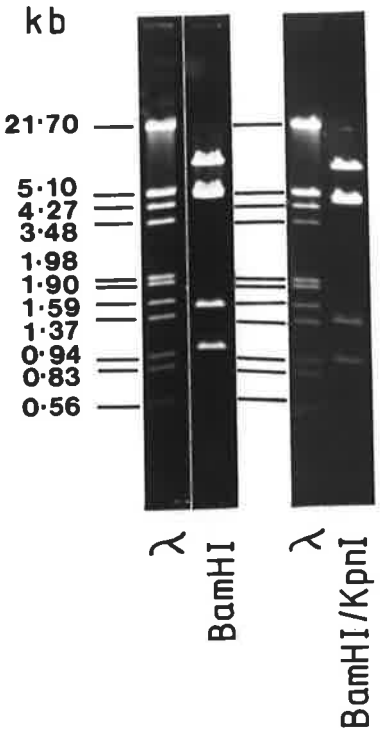
* indicates restriction fragments which hybridized with pSWK18 probe.

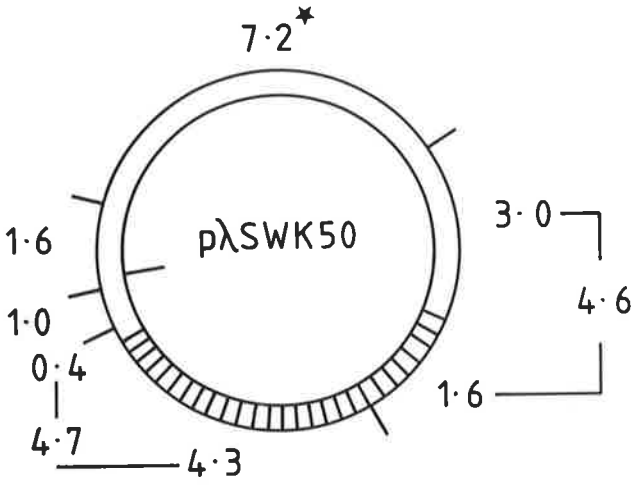
(a) BamHI and BamHI/KpnI digests of pλSWK50. The lines pointing away from the drawing represent the BamHI sites while the line directed inwards shows the location of the KpnI site.

(b) BglII/KpnI and BglII/BamHI digests of pλSWK50. The lines pointing away from the drawing, represent the BglII sites, K is the location of the KpnI site while the lines directed inward show the BamHI cleavage points.

(c) PstI, PstI/EcoRI, PstI/HindIII and PstI/KpnI digests of pλSWK50. The lines pointing away from the drawing represent the PstI sites, E and K are the locations of the EcoRI and KpnI sites respectively, while the lines directed inward show the HindIII cleavage points.

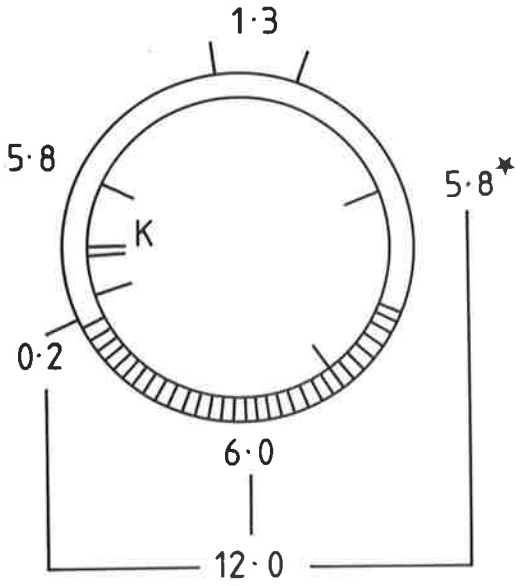
(a)


kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
λ
BamHI
λ
BamHI/KpnI

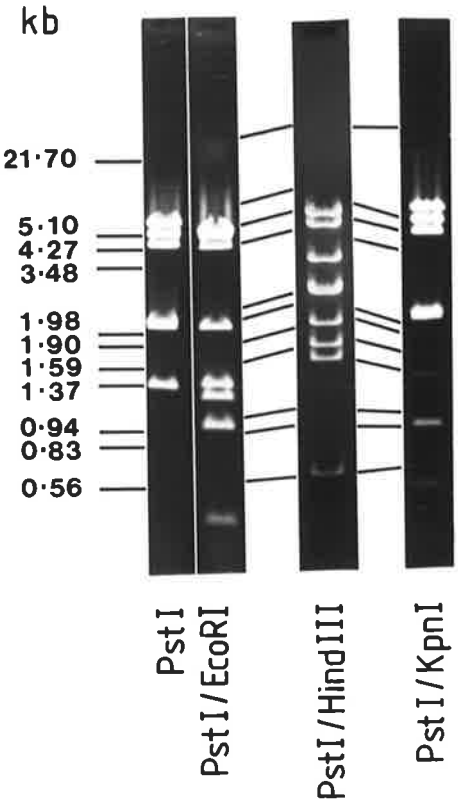

7·2
pλSWK50
1·6
1·0
0·4
4·7
4·3
3·0
4·6
1·6

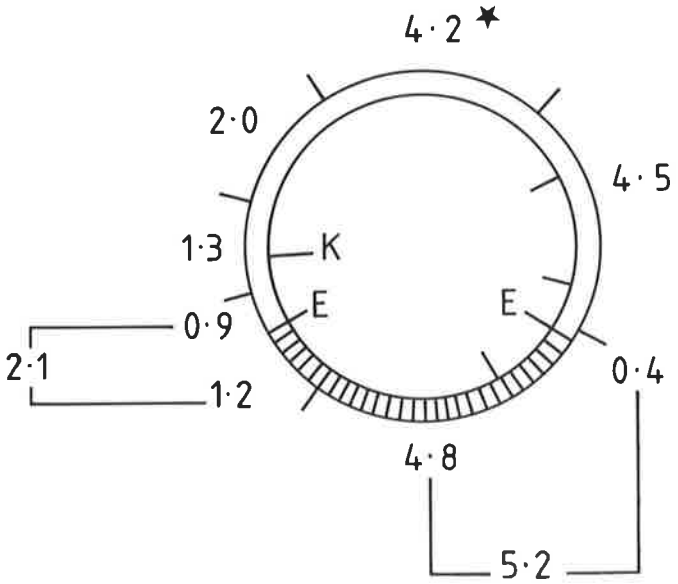
(b)


kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
BglII/KpnII
BglII/BamHI


1·3
5·8
K
0·2
6·0
12·0
5·8

(c)


kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
PstI
PstI/EcoRI
PstI/HindIII
PstI/KpnI


4·2
2·0
4·5
1·3
K
E
E
0·9
2·1
1·2
0·4
4·8
5·2

electrophoretic conditions used). If the 1.6 kb fragment was adjacent to the 0.4 kb fragment, KpnI restriction would have resulted in bands of 1.2 and 0.4 kb.

The orientation of the 1.3 and 5.8 kb BglII fragments (Fig.A.1c) was determined from a BglII/KpnI double-digest of pλSWK50. There is no KpnI site in pBR325 and only a single site in the genomic DNA insert. While the double digest of pλSWK50 is not complete (overdigestion did not improve the cleavage), there are bands at approximately 4.3 and 1.5 kb (Fig.A.2b). These bands are derived presumably from KpnI digestion of the 5.8 kb BglII fragment, and so orientates the BglII restriction sites as shown in Fig.A.2b. The BglII map was confirmed by a BglII/BamHI double-digest (Fig.A.2b). The 1.3 kb BglII and 1.0 and 1.6 kb BamHI fragments remain intact - this can only occur if the orientation of the BglII fragments is as shown in Fig.A.2b.

A PstI restriction map was determined for pλSWK50 (Fig.A.2c). Digestion with PstI alone results in six fragments : 5.2, 4.5, 4.2, 2.1, 2.0 and 1.3 kb. Further digestion in the presence of EcoRI cleaves the 5.2 and 2.1 kb fragments to generate bands of 4.8, 1.2, 0.9 and 0.4 kb in size. This indicates that the 0.9 kb fragment is adajacent to the 1.2 kb PstI/EcoRI segment of pBR325, while the 0.4 kb fragment is at the other end of the insert.

A PstI/HindIII double-digest orientates the 4.5 kb PstI fragment adjacent to the 0.4 kb fragment (Fig.A.2c). The important feature of the digest is that the 4.5 kb PstI fragment is cleaved by HindIII into 2.5, 1.4 and 0.6 kb. The other fragments are derived from either the pBR325 or the insert. This means that the 4.5 kb PstI and 2.5 kb HindIII fragments overlap, as shown in Fig.A.2c.

A PstI/KpnI double-digest of pλSWK50 cleaves the 1.3 kb PstI fragment into 0.5 and 0.8 kb pieces (Fig.A.2c). This positions the 1.3 kb PstI fragment adjacent to the 0.9 kb PstI/EcoRI end of insert.

Southern transfer of a λSWK50 PstI digest and probing with

pSWK18 showed that the BIIIB gene was located on the 4.2 kb fragment (data not shown). Since the gene was shown to lie between 3.5 kb and 5.8 kb from the left end of the insert in λSWK50 (Fig.A.1c), the 4.2 kb PstI fragment must be located next to the 4.5 kb PstI fragment, as shown in Fig.A.2c. Insertion of the 2.0 kb PstI fragment between the 4.5 kb and 4.2 kb pieces would place the gene too far away from the left end of the λSWK50 insert.

2. Subcloning of the λSWK50 BIIIB gene:

(a) Attempt to subclone a 2.8 kb BamHI/BglII fragment:

The restriction map of λSWK50 is shown in Fig.A.3a. The BIIIB gene is located on a 2.8 kb BamHI/BglII fragment (Fig.A.3b). Since BamHI and BglII restriction yield identical cohesive ends, attempts were made to subclone the 2.8 kb fragment into the BamHI site of pBR322 and M13mp8. The DNA fragment was prepared from λSWK50 using a number of procedures: (i) low-melting-temperature agarose, (ii) low-melting-temperature agarose followed by DEAE-cellulose chromatography, and (iii) electrophoresis onto DEAE-membrane (Chapter 4). All efforts to subclone the DNA were unsuccessful. There did not appear to be any inhibition of ligation since [^{32}P]-labelled aliquots of the fragment gave a ladder when treated with DNA ligase (data not shown). Control experiments in which pBR322 DNA was digested with EcoRI, electrophoresed, and then recovered from agarose as described above gave successful cloning results.

(b) Subcloning of a 2.3 kb HindIII/BglII fragment:

Digestion of λSWK50 with HindIII and BglII results in four fragments having both HindIII and BglII cohesive ends (Fig.A.3a and Charon 4A restriction map in de Wet et al., 1980) - 2.3 and 2.6 kb from the Charon 4A right arm, 20.3 kb from the left arm and the left end of the insert, and 2.3 kb containing the BIIIB gene. A HindIII/BglII digest of λSWK50 was ligated into HindIII/BamHI-digested M13mp8 or M13mp9. Following transformation of E.coli JM101, plaques were screened with pSWK18 probe as described in Chapter 4. Approximately 30% of the clear

FIGURE A.3. RESTRICTION MAP OF λSWK50 AND SUBCLONING STRATEGIES.

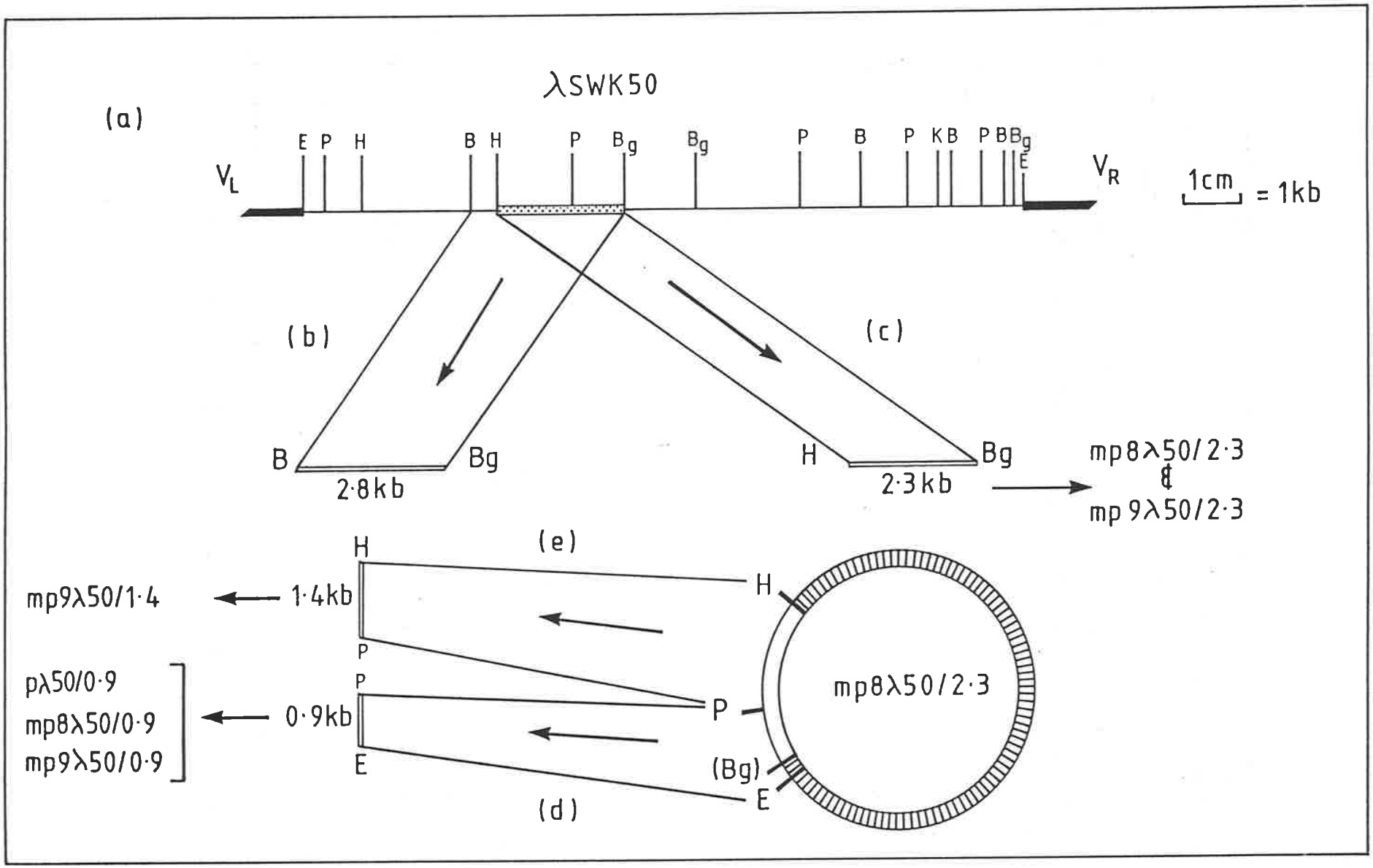
(a) Restriction map of the λSWK50 insert showing cleavage sites for EcoRI(E), PstI(P), HindIII(H), BamHI(B), BglII(Bg) and KpnI(K). V_L and V_R represent the Charon 4A left and right arms respectively. The striped bar shows the location of the BIIIB gene-containing fragment.

(b) The 2.8 kb fragment released by a BamHI/BglII double-digest of λSWK50.

(c) The 2.3 kb fragment released by a HindIII/BglII double-digest of λSWK50. This fragment was ligated to the HindIII and BamHI cohesive ends of M13mp8 and M13mp9 to produce the recombinants mp8λ50/2.3 and mp9λ50/2.3 respectively.

(d) The recombinant mp8λ50/2.3, obtained from the subcloning outlined in (c) was digested with PstI and EcoRI to release an 0.9 kb fragment. (The BglII site (Bg) was lost as a result of cloning into the BamHI site of M13mp8). The 0.9 kb fragment was ligated into the PstI and EcoRI cohesive ends of pBR322, M13mp8 and M13mp9 to produce pλ50/0.9, mp8λ50/0.9 and mp9λ50/0.9 respectively.

(e) The recombinant mp8λ50/2.3 was digested with PstI and HindIII and the excised 1.4 kb fragment was ligated to M13mp9 cleaved with the same enzymes.The resulting subclone was termed mp9λ50/1.4.


λSWK50
(a)
E P H B H P Bg Bg P B P K B P B Bg E
VL
VR
1cm = 1kb
(b)
(c)
B
Bg
2·8kb
H
Bg
2·3kb
mp8λ50/2·3
&
mp 9λ50/2·3
(e)
H
mp9λ50/1·4
1·4kb
P
pλ50/0·9
mp8λ50/0·9
mp9λ50/0·9
P
0·9kb
E
H
P
mp8λ50/2·3
(Bg)
E
(d)

plaques hybridized with the probe. The recombinants obtained from this subcloning, referred to as mp8λ50/2.3 and mp9λ50/2.3 (Fig.A.3c) were used for DNA sequencing and further subcloning studies.

(c) Subcloning of an 0.9 kb PstI/EcoRI fragment from mp8λ50/2.3:

The restriction map and hybridization data for λSWK50 (Figs.A.2 and A.3a) indicate that the BIIIB gene is contained on an 0.9 kb PstI/BglII fragment. The 4.5 kb PstI fragment to the left and the 1.3 kb BglII fragment to the right do not hybridize with pSWK18. The location of the gene was confirmed by Southern transfer and hybridization with pSWK18 of a PstI/EcoRI double-digest of mp8λ50/2.3. (Ligation of the HindIII/BglII fragment into HindIII/BamHI-digested M13mp8 and M13mp9 results in the loss of the BamHI and BglII recognition sequences. The EcoRI site of the vector is 10 bp 5' to the BamHI cleavage point).

Figure A.4 shows the result of a PstI/EcoRI digestion of mp8λ50/2.3 followed by hybridization with pSWK18. The digestion restricts an 0.9 kb fragment from the 9.5 kb recombinant bacteriophage. (A trace of uncut DNA is still visible). Hybridization with pSWK18 shows that only the 0.9 kb fragment has sequences homologous to the probe (Fig.A.4). On the basis of this data, the 0.9 kb fragment was subcloned into PstI/EcoRI-digested pBR322, M13mp8 and M13mp9 to give the recombinants pλ50/0.9, mp8λ50/0.9 and mp9λ50/0.9 respectively (Fig.A.3d). The recombinant plasmid was used for restriction mapping and subcloning while the M13 bacteriophage were used in DNA sequencing experiments.

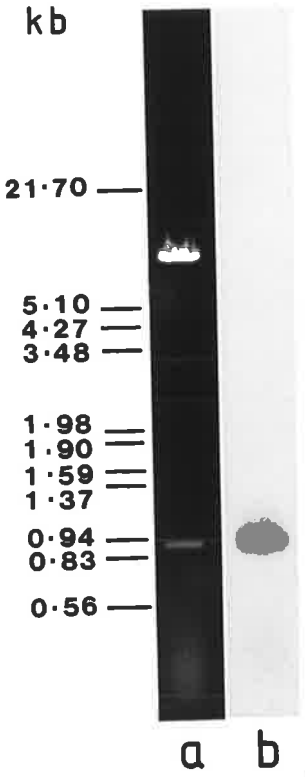
3. Restriction mapping of pλ50/0.9:

The recombinant plasmid pλ50/0.9 was digested with a number of enzymes and a restriction map was determined. It should be noted that the 752 bp PstI/EcoRI fragment of pBR322 was excised during the subcloning and restriction sites within this fragment are not present in the recombinant plasmid. The pBR322 restriction sites and fragment sizes are those published by Sutcliffe (1979).

FIGURE A.4. LOCATION OF THE BIIIB GENE IN mp8λ50/2.3.

(a) mp8λ50/2.3 DNA was digested with PstI and EcoRI, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight standard was a HindIII/EcoRI digest of λ DNA.

(b) Autoradiograph following Southern transfer of the agarose gel (a), hybridization with nick-translated pSWK18 for 20h at 67°C (Chapter 2), and washing in 0.1xSSPE/0.1% SDS at 65°C.


kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
a
b

The enzymes HhaI, RsaI, TaqI, Sau3A and HpaII did not cleave the pλ50/0.9 insert (data not shown). Digestion of λSWK50 with PvuII suggested that there was a restriction site within the 0.9 kb PstI/BglII fragment (data not shown). This was confirmed by digesting pλ50/0.9 with PstI and PvuII (Fig.A.5a). The small fragment observed in this electrophoretogram places the PvuII site approximately 120 bp from the PstI end of the insert (Fig.A.5a).

Restriction of pλ50/0.9 with HaeIII results in two bands derived from the insert - 690 and 450 bp in size (Fig.A.5b). Sequential digestion with HaeIII and PstI leaves the 690 bp fragment intact, but cleaves the 450 bp band into 120 bp (derived from the vector) and approximately 340 bp (Fig.A.5b). This places the single HaeIII site approximately 340 bp from the PstI end of the insert (Fig.A.5b).

Digestion with HinfI indicates that the pλ50/0.9 insert is cut twice (Fig.A.5c). The HinfI fragments containing insert DNA are approximately 1350, 300 and 200 bp in size. There are HinfI sites approximately 250 and 630 bp from the PstI and EcoRI sites respectively, which means that the 200 bp HinfI fragment is contained entirely within the insert. Cleavage of pλ50/0.9 with PstI and HinfI (Fig.A.5c) leaves the 1350 bp band intact while the 300 bp fragment is digested to 250 bp (derived from the vector) and 60 bp. Therefore there are HinfI sites within the insert approximately 60 and 260 bp from the PstI site (Fig.A.5c). EcoRI/HinfI digestion of pλ50/0.9 cleaves the 1350 bp fragment into 630 bp (derived from the vector) and 650 bp pieces (data not shown).

Restriction with AluI results in cleavage of the pλ50/0.9 insert at five sites. The AluI fragments containing insert DNA are approximately 400, 320, 130, 115, 50 and 45 bp (Fig.A.5d). Digestion with AluI and PstI digests the 130 bp fragment into bands of approximately 50 bp (derived from the vector) and 75 bp, while an AluI/EcoRI double-digest cleaves the 320 bp into fragments of 16 bp

FIGURE A.5. RESTRICTION ENZYME MAPPING OF pλ50/0.9.

pλ50/0.9 DNA was digested with the appropriate enzyme(s), electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was either a HaeIII digest of ϕX174 DNA or different digests of pBR322 DNA. The drawings show the pλ50/0.9 insert restriction maps for the appropriate enzymes.

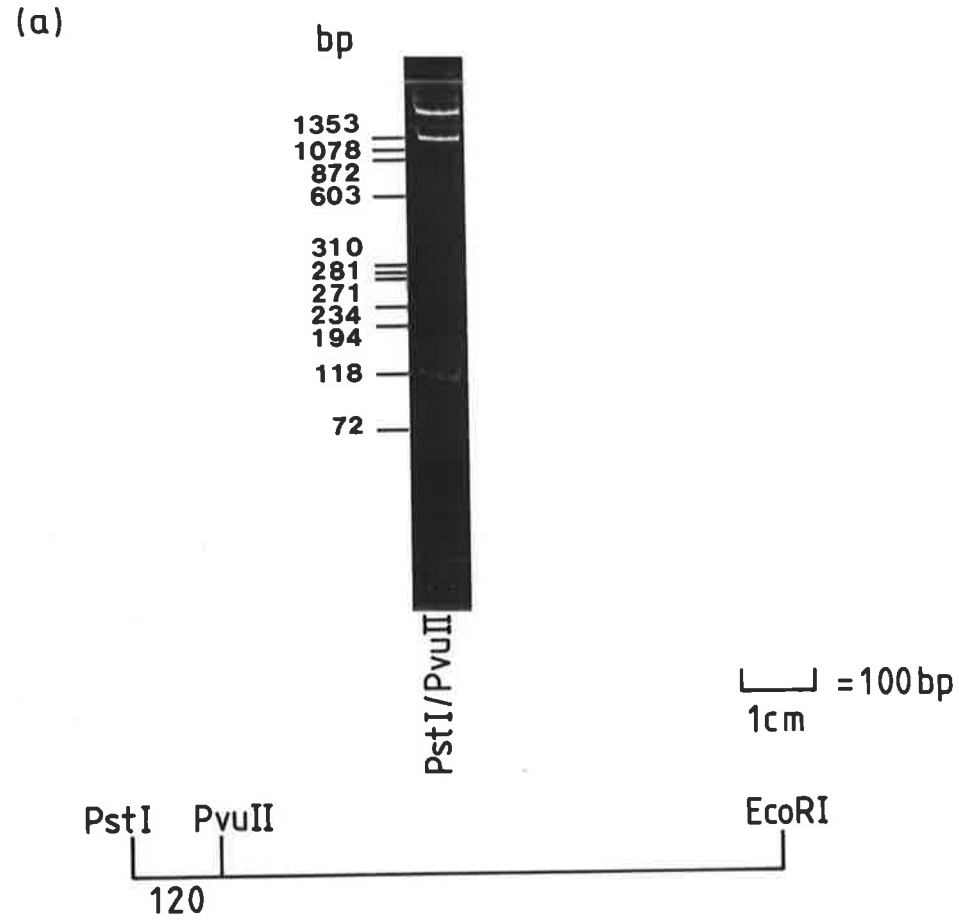
Fragment sizes are in base pairs (bp).

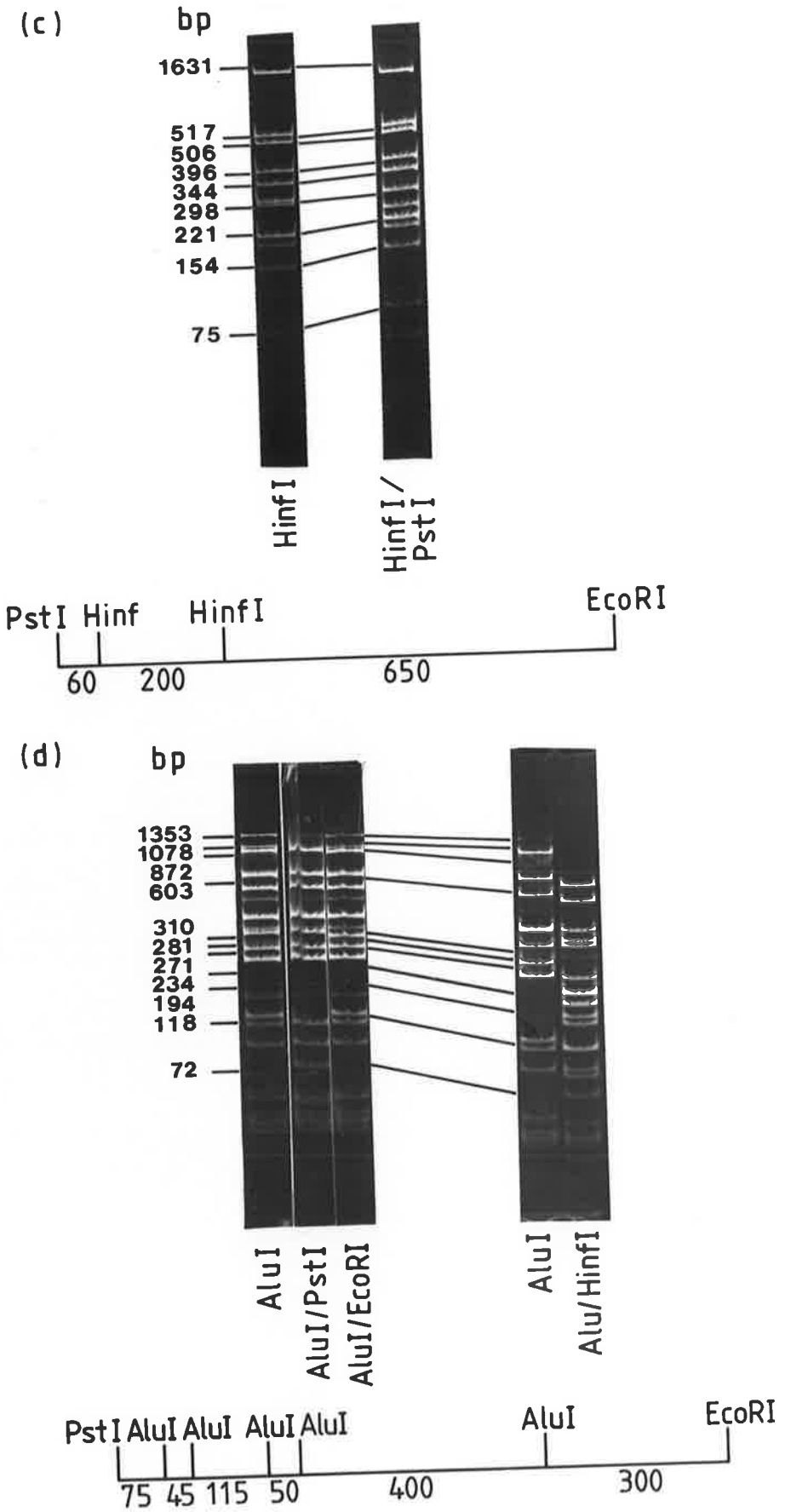
(a) PstI/PvuII digest.

(b) HaeIII and HaeIII/PstI digests.

(c) HinfI and HinfI/PstI digests.

(d) AluI, AluI/PstI, AluI/EcoRI and AluI/HinfI digests.


(a)
bp
1353
1078
872
603
310
281
271
234
194
118
72
PstI/PvuII
=100bp
1cm
PstI
PvuII
EcoRI
120
(b)
bp
1353
1078
872
603
310
281
271
234
194
118
72
HaeIII
HaeIII/
PstI
PstI
HaeIII
EcoRI
340


(c)
bp
1631
517
506
396
344
298
221
154
75
HinfI
HinfI/
PstI
PstI Hinf
HinfI
EcoRI
60
200
650
(d)
bp
1353
1078
872
603
310
281
271
234
194
118
72
AluI
AluI/PstI
AluI/EcoRI
AluI
Alu/HinfI
PstI AluI AluI AluI AluI
AluI
EcoRI
75 45 115 50
400
300

(vector DNA, and not visible on the electrophoretogram) and 300 bp (Fig.A.5d). These digests result in the assignment of the 75 bp and 300 bp fragments adjacent to the PstI and EcoRI sites respectively (Fig.A.5d).

As shown in figure A.5a, there is a PvuII site removed by 120 bp from the PstI site. AluI cleaves within the PvuII recognition sequence and it can be deduced that the 45 bp AluI fragment must lie adjacent to the 75 bp AluI/PstI fragment (Fig.A.5d). This leaves the 400, 115 and 50 bp fragments unordered. An AluI/HinfI double-digest resolves the location of these three insert fragments. If the 400 bp band was located adjacent to the PvuII site, or was separated from it by either the 50 or the 115 bp fragment, HinfI would cleave it. It can be seen (Fig.A.5d) that the 400 bp fragment remains intact following AluI/HinfI digestion (a 403 bp vector DNA band is cleaved in this digestion). Thus, the 400 bp band must lie adjacent to the 300 bp AluI/EcoRI fragment (Fig.A.5d). The 50 bp AluI fragment is cleaved by HinfI and so must be positioned between the 115 and 400 bp fragments (Fig.A.5d).

The AluI fragment sizes amount to an insert size of almost 1 kb, about 100 bp greater than indicated by the PstI/EcoRI digest of pλ50/0.9 (Fig.A.4) and the other restriction digests (Fig.A.5). Sequence data for the 400 bp AluI fragment showed that the size estimated from electrophoretic mobility was anomalously large (data presented in Chapter 5).

4. Subcloning of a 1.4 kb PstI/HindIII fragment from mp8λ50/2.3:

Preliminary DNA sequence data obtained from mp8λ50/0.9 showed that the PstI site of this subclone was located at the 5' end of the BIIIB coding region. To obtain DNA sequences 5' to the PstI site, the 1.4 kb PstI/HindIII fragment of mp8λ50/2.3 was subcloned into M13mp9 digested with the same enzymes (Fig.A.3e). Sequencing studies on the subclone mp9λ50/1.4 showed that there was an AvaII site located 116 bp 5'

to the initiation codon site. Further subcloning using restriction at this site allowed the sequencing studies on the BIIIB gene to be completed.

APPENDIX B

RESTRICTION MAPPING AND SUBCLONING OF λSWK61

APPENDIX B

RESTRICTION MAPPING AND SUBCLONING OF λSWK61

1. Restriction mapping of λSWK61:

The titre of the pSWK18-positive Charon 28 recombinants was very low compared with other bacteriophage in the library. This made DNA preparation difficult. The yield of DNA from a number of experiments was very low and so only limited restriction mapping was undertaken.

Figure B.1 shows the result of probing a BamHI digest of λSWK61 with pSWK18 and pλSWK50. The pSWK18 probe hybridized to a 7.2 kb fragment, while the pλSWK50 probe annealed with the 7.2 kb band as well as fragments at 1.0, 1.6 and approximately 22-23 kb. Charon 28 vector has a left arm of 22.7 kb and a right arm of 9.2 kb (R. Crawford, personal communication). Presumably, the large pλSWK50-positive band contains the vector left arm with a portion of insert DNA while the other fragments are derived from the genomic insert. There did not appear to be any insert DNA attached to the vector right arm following BamHI digestion (data not shown).

2. Subcloning and restriction mapping of the 7.2 kb BamHI fragment:

DNA from λSWK61 was digested with BamHI and the mixture of fragments were ligated to BamHI-cleaved pBR322. Of 130 recombinant plasmids screened with pSWK18 insert, only 5 were positive (data not shown). Each positive clone contained a 7.2 kb insert, and one, pλSWK61/7.2, was chosen for further study.

Figure B.2 shows the restriction fragments obtained following digestion of pλSWK61/7.2 with a number of enzymes. PstI digestion results in three fragments - 2.3, 4.2 and 5.0 kb in size (Fig.B.2a). The only orientation of PstI sites which enables the clone to have a 7.2 kb insert is for the 4.2 kb fragment to be within the insert, a 1.2 kb piece to be joined to the 1.1 kb PstI/BamHI pBR322 fragment and 1.8 kb to be attached to the remaining 3.2 kb of the vector (Fig.B.2e).

FIGURE B.1. RESTRICTION ENZYME MAPPING OF λSWK61.

λSWK61 DNA was digested with BamHI, electrophoresed in 1% agarose and transferred to nitrocellulose (Chapter 4).

(a) Autoradiograph following hybridization with pSWK18 probe for 20h at 42°C and washing in 0.5xSSC/0.1% SDS at 65°C (Chapter 2).

(b) Autoradiograph following hybridization with pλSWK50 probe using the same conditions as in (a). DNA molecular weight markers were either HindIII or HindIII/EcoRI digests of λ DNA.

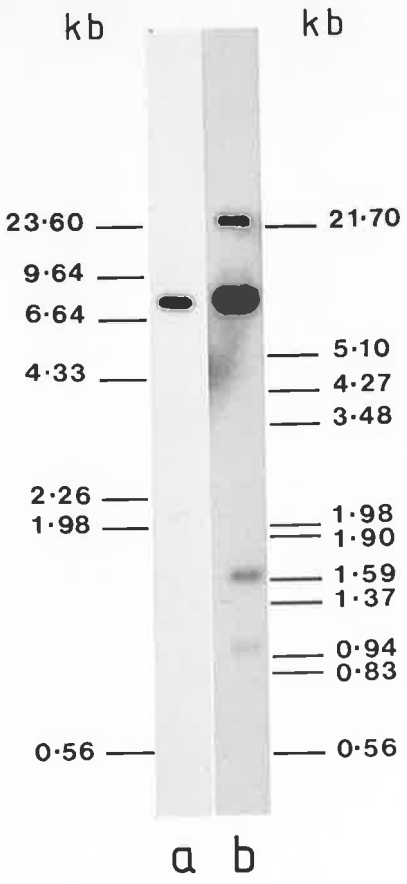

kb
kb
23·60
9·64
6·64
4·33
2·26
1·98
0·56
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
a
b

FIGURE B.2. RESTRICTION ENZYME MAPPING OF pλSWK61/7.2.

pλSWK61/7.2 DNA was digested with a number of enzymes, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was a HindIII/EcoRI digest of λ DNA.

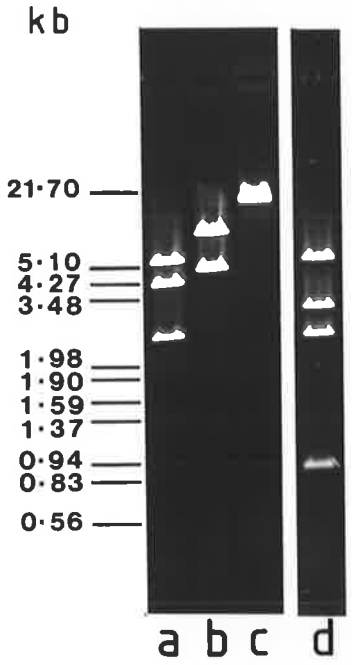
(a) PstI digestion.

(b) HindIII.

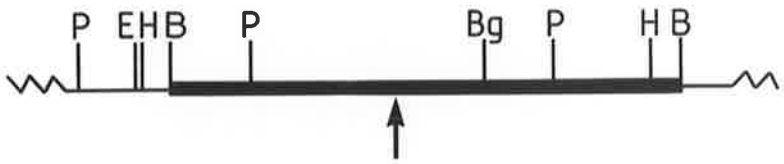
(c) BglII.

(d) PstI/BglII.

(e) Restriction map of pλSWK61/7.2. The thick line represents the 7.2 kb BamHI insert; the thin lines represent pBR322. The restriction sites shown are: P(PstI), E(EcoRI), H(HindIII), B(BamHI) and Bg(BglII). The arrow marks the position of a BglII site found in λSWK50 but absent in pλSWK61/7.2.


kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
a b c d

(e)


P E H B P Bg P H B

= 1kb
1cm

Digestion of pλSWK61/7.2 with HindIII yields two fragments - 4.4 and 7.1 kb in size (Fig.B.2b). The HindIII site in pλSWK61 must be located at the right-hand end of the insert, approximately 0.4 kb from the BamHI site (Fig.B.2e). Cleavage with BglII linearizes the plasmid DNA (Fig.B.2c). Since the vector does not have a BglII site, the result indicates that there is a single BglII cleavage within the insert. A double-digest with PstI and BglII results in the 4.2 kb PstI fragment being digested into pieces 3.3 and 0.9 kb in size (Fig.B.2d). The orientation of the BglII site within the 4.2 kb PstI fragment was inferred to be as shown in figure B.2e. The basis for this assignment was that the restriction data obtained for λSWK61 and pλSWK61/7.2 indicated that the genomic insert was essentially the same as that of λSWK50 other than it was approximately 3 kb shorter than λSWK50 and the 7.2 kb BamHI fragment contained only a single BglII site instead of two (compare Fig.B.2c,d and e with Fig.A.3a).

For the purpose of the work presented here, λSWK61 was assumed to be homologous to the portion of λSWK50 which contained the 7.2, 1.6 and 1.0 kb BamHI fragments (Figs.A.1c and A.3a). On the basis of this assumption, the single BglII site in pλSWK61/7.2 was assigned as shown in figure B.2e.

APPENDIX C

RESTRICTION MAPPING AND SUBCLONING OF λSWK96

APPENDIX C

RESTRICTION MAPPING AND SUBCLONING OF λSWK96

1. Restriction enzyme mapping of λSWK96:

The restriction map of the vector is based on the sites listed in Rimm et al. (1980) and was provided by Dr. R. Crawford, Howard Florey Institute of Experimental Physiology and Medicine, Melbourne. The Charon 28 stock used - 2221 - contains an EcoRI site in the left arm approximately 1.2 kb from the BamHI site.

Bacteriophage λSWK96 DNA was prepared as described in Chapter 4 and subjected to limited restriction enzyme mapping and Southern analysis (Fig.C.1). BamHI digestion cleaves the DNA at one site only (Fig.C.1a). Since both fragments are larger than the vector arms (22.7 and 9.2 kb for the left and right arms respectively), the BamHI sites of the Charon 28 have not been regenerated and the cleavage is within the genomic DNA insert. Southern analysis, using pSWK18 as probe indicates that the DNA linked to the right arm contains the BIIIB gene (Fig.C.1b).

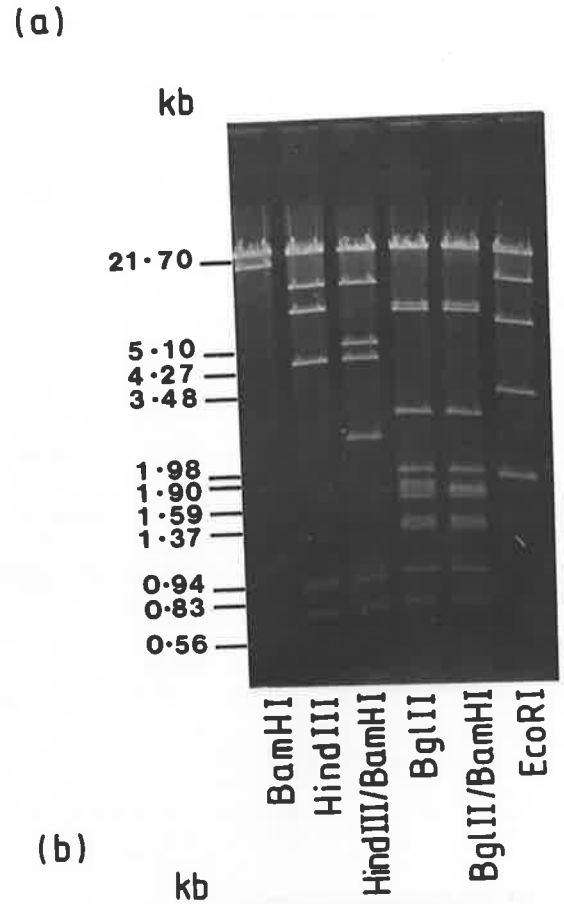
There are no HindIII sites in the vector arms. Cleavage of λSWK96 with HindIII in the presence or absence of BamHI indicates that the insert has 5 HindIII sites and that a 7.5 kb fragment is digested into components of 2.5 and 5.0 kb by BamHI (Fig.C.1a). A fragment of approximately 0.7 kb hybridized with pSWK18 (Fig.C.1b) indicating that the insert contains a single BIIIB gene.

Cleavage of λSWK96 with EcoRI results in 6 fragments (Fig.C.1a). The largest fragment is derived from the left arm of the vector while a band of approximately 10.6 kb contains 1.4 kb of insert DNA in addition to the 9.2 kb vector right arm. Southern analysis indicates that the BIIIB gene is located in the fragment linked to the right arm of the vector (Fig.C.1b). This result means that the 0.7 kb HindIII fragment which hybridized with pSWK18 must be located within 1.4 kb of the right end of the clone. The EcoRI digest resulted in other

FIGURE C.1. RESTRICTION ENZYME MAPPING OF λSWK96.

(a) λSWK96 DNA was digested with restriction enzymes as indicated, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was an EcoRI/HindIII digest of λ DNA.

(b) Autoradiograph following Southern transfer of the agarose gel shown in (a), hybridization with nick-translated pSWK18 insert for 20h at 42°C (Chapters 2 and 4), and washing in 0.5xSSPE/0.1% SDS at 65°C.


(a)
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
BamHI
HindIII
HindIII/BamHI
BglII
BglII/BamHI
EcoRI
(b)
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
BamHI
HindIII
HindIII/BamHI
BglII
BglII/BamHI
EcoRI

insert-containing fragments of 6.6, 3.4 and two of 1.8 kb in size (Fig.C.1a). Digestion of λSWK96 with EcoRI and BamHI cleaved the 3.4 kb fragment into 1.2 and 2.2 kb bands (Fig.C.2). Restriction analysis using PvuI with or without BamHI (see below) positioned the BamHI site 6.2 kb from the right end of the insert. This means that the two 1.8 kb bands and the 1.2 kb portion of the 3.4 kb BamHI-digested fragment must lie to the right of the BamHI site. Therefore, the 6.6 kb EcoRI band is located at the left end of the insert (1.2 kb vector, 5.4 kb insert). The EcoRI map of λSWK96 is shown in Fig.C.3. This map indicates that the insert is approximately 13.8 kb in length.

PvuI cleaves the Charon 28 arms at 2 sites. It can be seen (Fig.C.2) that λSWK96 has no internal PvuI sites. Double-digestion with PvuI and BamHI enables the location of the BamHI site in λSWK96 to be determined (Fig.C.2). The bands of approximately 12.0 and 8.0 kb are derived entirely from the vector arms. The largest fragment of the digest contains 11.5 kb of the vector left arm plus the adjoining portion of the insert. The band at approximately 7.5 kb contains 1.3 kb from the vector right arm and therefore 6.2 kb from the insert DNA. The BamHI site divides the λSWK96 insert into 6.2 and approximately 7.6 kb (Fig.C.3).

Digestion of λSWK96 with HpaI indicates that there are two HpaI sites in the insert (Fig.C.2). There is a doublet of approximately 3.4 kb, only one of which is derived entirely from the vector arms. In addition there are three fragments of approximately 2.2 kb, only two of which are derived from the Charon 28 arms. Double-digestion with HpaI and BamHI cleaves the 9.5 kb HpaI fragment into bands of 1.3 and approximately 8.1 kb (Fig.C.2). In order to accommodate the fragments obtained from the HpaI and HpaI/BamHI digests, the 8.1 kb band must be derived from the left end of the insert (0.5 kb vector arm, 7.6 kb insert) which means that either the 3.4 or 2.2 kb HpaI fragment is derived from the right end. A HpaI/EcoRI double-digest (Fig.C.2)

FIGURE C.2. RESTRICTION ENZYME DIGESTION OF λSWK96.

λSWK96 DNA was digested with restriction enzymes as indicated, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was an EcoRI/HindIII digest of λ DNA.

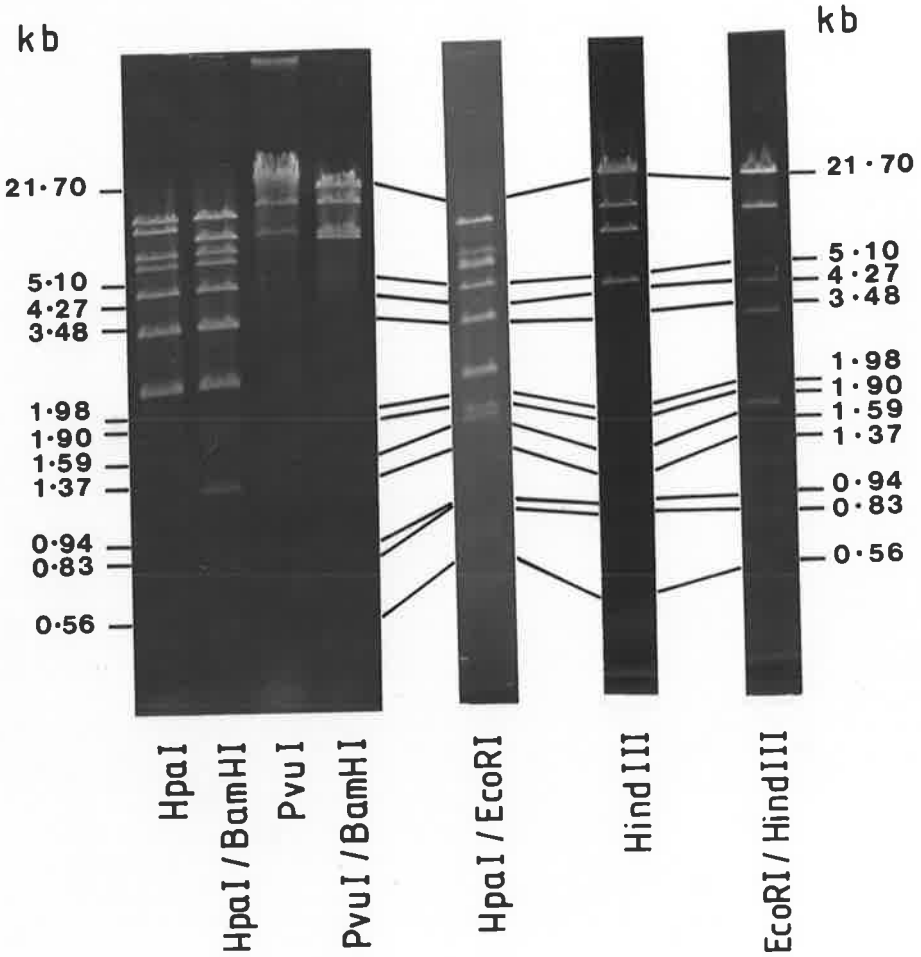
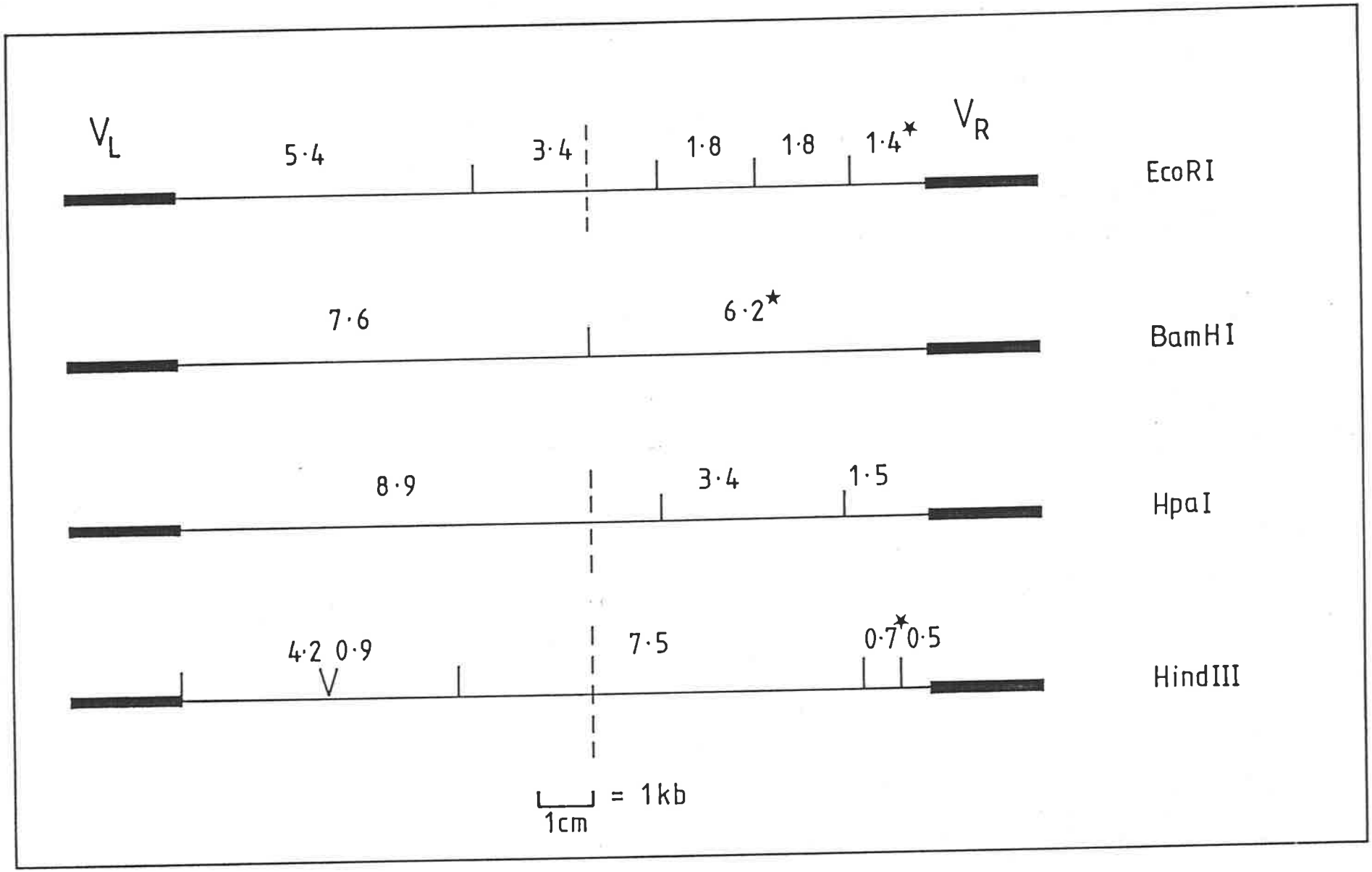

kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
HpaI
HpaI/BamHI
PvuI
PvuI/BamHI
HpaI/EcoRI
HindIII
EcoRI/HindIII

FIGURE C.3. RESTRICTION MAP OF λSWK96.

The restriction maps of λSWK96 for a number of enzymes based on the digestions shown in Figs.C.1 and C.2. Fragment sizes (kb) are shown. The unknown orientation of two HindIII fragments is indicated by sloped lines. The dashed, vertical line shows the location of the BamHI site.

V_L and V_R represent the Charon 28 left and right arms respectively.

* indicates restriction fragments which hybridized with pSWK18 probe (Fig.C.1b).


V_L
5·4
3·4
1·8
1·8
1·4
V_R
EcoRI
7·6
6·2
BamHI
8·9
3·4
1·5
HpaI
4·2
0·9
V
7·5
0·7
0·5
HindIII
= 1kb
1cm

indicates that the 2.2 kb HpaI fragment is at the right end of the insert. From the EcoRI map (Fig.C.3), if the 3.4 kb HpaI fragment was at the right end, it would have been cleaved with EcoRI to yield bands of 2.1 and 1.3 kb. There is no evidence for a 1.3 kb band in Fig.C.2. Instead, the EcoRI and HpaI sites at the right end of the insert almost overlap - the 2.2 kb HpaI fragment contains 1.5 kb of insert while the EcoRI site is 1.4 kb from the right end - thus maintaining a HpaI/EcoRI fragment of approximately 2.1 kb. The 3.4 kb HpaI fragment and the two 1.8 kb EcoRI fragments overlap and double-digestion results in bands of approximately 1.7 and 1.6 kb (Fig.C.2). These fragment sizes are in agreement with those expected from the restriction maps shown in Fig.C.3.

A double-digestion of λSWK96 using EcoRI and HindIII enables the orientation of a number of the HindIII fragments to be determined. As shown before, the 0.7 kb HindIII fragment, which contains the BIIIB gene, must lie close to the right end of the insert. The 0.7, 0.9 (both very faint in this digestion) and 4.2 kb HindIII fragments are not cleaved by EcoRI (Fig.C.2). This places the 4.2 kb fragment to the left of the BamHI site since there are three almost evenly spaced EcoRI sites to the right (Fig.C.3). It has been shown already (Fig.C.1) that BamHI cleaves the 7.5 kb HindIII fragment into 5.0 and 2.5 kb pieces. The 2.5 kb portion must lie to the left of the BamHI site since there is already a 4.2 kb HindIII fragment on this side and there is room for only 7.6 kb of insert DNA. The 7.5 kb HindIII fragment is digested by EcoRI into two pieces of 1.8 kb each, a fragment of 3.4 kb, and two small pieces not visible on the gel shown in Fig.C.2.

Electrophoretic comparison of a λSWK96 HindIII digest with λ HindIII DNA molecular weight markers indicated that the 9.2 kb vector right arm was linked to approximately 0.5 kb of insert DNA (data not shown). This means that the 0.9 kb HindIII fragment from the λSWK96 insert must be located to the left of the BamHI site, as shown in Fig.C.3. The orientation of the 4.2 and 0.9 kb HindIII fragments was not

determined.

The HindIII/EcoRI double-digest also shows a band of approximately 1.3 kb (Fig.C.2). This is derived from the left arm of the vector and a small portion of the insert - just less than 1.3 kb of this fragment is vector DNA. This means that there is a HindIIII site very close to the left end of the insert (Fig.C.3).

2. Subcloning and restriction mapping of an 0.7 kb HindIII fragment of λSWK96:

Restriction enzyme mapping of λSWK96 (Figs.C.1-C.3) indicated that the 0.7 kb HindIII fragment was the most suitable for subcloning the BIIIB gene. A HindIII digest of λSWK96 was ligated into the corresponding site in pBR322, and following transformation, recombinant plasmids were screened with pSWK18. One positive clone, pλ96/0.7, was used for restriction mapping and further subcloning.

Although pλ96/0.7 was shown to be cleaved by AluI, HinfI, HpaII, RsaI, HaeIII and Sau96I (data not shown) only 3 enzymes were required to enable the insert to be entirely sequenced - HindIII, PstI, DdeI and various combinations of these.

Digestion of pλ96/0.7 with PstI and electrophoresis on agarose revealed two bands approximately 1.1 and 3.9 kb in size (data not shown). Digestion with PstI and HindIII reduces the 1.1 kb band to fragments of 330 bp from the insert and 780 bp from the vector, while the 3.8 kb band gives rise to a large vector fragment and 250 bp from the insert. In addition, the digest reveals a 60 bp PstI fragment which had not been observed on the agarose gel (Fig.C.4).

DdeI digestion of pλ96/0.7, with or without HindIII, indicates that there are three DdeI sites in the insert (Fig.C.4). With DdeI alone there are, apart from the vector fragments, bands of 330, 250 and 90 bp and a high-molecular-weight band which contains 1550 bp of vector plus a portion of the insert. The double-digest gives rise to a reduction in size of the high-molecular-weight band, the loss of the 330 bp component

FIGURE C.4. RESTRICTION MAPPING OF pλ96/0.7.

(a) pλ96/0.7 DNA was digested with restriction enzymes, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was a HaeIII digest of ϕX174 DNA.

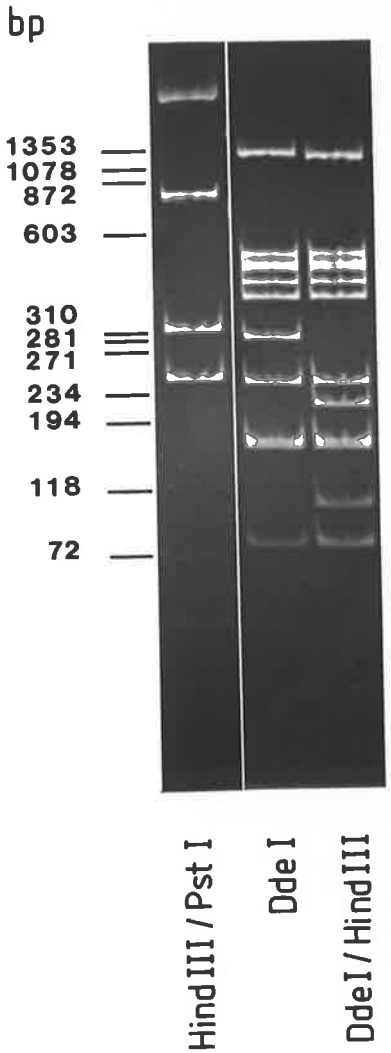
(b) Restriction map of the pλ96/0.7 insert based on results obtainedfrom (a). Fragment sizes (bp) for PstI digests are shown above the line and for DdeI digests, below the line. The orientation of the 250 and 90 bp DdeI fragments was obtained from DNA sequence analysis.

H: HindIII restriction sites.

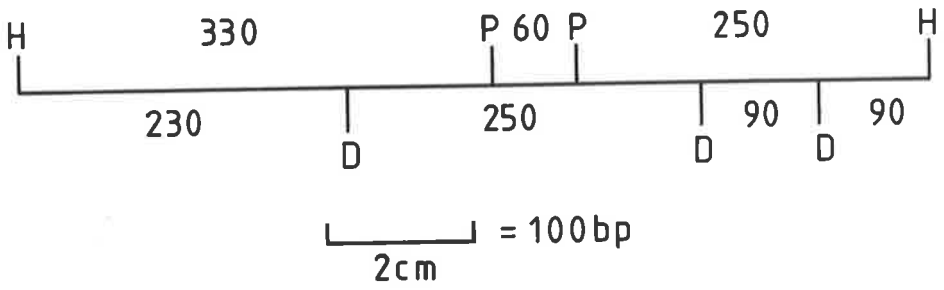
P: PstI.

D: DdeI.

(a)


bp
1353
1078
872
603
310
281
271
234
194
118
72
HindIII / Pst I
Dde I
DdeI / HindIII

(b)


H
330
P 60 P
250
H
230
D
250
D
90
D
90
= 100bp
2cm

and the appearance of bands of 230 and 100 bp with an additional fragment of 90 bp (Fig.C.4). The 230 bp fragment is adjacent to the 100 bp pBR322 DdeI fragment, while one of the 90 bp components is derived from the other end of the clone, adjacent to the 1550 bp vector fragment (Fig.C.4b). The orientation in the insert of the 250 and other 90 bp fragments was obtained from the nucleotide sequence data.

3. Further restriction mapping of λSWK96:

DNA sequence analysis of pλ96/0.7 (data presented in Chapter 6) indicated that only 50 bp of the BIIIB 3' non-coding region was present in the subclone. In order to complete the sequencing study, further restriction mapping of λSWK96 was needed.

The required mapping was simplified by two factors. Firstly, the 330 bp HindIII/PstI fragment of pλ96/0.7 (Fig.C.4) was comprised almost entirely of 5' flanking and non-coding sequence with only about 20 bp of coding region. Secondly, the 5' end of the pSWK18 insert coincided almost exactly with the start of the 250 bp HindIII/PstI fragment of pλ96/0.7. The 60 bp PstI fragment of pλ96/0.7, which is located between the 330 and 250 bp fragments, did not contain any sequence homologous to pSWK18. This meant that a PstI digestion of λSWK96 followed by Southern analysis using pSWK18 as probe should give hybridization only to the fragment containing the 250 bp HindIII/PstI sequence found in pλ96/0.7. Figure C.5 shows the result of such an experiment. The only λSWK96 PstI fragment which hybridized with pSWK18 was approximately 0.7 kb in size. This fragment extends the sequence of the gene by approximately 450 bp.

From the restriction map of λSWK96 (Fig.C.3), there is a HpaI site 300 bp away from the left HindIII site of the 0.7 kb fragment. If the pSWK18-positive 0.7 kb PstI fragment is orientated to the left, with respect to the λSWK96 map, then a PstI/HpaI digest should cleave approximately 150 bp from the positive band. If the orientation is towards the right arm of the vector, then double-digestion with PstI and HpaI should not alter the size of the pSWK18-positive band. Following

FIGURE C.5. FURTHER RESTRICTION MAPPING OF λSWK96.

(a) λSWK96 DNA was digested with restriction enzymes as indicated, electrophoresed in 1% agarose, stained with ethidium bromide and photographed under UV light. The DNA molecular weight marker was a HindIII/EcoRI digest of λ DNA.

(b) Autoradiograph following Southern transfer of the agarose gel shown in (a), hybridization with nick-translated pSWK18 insert for 18h at 42°C (Chapters 2 and 4), and washing in 0.5xSSPE/0.1% SDS at 65°C.


(a)
(b)
kb
21·70
5·10
4·27
3·48
1·98
1·90
1·59
1·37
0·94
0·83
0·56
PstI
PstI/HpaI
PstI
PstI/HpaI

PstI/HpaI digestion the positive band is reduced slightly in size (Fig.C.5) indicating that the sequence was extended towards the left arm of λSWK18.

The 0.7 kb PstI fragment was isolated from an acrylamide gel and subcloned into pBR322 (λ96/0.7A), and M13mp8 and 9. Subsequent subcloning was carried out as sequence data became available and restriction sites located.

APPENDIX D

PUBLICATION AND CONFERENCE PAPERS

Frenkel, M. J., Rogers, G. E. & Lock, R. A. (1979, February). Progress in the characterisation of ribonucleic acids which code for the keratin complex of hair. In D.A.D. Parry, L.K. Creamer (eds.), *Fibrous proteins: scientific, industrial and medical aspects. Based on the proceedings of the fourth International Conference on Fibrous Proteins.* (p. 69). Massey University, Palmerston North, New Zealand.

Rogers, G. E., Frenkel, M. J. & Lock, R. A. (1981). Ribonucleic acids coding for the keratin complex of hair. In C. E. Orfanos, W. Montagna & G. Stüttgen (Eds.*), Hair Research: Status and Future Aspects.* (pp. 84-93). Berlin, Springer-Verlag.

Frenkel, M. J. & Rogers, G. E. (1985). The isolation and characterization of two BIIIB high-sulphur keratin genes. In *Proceedings of the Australian Biochemical Society: abstracts of papers presented at the 29th Annual Meeting of the Society.* (p. 69). The Society, South Melbourne.

BIBLIOGRAPHY

BIBLIOGRAPHY

Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J.H. (1982) Cell 29, 319-328.

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

Baralle, F.E. and Brownlee, G.G. (1978) Nature 274, 84-87.

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

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

Berger, S.L. and Birkenmeier, C.S. (1979) Biochemistry 18, 5143-5149.

Berget, S.M. (1984) Nature 309, 179-182.

Bertolino, A.P., Gibbs, P.E.M. and Freedberg, I.M. (1982) J. Invest. Dermatol. 79, 173-177.

Blattner, F.R., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Richards, J.E., Slightom, J.L., Tucker, P.W. and Smithies, O. (1978) Science 202, 1279-1284.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L-A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) Science 196, 161-169.

Bollum, F.J. (1968) Methods Enzymol. XIIB, 169-173.

Bonner, W.M. and Laskey, R.A. (1974) Eur. J. Biochem. 46, 83-88.

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

Brunner, H. and Brunner, A. (1973) Eur. J. Biochem. 32, 350-355.

Case, S.T. (1982) Gene 20, 169-176.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.

Clarke, R.M. and Rogers, G.E. (1970) J. Invest. Dermatol. 55, 425-432.

Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. U.S.A. 62, 1159-1166.

Cohen, J.B. and Givol, D. (1983) EMBO J. 2, 1795-1800.

Collins, J. and Hohn, B. (1978) Proc. Natl. Acad. Sci U.S.A. 75, 4242-4246.

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

Crewther, W.G. (1976) Proc. Fifth Int. Wool Text. Res. Conf. Aachen 1, 1-101.

Crewther, W.G., Dowling, L.M. and Inglis, A.S. (1980) Proc. Sixth Int. Wool Text. Res. Conf. Pretoria 2, 79-91.

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

Crewther, W.G., Dowling, L.M., Gough, K.H., Marshall, R.C. and Sparrow, L.G. (1980) in "Fibrous Proteins: Scientific, Industrial and Medical Aspects". (D.A.D.Parry and L.K.Creamer, Eds. Academic Press, London) 2, 151-159.

Deng, G-r. and Wu, R. (1981) Nucleic Acids Res. 9, 4173-4188.

Denhardt, D.T. (1966) Biochem. Biophys. Res. Comm. 23, 641-646.

de Wet, J.R., Daniels, D.L., Schroeder, J.L., Williams, B.G., Denniston-Thompson, K., Moore, D.D. and Blattner, F.R. (1980) J. Virol. 33, 401-410.

Doney, J.M. and Evans, C.C. (1968) J. Agric. Sci. (Camb). 70, 111-116.

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

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

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

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

Downes, A.M., Chapman, R.E., Till, A.R. and Wilson, P.A. (1966a) Nature 212, 477-479.

Downes, A.M., Ferguson, K.A., Gillespie, J.M. and Harrap, B.S. (1966b) Aust. J. Biol. Sci. 19, 319-333.

Ebling, F.J. and Hale P.A. (1983) in "Biochemistry and Physiology of the Skin". (L.A.Goldsmith, Ed. Oxford University Press, New York) 1, 522-552.

Elleman, T.C., Lindley, H. and Rowlands, R.J. (1973) Nature 246, 530-531.

Ellis, W.J. (1948) Nature 162, 957.

Epstein, W.L. and Maibach, H.I. (1969) in "Advances in Biology of Skin: Hair Growth". (R.L.Dobson and W.Montagna, Eds. Pergamon Press, Oxford) 9, 83-97.

Fields, S. and Winter, G. (1981) Gene 15, 207-214.

Fraser, I.E.B. (1965) in "The Biology of Skin and Hair Growth". (A.G.Lyne and B.F.Short, Eds. Angus and Robertson, Sydney) 427-446.

Fraser, I.E.B. (1969a) Aust. J. Biol. Sci. 22, 213-229.

Fraser, I.E.B. (1969b) Aust. J. Biol. Sci. 22, 231-238.

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

Freedberg, I.M. (1970a) J. Invest. Dermatol. 54, 108-120.

Freedberg, I.M. (1970b) Biochem. Biophys. Acta 224, 219-231.

Freedberg, I.M. and Gilmartin, M.E. (1977) in "Biochemistry of Cutaneous Epidermal Differentiation". (M.Seiji and I.A.Berstein, Eds. University Park Press, Baltimore) 433-442.

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

Frenkel, M.J. and Blagrove, R.J. (1975) J.Chromatogr. 111, 397-402.

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.

Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) Cell 19, 959-972.

Fuchs, E. and Green, H. (1979) Cell 17, 573-582.

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

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.

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

Geisler, N., Plessman, U. and Weber, K. (1982) Nature 296, 448-450.

Gillespie, J.M. (1972) J. Chromatogr. 72, 319-324.

Gillespie, J.M. (1983) in "Biochemistry and Physiololgy of the Skin". (L.A.Goldsmith. Ed. Oxford University Press, New York) 1, 475-510.

Gillespie, J.M. and Frenkel, M.J. (1974a) Comp. Biochem. Physiol. 47B, 339-346.

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

Gillespie, J.M. and Frenkel, M.J. (1976) Proc. Fifth Text. Res. Conf. Aachen 2, 265-276.

Gillespie, J.M. and Marshall, R.C. (1981) in "Hair Research". (C.E.Orfanos, W.Montagna and G.Stuttgen, Eds. Springer-Verlag, Berlin) 76-83.

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

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

Gillespie, J.M., Haylett, T. and Lindley, H. (1968) Biochem. J. 110, 193-200.

Gillespie, J.M., Marshall, R.C. and Reis, P.J. (1981) Proc. 2nd Natl. Conf. Wool Harvesting Research and Development, Sydney. 111-115.

Gilmartin, M.E. and Freedberg, I.M. (1975) J. Invest. Dermatol. 64, 90-95.

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

Gough, N.M., Webb, E.A., Cory, S. and Adams, J.M. (1980) Biochemistry 19, 2702-2710.

Gray, J.C. (1974) Arch. Biochem. Biophys. 163, 343-348.

Gregg, K., Wilton, S.D., Parry, D.A.D. and Rogers, G.E. (1984) EMBO J. 3, 175-178.

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

Hanahan, D. and Meselson, M. (1983) Methods Enzymol. 100, 333-342.

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

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

Harris, S., Barrie, P.A., Weiss, M.L. and Jeffreys, A.J. (1984) J. Mol. Biol. 180, 785-801.

Hastings, K.E.M. and Emerson, C.P. Jr. (1983) J. Mol. Evol. 19, 214-218.

Haylett, T., Swart, L.S., Parris, D. and Joubert, F.J. (1971) Appl. Polym. Symp. 18, 37-44.

Haylett, T., Swart, L.S. and Parris, D. (1971) Biochem. J. 123, 191-200.

Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J. and Proudfoot, N.J. (1983) Nature 306, 398-400.

Hindley, J. and Phear, G.A. (1981) Biochem. J. 199, 819-823.

Hohn, B. and Collins, J. (1980) Gene 11, 291-298.

Hohn, B. and Murray, K. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 3259-3263.

Iatrou, K., Tsitilou, S.G. and Kafatos, F.C. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 4452-4456.

Jones, C.W. and Kafatos, F.C. (1982) J. Mol. Evol. 19, 87-103.

Jones, I.M., Primrose, S.B. and Ehrlich, S.D. (1982) Mol. Gen. Genet. 188, 486-489.

Joubert, F.J. and Burns, M.A.C. (1967) J. S. African Chem. Inst. 20, 161-173.

Joubert, F.J., de Jager, P.J. and Swart, L.S. (1968) in "Symposium on Fibrous Proteins. Australia 1967". (W.G.Crewther, Ed. Butterworths, Australia) 343-352.

Kao, F-T., Hartz, J.A., Law, M.L. and Davidson, J.N. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 865-869.

Kaplan, B.B., Bernstein, S.L. and Gioio A.E. (1979) Biochem. J. 183, 181-184.

Karin, M. and Richards, R.I. (1982) Nature 299, 797-802.

Kemp, D.J. (1975) Nature 254, 573-577.

Kemp, D.J., Partington, G.A. and Rogers, G.E. (1974) Biochem. Biophys. Res. Comm. 60, 1006-1014.

Kemp, D.J., Coppel, R.L., Cowman, A.F., Saint, R.B., Brown, G.V. and Anders, R.F. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3787-3791.

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

Kretschmer, P.J., Kaufman, R.E., Coon, H.C., Chen, M-J.Y., Geist, C.E. and Nienhuis, A.W. (1980) J. Biol. Chem. 255, 3204-3211.

Kuczek, E. and Rogers, G.E. (1985) Eur. J. Biochem. 146, 89-93.

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

Larson, R. (1982) University of Minnesota sequence analysis program version 2.1.

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

Lee, M.G-S., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) Cell 33, 477-487.

Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335.

Ley, K.F. and Crewther, W.G. (1980) Proc. Sixth Int. Wool Text. Res. Conf. Pretoria 2, 13-28.

Lienhard, G.E., Secemski, I.I., Koehler, K.A. and Lindquist, R.N. (1971) Cold Spring Harbour Symp. Quant. Biol. 36, 45-51.

Lindley, H. and Elleman, T.C. (1972) Biochem. J. 128, 859-867.

Lock, R.A. (1977) M.Sc. Thesis, University of Adelaide.

Lyne, A.G. and Short, B.F. (1965) "The Biology of Skin and Hair Growth". (Angus and Robertson, Sydney).

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) "Molecular Cloning. A laboratory manual". (Cold Spring Harbour Laboratory, New York).

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.

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

Marcu, K. and Dudock, B. (1974) Nucleic Acids Res. 1, 1385-1390.

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

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

Marshall, R.C., Gillespie, J.M. and Klement, V. (1985) J. Forens. Sci. Soc. 25, 57-66.

Marshall, R.C., Gillespie, J.M., Inglis, A.S. and Frenkel, M.J. (1980) Proc. Sixth Int. Wool Text. Res. Conf. Pretoria 2, 147-158.

Martin, S.L., Vincent, K.A. and Wilson, A.C. (1983) J. Mol. Biol. 164, 513-528.

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

Mercer, E.H. (1961) "Keratin and Keratinization: An essay in molecular biology". (Pergamon Press, Oxford).

Messing, J. and Vieira, J. (1983) Gene 19, 269-276.

Molloy, P.L., Powell, B.C., Gregg, K., Barone, E.D. and Rogers, G.E. (1982) Nucleic Acids Res. 10, 6007-6021.

Montagna, W. and Parakkal, P.F. (1974) "The Structure and Function of Skin". 3rd Ed. (Academic Press, New York).

Moore, G.P.M., Panaretto, B.A. and Robertson, D. (1982) Aust. J. Biol. Sci. 35, 163-172.

Murray, M.G., Kennard, W.C., Drong, R.F. and Slightom, J.L. (1984) Gene 30, 237-240.

McKnight, S.L. and Kingsbury, R. (1982) Science 217, 316-324.

O'Farrell, P.H., Kutter, E. and Nakanishi, M. (1980) Mol. Gen. Genet. 179, 421-435.

Orwin, D.F.G. (1976) J. Ultrastruct. Res. 55, 312-324.

Palmiter, R.D. (1974) Biochemistry 13, 3606-3615.

Parry, D.A.D., Fraser, R.D.B. and MacRae, T.P. (1979) Int. J. Biol. Macromol. 1, 17-22.

Partington, G.A., Kemp, D.J. and Rogers, G.E. (1973) Nature New Biol. 246, 33-36.

Perricaudet, M., le Moullee, J-M. and Tiollais, P. (1980) Nature 288, 174-176.

Pinder, J.C., Staynov, D.Z. and Gratzer, W.B. (1974) Biochemistry 13, 5373-5378.

Powell, B.C. and Rogers G.E. (1985) in "Biology of the Integument". (J.Bereiter-Hahn and A.G.Matoltsy, Eds. Springer-Verlag, Berlin). Vol.2. In Press.

Powell, B.C., Sleigh, M.J., Ward, K.A. and Rogers, G.E. (1983) Nucleic Acids Res. 11, 5327-5346.

Powell, B.C., Kemp, D.J., Partington, G.A., Gibbs, P.E.M. and Rogers, G.E. (1976) Biochem. Biophys. Res. Comm. 68, 1263-1271.

Presland, R.B. and Rogers, G.E. (1984) Proc. Aust. Biochem. Soc. 16, 35.

Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.

Quax, W., Egberts, W.V., Hendriks, W., Quax-Jeuken, Y. and Bloemendal, H. (1983) Cell 35, 215-223.

Radloff, R., Bauer, W. and Vinograd, J. (1967) Proc. Natl. Acad. Sci. U.S.A. 57, 1514-1521.

Reis, P.J. (1979) in "Physiological and Environmental Limitations to Wool Growth". (J.L.Black and P.J.Reis, Eds. The University of New England Publishing Unit). 223-242.

Respess, J., Erickson, J., Rushford, C., Jackson, D. and Schmickel, R. (1983) Gene 24, 331-334.

Rigby, P.W.J., Dieckmann, 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. U.S.A. 70, 2330-2334.

Rogers, G.E. (1964) in "The Epidermis". (W.Montagna and W.C.Lobitz, Eds. Academic Press, New York). 179-236.

Rogers, G.E. (1983) in "Biochemistry and Physiology of the Skin". (L.A.Goldsmith, Ed. Oxford University Press, New York). 1, 511-521.

Rogers, G.E. and Clarke, R.M. (1965) in "The Biolgy of Skin and Hair Growth". (A.G.Lyne and B.F.Short, Eds. Angus and Robertson, Sydney). 329-344.

Rogers, G.E., Frenkel, M.J. and Lock, R.A. (1981) in "Hair Research". (C.E.Orfanos, W.Montagna and G.Stuttgen, Eds. Springer-Verlag, Berlin). 84-93.

Rougeon, F. and Mach, B. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 3418-3422.

Rudall, K.M. (1956) Proc. First Int. Wool Text. Res. Conf. Australia F, 176-185.

Saint, R.B. (1979) Ph.D. Thesis, University of Adelaide.

Sanger, F. and Coulson, A.R. (1978) FEBS Letters 87, 107-110.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.

Schimenti, J.C. and Duncan, C.H. (1984) Nucleic Acids Res. 12, 1641-1655.

Shen, C-K.J. and Maniatis, T. (1980) Cell 19, 379-391.

Short, B.F., Wilson, P.A. and Schinckel, P.G. (1965) in "The Biology of Skin and Hair Growth". (A.G.Lyne and B.F.Short, Eds. Angus and Robertson, Sydney). 409-426.

Sim, G.K., Kafatos, F.C., Jones, C.W., Koehler, M.D., Efstratiadis, A. and Maniatis, T. (1979) Cell 18, 1303-1316.

Sims, R.T. (1967) Nature 213, 387-388.

Smith, A.J.H. (1979) Nucleic Acids Res. 6, 831-848.

Smith, G.E. and Summers, M.D. (1980) Anal. Biochem. 109, 123-129.

Southern, E. (1975) J. Mol. Biol. 98, 503-517.

Sparrow, L.G. and Crewther, W.G. (1972) J. Text. Inst. 63, 619-621.

Sparrow, L.G. and Inglis, A.S. (1980) Proc. Sixth Int. Wool Text. Res. Conf. Pretoria 2, 237-246.

Steinert, P.M. and Rogers, G.E. (1971a) Biochem. Biophys. Acta 232, 556-572.

Steinert, P.M. and Rogers, G.E. (1971b) Biochem. Biophys. Acta 238, 150-155.

Steinert, P.M. and Rogers, G.E. (1973a) Biochem. Biophys. Acta 312, 403-412.

Steinert, P.M. and Rogers, G.E. (1973b) Biochem. J. 135, 759-771.

Steinert, P.M., Idler, W.W. and Goldman, R.D. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 4534-4538.

Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.C. (1983) Nature 302, 794-800.

Steinert, P.M., Parry, D.A.D., Racoosin, E.L., Idler, W.W., Steven, A.C., Trus, B.L. and Roop, D.R. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713.

Sternberg, N., Tiemeier, D. and Enquist, L.(1977) Gene 1, 255-280.

Sutcliffe, J.G. (1978) Nucleic Acids Res. 5, 2721-2740.

Swart, L.S. (1973) Nature New Biol. 243, 27-29.

Swart, L.S. and Haylett, T. (1971) Biochem. J. 123, 201-210.

Swart, L.S. and Parris, D. (1974) Nature 249, 580-581.

Swart, L.S., Joubert, F.J. and Parris, D. (1976) Proc. Fifth Int. Wool Text. Res. Conf. Aachen 2, 254-263.

Taylor, J.M., Illmensee, R. and Summers, J. (1976) Biochem. Biophys. Acta 442, 324-330.

Thomas, P.S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5201-5205.

Townes, T.M., Shapiro, S.G., Wernke, S.M. and Lingrel, J.B. (1984) J. Biol. Chem. 259, 1896-1900.

Vanin, E.F. (1984) Biochem. Biophys. Acta 782, 231-241.

Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3683-3687.

Walker, M.D., Edlund, T., Boulet, A.M. and Rutter, W.J. (1983) Nature 306, 557-561.

Wallace, R.B., Johnson, P.F., Tanaka, S., Schold, M., Itakura, K. and Abelson, J. (1980) Science 209, 1396-1400.

Ward, K.A. and Kasmarik, S.E. (1980) J. Invest. Dermatol. 75, 244-248.

Watanabe, Y., Tsukada, T., Notake, M., Nakanishi, S. and Numa, S. (1982) Nucleic Acids Res. 10, 1459-1469.

Weaver, C.A., Gordon, D.F. and Kemper, B. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4073-4077.

Wilkinson, B.R. (1970a) Aust. J. Biol. Sci. 23, 127-138.

Wilkinson, B.R. (1970b) Aust. J. Biol. Sci. 23, 139-148.

Wilkinson, B.R. (1971) Biochem. J. 125, 371-373.

Woo, S.L.C. (1979) Methods Enzymol. 68, 389-395.

Woods, E.F. (1979) Aust. J. Biol. Sci. 32, 423-435.

Zahn, H. and Biela, M. (1968) Eur. J. Biochem. 5, 567-573.

Zehner, Z.E. and Paterson, B.M. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 911-915.