ORGANIZATION OF CHICKEN KERATIN GENES

A Thesis submitted to the University of Adelaide,
South Australia,
for the degree of Doctor of Philosophy,

by

RICHARD BRIAN PRESLAND B.Sc.Hons.(N.Z.),M.Sc.Hons.(N.Z.)

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

December, 1986

TABLE OF CONTENTS

APPENDICES

SUMMARY

The major gene product synthesized in the feather and other epidermal tissues of birds, such as scale and claw, is the intracellular protein, keratin. In the chicken, the feather keratins comprise a homologous family of about 20 proteins which are co-ordinately expressed during growth and differentiation in the embryonic feather. The work described in this thesis involves the analysis of the number and organization of feather keratin genes in the chick genome and sequence conservation outside the protein coding region amongst the gene family as a whole.

1. Previously a λ recombinant, λCFK1, had been isolated which contained four complete feather keratin genes and the 3' end of a fifth gene (Molloy et al., 1982). To extend these findings, the existing chick λ library was screened with 'walking' probes comprising DNA fragments from each end of the λCFK1 clone. One of the positive recombinants, λCFK9, extended λCFK1 rightwards by 2 kb. The clone λCFK9, as well as spanning part of λCFK1, contained the remainder of the gene which was only partially located in the λCFK1 isolate.

2. A chicken cosmid genomic library (provided by Dr. R. D'Andrea of this Department) was screened using one of the feather keratin genes from the λ clone as a probe. Fifteen positive cosmid recombinants were isolated and purified. Twelve of these clones were found to be derived from a single chromosomal locus, encompassing a continuous region 115 kb in length and including the genomic region contained in λCFK1 and λCFK9. Five

overlapping cosmid recombinants (Cosmids 4, 12, 25, 31 and 33), which together spanned 100 kb of this chromosomal region, were characterized in detail by restriction enzyme mapping and Southern blot analysis.

The feather keratin cluster, contained within these overlapping clones, consists of 18 genes spanning a segment of DNA 53 kb in length. The intergenic distance of about 3 kb centre-to-centre observed between the genes in λCFK1 (Molloy et al., 1982) is maintained throughout the entire gene cluster and at least 11 of the genes have the same transcriptional orientation.

Southern analysis using oligonucleotide probes made from highly conserved portions of the 5' non-coding, intron and 3' non-coding regions respectively, indicate that the gene family as a whole has been highly conserved, with one or two exceptions in each case.

Detailed restriction mapping of cosmids and derived plasmid subclones revealed regions containing symmetrical restriction enzyme sites; cosmid 4 contained 5 KpnI fragments of 3.4 kb and cosmid 31 5 HindIII fragments of 2.7 kb. These results are indicative of tandem duplication events in the recent history of this gene family.

3. The feather keratin gene locus is flanked on both sides by related types of keratin genes. On one side of the feather cluster are three feather-like genes which are located 5 kb away from the last feather keratin gene and are spaced 3-5 kb apart. The feather-like gene proximal to the feather keratin cluster was partially sequenced and encodes a protein of 115 amino acids, 18 amino acids

longer than the proteins encoded by feather keratin genes. The protein coding regions of the feather-like and feather keratin genes are about 75% homologous at the DNA and amino acid sequence levels. On the other side of the feather gene cluster, 21 kb distant from the last feather gene, is a cluster of three or four claw keratin genes. The claw genes are spaced 1-2 kb apart and two of them appear to form a divergently orientated gene pair (L. Whitbread, personal communication).

4. Genomic Southern analysis of chick DNA using a feather keratin gene as probe demonstrated that the 18 feather keratin genes contained in the characterized locus account for about three-quarters of the feather keratin genes present in the chick genome. The location of the remaining genes is at present unknown.

5. DNA sequence analysis of feather keratin genes and hybridization studies using feather keratin cDNA (Molloy et al., 1982; Gregg et al., manuscript in preparation; this work) have demonstrated that the feather keratin genes contain an intron of between 320-340 bp in the 5' non-coding region, which interrupts them 21 or 22 bp prior to the initiation codon.

To identify and partially characterize the putative precursor mRNA for feather keratin, whole cell and nuclear RNAs were isolated from keratinizing, embryonic feather tissue and Northern blot analysis performed with two intron-specific probes. However, attempts to detect the feather keratin pre-mRNA by these methods were not successful.

S T A T E M E N T

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 it contains no material that has been previously published or written by any other person, except where due reference is made in the text.

Signed.....

RICHARD B. PRESLAND

ACKNOWLEDGEMENTS

I would like to thank Professor W.H. Elliott for permission to work in the Department of Biochemistry, University of Adelaide.

I sincerely thank my supervisor, Professor George Rogers for his invaluable advice and support throughout the course of this work.

I would like to thank the following members of 'Keratin Korna', past and present, for their encouragement, criticisms and friendship:

Simon Bawden, Michael Calder, Dr. Graham Cam, Lesley Crocker, Dr. Richard D'Andrea, Cornelia Fieles, Michael Fietz, Dr. John Forrest, Dr. Keith Gregg, Dr. Anna Koltunow, Dr. Elizabeth Kuczek, Phil Mackinnon, Dr. Peter Molloy, Dr. Phil Morris, Phil Pain, Dr. Barry Powell, Dr. Joe Rothnagel, Imbi Semenov, Dr. A.V. Sivaprasad, Dr. Brandon Wainwright, Lel Whitbread and Dr. Steve Wilton.

I am especially grateful to Dr. Barry Powell for many helpful discussions and for reading drafts of this thesis; Dr. Richard D'Andrea for providing me with the chick cosmid library and also for reading part of this thesis; Imbi Semenov for preparing DNA from some of the cosmid recombinants and plasmid subclones and help in preparing some figures; Lesley Crocker for preparing and photographing many figures; Lel Whitbread for formatting some DNA sequence figures and chapter headings using the T_EX typesetting program and Dr. Graham Cam for help with proof reading.

I am deeply indebted to my wife, Hania, for drawing all of the line diagrams, for help with proof reading and for her constant support, patience and understanding.

I wish to thank my father-in-law, Dr. Richard Francki, for his advice and encouragement. I also thank Leanne Goodwin for typing this thesis.

I was supported during this work by a Postgraduate Scholarship from the Commonwealth Scholarship and Fellowship Plan. This work was supported by grants from the Australian Research Grants Scheme and the Australian Wool Corporation (to Professor G.E. Rogers).

ABBREVIATIONS

A_n	:	Absorbance of light as measured at wavelength n (1 cm path length).
bp	:	Nucleotide base pairs
BPB	:	Bromophenol blue
BSA	:	Bovine serum albumin
cDNA	:	Single-stranded DNA complementary to mRNA
Ci	:	Curie
cpm	:	Counts per minute
ddNTP	:	2',3'-dideoxynucleoside 5'-triphosphate
DMSO	:	Dimethyl sulphoxide
DNase	:	Deoxyribonuclease
dNTP	:	2'-deoxynucleoside 5'-triphosphate
DTT	:	Dithiothreitol
kb	:	Kilobase pairs
mRNA	:	Messenger ribonucleic acid
Myr	:	Million years
oligo (dT)	:	Oligodeoxythymedylic acid
pfu	:	Plaque forming units
poly A^+ RNA	:	Polyadenylated ribonucleic acid
RF	:	Replicative form of M13
RNase	:	Ribonuclease
rpm	:	Revolutions per minute
SDS	:	Sodium dodecyl sulphate
TCA	:	Trichloroacetic acid
TEMED	:	N,N,N',N'-tetramethyl ethylene diamine
UV	:	Ultra-violet

Other abbreviations are as listed in the Biochemical Journal (1976) 153, 1-21.

CHAPTER 1.

INTRODUCTION.

1.1 GENERAL INTRODUCTION

Keratins are a large group of structural proteins which are found in certain epithelia of higher organisms. Keratinized epithelial tissues are characteristically durable, pliable and insoluble and serve to protect the soft body tissues from environmental influences such as physical damage, bacterial infection, dehydration and heat loss. They include skin, hair, horns and nails in mammals and feather, scale, beak and claws in birds and reptiles and are likely to have played an important role in the colonization of land by vertebrates.

The formation of cornified epithelia is a terminal differentiation process and leads to the death of the keratin-synthesizing cells. In this laboratory we are attempting to elucidate the molecular basis of this process in the feather epidermis of developing chick embryos. This has proven to be a particularly amenable system for these studies because of the ready availability of feather tissue. The general aim of the research reported in this thesis was to determine the organization of feather keratin genes in the chicken genome and their spatial relationship to other keratin gene families, such as scale or claw keratins. Characterization of the organization of this, or any gene family, is an essential pre-requisite to studies of the mechanisms affecting their co-ordinate expression. It was also hoped that some clues pertaining to how the avian keratin gene families evolved might be obtained.

This chapter is intended to provide some background information on the processes of chick embryonic feather development and keratinization at both the morphological and molecular levels. These processes will be compared with those occurring

in other avian epidermal tissues such as scale and in mammalian tissues. A brief summary of current knowledge of the structure and regulation of the genes expressed in the developing avian feather will also be included.

1.2 THE BIOLOGY OF FEATHER DEVELOPMENT

1.2.1. FEATHER MORPHOGENESIS

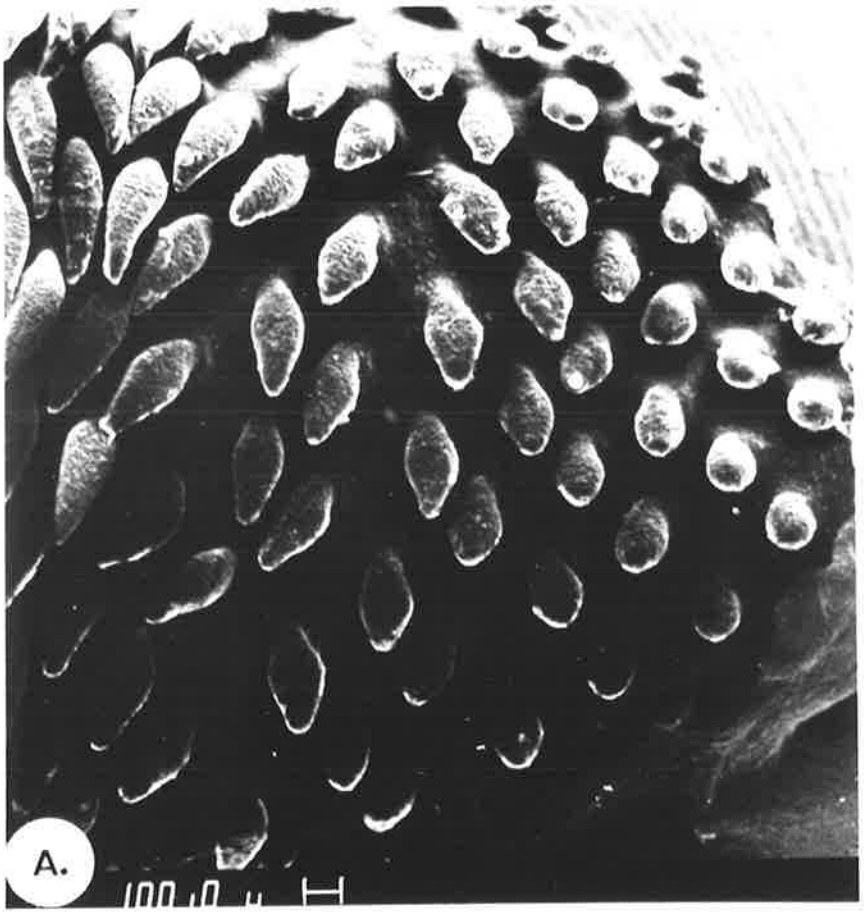
The development of embryonic chick feathers has been studied extensively using both light microscopy (for reviews see Romanoff, 1960; Lillie, 1965; Wessells, 1965; Voitkevich, 1966; Lucas and Stettenheim, 1972) and electron microscopy (Kischer, 1963; Matulionis, 1970). The following discussion will summarize the findings of these studies.

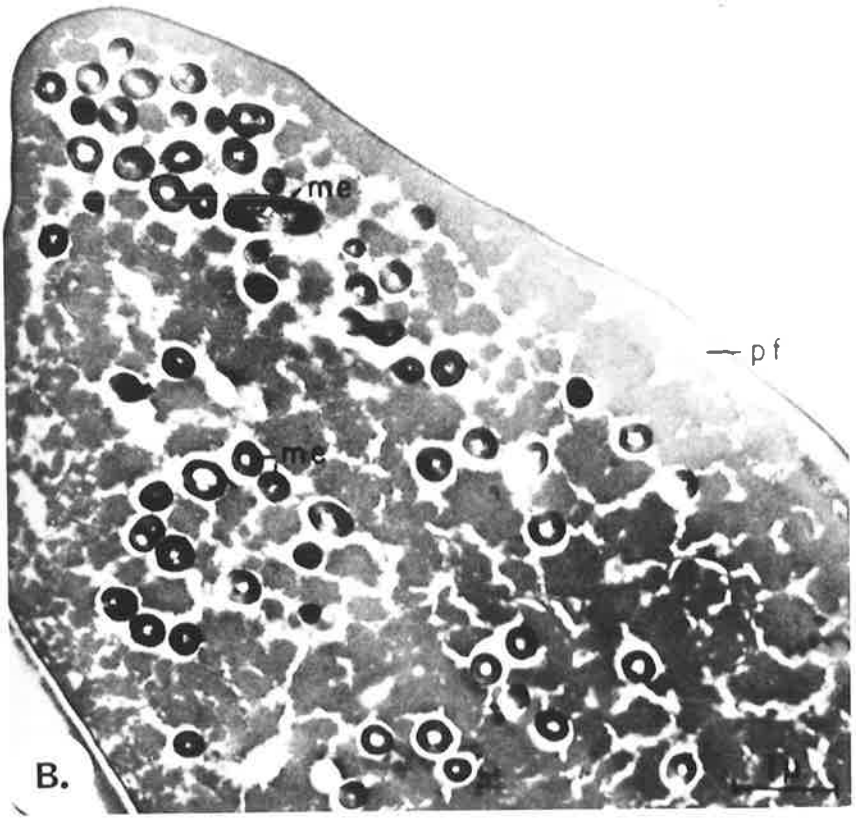
At 5 days of embryonic growth in the chick, the skin consists of a layer of epithelial cells overlain by peridermal cells and underlain by mesenchyme (mesoderm). About this time the primordia of feathers appear. The process begins with the condensation of mesenchymal cells into a layer of dermis (Wessells, 1961a) and concurrent rapid proliferation of epidermis, producing clusters of epidermal cells known as placodes. During the next two to three days, dermal cells begin to congregate beneath these epidermal thickenings giving rise to rudimentary feathers. These 'feather germs' elongate by rapid cell division and become macroscopically visible by day 9 of development as epidermal cylinders of 0.5 to 1.0 mm in size arranged in well defined tracts (Figure 1.1.a). Lucas and Stettenheim (1972) have defined three epidermal layers at this stage: an outer epidermally derived layer which is destined to become the feather sheath; an intermediate layer, also epidermally derived, which gives rise to the barbs and barbules of the

FIGURE 1.1

A. Scanning electron micrograph of chick feathers from a 9-day old embryo, showing the well-defined feather tracts (photograph courtesy of G.E. Rogers).

B. Electron micrograph of a developing feather barb cell in cross-section at an advanced stage of development. The cell contains a large number of keratin fibrils, some of which form a dense layer (pf) beneath the membrane enclosing the barb cell. Melanin granules (me) are present.
Magnification, x 15,000.
(from Filshie and Rogers, 1962).


A.


me
me
pf
B.

feather; and a dermally derived inner layer which forms the pulp of the growing feather (Figure 1.2).

By about the 10th day of incubation, the cells of the intermediate layer organize into a pair of ridges, called barb ridges, that are parallel to the long axis of the feather germ (Bell and Thathachari, 1963); these barb ridges are destined to become the barbs and barbules of the completed down feather (Figure 1.2). On the 11th day the epidermis at the base of the feather begins to push downward into the dermis to form the feather follicle, which by day 12 is composed of a series of barb ridges organized into a cylindrical arrangement. During this period feather growth proceeds chiefly by cell division at the base of the follicle which produces a feather more differentiated at its tip than its base.

By day 13, when keratinization commences, cell division has essentially ceased and feather growth is due to cell elongation and movement (Bell and Thathachari, 1963; Kischer and Furlong, 1967; Kemp et al., 1974a). Over the next few days the downy barbules form as columns of cells in the barb ridges (Figure 1.2). At the onset of keratinization fibrils of ~ 3 nm are visible which extend throughout the length of the barb ridge cells (Filshie and Rogers, 1962; Bell and Thathachari, 1963) and coincides with the appearance of feather keratin proteins (Kemp et al., 1974a). This terminal differentiation process commences from the tip of the feather and proceeds towards the proximal end in both the barb and barbule cells as well as the sheath cells. During the next few days keratin synthesis proceeds rapidly, until day 18 or 19 when the feather cells have filled with keratin fibrils (Figure 1.1.b)

FIGURE 1.2 : SCHEMATIC DRAWING OF A LONGITUDINAL SECTION OF AN 11-DAY EMBRYONIC CHICK FEATHER

ax. pl. - Barb ridge axial plate

B'bl. - Barbule

B'bl. pl. - Barbule plate

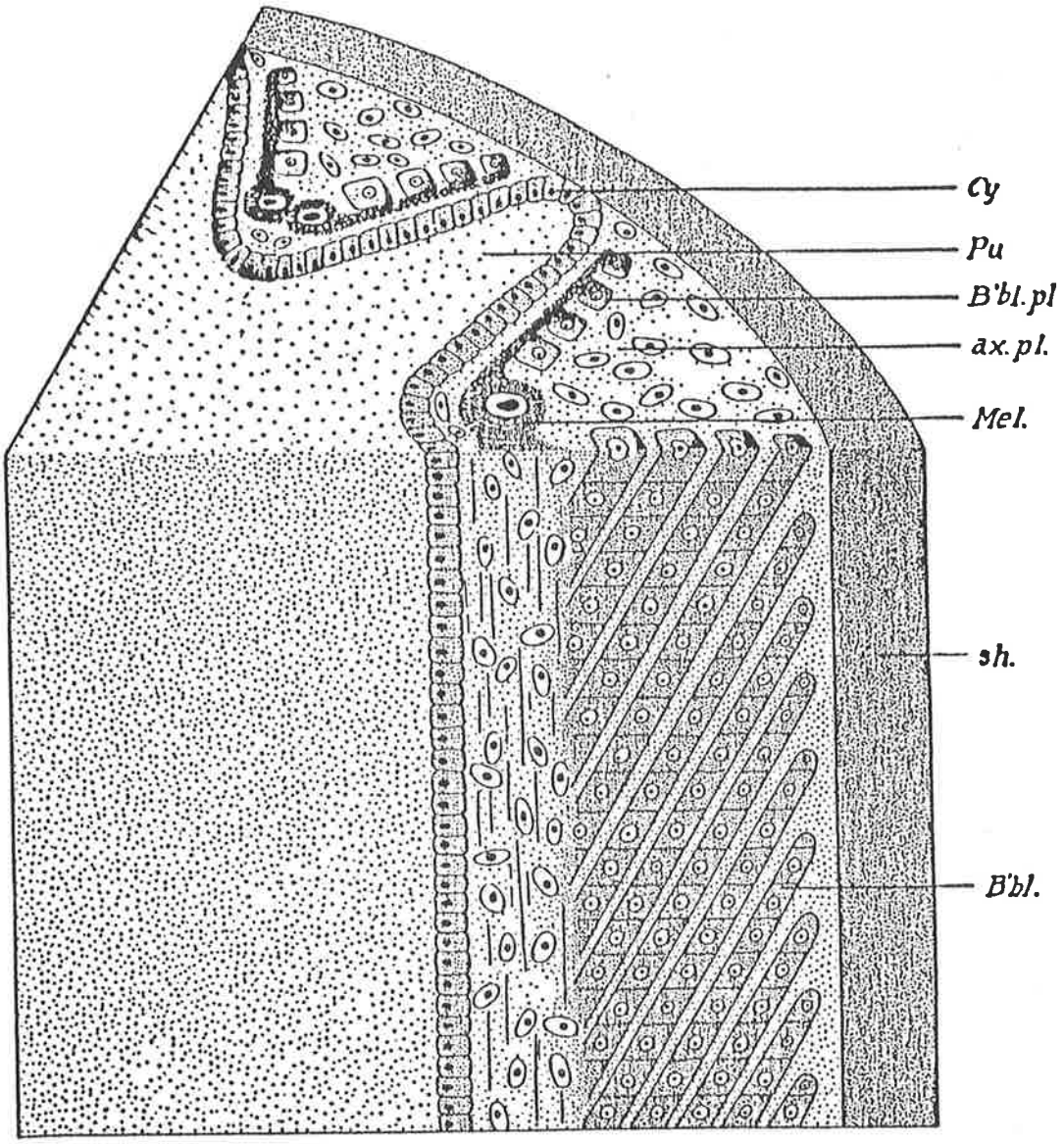
Cy. - Cylinder cells

Mel. - Melanophore

Pu. - Pulp

sh. - Feather sheath

(From Lillie, 1965).


Cy
Pu
B'bl.pl
ax.pl.
Mel.
sh.
B'bl.

which coalesce to become cross-linked into a tough lattice (Fraser et al., 1972). The organelles and cell boundaries of the feather cells are resorbed, culminating in cell death (Matulionis, 1970).

After the chicken has hatched at 21 days, the feathers dry out and the sheath of each feather dries and flakes away from the bundle of barbs. The barbs spread out and the new down feathers become fluffy. In the domestic chicken there is a single downy plumage which is replaced early in the life of the adult by the adult plumage. These new feathers grow from the established follicles, displacing the original down feather which initially remains attached to the distal tip of the growing feather.

1.2.2. THE ROLE OF THE DERMIS

The most immediate control over keratinizing tissues, from outside the epidermis, is exerted by the underlying dermis or mesenchyme. This has been demonstrated using embryonic skin which has been separated into its component dermis and epidermis and cultured after recombining dermis and epidermis components from different tissues or even species. (Sengel, 1958, 1971, 1986; Wessells, 1962, 1965; Rawles, 1963, 1965; Kato, 1969; Dhouailly et al., 1978; Sawyer, 1979). Studies using these heterotopic recombinants have revealed that the epidermis develops structures characteristic of those normally associated with the dermis used, irrespective of the origin of the epidermis. For example, the recombination of presumptive chick dorsal feather dermis with tarsometatarsal scale-forming chick epidermis induces the formation of feathers (Sengel, 1958; Rawles, 1963; Dhouailly et al., 1978). Conversely, the

culturing of presumptive scale-forming dermis with feather-forming epidermis produces scales. Dhouailly et al. (1978) demonstrated that the keratin polypeptides and other proteins synthesized in these heterotopic recombinants were identical to those found in normal epidermal structures. Thus, for example, the recombination of presumptive feather dermis from the back of the chick embryo with presumptive scale epidermis causes the epidermis to form feathers and express the set of feather keratin proteins, rather than the set of scale proteins that would normally be expressed in this epidermal tissue.

The age of the dermis is critical for directing epidermal development. In the chick, the alterations in developmental fate can only be induced if the feather epidermis is from an embryonic chick aged 7-10 days and the scale dermis is from a 13-16 day old chick (Rawles, 1963, 1965). The difference in ages at which scale and feather components are receptive to these influences is probably due to the observed differences in the time course of keratin protein synthesis in these two epidermal tissues (Kemp et al., 1974a; Wilton et al., 1985). In summary, the chick dermis reaches a stage at which it begins to exert an influence over the epidermis and the epidermis passes through a phase during which it is receptive to that influence. Once keratinization begins, the direction of epidermal differentiation is not reversible by dermal factors.

Studies by Dhouailly (1967) using heterospecific dermal-epidermal recombinants have shown that duck dermis could direct chick epidermis to produce feathers which resembled those of duck. Furthermore, Dhouailly et al. (1978) demonstrated that mouse footpad dermis could direct the formation of a footpad-

like structure in presumptive chick feather epidermis; the footpad-like structure synthesized both α- and β-keratins which were electrophoretically indistinguishable from normal chick scale epidermis. It appeared that the dermal signal from the mouse footpad was recognized by the chick feather epidermis since the proteins produced were of the scale type, implying that this dermal signal is not species specific and has been evolutionarily conserved.

It is not clear whether the dermally-derived signal is mediated by a diffusible extracellular substance, such as a protein, or by direct cell to cell contact, although Wessells (1962) demonstrated that isolated epidermis would grow and differentiate to some extent, even if a millipore filter was interposed between the dermis and epidermis prior to culturing. The dermis is also essential for the maintenance of epidermal cell proliferation, however, artificial substrates with appropriate nutrients can substitute for this function of the dermis. It has also been shown that dermis killed by freeze thawing is able to fulfil the functions of maintaining the mitosis, orientation and spreading of the epidermis (Dodson, 1963).

The relationship between dermal and epidermal development is complicated by other observations (Linsenmayer, 1972) which indicate that the epidermis may influence the dermis in a positive way.

1.2.3 HORMONES IMPLICATED IN FEATHER DEVELOPMENT

A number of hormones of the pituitary and thyroid glands as well as vitamin A are required for normal feather development but little is known of the molecular mechanisms involved (Voitkevich, 1966).

Vitamin A seems to be required for normal keratinization, since a deficiency of vitamin A may lead to squamous metaplasia and keratinization of epithelia which do not usually undergo such differentiation. Conversely, when undifferentiated embryonic chick skin was cultured with excess vitamin A, it underwent a mucous metaplasia and in some cases formed tracts of actively beating ciliated cells (Fell and Mellanby, 1953); the synthesis of keratin proteins is completely repressed under these conditions (Beckingham-Smith, 1973). Using the vitamin A derivative, retinoic acid, Dhouailly et al. (1980) were able to induce the synthesis of feathers on the scaled epidermis of the chick foot.

Thyroxine has been shown to accelerate epidermal keratinization (Bartels, 1943; Wessells, 1961b; Kitano and Kuroda, 1967) and coincidentally, the thyroid gland attains maximum thyroxine secretion at about the same time as the onset of keratin synthesis in the feather (Shain et al., 1972). Removal of the pituitary abolishes keratin synthesis apparently by preventing polysome formation (Yatvin, 1966a). Normal development can be restored by injecting pituitary gland extracts (Yatvin, 1966b). Hydrocortisone has been reported to stimulate α-keratin synthesis in skin cultures (Fell, 1962; Sugimoto and Endo, 1969) and causes feather primordia destined to synthesize β-keratins to abort. This hormone, when added to cultures of scale epidermis has been reported to stimulate α-keratin synthesis while abolishing the synthesis of scale β-keratins (Sugimoto et al., 1974).

1.3 STRUCTURAL PROTEINS OF THE EMBRYONIC CHICK FEATHER

In the developing feather of chick embryos there are two major families of structural proteins synthesized; the β-keratins and fast proteins (or histidine-rich proteins). The properties of these two protein families are presented below. Current knowledge of other types of β-keratin protein families, as well as α-keratins, are also discussed briefly.

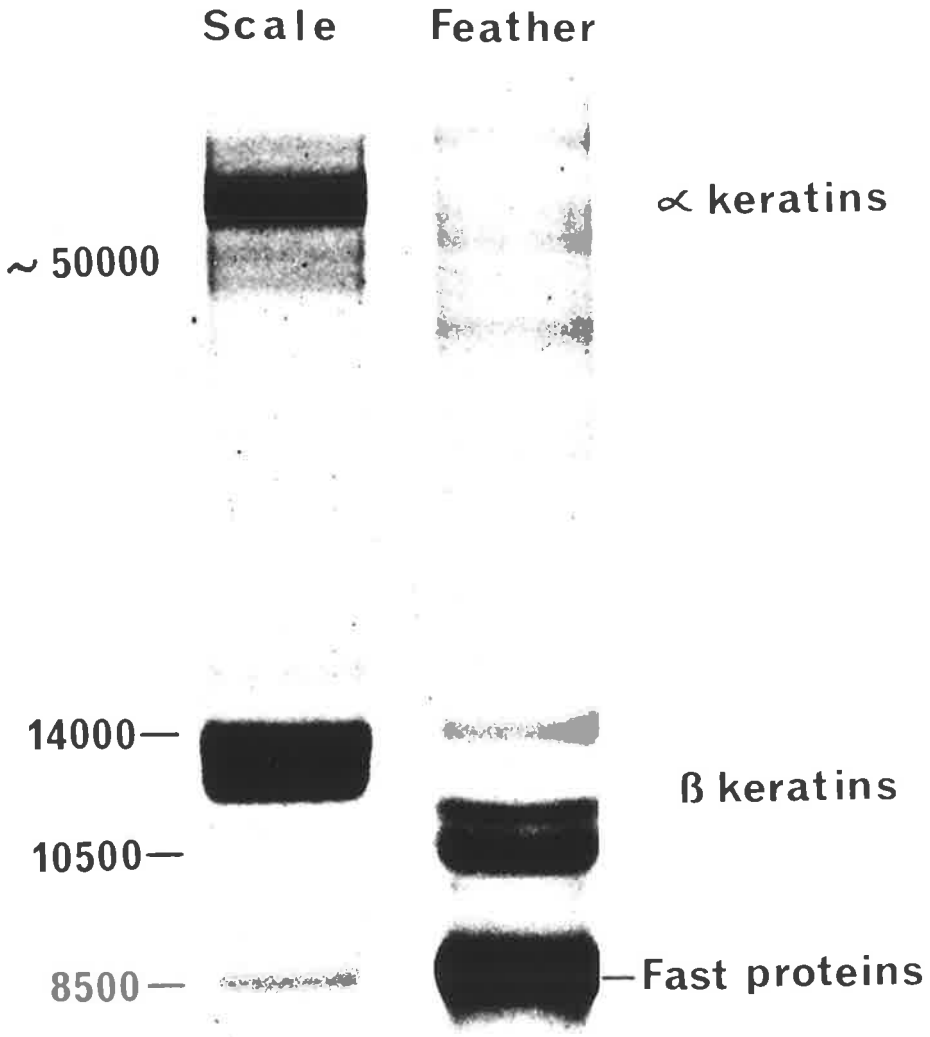
1.3.1. THE FEATHER β-KERATINS

The feather β-keratins are fibrous, insoluble intracellular proteins which are characterized by their high molar content of cysteine, serine, glycine and aliphatic amino acids with little, if any, lysine, histidine or tyrosine (Harrap and Woods, 1964a; Walker and Rogers, 1976a). Due to the stabilization of these proteins by intermolecular disulphide bonding (Goddard and Michaelis, 1934), their study requires the cleavage of the disulphide bonds by reducing agents followed by stabilization by alkylation, e.g. carboxymethylation (Harrap and Woods, 1964a; Kemp and Rogers, 1972). Using these methods feather keratins can be readily isolated from newly hatched chicks in which they comprise about 90% of the extractable feather protein (Walker and Rogers, 1976a). The feather keratins are homogenous in molecular weight, with published estimates varying between 10,500 and 11,500 daltons (Figure 1.3; Harrap and Woods, 1964b; Walker and Rogers, 1976a).

Protein sequencing data of keratins from adult feather tissue (calamus) of emu (O'Donnell, 1973) and silver gull (O'Donnell and Inglis, 1974), and from barbs of adult chick feather (Arai et al., 1983) all revealed protein chains of

FIGURE 1.3 : ELECTROPHORETIC ANALYSIS OF CHICK SCALE AND FEATHER PROTEINS ON AN SDS POLYACRYLAMIDE GEL

S-carboxymethylated proteins (Section 1.3.1) from 17-day feathers and 18-day anterior shank scales fractionated by electrophoresis on a 10% SDS polyacrylamide gel. The position and molecular weights of α-keratins, β-keratins and fast proteins are indicated at the appropriate position on the gel. (Modified from Dhouailly et al., 1978).


Scale
Feather
~ 50000
∝ keratins
14000
10500
8500
ß keratins
Fast proteins

about 100 residues (or a molecular size of ~ 10,000 daltons) and amino acid sequences that were highly homologous to each other. The proteins, particularly the adult chick feather keratin chain, were remarkably similar to embryonic chick feather keratins in both protein size and sequence (see Figure 1.6). In an earlier study, Kemp and Rogers (1972) reported that embryonic feather tissue contained some keratin chains that appeared to be present in much reduced amounts or even absent from adult feather tissue. At present, therefore, the question of whether there are separate sets of keratin genes expressed in embryonic and adult feather tissue is not completely resolved.

Walker and Rogers (1976a) separated protein extracts from feather tissue of newly hatched chickens into 22 different polypeptide species using ion exchange chromatography and polyacrylamide gel electrophoresis under acidic and alkaline pH conditions. Amino acid analysis demonstrated that all except one of the 9 crude fractions obtained had very similar amino acid compositions which were typical of total keratin, e.g. high in cysteine and serine. The remaining fraction had a markedly different composition and migrated differently on polyacrylamide gels (Walker and Rogers, 1976a). From this, and their other findings, they suggested that there are about 19 keratin proteins synthesized in feather tissue and that they are the products of non-allelic genes. Partial sequence analysis of some of these purified fractions indicated that most, and probably all of these feather chains were the products of highly homologous but non-identical genes (Walker and Rogers, 1976b).

All avian β-keratins (e.g. embryonic and adult feather keratins, scale and claw keratins) are able to form oriented filaments *in vivo* about 3 nm in diameter (Fraser *et al.*, 1972) and, when analysed by X-ray diffraction techniques, gave a characteristic β-type diffraction pattern (Astbury and Marwick, 1932; Stewart, 1977). The currently accepted model proposed by Fraser *et al.* (1971) which accounts for the X-ray spacings and the electron microscopy data (Filshie and Rogers, 1962) is shown in Figure 1.4a. The unit of structure is a pleated sheet of four peptides each of eight amino acids in length. A linear array of these units follows a right-handed helical path and two sets of these, running in opposite senses (i.e., anti-parallel), make up the core filament (Figure 1.4b). It is thought that each sheet of four units is derived from a single polypeptide chain. In support of this proposal, it can be demonstrated by using Chou and Fasman (1978) rules that the feather sequence between residues ~ 27-62 would adopt a highly regular conformation of 4-5 strands of the β-sheet type (see Figure 1.9a, Gregg *et al.*, 1984). Thus, with each unit constructed in this fashion, the amino and carboxy ends of each molecule containing the half-cysteine residues would probably be able to produce covalent and non-covalent interactions between adjacent filaments.

1.3.2 OTHER AVIAN β-KERATIN FAMILIES

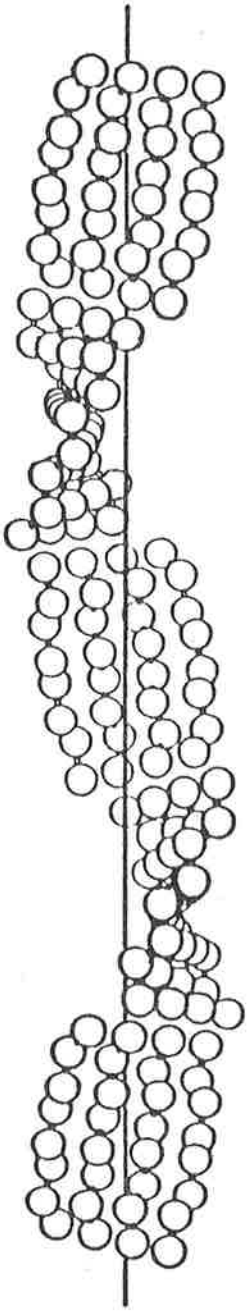
Avian species have three other β-keratin protein families, viz. the scale, beak and claw keratins which comprise the major structural proteins of these epidermal appendages. The β-keratins of chick scale have been extensively studied at the protein and DNA levels (Harrap and Woods, 1964a; Walker and

FIGURE 1.4 : MOLECULAR MODEL FOR THE CORE FILAMENT OF FEATHER KERATIN

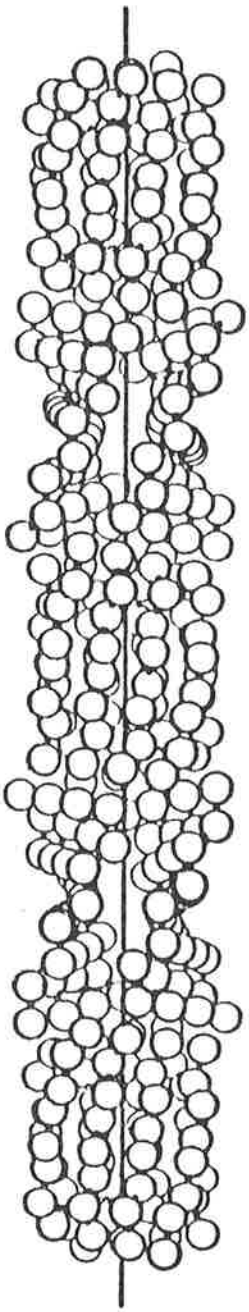
(a) A right-handed strand of four twisted β-pleated sheets, each sheet being made up of units of four anti-parallel chains of eight residues.

(b) Two of the strands shown in (a) assembled to form the core filament.

(After Fraser et al., 1971).



a



b

Bridgen, 1976; Wilton, 1978, 1983; Wilton et al., 1985). The scale keratins comprise a family of 9-12 closely related proteins with a molecular weight of about 14,000 daltons (Figure 1.3), about 40% bigger than the feather keratins (Walker and Bridgen, 1976; Wilton, 1983; Wilton et al., 1985). The difference in molecular weight between the scale and feather keratins can be accounted for almost entirely by the presence in the central portion of all scale proteins of a repeating tripeptide of the form gly-gly-x. This was originally demonstrated by partial amino acid sequencing of scale proteins (Walker and Bridgen, 1976) and more recently by sequencing of scale cDNA clones and genes isolated from a chicken λ library (see Section 1.4.6; Wilton, 1983; Gregg et al., 1984; Wilton et al., 1985). The protein sequencing data also demonstrated that the amino- and carboxy-terminal regions of the feather and scale proteins were very similar in sequence, e.g. rich in cysteine, proline and serine, indicating that feather and scale genes may have had a common ancestry (see Section 1.4.6).

Until recently, very little was known about the beak and claw keratin proteins. Analysis of the keratins isolated from beak and claw tissue of chickens have shown that they have a similar molecular weight (about 14,500) and amino acid composition to chick scale keratins (Frenkel, 1975; Frenkel and Gillespie, 1976; Gibbs, P.E.M., Walker, I.D. and Rogers, G.E., unpublished results), supporting the view that the beak and claw keratins are more closely related to scale than feather keratins. Recently, a cDNA clone (CCK22) has been isolated and sequenced in this laboratory which was identified as a claw

keratin (Whitbread, 1985). As suggested by the protein chemical data, DNA sequence analysis of a cDNA clone and a gene coding for claw keratins showed that they contain a gly-gly-x repeat in the central portion of the protein coding region (L. Whitbread and K. Gregg, unpublished results).

1.3.3. THE FAST PROTEINS (HISTIDINE-RICH PROTEINS)

These proteins, originally named 'fast' proteins because of their high electrophoretic mobility on low -pH polyacrylamide gels (Figure 1.3), were first reported by Walker and Rogers (1976a) to be present in embryonic but not adult feathers of the chicken. In newly hatched chicks they have variously been reported to comprise between 10% to 40% of total feather protein (Walker and Rogers, 1976a; Morris, 1984).

They appear to be a family of 3-5 related polypeptide chains and have an amino acid composition which is strikingly different from keratins, being rich in histidine, tyrosine and phenylalanine but deficient in cysteine, alanine and isoleucine in contrast to keratins. The molecular weight estimate of fast protein, obtained by the conventional technique of polyacrylamide gel electrophoresis, was found to be 8,400 daltons (Figure 1.3; Walker and Rogers, 1976a; Powell and Rogers, 1979). Recently, however, sequencing of a fast protein gene isolated from a chick λ library has demonstrated that at least one, and probably all of the fast proteins are polypeptides of 119 amino acids with a molecular weight of 14,000 (see Section 1.4.5). Fast proteins are co-ordinately synthesized with the keratins in the embryonic feather (Powell and Rogers, 1979); they are also present in some other keratinizing tissues, but at much lower levels (Wilton, 1983; Wilton et al.,

1985; Gibbs, P.E.M., Walker, I.D. and Rogers, G.E., unpublished data). Fast proteins are absent from all adult tissues (Walker and Rogers, 1976a; Wilton et al., 1985) which implies that, whatever their function, they are no longer required in the hardened dry keratinized tissue of adult chickens.

1.3.4 THE α-KERATINS

Powell (1979), in studies on proteins synthesized in feather tissue during the early stages of feather development, observed that cysteine-containing proteins in the molecular weight range 50 to 70 kD are present as major species. As differentiation proceeds, these proteins comprise less and less of the total feather protein until they are almost undetectable by day 21. Dhouailly et al. (1978) showed that these high molecular weight proteins isolated from feather tissue co-migrated with the α-keratin proteins of scale (Figure 1.3) and they speculated that they may have originated from the feather sheath, which is sloughed off at hatching and conceivably has the same composition as interplumar stratum corneum. It is interesting that the level of α-keratins present in scale tissue is significantly higher than that found in the feather (Dhouailly et al., 1978; Sawyer et al., 1986). These proteins isolated from scale tissue give an α-type X-ray diffraction pattern (MacRae, T.P., personal communication). Moreover, it has been demonstrated that one type of metatarsal chick scales, the reticulate scales, are devoid of β-keratins and only the α-keratin genes are expressed (Sawyer et al., 1986).

These high molecular weight proteins which are present in significant amounts in early feather development, and in scale tissues, have a similar molecular weight range to the mammalian

intermediate filament (IF) proteins (Steinert et al., 1985) and antibodies prepared against human IF keratins cross-reacted with a group of chicken epidermal proteins which had molecular weights of between 50 and 70 kD (Fuchs and Marchuk, 1983). Taken together, these results strongly suggest that these 50 to 70 kD proteins are probably the α-keratins.

1.4 MOLECULAR BIOLOGY OF FEATHER KERATINS IN THE DEVELOPING CHICK EMBRYO

The abundant synthesis of keratins in the developing feather of chick embryos facilitated the purification of feather keratin mRNA and subsequent studies on the structure and expression of these mRNAs, as well as estimates of the gene reiteration frequency in the chicken genome. More recently, the structure of the feather keratin gene has been determined by the DNA sequence analysis of several genes from a λ Charon 4A clone isolated from a chicken genomic library. A summary of these findings is presented in this section.

1.4.1 THE TIME COURSE OF PROTEIN SYNTHESIS

The synthesis of feather keratins begins at about the 12th day of embryonic development, as judged by the appearance of keratin fibrils in feather cells (Matulionis, 1970) and of keratins and fast proteins, as detected by polyacrylamide gel electrophoresis (Kemp et al., 1974a; Powell, 1979). The major keratin protein bands which were resolved on alkaline polyacrylamide gels (Kemp et al., 1974a) and the fast protein band, detected on SDS gels (Powell, 1979), were present in trace amounts in 11 and 12-day feather extracts but rapidly increased in quantity after day 12 to represent over 90% of the total proteins detected by day 15. By the 18th day of embryonic

development, most protein synthesis has essentially ceased and the feather completely cornified. Pulse labelling studies revealed that the different keratin bands resolved on acrylamide gels were synthesized co-ordinately (Kemp et al., 1974a). Other evidence suggested that different keratin polypeptide chains may be synthesized in different parts of the adult feather (Kemp and Rogers, 1972).

1.4.2. KERATIN mRNA: STRUCTURE AND TIME COURSE OF ITS SYNTHESIS DURING EMBRYONIC DEVELOPMENT

Feather keratin mRNAs were first purified over ten years ago by sucrose gradient centrifugation of polysomal RNA and the translatable mRNAs sediment at about 12S (Partington et al,. 1973; Kemp et al., 1974b). They have a molecular weight of 250,000 daltons (or 750-800 nucleotides) as determined by polyacrylamide gel electrophoresis in the presence of 98% formamide (Kemp et al., 1974b). Purified feather keratin mRNA contains a 7-methyl guanosine 'cap' structure as the 5' end residue of the mRNA and a poly (A) tail with an average length of 65 nucleotides at the 3' end (Morris and Rogers, 1979). These two characteristics are common features of eukaryotic mRNAs (for reviews see Brawerman, 1974; Revel and Groner, 1978).

The detailed structure of feather keratin mRNA remained, for the most part, an enigma until the advent of recombinant DNA techniques. However, Kemp (1975) and later, Lockett et al. (1979) used cDNA prepared from purified feather keratin mRNA to show that it is comprised of a unique sequence at the 3' end which is attached to a reiterated portion. They concluded that these sequences comprised the 3' non-coding and coding regions,

respectively. This model was later confirmed by DNA sequencing (Section 1.4.4.).

Powell et al. (1976) used labelled cDNA prepared from feather keratin mRNA to measure the amounts of homologous mRNAs in feather tissue of various ages. They observed that a very rapid increase in the amount of feather keratin mRNA occurred between day 11 and day 14 of development, from very low levels on day 11 to maximum levels (more than 3×10^5 molecules keratin mRNA/cell) at day 15. They proposed that, since the levels of keratin mRNA in polysomal RNA and the total number of keratin mRNA sequences were similar, the onset of keratin protein synthesis after day 12 is due to an increased rate of synthesis of keratin mRNA (i.e. transcription) rather than translation of a pre-formed mRNA pool.

1.4.3 NUMBER AND GENOMIC ARRANGEMENT OF FEATHER KERATIN GENES IN THE CHICK GENOME

In early studies, Kemp (1975) estimated that there were between 100-240 keratin genes in the chick genome, as determined by the kinetics of hybridization and re-annealing of feather keratin cDNA with chicken DNA. (Cot analysis). Because of the considerable sequence homology between scale, claw and feather keratin genes (see Section 1.4.6 and Chapter 4) it is likely that this total number would include keratin genes other than those of the feather type (Gregg et al., 1984). This hypothesis is supported by the observation that only about 25 to 35 of these keratin genes are expressed in the developing embryonic feather (Kemp, 1975). More recently, Southern blot hybridization of feather keratin cDNA clones (Saint, 1979) and feather gene probes (this work) to

chicken DNA restricted with Hind III, both detected about 20 bands. At present, therefore, the total number of feather keratin genes remains unknown.

The idea that keratin genes may be clustered in the chicken genome was first suggested by Lockett et al. (1979). These workers reported that large fragments of the chicken genome (up to 140 kb in length) with a base content similar to keratin cDNA were able to hybridize to feather keratin cDNA, indicating that the feather keratin genes may be closely linked in the chicken genome.

Recent studies on a λ-recombinant genomic clone have demonstrated that some of the feather keratin genes are indeed clustered in the chicken genome. This clone, λCFK1, was isolated from a chicken genomic library using cDNA prepared from embryonic feather keratin mRNA as a probe (Figure 1.5a; Molloy et al., 1982). It contains four complete feather keratin genes and the 3' half of a fifth gene which have all been sequenced (Molloy et al., 1982; Gregg and Rogers, 1986; Gregg et al., manuscript in preparation). The genes are evenly spaced, with a centre-to-centre separation between the genes of 3.3 kb, and are all transcribed in the same direction (Figure 1.5a). The arrangement of the gene cluster suggests that it has arisen by a series of tandem duplications.

1.4.4 STRUCTURE OF THE GENOMIC FEATHER KERATIN GENES

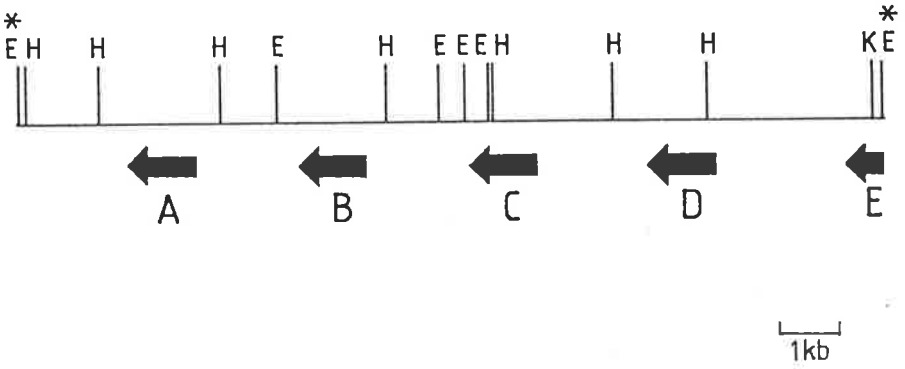
All of the feather keratin genes which have been sequenced from λCFK1 have the same general structure which is depicted in Figure 1.5 b (Molloy et al., 1982; Gregg and Rogers, 1986; Gregg et al., manuscript in preparation). The protein coding regions are all precisely 297 bp long, including the 'start'

FIGURE 1.5

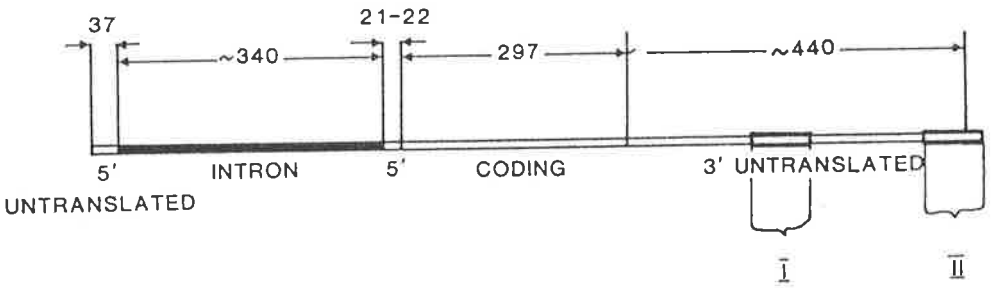
A. Organization of feather keratin genes in the genomic clone λCFK1. The position and orientation of the genes A-E contained in the λ recombinant was determined by Southern hybridization and DNA sequencing (Molloy et al., 1982; Gregg et al., manuscript in preparation). The position of the HindIII (H) and EcoRI (E) restriction enzyme sites are indicated. The EcoRI sites at each end of the clone (E*) are artificial, i.e. they were introduced during the construction of the library (redrawn from Molloy et al., 1982).

B. General structure of embryonic feather keratin genes (Gregg et al., 1983). The marked portions of the 3' non-coding region represent conserved sequences (see Figure 1.8d,e). Numbers indicate length of gene segments in base pairs.

A.


* E H H H E H E E H H H * K E
A
B
C
D
E
1kb

B.


37
21-22
~340
297
~440
5' UNTRANSLATED
INTRON
5'
CODING
3' UNTRANSLATED
I
II

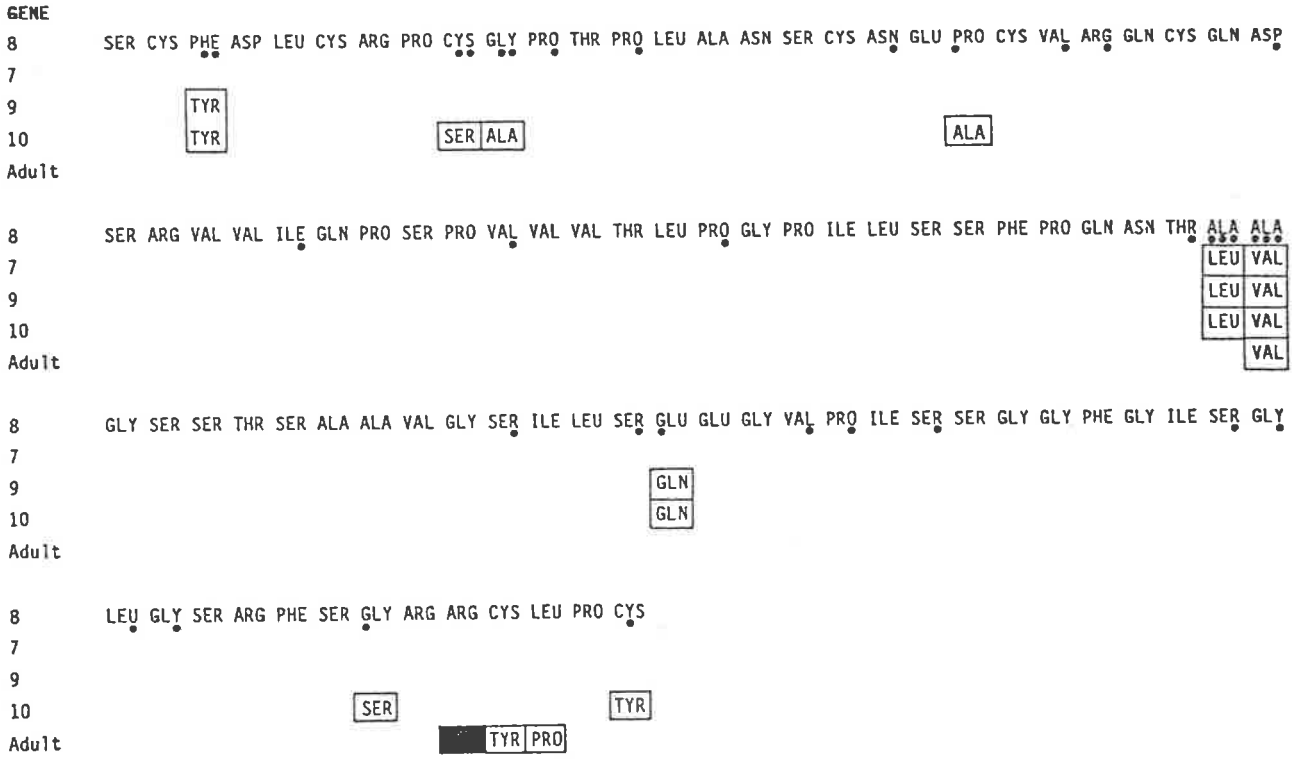
and 'stop' signal codons, producing a polypeptide of 97 amino acids. Among the genes (A-D) the DNA sequence of the protein coding regions show greater than 90% homology which is also reflected in the protein sequence (Figure 1.6, 1.7). The single adult fowl feather keratin that has been sequenced (Arai et al., 1983) is very similar indeed to the embryonic proteins, differing principally in the absence of a single amino acid residue near the C-terminus (Figure 1.6). As observed in the early protein sequencing study of Walker and Rogers (1976b), the amino and carboxy terminal regions are rich in cysteine, proline and serine.

The feather keratin genes have a 5' non-coding region of 58-59 bp which is interrupted by an intron, 37 bp from the transcription initiation site (Figure 1.5b). The sequence of this 37 base exon is rigidly conserved with only one base change throughout the four known sequences (Figure 1.8 b). This conservation is greater than that seen in the protein coding regions which indicates that this region of the gene is of important functional significance, either at the transcriptional and/or translational levels. The remaining 21-22 bases between the 3' splice site and the ATG initiation codon are less strongly conserved, implying that the sequence immediately adjacent to the protein coding region may be of less functional significance than the preceding part.

The four feather keratin genes which have been completely sequenced all contain a TATA or 'Goldberg-Hogness' box although gene C has the slightly altered sequence CATA (Molloy et al., 1982; Gregg and Rogers, 1986). The TATA sequence, which is usually about 30 bases 5' to the cap site, has been implicated

FIGURE 1.6

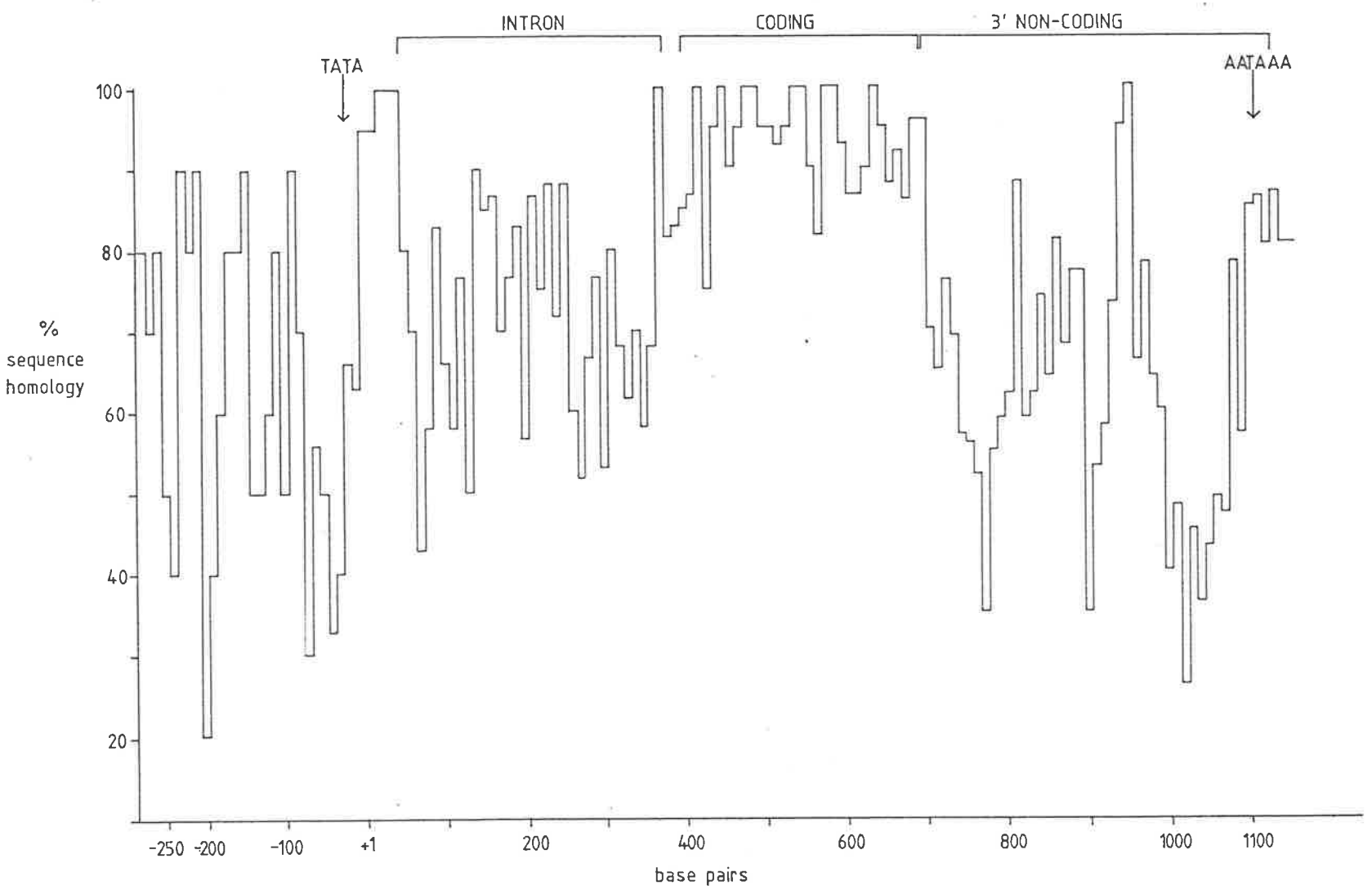
Comparison of the known embryonic feather keratin genes (Rogers, 1984) and an adult chicken feather keratin (Arai et al., 1983). Boxed amino acids represent changes from the illustrated sequence. The black box represents an apparent deletion or insertion difference between adult and embryonic proteins. Dots indicate base changes within the gene sequence - unknown for the adult feather keratin.


GENE
8
7
9
10
Adult
SER CYS PHE ASP LEU CYS ARG PRO CYS GLY PRO THR PRO LEU ALA ASN SER CYS ASN GLU PRO CYS VAL ARG GLN CYS GLN ASP
TYR
TYR
SER
ALA
ALA
SER ARG VAL VAL ILE GLN PRO SER PRO VAL VAL VAL THR LEU PRO GLY PRO ILE LEU SER SER PHE PRO GLN ASN THR ALA ALA
LEU
VAL
LEU
VAL
LEU
VAL
VAL
GLY SER SER THR SER ALA ALA VAL GLY SER ILE LEU SER GLU GLU GLY VAL PRO ILE SER SER GLY GLY PHE GLY ILE SER GLY
GLN
GLN
LEU GLY SER ARG PHE SER GLY ARG ARG CYS LEU PRO CYS
SER
TYR
TYR
PRO

in determining the accuracy of transcriptional initiation (Benoist and Chambon, 1981; Mathis and Chambon, 1981). Only two of the feather genes in the λ recombinant (A and B) have been sequenced upstream of the TATA box. Comparison of these 5' flanking sequences (Figure 1.7, 1.8a) demonstrates that there is considerable DNA sequence homology between the genes. However, it is difficult to determine whether the observed homology is a result of tandem gene duplication (Molloy et al., 1982) or due to conservation of sequences which are important for tissue-specific expression of the keratin genes, as is the case for a growing number of eukaryotic genes (Karin et al., 1984; Parker and Topol, 1984; Staudt et al., 1986). Nevertheless, there are two blocks of homology 5' to the TATA box which may be important. Between bases -158 and -178 is a purine-rich sequence which may represent a hypersensitive site that are often associated with the 5' flanking regions of eukaryotic genes (Figure 1.8a; Gregg and Rogers, 1986). A second region of homology, the so-called 'CAAT box', has been observed about 50 bases upstream from the TATA box in a number of genes (Benoist et al., 1980; Efstratiadis et al., 1980). The CAAT sequence has been specifically implicated in the in vivo expression of a number of eukaryotic genes; it is thought to modulate transcription by binding a specific protein known as CTF (CAAT transcription factor) (Jones et al., 1985). Of the two genes in the λ recombinant for which flanking sequences are known, only gene B has been found to contain a canonical CAAT sequence with gene A having a somewhat altered structure (Figure 1.8a).

FIGURE 1.7 : DNA SEQUENCE HOMOLOGY AMONG THE FEATHER KERATIN GENES

The figure illustrates DNA sequence homology among the four completely sequenced genes from the clone λCFK1 (A-D, Figure 1.5a). Pairs of genes (e.g. A, B; A, C; B, C, etc.) were compared in segments of 10 bases and the sequence homology among the different pairs was averaged and plotted. The figure was drawn from data supplied by Dr. K. Gregg.


INTRON
CODING
3' NON-CODING
TATA
AATAAA
100
80
60
40
20
% sequence homology
-250
-200
-100
+1
200
400
600
800
1000
1100
base pairs

The introns, which vary in size between 324 and 341 bases in the four genes characterized, generally show a lower degree of sequence homology between the four genes than the coding or 5' non-coding regions (Figure 1.7). However, there is a strict conservation of the DNA sequence at each end of the intron adjacent to the splice sites which extends inside the intron by 9 bases at the 5' end and 14 bases at the 3' end (Figure 1.8c). Beyond these points there is a sharp drop in sequence similarities to between 50 to 80%. There are also several blocks of sequence in the middle portion of the intron which show 80-90% conservation between the four genes (Figure 1.7).

Recently, Koltunow et al. (1986) have observed that removal of the intron from a feather keratin gene increased transcription in Xenopus oocytes significantly over the levels observed with an intact feather keratin gene. This repression of feather keratin gene transcription appeared to be caused either directly or indirectly by DNA sequences within the intron, since no inhibitive effect was observed when the intron was replaced with a similar sized fragment from pBR322.

The 3' non-coding regions of the feather genes in λCFK1 are 440-460 bases long and over most of that length show a considerable degree of sequence divergence (Figure 1.7). However, there are two portions of the 3' non-coding region which are very well conserved between the known gene sequences; a region of about 40 bases around the polyadenylation signal (Figure 1.8d) and an 18 base sequence in the middle of the 3' non-coding sequence (Figure 1.8e).

FIGURE 1.8 : COMPARISONS OF FEATHER KERATIN GENE SEQUENCES

A. DNA sequence flanking the 5' side of genes 9 and 10. Identical sequence is illustrated by the dark line.

B. DNA sequence of the 5' non-coding regions. Arrow indicates the splice junction. Gene 8-10 sequences are identical to the sequence of gene 7 except where shown. Gene 10 has an extra base between the 3' splice site and initiation codon.

C. The conserved sequences surrounding the splice junctions of the feather keratin gene intron. Conserved regions are identical in the four known sequences (genes 7-10).

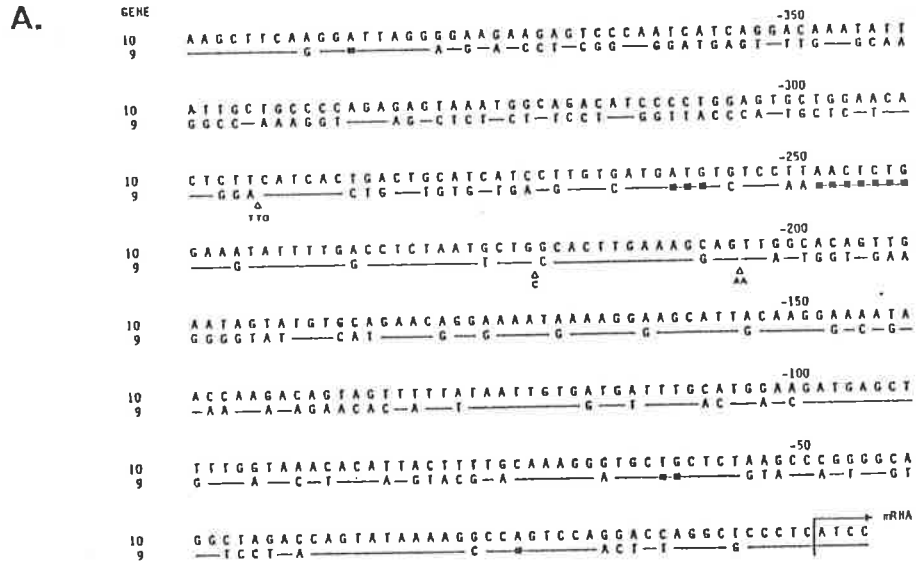
D. Gene sequences in the region of the mRNA 3' terminus location II in Figure 1.5b. The vertical line indicates the point of polyadenylation of the mRNA. Boxed regions are those consensus sequences which are believed to be associated with the polyadenylation process (see Text).

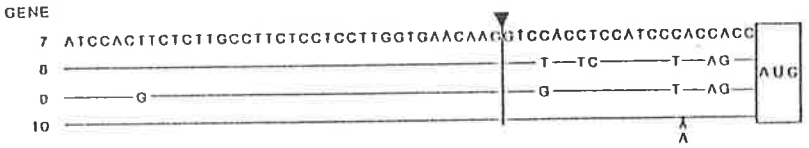
E. The rigidly conserved sequences from location I of the 3' non-coding region (Figure 1.5b). The gene listed as pCKF17 is a cDNA clone isolated in this laboratory (Kuczek, 1980).

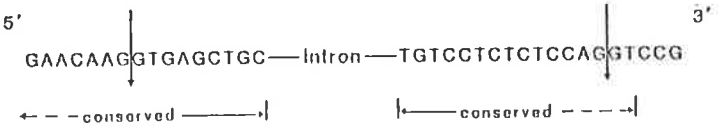
Following the isolation of cosmid 4, which encompasses the clone λCFK1 and contains 11 feather keratin genes (see Section 4.3.2.1), the genes A-E were renamed as follows:

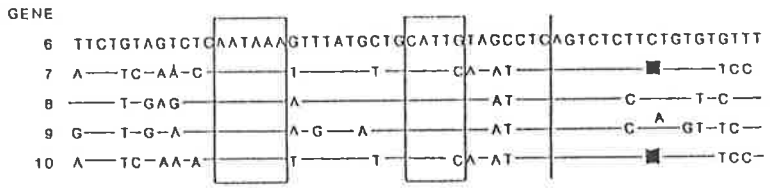
Gene A - 10
Gene B - 9
Gene C - 8
Gene D - 7
Gene E - 6

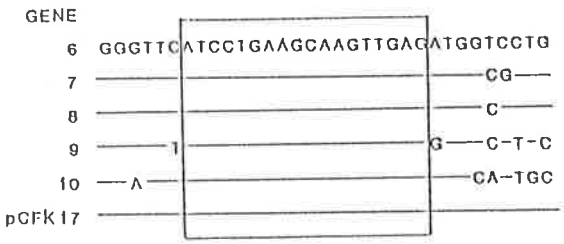
It is this numerical nomenclature which is used in this figure. All diagrams were taken from Gregg and Rogers (1986).


A.
GENE
-350
-300
-250
-200
-150
-100
-50
mRNA


B.
GENE
AUG


C.
5'
3'
GAACAAGGTGAGCTGC — Intron — TGTCCTCTCTCCAGGTCCG
conserved
conserved


D.
GENE


E.
GENE
pCFK17

The five genomic keratin genes and three cDNA clones sequenced all contain the AATAAA motif (Figure 1.8d, Kuczek, 1980; Morris, 1984; Gregg and Rogers, 1986) which is involved in the 3'-end processing of eukaryotic mRNA (Fitzgerald and Shenk, 1981; Montell et al., 1983). A second recognition sequence element, CAYTG (where Y is a pyrimidine), which has been found adjacent to the site of poly (A) addition in a number of eukaryotic genes (Benoist et al., 1980; Berget, 1984) is also present in feather keratin genes a few bases upstream of the poly (A) addition site (Figure 1.8d). This pentanucleotide sequence and the poly (A) signal are complementary to regions of the small nuclear RNA U4 (Reddy et al., 1981; Berget, 1984) and it has been proposed that this molecule is involved in the correct processing of the 3' terminus of mRNA (Berget, 1984; Gil and Proudfoot, 1984). In all feather keratin genes sequenced to date the 5' A residue of the polyadenylation signal is likely to be 27 bases from the point of polyadenylation (Molloy et al., 1982). This 27 base pair region is very well conserved between the five genomic feather keratin sequences and the six bases following the point of polyadenylation appear invariant. Presumably, at least part of this conserved region downstream of the polyadenylation signal, including the CAYTG sequence in particular, are required for correct 3'-end formation as is the case for β-globin mRNA (Gil and Proudfoot, 1984).

Near the middle of the 3' non-coding region is an 18 bp stretch of sequence which is perfectly conserved in six known sequences (Figure 1.7, 1.8e; Gregg and Rogers, 1986). The dramatic conservation of this region strongly suggests that it

has some functional significance, the nature of which is at present unknown. However, deletion of this 18 bp sequence from an intact feather keratin gene did not alter the levels of transcription detected in Xenopus oocytes (Koltunow, 1986). It is possible that this sequence is required for transcription regulation in vivo, e.g., with a protein factor present in feather extracts or alternatively, it is important for the translation of feather keratin mRNA.

1.4.5 STRUCTURE AND ORGANIZATION OF FAST PROTEIN GENES

A cDNA clone and a chick genomic λ clone containing fast protein genes have been isolated and sequenced (Morris, 1984). The DNA sequence of the fast protein gene revealed that, including the initiating methionine, it encoded a protein of 120 amino acids. The amino acid composition of this protein closely matched that of fast protein isolated from embryonic feathers (Powell, 1979), but the translated gene sequence was calculated to have a molecular weight of 14,000 whereas fast proteins isolated from embryonic feather tissue were estimated to have a molecular weight of 8,400 (Walker and Rogers, 1976a). However, the N-terminal sequence of a fast protein polypeptide purified by high pressure liquid chromatography was identical to the N-terminal portion of the gene sequence, demonstrating that the characterized gene did indeed code for a fast protein (Morris, 1984).

The fast protein gene was found to contain an intron of 670 bp in the 5' non-coding region at the same position as in feather keratin genes, i.e. 37 bp from the cap site. The total length of the 5' non-coding region of the fast protein gene was 58 bp which is the same as in feather keratin genes. Further-

more, the sequence of the small 37 bp exon was found to be very similar (75% homologous) to that of feather keratin genes. The sequence of the 21 bases between the 3' splice site and the initiation codon of the fast protein gene also showed considerable homology to that region of feather keratin sequences.

During the characterization of the fast protein genomic λ clone, Morris (1984) noted that it contained fragments flanking the sequenced fast protein gene which only weakly hybridized to the fast protein coding probe. He concluded that the data were consistent with the presence of at least two other fast protein-like genes in close proximity to the original fast protein gene.

1.4.6 COMPARISON OF SCALE AND FEATHER KERATIN GENE SEQUENCES

The comparatively recent isolation of a number of λ recombinants containing scale keratin genes from a chicken genomic library (Wilton, 1983) has enabled a direct comparison of both DNA and protein sequences for scale and feather keratins to be carried out (Gregg et al., 1984). These recombinants contained a total of four closely-linked scale keratin genes and each of these genes have been at least partially sequenced. They are spaced varying distances apart (2-4 kb) and are all transcribed in the same direction (Wilton, 1983).

A comparison of chick feather and scale keratin sequences is shown in Figure 1.9a. As observed previously at the protein level by Walker and Bridgen (1976), it can be seen from Figure 1.9a that there is a high degree of homology between the two derived protein sequences with the outstanding difference being a glycine-rich region of 52 amino acid residues. In this particular scale keratin a 13 amino acid sequence repeats four

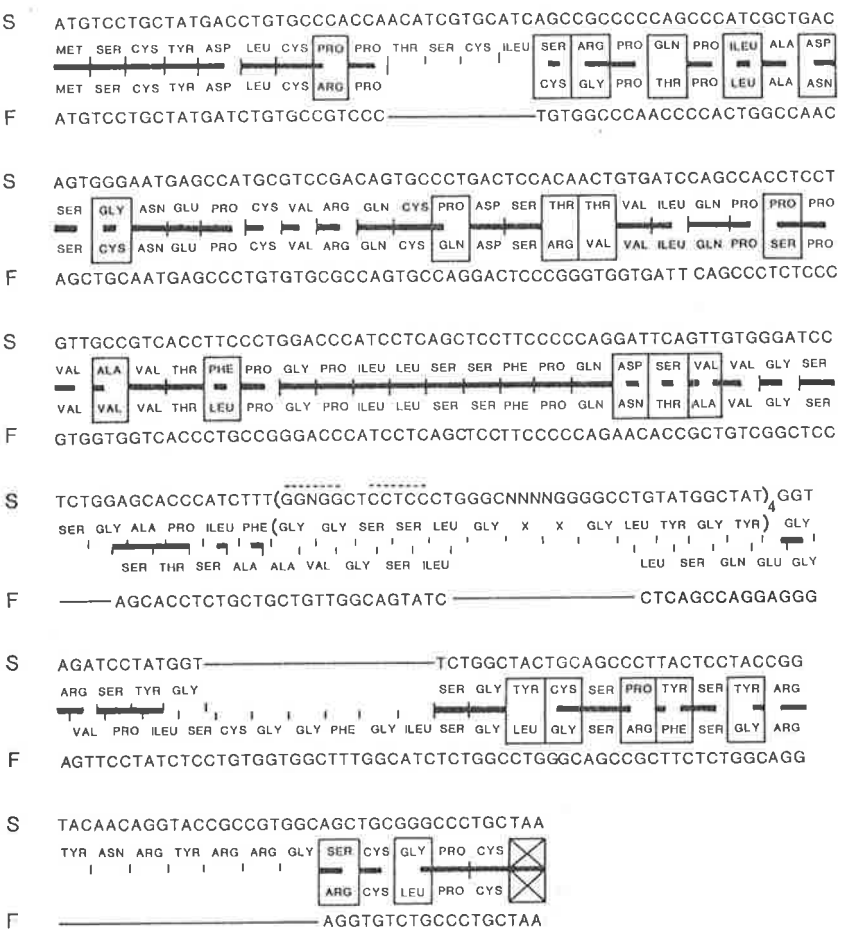
FIGURE 1.9

A. DNA and protein sequences of chick feather (F) and scale (S) keratin genes. The dark line between sequences indicates base sequence homology, fine lines within a sequence indicate an apparent deletion. Boxed amino acids highlight those which differ between the two deduced protein sequences. The region in brackets is the 39 bp (13 amino acid) repeat of the scale keratin gene which is repeated four times.

B. Schematic representation of the similarities between the protein coding regions of feather and scale keratin genes. The major difference is the region coding for a 52 amino acid sequence which is absent in feather keratin. The regions denoted as homologous have > 60% base homology. The non-homologous region (dotted) shows no homology above that expected from random sequence.

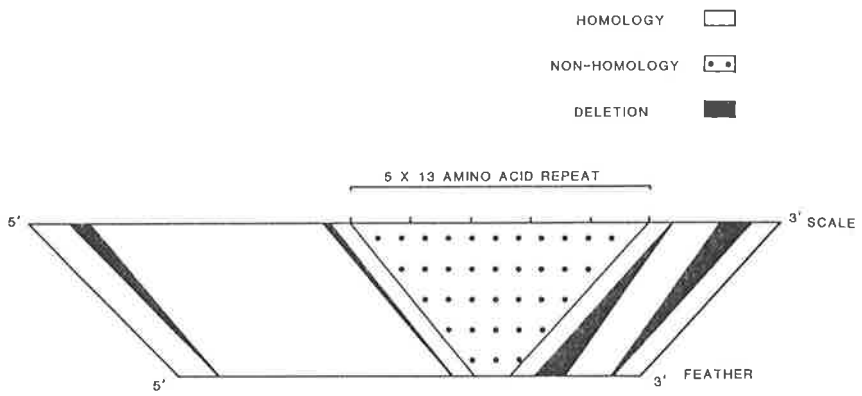
(From Gregg et al., 1984).

A.



B.

COMPARISON OF NUCLEOTIDE SEQUENCES OF CODING REGIONS OF AVIAN FEATHER AND SCALE GENES



times to give the 52 residue addition to the basic feather sequence containing the half-cystine rich C- and N- terminal regions. With some smaller deletions in the scale sequence the difference in molecular size between the two protein families can be accounted for (Walker and Bridgen, 1976). From the fossil record it is known that the development of scales chronologically preceded the appearance of feathers and lead to the suggestion (Spearman, 1966) that scale genes may have given rise to feather genes during vertebrate evolution. From this information, and the available DNA sequences of feather and scale genes, Gregg et al. (1984) further proposed that the two gene families have a common ancestry and the original (primordial) feather keratin gene evolved from an ancestral scale keratin gene by the deletion mainly of the glycine-rich repeat peptide sequence (Figure 1.9b). Following that deletion event, both types of keratin gene underwent a series of duplication events to give rise to the current repertoire of feather and scale keratin genes.

It is interesting to note that scale keratin genes, like their feather counterparts, also appear to contain an intron in the 5' non-coding region 20 or 21 bp upstream of the ATG codon (Wilton, 1983). This 20 bp region showed some homology with the same region of feather keratin genes. Limited sequence data from one scale gene suggested that its intron is about 550 bp long and that the rest of the 5' non-coding region upstream of the 5' splice site is a similar size to feather keratin (Wilton, 1983).

The total number of homologous chains making up the family of chick scale keratins is not yet known although a minimal

number of 9 chains was proposed by Walker and Bridgen (1976). Further DNA studies should resolve this question.

1.5 THE MAMMALIAN EPIDERMIS: SIMILARITIES AND DIFFERENCES WITH AVIAN FEATHER

The proteins of mammalian hair are considerably more complex than that of bird feathers. For example, the keratin proteins of wool which are the most extensively studied of any mammal are comprised of three different classes probably totalling more than 100 proteins (Gillespie, 1983; Powell and Rogers, 1986). By contrast, the feather keratins comprise about 20 different proteins (Walker and Rogers, 1976 a,b). Ultrastructural studies of hair (Rogers 1959 a,b; Fraser et al., 1972), have shown that it is comprised of a cortex made up of microfibrils 8 nm in diameter which are embedded in a non-filamentous matrix.

The microfibrils are composed of three families of proteins collectively known as low sulphur proteins (for reviews see Crewther, 1976; Gillespie, 1983) and thus they perform a similar structural role as do the avian keratins in feather (which form 3 nm microfibrils). However, their chemistry is quite different in that the feather keratins contain predominantly β-sheet structure (Fraser et al., 1972) whereas the mammalian keratins are mostly α-helical (Gillespie, 1983). The low sulphur proteins of wool have recently been shown to be members of the family of proteins known as intermediate filaments (IF) (for reviews see Steinert et al., 1985; Powell and Rogers, 1986).

The IF keratins have been highly conserved throughout vertebrate evolution from fish to humans (Fuchs and Marchuk,

1983), demonstrating their important structural role. The IF keratins of chicken are expressed at a low level in feather and at somewhat higher levels in the leg scale tissue of the developing embryo (Dhouailly et al., 1978; Wilton et al., 1985; Sawyer et al., 1986). The significance of this expression in these β-keratinizing tissues is at present not understood, although various ideas have been put forward (Dhouailly et al., 1978).

The non-filamentous matrix of mammalian hair is composed of two classes of non-helical proteins, the high sulphur proteins and the high glycine/tyrosine proteins (Gillespie, 1983; Powell and Rogers, 1986). It is thought that the high sulphur proteins form disulphide cross-links with the cysteine-rich non-helical domains of the IF keratins in hair producing a rigid matrix of insoluble protein (Gillespie, 1983). The similarity in amino acid composition of the feather keratins and high sulphur wool proteins lead to the suggestion (Fraser et al., 1972), that they may have had a common origin. Although there is no significant homology between the two protein types (Molloy et al., 1982; Powell et al., 1983) this does not rule out the possibility that they did have a common origin and have merely diverged in both sequence and function since the avian/mammalian radiation which occurred about 300 million years ago (Fraser et al., 1972; Efstratiadis et al., 1980).

Genes encoding members of all three wool keratin protein classes have been isolated in this laboratory from sheep genomic libraries (Powell et al., 1983, 1986; Frenkel, 1985; Kuczek, 1985; Kuczek and Rogers, 1985). Powell et al. (1983)

isolated two genomic clones from a λ Charon 4A library each of which contained two keratin genes from the B2 high sulphur family (see also Rogers, 1984). The high sulphur genes contain no introns and were found to be highly homologous in the 5' non-coding regions. Two chromosomal sheep high glycine-tyrosine genes have been cloned and sequenced (Kuczek, 1985; Kuczek and Rogers, manuscript in preparation). They were isolated from different genomic clones and thus are at least 14 kb apart (the size of the λ insert) in the sheep genome. Interestingly, an 18 bp sequence which is immediately upstream of the initiation codon in the 5' non-coding region is conserved between a number of high sulphur genes and the high glycine-tyrosine genes (Powell and Rogers, 1986), which suggests it may play a role in the co-ordinate expression of these two gene classes in the wool matrix. Recently, a number of cosmid clones containing low sulphur keratin genes have been isolated (Powell et al., 1986). The three genomic clones that have been studied in detail contain a total of seven genes which are spaced varying distances apart (10-30 kb) and are mostly transcribed in the same direction, with the exception of one cosmid. Recently, it has been reported that IF-keratin genes are also clustered to some extent in the human genome (RayChaudhury et al., 1986). The extent of gene linkage both within each and between the three classes of wool keratins is at present unknown.

1.6 AIMS OF THIS THESIS

At the time this research project was initiated, it was known from studies on the λ recombinant λCFK1 that the genes coding for some of the feather keratin proteins were clustered

in the chick genome. The cluster of four complete genes isolated in this λ clone are spaced evenly apart and the genes are all transcribed in the same direction. However, protein chemical and DNA hybridization data discussed in this chapter demonstrate that these genes are only part of a larger feather keratin gene family, comprising perhaps 20 genes.

The aims of this work were therefore:

1. To isolate the total complement of genes comprising the feather keratin gene family and determine the overall organization of feather keratin genes in the chick genome.
2. To determine if the DNA sequences, particularly those outside the coding region which have been found to be conserved in the genes present in the λ clone, are maintained in all the members of the feather gene family. From these studies it was also envisaged that some clues as to the evolution of the feather gene family and the importance of these sequences might be obtained.
3. To attempt to address the question of whether or not the different keratin gene families such as feather, scale and claw are linked in the genome and, if so, if this data can shed any light on the evolution of these gene families which are thought to have evolved from a common ancestral gene.
4. Lastly, this thesis describes studies which attempted to identify and partially characterize the putative precursor for feather keratin mRNA.

CHAPTER 2.

MATERIALS AND

GENERAL METHODS.

2.1 MATERIALS

2.1.1 TISSUE

Fertilized eggs of White Leghorn Fowls (Gallus domesticus) were obtained from the Parafield Poultry Research Station of the Department of Agriculture, Parafield, South Australia. The eggs were stored at 10°C for no more than seven days and incubated at 37°C, 54% humidity in a forced draught incubator (Saunders Products Pty. Ltd., Adelaide). After incubating the eggs for a total of 14 days, the embryos were removed and washed with sterile 0.9% saline. The body feathers were then removed by plucking them with KOH-washed jewellers' forceps.

2.1.2 BACTERIAL STRAINS

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

HB101: pro^-, leu^-, thi^-, lac Y^-, $hsdR^-$, $endA^-$, $recA^-$, rpsL20, ara-14, galK2, xyl-5, mtl-1, supE44 (Bolivar and Backman, 1979).

JM101: lac, pro. supE, thi. F' tra D36, pro AB, lac I^q Z ΔM15. (Messing, 1979).

LE392: F^-, hsd R514 ($r_K 4^-$, m_K^+), sup E44, sup F58, lac Y_1 or Δ(lac IZY)$_6$, gal K2, gal T22, met B1, trp R55, λ^-. (Murray et al., 1977).

Mc1061: ara D 139, Δ(ara, leu) 7697, ΔlacX74, gal U^-, gal K^-, hsr^-, hsm^+, str A. (Casadaban and Cohen, 1980).

2.1.3 BACTERIOPHAGE STRAINS

The bacteriophage strains used in this study were:

λ Charon 4A: A am 32, Bam 1, lac 5, bio 256, ∇ KH 54, ∇ NIN 5, ϕ 80 QSR. (Williams and Blattner, 1979). Vector used for cloning chicken genomic DNA.

M13mp83, M13mp93 (Messing and Vieira, 1982), M13mp18, M13mp19

(Norrander et al., 1983). Vectors used for cloning restriction fragments prior to sequencing. The replicative forms of the different M13 vectors were gifts from L. Crocker, C.P. Morris or purchased from Biotechnology Research Enterprises S.A. Pty. Ltd. (BRESA), Adelaide.

2.1.4 ENZYMES

All restriction endonucleases used during the course of this study were purchased from either New England Biolabs, Beverly, Massachusetts or Boehringer, Mannheim.

The remaining enzymes were obtained from the sources listed below.

Calf Intestinal Phosphatase: Boehringer, Mannheim.

Creatine Phosphokinase: Sigma

E. coli Deoxyribonuclease I (DNase I): Initially purchased from Sigma but later from BRESA.

E. coli DNA Polymerase I: Initially obtained as a gift from C.P. Morris but later purchased from BRESA.

E. coli DNA Polymerase I (Klenow Fragment): Initially purchased from Boehringer, Mannheim but later from BRESA.

Lysozyme (from chicken egg white): Sigma.

Polynucleotide Kinase: Initially obtained from Boehringer, Mannheim but later from BRESA.

Proteinase K: Boehringer, Mannheim.

Ribonuclease A (from bovine pancreas): Sigma.

RNA-Dependent DNA Polymerase (reverse transcriptase) from Avian Myeloblastosis Virus (AMV): A gift from Dr. J.R.E. Wells, originally donated by Dr. J.W. Beard, Life Sciences Inc., St. Petersburg, Florida.

T_4 DNA Ligase: Initially purchased from Boehringer, Mannheim but later from BRESA.

2.1.5 RADIOCHEMICALS

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

L-[4,5-^{3}H]leucine (specific activity 110 Ci/mmol) was purchased from The Radiochemical Centre, Amersham.

2.1.6 SYNTHETIC DNA OLIGONUCLEOTIDES

The various synthetic oligonucleotides used as probes and primers were obtained from BRESA. Synthesis of some of the oligomers was performed manually by Dr. D. Skingle and Mr. S. Rogers using the solid-phase phosphite method with morphilinoamidites (Beaucage and Caruthers, 1981; McBride and Caruthers, 1983) and then purified by reverse-phase high performance liquid chromatography. In the later stages of this work, oligomers were synthesized using an Applied Biosystems Model 380B DNA Synthesizer.

The following is a list of the oligonucleotide primers used in this study.

1. Hybridization probe - primer (16-mer) 5'-dCACAATTCCACACAAC-3' used as a primer for making M13 probes. It hybridizes to the complementary sequence in M13 DNA on the 5' side of the cloned insert.
2. Feather keratin gene oligonucleotides.
   (a) 5' non-coding 20-mer 5'-dTCACCTTGTTCACCAAGGAG-3'
   (b) Intron 19-mer 5'-dACCTGGAGAGAGGACA$^{A}_{G}$G-3'
   (c) Coding 25-mer 5'-dTTGGGCCACAGGGACGGCACAGATC-3'
   (d) 3' Non-coding 19-mer 5'-dATCCTGAAGCAAGTTGAGA-3'

3. Universal sequencing primer (17-mer) 5'-dGTAAAACGACGGCCAGT-3' for dideoxy sequencing of M13 clones.

2.1.7 BUFFERS

Buffers commonly used in this study were:

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

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

SSC: 150 mM NaCl, 15 mM trisodium citrate pH 7.5.

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

TBE (pH 8.3): 90 mM Tris, 90 mM boric acid, 2.5 mM EDTA

TBE (pH 8.8): 130 mM Tris, 45 mM boric acid, 2.5 mM EDTA.

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

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

2.1.8 BACTERIAL MEDIA

2.1.8.1 Growth Media for **E. coli** MC1061

Luria broth (L-broth) contained 1% Bacto-tryptone (Difco), 0.5% Bacto-yeast extract (Difco), 1% NaCl and the pH was adjusted to 7.0. Where appropriate, the media was supplemented with ampicillin (50 µg/ml) or tetracycline (20 µg/ml).

2.1.8.2 Growth Media for **E. coli** JM101

M13 minimal media contained 1.05% K_2HPO_4, 0.45% KH_2PO_4, 0.1% $(NH_4)_2\ SO_4$, 0.05% Na_3 citrate which was supplemented after autoclaving with 0.02% $MgSO_4$, 0.0005% thiamine-HCl and 0.2% glucose.

2 x YT broth contained 1.6% Bacto-tryptone, 1% Bacto-yeast, 0.5% NaCl and the pH was adjusted to 7.0.

2.1.8.3 Growth Medias for λ Charon 4A Clones in **E. coli** LE392

Charon 4A plates contained 1% Amine A, 0.5% Bacto-yeast, 0.5% NaCl, 0.2% glucose, 1 mM $MgCl_2$, 10 mM Tris-HCl pH 7.5 and 1.5% Bacto-agar.

NZCYM broth for the propagation of λ Charon 4A clones contained 1% Amine A, 0.5% Bacto-yeast, 0.5% NaCl, 0.1% casamino acids, 0.25% $MgSO_4$ and the pH adjusted to 7.5.

Agar plates were prepared by supplementing the above media with 1.5% Bacto-agar (Difco). 0.7% Agar was used for preparing soft agar overlays. All media was made sterile by autoclaving.

2.1.9 CHEMICALS AND REAGENTS

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

Acrylamide : Aldrich Chemical Company

Agarose Type I : Bethesda Research Laboratories Inc. (BRL)

Agarose, Low Gelling Temperature : BRL

Ampicillin : Sigma

BClG (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside) : Sigma

Bromophenol Blue : BDH Chemicals Ltd.

Caesium Chloride, optical grade : Cabot Corporation

Chloramphenicol : Sigma

Deoxynucleoside Triphosphates (dNTPs) : Sigma

Dextran Sulphate, Sodium Salt : Pharmacia Fine Chemicals Ltd.

DMSO (Dimethyl Sulphoxide) : Ajax Chemicals, Sydney

DTT (Dithiothreitol) : Sigma

Ethidium bromide : Sigma

Ficoll 400 : Pharmacia Fine Chemicals Ltd.

Formamide : BDH Chemicals Ltd. or Ajax Chemicals

It was deionized with a mixed bed resin before use (except in the case of hybridization solutions).

Glyoxal (deionized) : Obtained from Dr. J.L. McInnes.

Guanidine Hydrochloride (Grade I) : Sigma

IPTG (isopropyl β-D-thiogalactoside) : Sigma

Nonidet (NP-40) : Shell Chemicals Ltd.

N',N'-Methylene Bisacrylamide : Sigma

Oligo$(dT)_{12-18}$: P-L Biochemicals

Oligo(dT)-cellulose : BRL

PEG (Polyethylene Glycol) 8000 : Sigma

Phenol : BDH Chemicals Ltd. or Wako Pure Chemical Industries (China).

Polyvinylpyrrolidone : Sigma

PPO (2,5-diphenyloxazole) : Sigma

Sephadex G-50 (medium or fine) : Pharmacia Fine Chemicals Ltd.

SDS (sodium dodecyl sulphate) : Sigma

Sodium Azide : Sigma

Sucrose (ultra-pure) : BRL

TEMED (N,N,N',N'-tetramethylethylenediamine) : Eastman Kodak Company

Urea : Merck, Darmstadt

Xylene Cyanol FF : Tokyo Kasei, Tokyo

2.1.10 NUCLEIC ACIDS AND GENOMIC LIBRARIES

Chicken Cosmid Library : generously donated by Dr. R. D'Andrea.

Chicken Genomic DNA : Obtained from Dr. K. Gregg or Mr. S. Dalton.

E. coli DNA (sonicated and denatured) : Gift from Dr. B.C. Powell.

λ Charon 4A Library of Chicken DNA : Obtained from J. Dodgson, J. Engel and R. Axel (Caltech).

pBR322 DNA : Gift from L. Crocker.

RNA, 18S and 28S from Sheep Wool : Gift from Dr. E.S. Kuczek.

Salmon Sperm DNA : Sigma.

2.1.11 MISCELLANEOUS MATERIALS

GF/A Glass Fibre Discs : Whatman Ltd.

Nitrocellulose Filter Paper : Sartorius, Göttingen and Schleicher and Schüll (BA85), Dassel, West Germany.

X-ray Film : Fujï Rx X-ray Film. Fujï Photo Film Company Ltd.

2.2 GENERAL METHODS

2.2.1 PREPARATION OF GLASSWARE AND SOLUTIONS

To limit the activity of exogenous nucleases, all solutions were prepared using glass-distilled or double-deionized water and sterilized either by autoclaving at 100 kPA for 25 minutes at 120°C or, in cases where the solution was unable to withstand autoclaving, filtering through a Millipore filter under suction. Glassware and plasticware were sterilized by autoclaving. Spatulas, etc. were washed with alkali and rinsed in sterile water.

Tris-containing buffers were adjusted to the required pH at room temperature with HCl except where otherwise stated. Buffer-saturated phenol (pH 7.5 - 8.0) was prepared as described by Maniatis et al. (1982).

2.2.2 ETHANOL PRECIPITATION

In all cases samples containing either DNA or RNA were made 300 mM with respect to Na acetate using a 3 M stock solution at pH 5.5. About 2 to 3 volumes of redistilled ethanol were added to each sample and after mixing, the samples were chilled for either 15 minutes in a dry ice-ethanol bath or overnight at -20°C. Precipitates were collected by centrifugation at 12,000 r.p.m. for between 5-10 minutes in an Eppendorf centrifuge. The supernatant was poured off and any remaining ethanol was removed with a drawn-out pasteur pipette. The

pellet was washed with 1 ml of cold 70% ethanol, dried briefly *in vacuo* and resuspended in the appropriate solution.

2.2.3 PREPARATION OF PLASMID DNA

Recombinant plasmid DNA was prepared as described by Birnboim and Doly (1979), with some modifications.

The recombinant clones were grown overnight in 5 ml of L-broth containing ampicillin (40 µg/ml), diluted 100-fold into 500 ml of fresh L-broth and grown with aeration to stationary phase (OD_{600} of 1.0). Chloramphenicol (150 µg/ml) was added and the cultures grown overnight to amplify the plasmid. The cells were pelleted by centrifugation and resuspended in 4 ml of ice-cold 25 mM Tris-HCl pH 8.0, 10 mM EDTA, 15% sucrose and lysozyme was added to a final concentration of 2 mg/ml.

The solution was kept on ice for 40 minutes and then 8 ml of freshly made 0.2 M NaOH, 1.0% SDS was added and the mixture left on ice for another 10 minutes. Following these lysis steps, 5 ml of 3 M Na acetate pH 4.6 was added, mixed gently and after a 40 minute incubation on ice, the cellular debris and chromosomal DNA was pelleted by centrifugation at 18,000 r.p.m. for 30 minutes at 4°C. The supernatant was carefully decanted and treated with 50 µg RNase A (heat denatured at 80°C for 20 minutes to inactivate any contaminating DNases) for 1 hour at 37°C to remove the contaminating *E. coli* RNA. The solution was extracted once with an equal volume of phenol/chloroform (1:1), then with ether and the nucleic acids precipitated by the addition of 2.5 volumes of redistilled ethanol and chilling at -20°C overnight. The pellet was resuspended in 1.6 ml H_2O, to which was added 0.4 ml of 4 M NaCl and 2 ml of 13% PEG 6000. The mixture was left on

ice for at least 1 hour and the DNA was pelleted by centrifugation (10,000 r.p.m., 10 minutes, 4°C), washed with cold 70% ethanol, dried briefly *in vacuo* and resuspended in 0.5 ml H_2O.

This method yielded between 200-500 µg of plasmid DNA from a 500 ml culture.

2.2.4 RESTRICTION ENZYME DIGESTION

Restriction endonuclease digestion of DNA was carried out using the conditions appropriate for each enzyme detailed in the New England Biolabs catalogue as modified by Maniatis *et al*. (1982). Generally, two units of enzyme was added for each microgram of DNA to be digested and the reaction mix was incubated for at least 2 hours to ensure complete digestion. Reactions were terminated either by ethanol precipitation or by the addition of gel loading mix (see Sections 2.2.5 and 2.2.7).

2.2.5 POLYACRYLAMIDE GEL ELECTROPHORESIS OF DNA

Electrophoresis of DNA of less than about 1 kb in length was carried out on vertical 16 cm x 16 cm x 1.1 mm gels. The 50% acrylamide stock solution (acrylamide to bisacrylamide ratio of 20:1) was deionized using Amberlite MB-1 mixed bed resin and kept in a dark bottle at 4°C. The TBE buffer pH 8.3 (Section 2.1.7) was used routinely in analytical and preparative polyacrylamide gels.

Prior to loading, the DNA samples wre mixed with an appropriate volume of 5 x acrylamide gel loading mix (25% glycerol, 0.1 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol). Electrophoresis was performed at 150-200 V for 2-3 hours and the DNA visualized under UV light after the gel had been briefly stained in a 1 µg/ml solution of ethidium bromide.

2.2.6 ISOLATION OF RESTRICTION FRAGMENTS FROM POLYACRYLAMIDE GELS

DNA fragments that had been fractionated for preparative purposes on a polyacrylamide gel were excised using a scalpel blade and the DNA eluted using the elution buffer described by Maxam and Gilbert (1980). (0.5 M NH_4 acetate, 10 mM $MgCl_2$, 1 mM EDTA, 0.1% SDS). Later, the elution buffer was modified to 0.3 M Na acetate, 0.1% SDS. After incubating overnight at 37°C, the eluate was transferred to an Eppendorf tube and ethanol precipitated. The pellet was resuspended in 200 µl of 300 mM Na acetate pH 5.5 and reprecipitated with ethanol (Section 2.2.2). To remove any residual salt, the precipitated DNA was washed with 70% ethanol and resuspended in sterile water.

2.2.7 AGAROSE GEL ELECTROPHORESIS OF DNA

Electrophoresis of DNA for analytical purposes (checking digests, etc.) was generally carried out either on small 10 x 7.5 x 0.2 cm plastic gel trays (Pharmacia GNA 100) or microscope slides; DNA samples destined for Southern transfer were electrophoresed on either 20 x 20 x 0.15 or 0.3 cm horizontal slab agarose gels or on vertical 15 x 15 x 0.3 cm vertical agarose gels.

The agarose (0.8-2%) was dissolved in 1 x TAE buffer (Section 2.1.7) and cast in the gel apparatus. Prior to loading, the DNA samples were mixed with an appropriate volume of 10 x agarose gel loading mix (25% Ficoll 400, 0.2% bromophenol blue, 0.2% xylene cyanol-FF). Electrophoresis was performed at 50-200 mA until the bromophenol blue had migrated an appropriate distance to ensure that adequate separation of the

fragments had taken place. The agarose gel was stained with ethidium bromide (1 µg/ml) and visualized under UV light.

2.2.8 ISOLATION OF DNA FRAGMENTS FROM AGAROSE GELS

Low gelling temperature (LGT) agarose was dissolved in 1 x TAE buffer and cast in large or small horizontal gel trays as described above (Section 2.2.7). Electrophoresis was performed as described above except that the current was usually not greater than 100 mA and that it was generally carried out at 4°C. Restriction fragments were detected under UV light after brief ethidium bromide staining and the desired bands excised from the gel.

The gel slice containing the DNA fragment was melted at 65°C for 15 minutes after the addition of several volumes of 0.1 M Tris-HCl pH 8.0, 0.5 mM EDTA. An equal volume of buffer-saturated phenol at room temperature was added, the phases mixed thoroughly by vortexing and then re-separated by centrifugation in an Eppendorf centrifuge. The aqueous phase was removed and re-extracted with phenol at least once more (or until there was no visible interphase), then with ether, and the DNA ethanol precipitated (Section 2.2.2).

Typically, at least 50% of the DNA present in any band was recovered.

2.2.9 TRANSFER OF DNA TO NITROCELLULOSE (SOUTHERN BLOTTING)

Restricted DNA fractionated on agarose slab gels was transferred to nitrocellulose using the method of Southern (1975), as modified by Wahl et al. (1979). In this procedure, the rapid transfer of DNA from the gel is facilitated by the partial hydrolysis of the DNA with 0.25 M HCl. Bidirectional transfer of DNA from agarose gels was performed as described by

Smith and Summers (1980) in which the gel is neutralized with 1 M NH_4 acetate, 20 mM NaOH in the final washing step. After allowing the transfer to occur (12-18 hours), the filter(s) were separated from the agarose gel, and were then air dried and baked for 1-2 hours at 80°C in vacuo. The filters were stored in a sealed plastic bag at room temperature until used.

2.2.10 KERATIN PROBES USED IN HYBRIDIZATION STUDIES ON λ AND COSMID GENOMIC CLONES

1. The feather keratin gene probe comprised the 2 kb EcoRI/HindIII fragment from the subclone p λCFK1-4 which contains a complete feather keratin gene (gene B from λCFK1, Molloy et al., 1982) and some flanking DNA (Figure 2.1). p λCFK1-4 DNA was digested with EcoRI and HindIII, fractionated on a 1% LGT agarose gel and the 2 kb fragment recovered by phenol extraction (Section 2.2.8). About 0.5-1 μg of the fragment was labelled by nick translation (Section 2.2.11.2) and hybridized to Southern filters (Section 2.2.12).

2. The scale keratin gene probe comprised an M13 clone (1.8 kb) containing a complete scale gene (Wilton, 1983). The cloned DNA was labelled as described in Section 2.2.11.4 and used in hybridization experiments (Section 2.2.12).

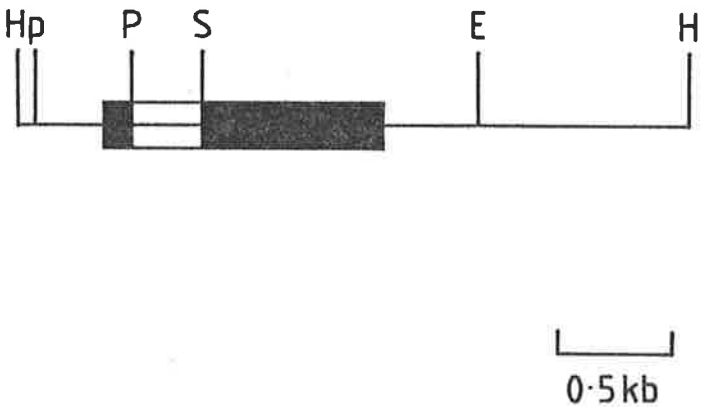
3. The claw keratin probe (CCK22) comprised an M13 clone which was a cDNA clone containing the 3' non-coding region and the 3' half of the protein coding region of a claw gene (Whitbread, 1985). The M13 DNA was labelled with ^{32}P (Section 2.2.11.4) and hybridized to Southern blots of cosmid recombinants.

4. Oligonucleotide probes. Four oligonucleotides were used as hybridization probes in this study. They were designed

FIGURE 2.1 : RESTRICTION MAP OF THE GENE B - CONTAINING SUBCLONE P λCFK1-4

The arrangement of gene B in the subclone p λCFK1-4 (Molloy et al., 1982), encoding the mRNA (■) and intronic (□) sequences, is shown. The complete feather gene probe routinely used in some of the hybridization studies described in this thesis (including the screening of the cosmid library) was the 2 kb EcoRI/Hind III restriction fragment of this subclone which contains gene B of λCFK1 and some flanking DNA (Figure 3.2).

The restriction sites shown are E - EcoRI, H - Hind III, P - Pst I (only two of the Pst sites are shown), S - Sal I. The scale represents 0.5 kb.


Hp
P
S
E
H
0·5 kb

from highly conserved segments of the 5' non-coding, intron, coding and 3' non-coding regions and are based on the sequences of the four fully characterized feather keratin genes of the recombinant λCFK1 (see Section 1.4.4 and Figure 1.8).

The four oligomers were (1) a 20-mer which includes part of the 37 bp 5' non-coding exon and the 5' splice junction (a gift from A. Koltunow), (2) a 19-mer, from the intron of the feather keratin gene, containing the conserved sequence at the 3' end of the intron and the 3' splice site. (3) A 25-mer from the N-terminal end of the protein coding region (a gift from A. Koltunow). (4) a 19-mer prepared from the middle of the 3' non-coding region.

The location of these conserved oligonucleotides in a feather keratin gene is displayed in Figure 2.2. The actual sequence of these oligonucleotides and their method of preparation is described in Section 2.1.6. The oligonucleotides were labelled by kinasing (Section 2.2.11.3) and hybridized to Southern filters as described in Section 2.2.13.

5. In some experiments, an alternative 5' end probe to the 5' non-coding oligonucleotide was used. This probe, originally obtained from Dr. K. Gregg, was an M13 clone with a 400 bp PstI insert which contained the 5' non-coding exon and 5' flanking sequences of feather keratin gene B. The genomic region located in this M13 clone is shown in Figures 2.1 and 2.2.

2.2.11 PREPARATION OF RADIOACTIVE DNA PROBES

2.2.11.1 Synthesis of cDNA

High specific activity ^{32}P-labelled DNA complementary to poly A^+ RNA from 14-day embryonic feathers was synthesized

FIGURE 2.2 : LOCATION OF FEATHER KERATIN OLIGONUCLEOTIDES IN GENOMIC SEQUENCES

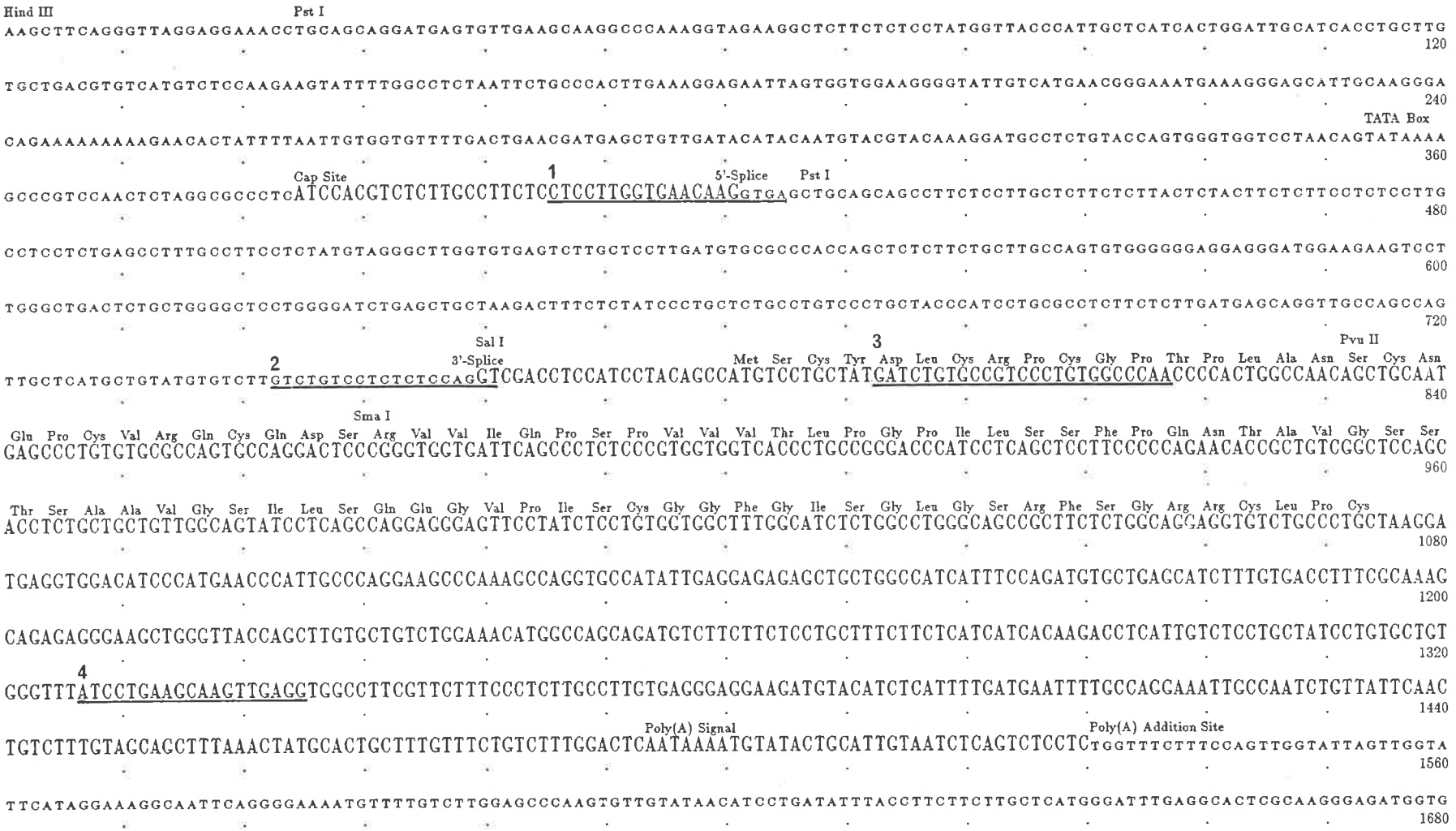
The position of the oligonucleotides are displayed in relation to feather keratin gene B. The oligonucleotides were based on the sequences of the four complete feather keratin genes of λCFK1 (Gregg and Rogers, 1986; Gregg et al., manuscript in preparation).

The oligonucleotides were:

1. A 5' non-coding 20-mer
2. A intron 19-mer
3. A coding 25-mer
4. A 3' non-coding 19-mer

The segment that encodes the mRNA of gene B begins at nucleotide 385 and ends at nucleotide 1530. The gene has one intron of 337 bp which splits the 5' non-coding region into two sections, 37 and 21 bases respectively. The large capital letters represent the mRNA sequence and the small capital letters the intron and flanking sequences. Other features of the genes, including the cap site, TATA box and poly(A) signal are indicated. Selected restriction enzyme sites are shown above the recognized DNA sequence.

Gene B


Hind III
Pst I
AAGCTTCAGGGTTAGGAGGAAACCTGCAGCAGGATGAGTGTTGAAGCAAGGCCCAAAGGTAGAAGGCTCTTCTCTCCTATGGTTACCCATTGCTCATCACTGGATTGCATCACCTGCTTG
120
TGCTGACGTGTCATGTCTCCAAGAAGTATTTTGGCCTCTAATTCTGCCCACTTGAAAGGAGAATTAGTGGTGGAAGGGGTATTGTCATGAACGGGAAATGAAAGGGAGCATTGCAAGGGA
240
TATA Box
CAGAAAAAAAAAGAACACTATTTTAATTGTGGTGTTTTGACTGAACGATGAGCTGTTGATACATACAATGTACGTACAAAGGATGCCTCTGTACCAGTGGGTGGTCCTAACAGTATAAAA
360
Cap Site
1
5'-Splice
Pst I
GCCCGTCCAACTCTAGGCGCCCTCATCCACGTCTCTTGCCTTCTCCTCCTTGGTGAACAAGGTGAGCTGCAGCAGCCTTCTCCTTGCTCTTCTCTTACTCTACTTCTCTTCCTCTCCTTG
480
CCTCCTCTGAGCCTTTGCCTTCCTCTATGTAGGGCTTGGTGTGAGTCTTGCTCCTTGATGTGCGCCCACCAGCTCTCTTCTGCTTGCCAGTGTGGGGGGAGGAGGGATGGAAGAAGTCCT
600
TGGGCTGACTCTGCTGGGGCTCCTGGGGATCTGAGCTGCTAAGACTTTCTCTATCCCTGCTCTGCCTGTCCCTGCTACCCATCCTGCGCCTCTTCTCTTGATGAGCAGGTTGCCAGCCAG
720
Sal I
Pvu II
2
3'-Splice
3
TTGCTCATGCTGTATGTGTCTTGTCTGTCCTCTCTCCAGGTCGACCTCCATCCTACAGCCATGTCCTGCTATGATCTGTGCCGTCCCTGTGGCCCAACCCCACTGGCCAACAGCTGCAAT
840
Sma I
Met Ser Cys Tyr Asp Leu Cys Arg Pro Cys Gly Pro Thr Pro Leu Ala Asn Ser Cys Asn
Glu Pro Cys Val Arg Gln Cys Gln Asp Ser Arg Val Val Ile Gln Pro Ser Pro Val Val Val Thr Leu Pro Gly Pro Ile Leu Ser Ser Phe Pro Gln Asn Thr Ala Val Gly Ser Ser
GAGCCCTGTGTGCGCCAGTGCCAGGACTCCCGGGTGGTCATTCAGCCCTCTCCCGTCGTGGTCACCCTGCCCGCCACCCATCCTCAGCTCCTTCCCCCAGAACACCGCTGTCGGCTCCAGC
960
Thr Ser Ala Ala Val Gly Ser Ile Leu Ser Gln Glu Gly Val Pro Ile Ser Cys Gly Gly Phe Gly Ile Ser Gly Leu Gly Ser Arg Phe Ser Gly Arg Arg Cys Leu Pro Cys
ACCTCTGCTGCTGTTGCCAGTATCCTCAGCCAGGAGGGAGTTCCTATCTCCTGTGGTGGCTTTGGCATCTCTGGCCTGGGCAGCCCGCTTCTCTGGCAGGAGGTGTCTGCCCTGCTAAGGA
1080
TGACGTGGACATCCCATGAACCCATTGCCCAGGAAGCCCAAAGCCAGGTGCCATATTGAGGAGAGAGCTGCTGGCCATCATTTCCAGATGTGCTGAGCATCTTTGTGACCTTTCGCAAAAG
1200
CACAGAGGCAAGCTGGGTTACCAGCTTGTGCTGTCTGCAAACATGGCCAGCAGATGTCTTCTTCTCCTGCTTTCTTCTCATCATCACAAGACCTCATTGTCTCCTGCTATCCTGTGCTGT
1320
4
GCGTTTATCCTGAAGCAAGTTCAGGTGGCCTTCGTTCTTTCCCTCTTGCCTTGTGAGGCAGCAAGATGTACATCTCATTTTGATGAATTTTGCCAGCAAATTGCCAATCTGTTATTCAAC
1440
Poly(A) Signal
Poly(A) Addition Site
TGTCTTTGTAGCAGCTTTAAACTATGCACTGCTTTGTTTCTGTCTTTGGACTCAATAAAATGTATACTGCATTGTAATCTCAGTCTCCTCTGGTTTCTTTCCAGTTGGTATTAGTTGGTA
1560
TTCATAGGAAAGGCAATTCAGGGGAAAATGTTTTGTCTTGGAGCCCAAGTGTTGTATAACATCCTGATATTTACCTTCTTCTTGCTCATGGGATTTGAGGCACTCGCAAGGGAGATGGTG
1680

essentially as described previously (see Maniatis et al., 1982). Synthesis was performed at 42°C for 1 hour in a final volume of 40 µl and contained 50 mM Tris-HCl pH 8.3, 50 mM KCl, 10 mM DTT, 10 mM $MgCl_2$, 0.8 mM of each of dATP, dGTP and dTTP, 90 µM [α-^{32}P]dCTP (1800 Ci/mmol), 15 µg/ml oligo(dT), 1-2 µg of polyA$^+$ RNA and 20 U reverse transcriptase. The reaction was stopped by the addition of an equal volume of H_2O and an equal volume of 1% SDS.

Following a 2 minute incubation to disrupt the protein-nucleic acid complexes, the RNA template was removed by alkaline hydrolysis in 0.3 M NaOH for at least 1 hour at 37°C. After the incubation the mixture was cooled, neutralized by the addition of HCl to 0.3 M and Tris-HCl pH 7.5 to 0.1 M and the unincorporated nucleotides removed by gel filtration using a Sephadex G-50 column (20 x 0.6 cm) equilibrated in TE pH 8.0.

2.2.11.2 Nick Translation of Double-Stranded DNA

The labelling of double-stranded DNA using E. coli DNA polymerase I (Rigby et al., 1977) was carried out using the nick translation kit supplied by BRESA. The 20 µl reaction mix contained 50 mM Tris-HCl pH 7.6, 10 mM $MgSO_4$, 0.1 mM DTT, 100 µg/ml gelatin, 0.5-1 µg DNA, 25 µM of unlabelled dATP, dGTP and dTTP, 10 µM [α-^{32}P]dCTP, 40 pg DNase I and 5 units of DNA polymerase I. The reaction was incubated at 15°C for 90 minutes after which it was terminated by the addition of EDTA to 20 mM. The mixture was passaged through a Sephadex G-50 column to remove unincorporated nucleotides.

2.2.11.3 Preparation of Oligonucleotide Probes

Oligonucleotides obtained from BRESA (Section 2.1.6) were labelled with [γ-^{32}P]ATP using polynucleotide kinase (Maniatis

et al., 1982). The 20 µl reaction mixture contained 50 mM Tris-HCl pH 9.0, 10 mM $MgCl_2$, 10 mM DTT, 200-500 ng DNA, 100 µCi [γ-^{32}P]ATP (2000 Ci/mmol) and 1-2 U polynucleotide kinase. The reactions were incubated for 1 hour at 37°C after which the products were fractionated on a 20% polyacrylamide gel (acrylamide:bisacrylamide ratio of 20:1) to remove unincorporated label (see Section 2.2.5). The kinased DNA was detected by autoradiography (< 1 minute at room temperature) and the band excised and eluted as described in Section 2.2.6.

2.2.11.4 Preparation of M13 Probes

DNA fragments subcloned into M13 (see Section 5.2) were labelled using a primer purchased from BRESA which is complementary to a region of M13 on the 5' side of the insert. Transcription from this primer does not extend through the cloned DNA which remains single stranded (Hu and Messing, 1982).

The primer and single-stranded M13 DNA were annealed by combining 5 µl of M13 DNA (~0.5 µg), 1 µl (15 ng) primer, 2 µl of 100 mM Tris-HCl pH 8.0, 50 mM $MgCl_2$ and 9 µl of H_2O and heating in a boiling water bath for 2 minutes. After allowing the mixture to cool slowly to room temperature, the annealed DNA was added to a tube containing 2 µl of a mixture containing 2.5 mM each of dATP, dGTP and dTTP, 90 µM [α-^{32}P]dCTP (1800 Ci/mmol), 1 µl of 0.1 M DTT and 2 µl Klenow (4 U/µl) and incubated at 37°C for 30 minutes. The labelled DNA was chased for 20 minutes at 37°C after adding 2 µl of a mixture containing 2.5 mM each of the four dNTPs. The reaction was terminated by adding EDTA pH 7.5 to 20 mM and the labelled DNA was purified by passage through a Sephadex G-50 column (20 x 0.6 cm) to remove unincorporated nucleotides.

2.2.11.5 3' End-Labelling of DNA

Both restricted pBR322 (used as a molecular weight marker) and recombinant plasmid DNA were end-labelled using the Klenow fragment of E. coli DNA polymerase I and either [α-^{32}P]dATP or [α-^{32}P]dCTP as described below. This method facilitates the detection of short fragments, which could not be easily seen in ethidium bromide-stained gels.

The plasmid DNA (up to ~ 1 µg) was digested with the appropriate restriction enzymes (Section 2.2.4), the NaCl concentration adjusted to 50 mM where necessary and the mixture transferred to a tube in which 10-20 µCi of ^{32}P-labelled nucleotide had been dried in vacuo. To this tube was added 5 µM of each of the three cold nucleotides and 1 U of Klenow and the reaction carried out at 37°C for 30 minutes. The radiolabelled DNA was fractionated on a polyacrylamide gel (Section 2.2.5) and the separated DNA fragments visualized by autoradiography.

2.2.12 HYBRIDIZATION OF RADIOACTIVE DNA PROBES TO SOUTHERN FILTERS

The prehybridization and hybridization of radioactive probes to DNA immobilized on nitrocellulose filters was carried out exactly as described previously (Wahl et al., 1979; Meinkoth and Wahl, 1984). Generally, the filters were prehybridized for 3-8 hours and hybridized with the probes for 15-24 hours at 42°C. The filters were washed at a stringency appropriate for

the particular experiment (see Footnote below), dried, sealed in a plastic bag and exposed to X-ray film in contact with a tungsten intensifying screen (Cronex DuPont Lightning Plus) at -80°C for the appropriate length of time (a few hours - 3 weeks).

2.2.13 HYBRIDIZATION OF OLIGONUCLEOTIDE PROBES TO SOUTHERN FILTERS

The filters were prehybridized for 3-8 hours in 6 x NET (1 x NET = 0.15 M NaCl, 15 mM Tris-HCl pH 7.5, 1 mM EDTA), 5 x Denhardts (Maniatis et al., 1982), 0.5% NP-40 and 100 µg/ml salmon sperm DNA. Hybridizations were carried out for 15-24 hours usually at 42°C in 6 x NET, 1 x Denhardts, 0.5% NP-40 and 100 µg/ml salmon sperm DNA and the blots washed in 4 x SSC at 37°C-65°C before autoradiography at -80°C.

2.2.14 DOT BLOT ANALYSIS OF DNA OR RNA

Samples of double-stranded DNA (up to 5 µg) were denatured by heating at 100°C for 5 minutes followed by rapid cooling in wet ice. In the case of single-stranded nucleic acids

---

Footnote: Following the hybridization step, filters were removed from their hybridization bags and washed for at least 1 hour in solutions containing 0.1% SDS and varying concentrations of SSC solution (Section 2.1.7), depending on the 'stringency' of the particular hybridization experiment as follows.

Low stringency conditions: 2 x SSC, 0.1% SDS, 65°C
Moderate stringency conditions: 0.5-1.0 x SSC, 0.1% SDS, 65°C.
High stringency conditions: 0.05-0.1 x SSC, 0.1% SDS, 65°C.

(e.g. M13 single-stranded clones, RNA), no denaturation step is necessary. The nucleic acid samples were then spotted on to a nitrocellulose filter which had previously been soaked in either 20 x SSC or 1 M NH_4 acetate, 20 mM NaOH. The filters were then processed as described for Southern blots (Section 2.2.9).

2.2.15 RNA AND DNA CONCENTRATIONS

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

2.2.16 CONTAINMENT FACILITIES

All work involving recombinant DNA material was carried out in accordance with the regulations and approval of the Australian Academy of Science Committee on Recombinant DNA and the University Council of the University of Adelaide.

CHAPTER 3.

ISOLATION OF

CHICKEN KERATIN GENE RECOMBINANTS

3.1 INTRODUCTION

The development of techniques for *in vitro* packaging of foreign DNA into phage particles (Hohn and Murray, 1977; Sternberg *et al.*, 1977) together with the development of bacteriophage λ and λ-pBR322 derived cosmid cloning vectors (Blattner *et al.*, 1977; Collins and Hohn, 1978) has enabled genomic libraries of eukaryote DNA to be constructed (Blattner *et al.*, 1978; Maniatis *et al.*, 1978; Meyerowitz *et al.*, 1980; Grosveld *et al.*, 1981). Such cloned libraries have been used to isolate single copy genes and gene families from many different organisms and has enabled rapid progress to be made in our knowledge of gene evolution and the control of gene activity. Since essentially all the data reported in this thesis was obtained from cosmid clones, I shall limit the present discussion to this technique of library construction.

Cosmids are plasmid vectors that contain a λ-derived cohesive-end (cos) site enabling recombinants to be packaged *in vitro* and propagated inside phage heads (Collins and Hohn, 1978). A number of small (6-10 kb) cosmid vectors have been constructed (Hohn and Collins, 1980; Meyerowitz *et al.*, 1980; Ish-Horowicz and Burke, 1981; Lund *et al.*, 1982) containing this cos sequence as well as a plasmid origin of replication, drug resistance markers and unique restriction sites for cloning (Figure 3.1). Because a cosmid clone mimicks a λ molecule (since it contains a λ cos site), it allows fragments of eukaryotic DNA up to 45 kb to be cloned. Once transduced into bacteria the cosmid clones replicate as plasmids and are selected for by exploiting the antibiotic resistance carried by them.

FIGURE 3.1 : COSMID VECTOR pHC79

A restriction map of the cosmid vector pHC79. Data from Hohn and Collins (1980).

The position of the ampicillin (amp) and tetracycline (tet) resistance genes and the λ cos sequence in the vector are shown.

Restriction sites are indicated as follows:

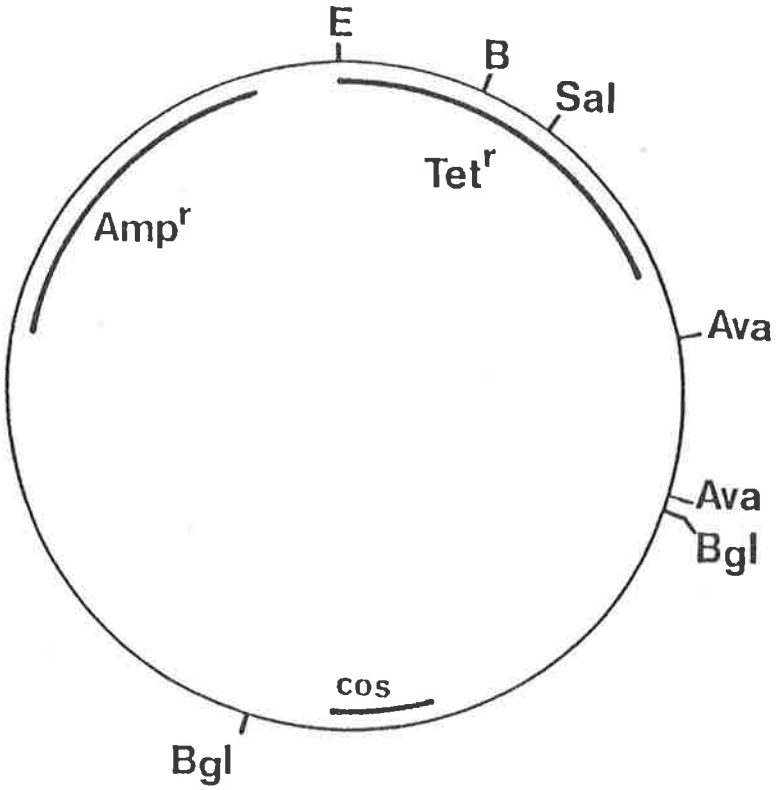
E : EcoRI

B : Bam HI

Sal : Sal I

Ava : Ava I

Bgl : Bgl II


E
B
Sal
Tet^r
Amp^r
Ava
Ava
Bgl
cos
Bgl

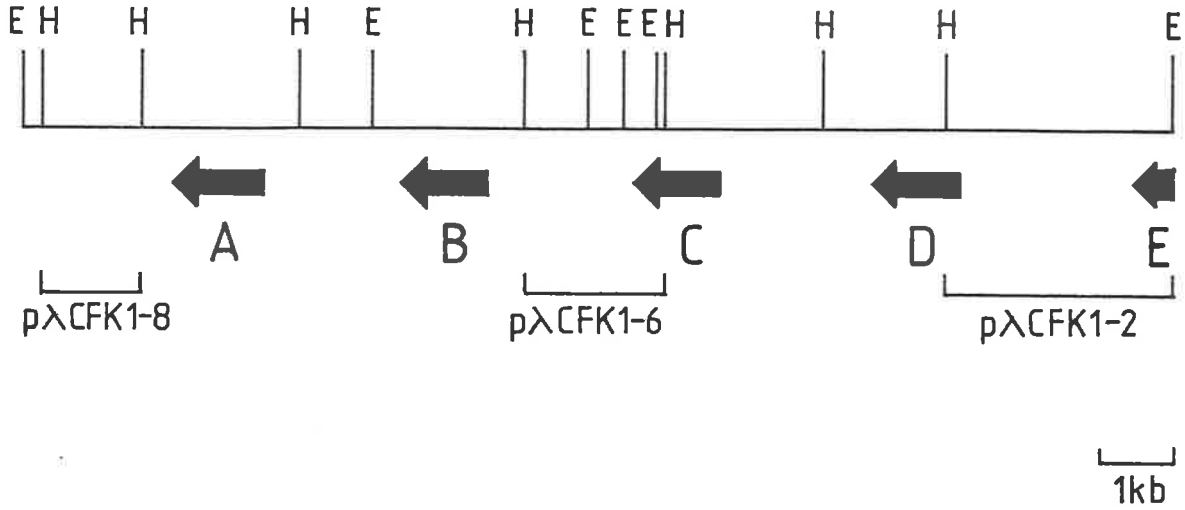
Libraries constructed using cosmid vectors have a number of advantages over the more conventional λ libraries. Firstly, because the clones generally have much larger inserts, they facilitate the study of clustered multigene families (Cattaneo et al., 1981; Grosveld et al., 1981; Steinmetz et al., 1982, 1986; D'Andrea et al., 1985) and the isolation of large genes (Weiss et al., 1982; Gitschier et al., 1984). Secondly, since less recombinant molecules are required to represent the entire genome, fewer colonies have to be screened to isolate the gene(s) of interest which therefore reduces the workload required. Thirdly, genes can be isolated together with their surrounding sequences. This is important in view of recent findings that regulatory sequences for some genes may be located many kilobases away from them (Rusconi, 1986).

The feather keratins comprise a family of about 20 proteins which are highly homologous in sequence and are probably encoded by a similar number of genes (for reviews see Rogers, 1984; Gregg and Rogers, 1986). To date, our knowledge of the organization and structure of feather keratin genes has come from studies on the Charon 4A clone λCFK1 (Molloy et al., 1982). This clone, which was isolated from a chick λ library (Dodgson et al., 1979) using embryonic feather keratin cDNA as a probe, contains five feather keratin genes which are evenly spaced in a tandem array with a mean distance apart of 3.3 kb centre to centre (Figure 3.2; Molloy et al., 1982). All five genes are transcribed in the same direction. The data suggested that this gene family arose by a series of tandem duplications and it was speculated that this cluster was part of a longer cluster of tandemly arranged genes. In order to obtain

FIGURE 3.2 : MAP OF THE CHICKEN GENOMIC CLONE λCFK1

The restriction sites indicated are E - EcoRI, H - Hind III. Below the map is shown the location of each of the five genes A - E, the arrows indicating their orientation, 5' to 3'. Gene E is only partially contained in this recombinant.

The portions of λCFK1 contained in the subclones p λCFK1-2 and p λCFK1-8, which were used as 'walking probes' to screen the λ chicken genomic library, are shown. The subclone p λCFK1-6 from which a λCFK1-specific probe was isolated is also indicated (see Section 3.3.3). The scale represents 1 kb (redrawn from Molloy et al., 1982).


E H H H E H E E E H H H E
A
B
C
D
E
pλCFK1-8
pλCFK1-6
pλCFK1-2
1kb

a more complete picture of keratin gene organization, the available λ library was rescreened with 'walking' probes isolated from both ends of λCFK1. Later a chick cosmid library was probed with a feather keratin gene from λCFK1. This chapter describes the screening and subsequent isolation of a number of λ and cosmid clones containing keratin genes.

3.2 METHODS

3.2.1 Screening the λ Genomic Library

The λ library was screened with 'walking' probes from either end of the clone λCFK1. The rationale of this approach was that overlapping clones as well as clones identical to λCFK1 would be selected. The two restriction fragments from each end of λCFK1 which were used as hybridization probes had previously been subcloned into pBR322 (p λCFK1-2 and p λCFK1-8; Molloy et al., 1982). The 3.0 kb EcoRI/Hind III subclone p λCFK1-2 contains the 5' non-coding region of gene D, the intergenic region between genes D and E and the 3' non-coding region of gene E; p λCFK1-8 contains the 1.3 kb Hind III fragment which is located downstream of gene A and contains no feather keratin genes (Figure 3.2).

The λ Charon 4A chick genomic library was screened by plating approximately 5 x 10^4 plaque forming units (pfu) on each of four fresh Charon 4A plates (Section 2.1.8.3) cast in 150 mm petri dishes and incubated overnight at 37°C. Duplicate nitrocellulose filters were prepared from each plate essentially as described by Benton and Davis (1977). The phage were absorbed to nitrocellulose filters (1 minute for the first filter, 2 minutes for the second), denatured by soaking in 0.5 M NaOH, 1.5 M NaCl for 1 minute, neutralized with two

1 minute washes with 0.5 M Tris-HCl pH 7.5, 1.5 M NaCl, blotted dry and baked at 80°C for 2 hours *in vacuo*.

The hybridization of radioactive probes to these filters and detection of positive recombinants was carried out as described in Section 2.2.12. Plaques which gave positive signals with the probes were picked into 0.5 ml PSB (Section 2.1.7) and were purified by two further rounds of plating and rescreening at lower plaque density.

3.2.2 Preparation of Recombinant Phage DNA

Bacteriophage DNA was prepared using a method based on that of Kao *et al*. (1982). This method was found to give good yields of high quality DNA which obviated the need for purification on CsCl gradients (Frenkel, 1985). Approximately 10^8 phage were added 0.25 ml of PSB, 0.25 ml of 10 mM $CaCl_2$, 10 mM $MgCl_2$ and 0.25 ml of a fresh overnight of *E. coli* LE392 and incubated at 37°C for 15 minutes with gentle shaking. The mixture was then used to inoculate 50 ml of warm NZCYM broth (Section 2.1.8.3) and this was incubated at 37°C with vigorous shaking until lysis occurred - generally about 6 hours. Chloroform (2 to 3) drops were added, the incubation continued for 15 minutes, and the lysate stored overnight at 4°C. The bacterial nucleic acids were then removed by incubation for 1 hour on ice in the presence of 50 µg of RNase A and DNase I and the cellular debris removed by centrifugation (10,000 rpm, 10 minutes, 4°C). The resulting supernatant was centrifuged at 20,000 rpm for 3 hours at 4°C to pellet the bacteriophage which were then allowed to resuspend overnight in 0.4 ml of 0.3 M NaCl, 0.1 M Tris-HCl pH 8.0 at 4°C. The phage were then gently extracted three times in phenol saturated with 20 mM Tris-HCl

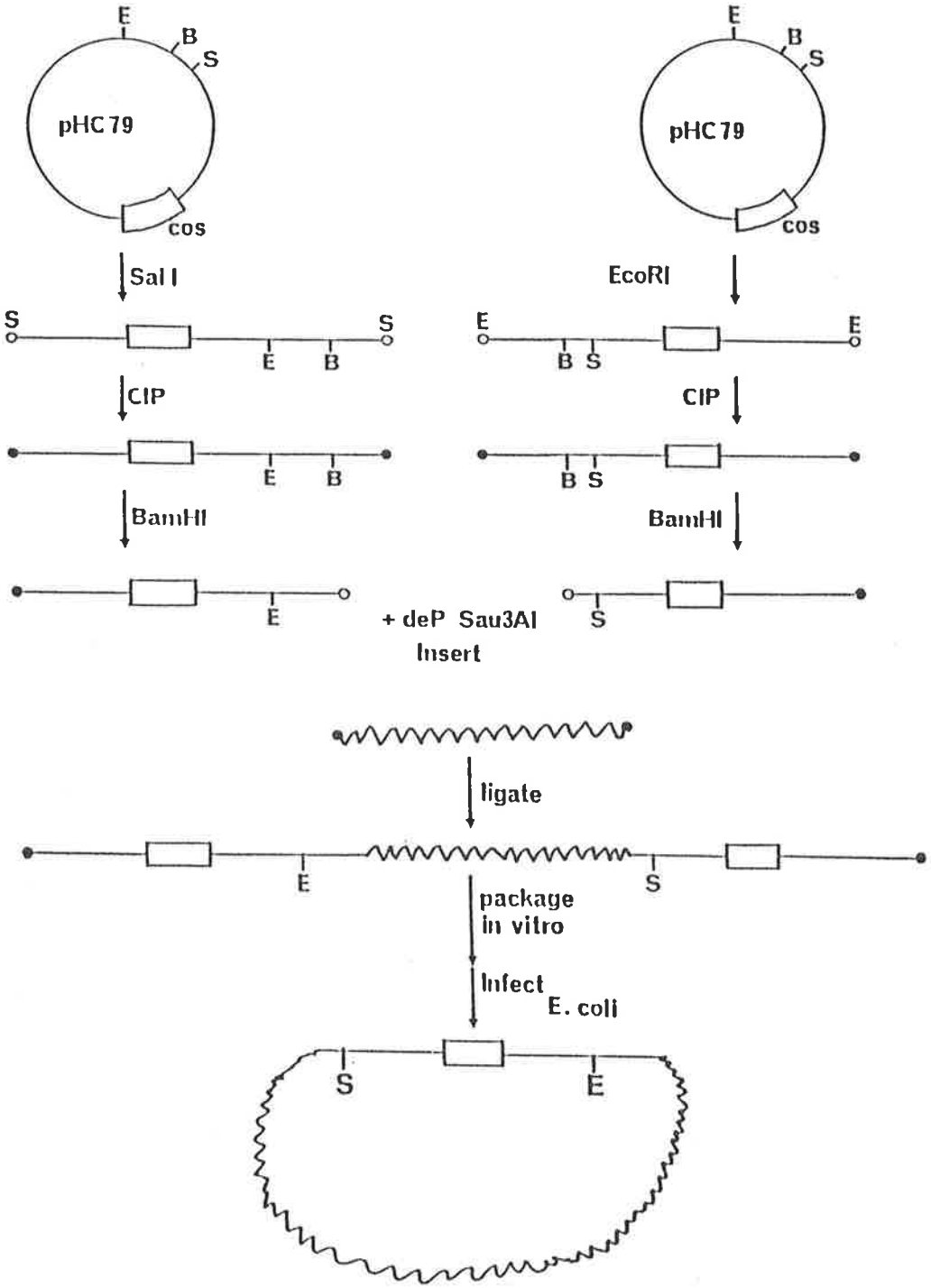
pH 8.0, 1 M NaCl, 1 mM EDTA, once with chloroform and 3 times with diethyl ether before ethanol precipitation at -20°C. The DNA was resuspended in 1 mM EDTA pH 7.5 and stored at 4°C. The procedure yielded up to 200 µg from a 50 ml preparation.

3.2.3 Construction of the Cosmid Library (Figure 3.3)

The chick cosmid library used in this study was constructed by R. D'Andrea and L. Tabe using a modified procedure to that described by Ish-Horowicz and Burke (1981) (Figure 3.3). This protocol was designed to overcome two major problems which had limited the usefulness of this cosmid cloning technique; (1) formation of vector concatamers leading to recombinant molecules with small genomic inserts and (2) ligation of smaller non-contiguous genomic fragments into a single insert, leading to a distorted view of genomic organization ('scrambling'). Ish-Horowicz and Burke (1981) overcame these two problems by using equimolar amounts of vector cut with two different enzymes (to prevent vector-vector ligation) and by dephosphorylating both insert and vector DNA to prevent self ligation of the vector and scrambling of genomic sequences (see Figure 3.3). In order to increase the ligation efficiency and further decrease the chances of scrambling of genomic sequences (Cattaneo et al., 1981; Grosveld et al., 1981), D'Andrea (1985) added a sizing step to select insert DNA of 35-45 kb in length. Using this procedure, they generated a library of about 50,000 recombinants which represents about 1 ½ chick genome equivalents (the chick genome is approximately 1.2×10^6 kb in size).

FIGURE 3.3. : PROTOCOL FOR COSMID LIBRARY CONSTRUCTION

This procedure is a modification of the protocol described by Ish-Horowicz and Burke (1981). Following digestion of pHC79 DNA with EcoRI or Sal I and inactivation of the protruding termini by calf intestinal phosphatase, CIP, the linear vector DNA is cleaved by Bam HI. The two vector fragments containing the cos sequences are isolated and ligated to genomic fragments (size fractionated; 35-45 kb) of eukaryotic DNA generated by partial digestion with Sau 3AI. An entire complement of plasmid DNA is contained between the two cos sequences. The concatamers formed are used as substrates for *in vitro* packaging. On introduction into *E. coli* host HB101 (Section 2.1.2) the DNA is circularized and replicates in the form of a large plasmid, conferring resistance to ampicillin (from D'Andrea (1985), with permission).


E
B
S
pHC79
cos
Sal I
S
S
E
B
CIP
BamHI
E
EcoRI
B S
+ deP Sau3AI
Insert
ligate
package
in vitro
Infect
E. coli

3.2.4 Screening the Cosmid Library

The cosmid library was replica plated and screened using the procedure of Hanahan and Meselson (1980).

The library, which was stored as eight sandwiches of duplicate nitrocellulose filters, was thawed and the sandwiches peeled apart. The eight original master nitrocellulose filters were grown for 2 hours on L-agar plates containing ampicillin and 25% glycerol, replicated as described (Hanahan and Meselson, 1980; Maniatis et al., 1982) and the sandwiches stored at -80°C between sterile Whatman 3 MM filter paper. (These filters would be used for future screenings).

The other eight duplicate nitrocellulose filters were placed on L-agar plates containing ampicillin and incubated at 37°C until the colonies had reached 0.5-1.0 mm in size (this usually took 6-8 hours). These filters were sandwiched and the replicate filters were grown as above. Following incubation, one set of plates was sealed with parafilm and stored at 4°C for picking positive recombinants. The duplicate set of filters were incubated overnight on L-agar plates containing 150 µg/ml chloramphenicol to amplify the plasmid (Hanahan and Meselson, 1980). The filters were then sandwiched with fresh nitrocellulose filters to provide duplicate filters for hybridization experiments and the colonies were then lysed (while the filters were sandwiched together) using a modification of the Grunstein and Hogness (1975) procedure (Section 4.2.5).

The hybridization of radioactive probes to these filters and detection of positive recombinants was carried out as described in Section 2.2.12. Cosmids which hybridized with the feather gene probe were picked into L-broth containing ampic-

illin and purified by a further one or two screenings (see Results).

3.2.5 Preparation of Cosmid DNA

Small scale 'miniscreen' cultures of cosmid DNA were grown and the DNA prepared as described by Grosveld et al. (1981) except that the isopropanol step was omitted. Large scale (500 ml) cultures were grown and the cosmid DNA isolated exactly as described for plasmids (Section 2.2.3) except that the large cultures were grown in the presence of 40 µg/ml ampicillin.

3.3 RESULTS

3.3.1 Screening of the λ Genomic Library

The two 'walking' probes which were used to screen the λ library had previously been subcloned into pBR322 (p λCFK1-2 and p λCFK1-8; Figure 3.2). The subclone p λCFK1-2 contains the 5' non-coding region of gene D and the 3' non-coding region of gene E as well as the intergenic region between the two genes; p λCFK1-8 is located downstream of gene A and contains no feather keratin genes (Figure 3.2).

The screening of about 2 x 10^5 phage (or about 2 ½ genome equivalents) using p λCFK1-2 as a probe resulted in the detection of four strongly positive clones (Figure 3.4). In contrast, screening of an identical number of phage with p λCFK1-8 resulted in the detection of several hundred positive clones (data not shown). In view of the fact that p λCFK1-8 does not contain any keratin genes (Figure 3.2; Molloy et al., 1982), this result was unexpected and was probably due to the fact that this subclone contains a repeat sequence. Neither probe was assessed for the presence of repeated sequences by hybridization to restricted chicken DNA.

FIGURE 3.4 : PRIMARY SCREENING OF A CHICKEN λ GENOMIC LIBRARY

The chicken λ library was plated at a density of 5 x 10^4 pfu/plate on four 150 mm plates. The plaques were transferred to duplicate nitrocellulose filters and the phage DNA denatured and immobilized on the filters as described in Section 3.2.1. The filters were probed with the insert of pλCFK1-2 (Figure 3.2) after labelling by nick-translation (Section 2.2.11.2), washed with 1.0 x SSC/0.1% SDS at 65°C and autoradiographed for 2 days at -80°C. The arrow indicates the single strong positive signal obtained on each of these duplicate filters.

FIGURE 3.5 : PURIFICATION OF POSITIVE λ RECOMBINANTS OBTAINED FROM THE PRIMARY SCREENING

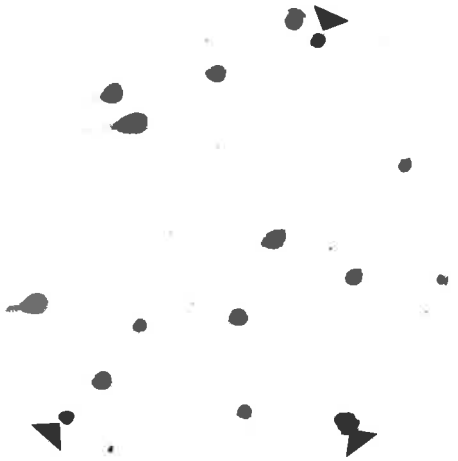
The positive recombinant phage obtained from the primary screening of the λ library with the p λCFK1-2 probe (see Figure 3.4) were picked into 0.5 ml PSB (Section 2.1.7), replated on to small 90 mm plates and the plaques rescreened with p λCFK1-2 probe (Figure 3.4).

a. Second round screening: About 2 x 10^3 phage were plated, filters prepared and the phage DNA hybridized with p λCFK1-2 probe.

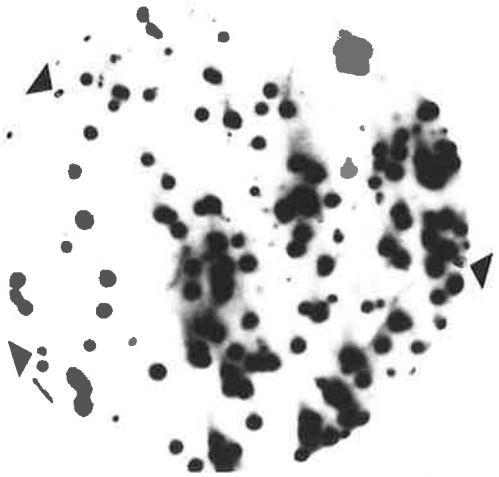
b. Third round screening: About 200-300 phage were plated, filters prepared and the phage DNA hybridized with p λCFK1-2 probe.

Arrows indicate the position of holes used for orientation of filters.

a.



b.



The four λ recombinants which hybridized to the p λCFK1-2 insert were picked into 0.5 ml PSB and purified by two further rounds of plating and rescreening (Figure 3.5). All four clones gave positive signals in subsequent hybridization steps. The phage DNA was prepared as described in Section 3.2.2. The characterization of these clones is described in Chapter 4 (Section 4.3.1). Only one of these clones (λCFK9) extended λCFK1.

3.3.2 Screening of the Cosmid Library

The hybridization probe used to screen the cosmid library was a 2 kb EcoRI/Hind III fragment which contains a feather keratin gene and some flanking DNA (Section 2.2.10). In a control experiment to check that the probe would only hybridize to keratin gene sequences and not to the cosmid vector, the probe was first annealed to pBR322 as well as λCFK1 and the subclone p λCFK1-4. Figure 3.6 clearly shows that, as expected, the probe strongly annealed to both λCFK1 and p λCFK1-4 but not to pBR322 DNA even at the highest amount of DNA spotted.

An example of an autoradiograph of the first round of library screening is shown (Figure 3.7). Thirty-five regions corresponding to positive signals of various strength were identified of which most but not all were present on duplicate filters. All positives were picked into a small volume of L-broth containing ampicillin (40 µg/ml) and replated on to small ampicillin plates. Fifteen of the original isolates gave positive hybridization signals in the rescreen (Figure 3.8). In most cases a third screening was performed at low plating density (100 colonies per plate) from which single colonies were picked.

FIGURE 3.6 : DOT BLOT ANALYSIS OF CONTROL DNAS WITH THE COMPLETE FEATHER GENE PROBE

A dot blot filter was prepared (Section 2.2.14) containing varying amounts of λCFK1, pλCFK1-4 and pBR322 DNAs (see below). The 2 kb EcoRI/Hind III fragment of pλCFK1-4 containing feather keratin gene B (Section 2.2.10) was isolated from agarose (Section 2.2.8) and labelled by nick translation (Section 2.2.11.2). The filter-bound DNA was hybridized with the gene B probe, washed in 1.0 x SSC, 0.1% SDS and exposed overnight at -80°C (Section 2.2.12). The amount of DNA (µg) spotted on to the filter from left to right was as follows.

pBR322 (top panel): 0.1, 0.5, 1.5 and 3.0 µg.

λCFK1 (bottom panel): 0.25, 0.75 and 1.5 µg.

pλCFK1-4 (bottom panel): 0.25 µg.

pBR322

λ CFK1

pλ CFK1-4

FIGURE 3.7 : PRIMARY SCREENING OF THE CHICKEN COSMID LIBRARY FOR KERATIN SEQUENCES

About 6000 recombinants grown on a 150 mm nitrocellulose filter were lysed, the cosmid DNA immobilized and screened in duplicate by hybridization using a complete feather keratin gene (see Figure 3.6) as a probe (Section 3.2.4).

The filters were washed at moderate stringency (1.0 x SSC, 0.1% SDS at 65°C) and autoradiographed at -80°C for 1-3 days with an intensifying screen.

The Figure shows autoradiographs of two colony filters.

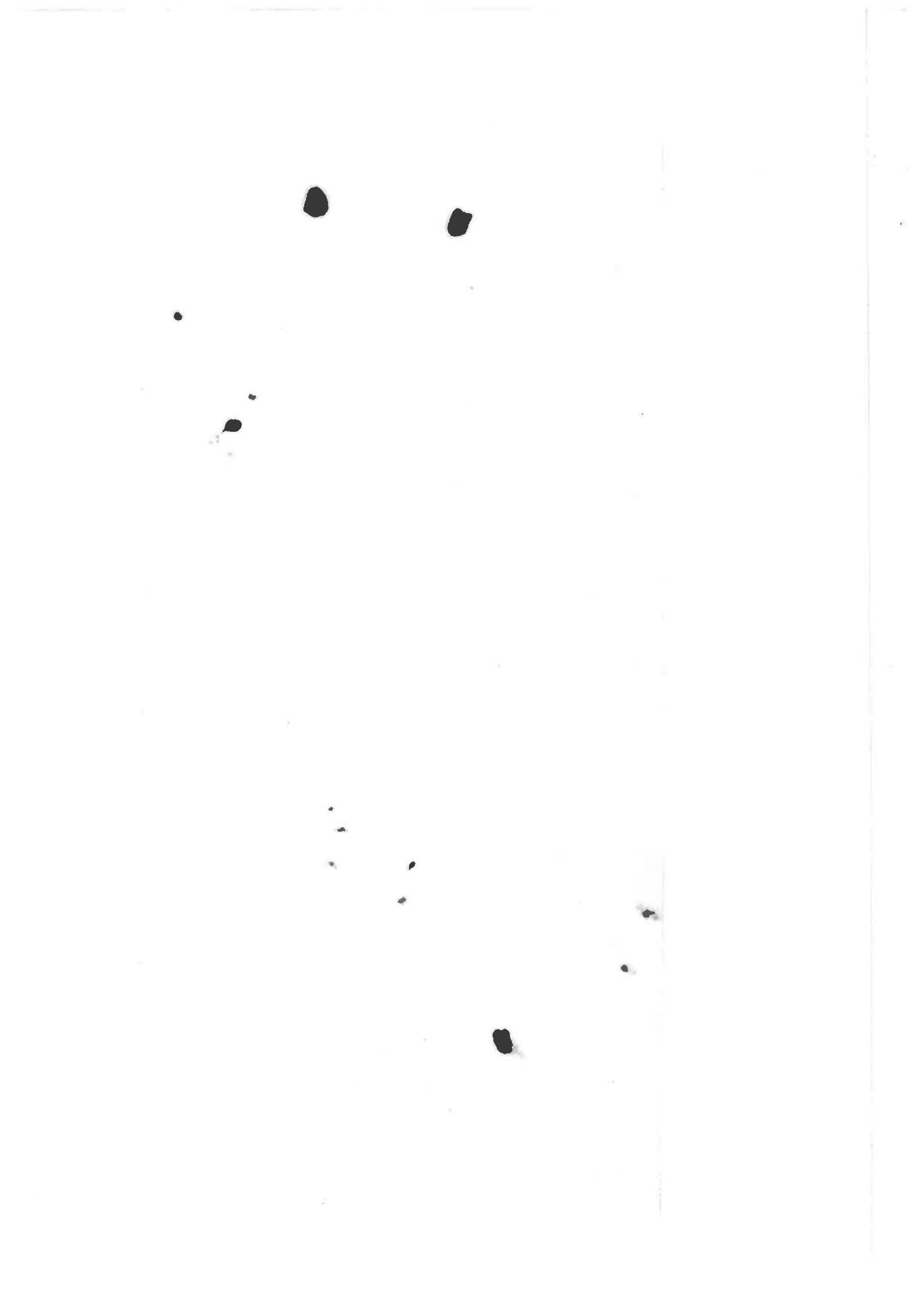
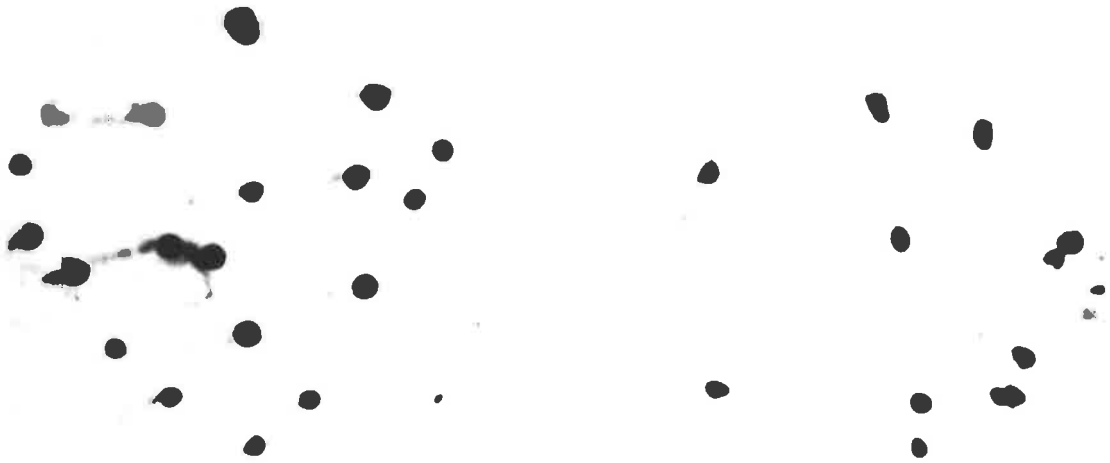


FIGURE 3.8 : PURIFICATION OF POSITIVE RECOMBINANTS OBTAINED FROM THE PRIMARY SCREENING OF THE CHICKEN COSMID LIBRARY

The positive clones obtained from the primary screening of the chicken cosmid library (Figure 3.7) were picked into L-broth + ampicillin (40 µg/ml), replated on to 90 mm plates and the resulting cosmid filters hybridized and washed as described in the legend of Figure 3.7.

The Figure shows autoradiographs of two cosmid isolates after a second round screening. Density was generally about 1000 colonies/plate.



3.3.3 Preliminary Characterization of Cosmid Recombinants by Dot Blot Hybridization

Cosmid DNA was prepared from 10 ml overnight cultures of each of the purified recombinants using the small scale cosmid DNA isolation method described in Section 3.2.5.

Samples (0.25 µg) of each clone were spotted on to a series of nitrocellulose filters (Section 2.2.14) and hybridized with four different DNA probes. The following probes were used in the preliminary analysis of the cosmid clones:

(a) The feather keratin gene probe used to screen the library (Section 2.2.10).

(b) A synthetic 19 base sequence which is present in the 3' non-coding region of the feather keratin genes in λCFK1 (Gregg and Rogers, 1986; Gregg et al., in preparation). This sequence has not been found in the 3' non-coding region of other keratin gene families or fast protein genes (Wilton, 1983; Morris, 1984; Whitbread, 1985) and has therefore been used as a feather-specific probe (see Section 2.2.10).

(c) A 0.85 kb EcoRI/Hind III fragment isolated from the subclone pλCFK1-6, which contains the 1.84 kb Hind III fragment from λCFK1 (Figure 3.2). This fragment does not contain any feather keratin genes and furthermore, a Southern blot hybridization to chicken DNA restricted with EcoRI or Hind III demonstrated that it is unique in the genome (data not shown). Therefore, only those cosmid clones which span the probe should hybridize to it.

(d) A M13 clone (1.8 kb) which contains a complete scale gene (Section 2.2.10).

The results of the dot blot hybridizations are shown in Figure 3.9 and are summarized in Table 3.1. As observed during the screening process (Figure 3.7), there is a considerable variation in the intensity of the hybridization signal between clones with both the feather probes (Figure 3.9a,b). For example, Cosmids 4, 11, 12, 25 and 33 hybridize more strongly with the complete feather gene probe than do cosmids 2, 5, 16, 28, 31, 34 and 35 which in turn give stronger signals than cosmids 1 and 3 (Table 3.1). The pattern of hybridization intensity observed with the feather-specific oligonucleotide probe (Figure 3.9b) is mostly similar to that found with the complete feather gene probe with four exceptions; cosmids 1 and 5 which both hybridized with the complete feather gene probe, failed to hybridize to the feather-specific probe and cosmids 31 and 35 which only hybridize weakly to the feather-specific probe even though they bound to the complete feather gene probe quite strongly. It is also interesting to compare the filters probed with the complete feather and scale gene probes (Figure 3.9a,d; Table 3.1). For example, cosmid 4, 11, 12, 25, 28 and 33 hybridize more strongly with the feather probe than the scale probe; in the case of 2 and 35 the reverse is true. Cosmids 5, 16, 31 and 34 appear to bind feather and scale probes equally as well; cosmid 5, however, does not hybridize at all with the feather-specific oligonucleotide probe (Figure 3.9b).

FIGURE 3.9 : DOT BLOT ANALYSIS OF COSMID RECOMBINANTS CONTAINING KERATIN-RELATED GENES

Cosmid DNA was prepared from 10 ml overnight cultures of each of the purified recombinants (Section 3.2.5) and 0.25 µg aliquots of each clone were spotted on to a series of nitrocellulose filters (Section 2.2.14). The four DNA probes (Section 3.3.3), prepared as detailed below, were then hybridized to the dot filters, washed and autoradiographed at -80°C (Section 2.2.12).

	PROBE	PREPARATION OF PROBE (REFERENCE)	WASHING CONDITIONS
a.	Complete feather gene probe	Section 2.2.10	1.0 x SSC/0.1% SDS, 65°C
b.	Feather-specific 19-mer	Section 2.2.10	4 x SSC, 65°C
c.	λCFK1-specific probe	Sections 2.2.8, 2.2.11.2 & 3.3.3	0.1 x SSC/0.1% SDS, 65°C
d.	Complete scale gene probe (M13 clone)	Section 2.2.10	1.0 x SSC/0.1% SDS, 65°C

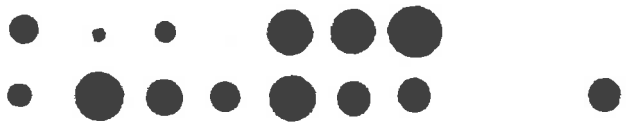
The pattern of cosmids on the filters was:

5	1	2	3	4	11	12		
16	25	28	31	33	34	35		λCFK1

One recombinant (cosmid 29) was not included in the dot blot experiments reported here.

PROBE

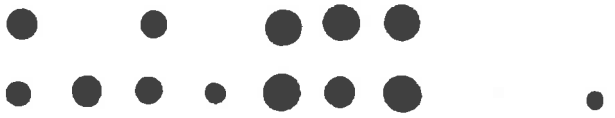
(a) complete feather gene



(b) feather-specific 19 mer

(c) λ CFK1-specific probe

(d) complete scale gene



Four cosmid clones (4, 11, 12 and 25) hybridized to the λCFK1-specific intergenic probe (Figure 3.9c) and thus presumably span at least this portion of the genomic clone λCFK1. No hybridization was observed to pBR322 DNA by the complete feather gene probe (Figure 3.6) or by any of the other three probes (data not shown).

3.4 DISCUSSION

This chapter described the isolation of several λ clones selected using a 'walking' probe and the isolation and preliminary characterization of a bank of cosmid clones selected using a chick feather keratin gene as a probe. Initial attempts to 'walk' from the previously isolated clone λCFK1 were made by screening the existing λ Charon 4A library with probes from both ends of λCFK1 (Figure 3.2). Four positive recombinants were isolated by such an approach, only one of which was found to extend λCFK1 (see Section 4.3.1). At about this time a chick cosmid library was constructed in this Department from which clones bearing histone genes were isolated (D'Andrea, 1985; D'Andrea et al., 1985). In view of the advantages of using cosmid libraries for chromosome walks (outlined in Section 3.1), it was decided to screen this library using a complete avian feather keratin gene as a probe.

The primary screening of the cosmid library with a feather keratin probe resulted in the detection of 35 positive spots (Figure 3.7) of which 15 rehybridized upon rescreening (Figure 3.8). Most of the original positives were present on duplicate autoradiographs, and great care was taken to align the autoradiographs with the master filters when the colonies were being picked for rescreening, and so it was somewhat

surprising that such a decrease in the number of positive signals should occur. Similar decreases in the number of positives obtained during the purification of cosmid recombinants from libraries have been reported by other groups (Cattaneo et al., 1981; Grosveld et al., 1981).

The cosmid clones, once purified, were analysed by dot blot hybridization with a series of DNA probes (Figure 3.9), the results of which are summarized in Table 3.1. There was a considerable variation in the intensity of the hybridization signal between clones which had been probed with the two feather probes (Figure 3.9a,b; Table 3.1). This could be caused by a number of factors. Firstly, it could be due to variations in the yield of cosmid DNA. However, the yields of DNA obtained from the small scale (10 ml) overnight cultures were remarkably constant as judged by agarose gel electrophoresis with the exception of cosmid 3 from which very small amounts of DNA were obtained (data not shown). This is consistent with the very low hybridization signals observed with this recombinant. Cosmid 3 was later shown to be identical to cosmid 4 (see Chapter 4). Secondly, it could be due to different numbers of keratin genes present in the recombinants; detailed analysis of some of the cosmid recombinants demonstrated that there is indeed a correlation between the intensity of hybridization with the feather gene probes (Figure 3.9a,b; Table 3.1) and the number of feather keratin genes present in the clone (see Section 4.4.1).

Careful examination of the dot blot hybridization data suggests that some of these recombinants may contain keratin genes other than those of the feather type. It is important to

TABLE 3.1 : ANALYSIS OF COSMID CLONES BY DOT BLOT HYBRIDIZATION WITH FOUR DNA PROBES

The results obtained in Figure 3.9 with each of the DNA probes is presented, showing the relative signal intensities. The symbols used were as follows:

+ + +	strong hybridization
+ +	moderate hybridization
+	weak hybridization
-	no hybridization

COSMID CLONE	(a) COMPLETE FEATHER GENE	(b) FEATHER-SPECIFIC PROBE	(c) λCFK1-SPECIFIC PROBE	(d) COMPLETE SCALE GENE
1	+	-	-	-
2	+	+	-	++
3	+/-	-	-	-
4	+++	+++	+++	++
5	++	-	-	++
11	+++	+++	+	++
12	+++	+++	+++	++
16	++	+	-	++
25	+++	+++	+++	++
28	++	++	-	++
31	++	++	-	++
33	+++	++	-	++
34	++	++	-	++
35	++	+	-	+++
Controls				
λCFK1	++	++	++	+
pBR322	-	-	-	-

note that the scale probe used in this study does hybridize to feather keratin genes in λCFK1 (Figure 3.9d) but the signal is considerably less intense than that observed with the feather keratin gene probe (Figure 3.9a). This is due to the considerable DNA sequence homology between scale and feather keratin genes particularly in the N-terminal half of their coding regions (see Figure 1.9a; Gregg et al., 1984). Cosmid 2 hybridizes more strongly with the scale gene probe than the feather gene probe but must contain at least one feather gene since it hybridizes with the feather-specific probe (Figure 3.9b). Therefore it seems possible that this recombinant contains both scale and feather genes.

A similar pattern of hybridization was observed with cosmid 35, i.e. this recombinant hybridized strongly with scale and feather probes. Recent studies suggest that this clone may contain feather, scale and claw keratin genes (L. Whitbread, personal communication). Cosmid 5, although it apparently bound complete feather and scale gene probes equally as well, failed to hybridize with the feather-specific probe (Figure 3.9b) indicating that either it contains feather keratin genes variant in the region spanned by the feather-specific oligonucleotide or that its keratin genes are of a different type, such as scale.

Four cosmid clones hybridized to the λCFK1-specific probe (Figure 3.9c) and as the cosmid insert should be at least twice as large as that of the λ clone it should be possible to determine whether λCFK1 is indeed part of a longer cluster of tandemly arranged genes (Molloy et al., 1982).

CHAPTER 4.

CHARACTERIZATION OF

CHICKEN KERATIN GENOMIC CLONES.

4.1 INTRODUCTION

In the preceding chapter, four λ Charon 4A and 15 cosmid recombinants were selected from chicken λ and cosmid genomic libraries using a 'walking' fragment from λCFK1 (which contains the 5' non-coding region of gene D, the intergenic region between genes D and E and the 3' non-coding region of gene E, Figure 3.2) and a feather keratin gene, respectively, as probes. This chapter describes the characterization of most of those cosmid recombinants and one of the λ clones. This analysis involved extensive mapping using several six base restriction enzymes, Southern blot hybridization and the sub-cloning of restriction fragments into pBR322 plasmids to determine the location and arrangement of the keratin genes. Using these techniques, it was shown that most of the clones were derived from a single chromosomal locus which included the genomic region contained in the recombinant λCFK1 (Molloy et al., 1982). These studies have enabled a complete analysis of this feather keratin gene locus.

4.2 METHODS

4.2.1 PREPARATION OF pBR322 VECTORS

pBR322 DNA (5 µg) was restricted with the appropriate enzyme(s) (Section 2.2.4) and the linearized vector dephosphorylated to prevent self ligation. Two µl of Calf Intestinal Phosphatase (7 U/µl) were added to the restriction digest and the reaction incubated for 2 hours at 37°C. The reaction was then terminated by adding EDTA (pH 7.5) to 10 mM and SDS to 0.5% and heating at 70°C for 15 minutes. The vector DNA was then extracted twice with phenol/chloroform (1:1), once with chloroform and once with ether before ethanol precipitating (Section 2.2.2).

The vector DNA was resuspended in 1 mM EDTA pH 7.5 at a concentration of 50 ng/ µl.

4.2.2 PREPARATION OF INSERT DNA

DNA to be subcloned was restricted with the appropriate enzyme(s) (Section 2.2.4). The products were fractionated on a LGT agarose gel and the appropriate fragments excised and isolated from the agarose as described in Sections 2.2.8.

4.2.3 DNA LIGATIONS

The DNA fragment and appropriate pBR322 vector were combined in a molar ratio of approximately 3:1 in a 20 µl reaction mixture containing 50 ng (1 µl) vector, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM rATP, 1 mM DTT, 0.5 - 1 U ligase (1 U/ µl) and the DNA fragment to be cloned. Ligations were performed for 4-8 hours at 15°C or, more commonly, overnight at 4°C.

4.2.4 BACTERIAL TRANSFORMATION

E. coli strain LE392 or MC1061 cells were made competent and transformed using a modification of the method described by Dagert and Ehrlich (1979). *E. coli* cells were grown overnight at 37°C in L-broth and then diluted 1:50 into fresh L-broth and grown at 37°C to an A_{600} of 0.4-0.6. The cells were chilled on ice for 15 minutes, harvested by low speed centrifugation at 4°C and gently resuspended in a ½ volume of ice-cold 0.1 M $CaCl_2$. The cells were kept on ice for a further 20-30 minutes before repelleting and resuspending in 0.1 volumes of cold 0.1 M $CaCl_2$. The *E. coli* cells were kept on ice for at least 2 hours before use.

To transform the cells, the ligated DNA, diluted to 0.1 ml in 10 mM Tris-HCl pH 7.5, was added to 0.2 ml cells and kept on ice for 30 minutes with occasional mixing. The mixture was

warmed at 37°C for 5 minutes after which 2 ml of L-broth was added and the transformed cells incubated at 37°C with shaking for 1 hour to allow expression of the plasmid's antibiotic resistance. Aliquots of each tube were then plated on to L-agar plates containing ampicillin and incubated for 12-15 hours at 37°C.

4.2.5 DETECTION OF RECOMBINANT PLASMIDS

Initially, detection of pBR322 plasmids carrying DNA complementary to various keratin genes was carried out using the procedure of Grunstein and Hogness (1975) as modified by Maniatis et al. (1982). Amp^r tet^s recombinants (i.e. recombinants containing some form of cloned DNA) were transferred using sterile toothpicks to nitrocellulose sheets that had been boiled three times in distilled water and overlaid on to L-agar plates containing ampicillin. The duplicate plates were incubated overnight at 37°C. DNA from the colonies grown on the filters was liberated using the lysis procedure of Maniatis et al. (1982).

Radioactive DNA probes were hybridized to the colonies as described in Section 2.2.12.

4.2.6 PLASMID MINISCREENS

The isolation of plasmid DNA from small overnight cultures for size estimation of the inserted sequence and preliminary restriction analysis was carried out as follows. The method is a modification of that described by Holmes and Quigley (1981).

The recombinants were grown overnight at 37°C in L-broth (3 ml) containing 40 µg/ml ampicillin in a rotary shaker. The cells were pelleted by centrifugation at 5,000 rpm in a bench-top centrifuge, washed once with 1 ml of 50 mM Tris-HCl pH 9.0,

50 mM EDTA and resuspended in 200 µl of 15% sucrose, 50 mM Tris-HCl pH 9.0 and 50 mM EDTA. Lysozyme (200 µg in 25 µl H_2O) was added and the mixture incubated at room temperature for 15 minutes and then on ice for 30 minutes. Ice-cold water (175 µl) was added and the samples heated at 70°C for 15 minutes. Cellular debris was removed by centrifugation at 20,000 rpm for 45 minutes at 4°C. The supernatant was carefully decanted, extracted once with buffer-saturated phenol, twice with ether and ethanol precipitated.

About 2-3 µg of plasmid DNA was generally obtained by this method, and providing the contaminating RNA was removed using RNase A, the DNA generally cut well with most restriction enzymes.

4.3 RESULTS

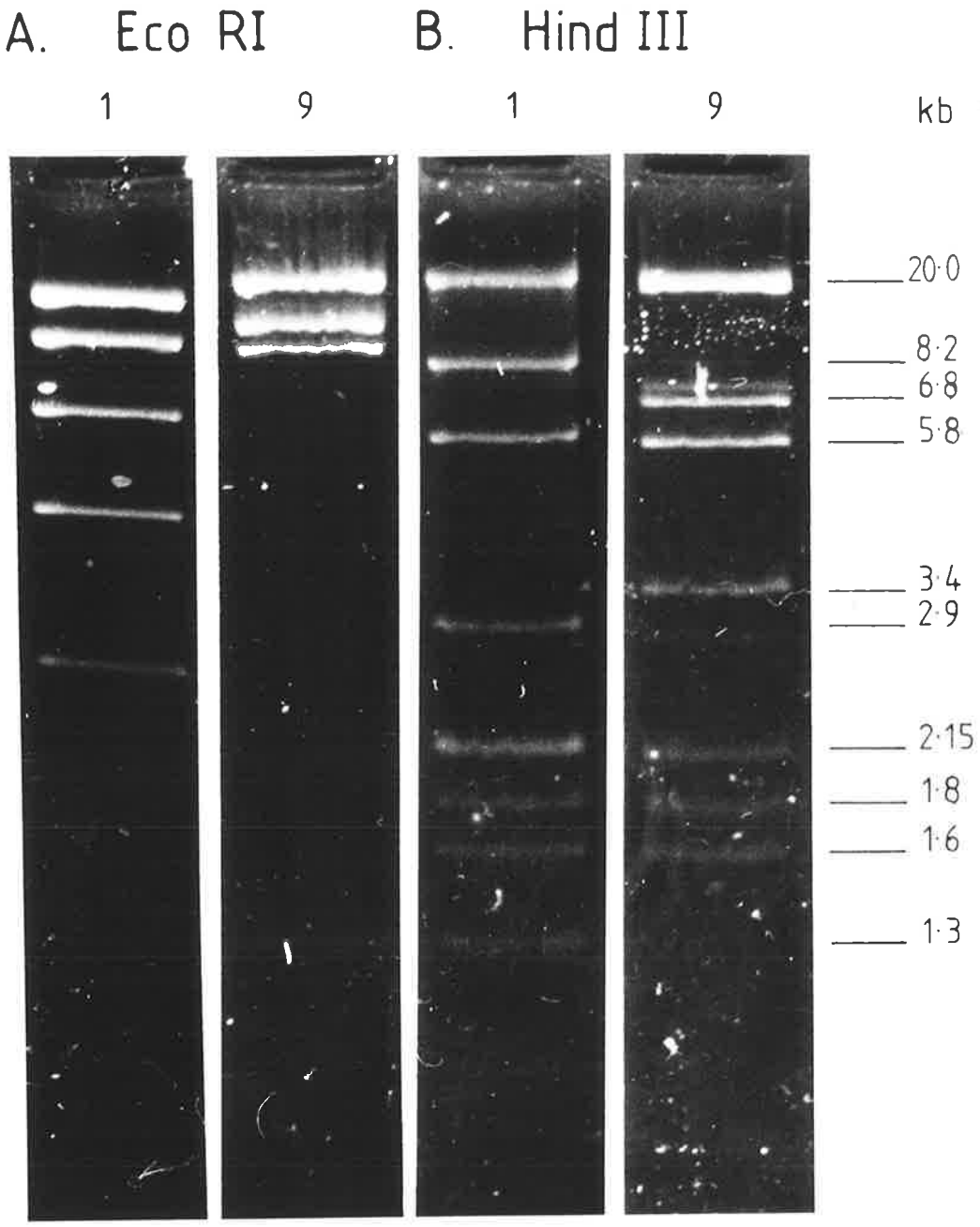
4.3.1 CHARACTERIZATION OF λ RECOMBINANTS

As described in Chapter 3 (Section 3.3.1), the screening of approximately 2 x 10^5 phage (2 ½ genome equivalents) using the 'walking' probe pλCFK1-2 (Figure 3.2) resulted in the detection of four strongly positive clones (Figure 3.4) which were picked and purified (Figure 3.5). Of the four clones, two were identical to λCFK1 and a further one (λCFK8) was judged to be a rearrangement of λCFK1 (data not shown).

The fourth isolate (λCFK9) was found to overlap extensively with λCFK1. Figure 4.1 shows the restriction pattern of λCFK9 and λCFK1 obtained after digestion with EcoRI or HindIII followed by agarose gel electrophoresis. The HindIII fragments of sizes ~ 20 kb and 5.8 kb which are present in both recombinants (Figure 4.1b) contain the left and part of the right arms

FIGURE 4.1 : RESTRICTION ANALYSIS OF λCFK1 AND λCFK9

Phage DNA prepared from the two λ clones, λCFK1 and λCFK9 (denoted as 1 and 9, respectively), was digested with either EcoRI (A) or HindIII (B) and fractionated on a 1% agarose gel (Section 2.2.7). DNA was visualized under UV light after ethidium bromide staining. The sizes of the HindIII fragments produced from the HindIII digests of λCFK1 (Molloy et al. , 1982) and λCFK9 are shown.


A. Eco RI
B. Hind III
1
9
1
9
kb
20·0
8·2
6·8
5·8
3·4
2·9
2·15
1·8
1·6
1·3

of the λ Charon 4A vector, respectively (see Figure 4.3; Blattner et al., 1977; de Wet et al., 1980). Of the remaining HindIII fragments, it can be seen that λCFK9 contains one of the 2.15, the 1.8 and 1.6 kb bands present in λCFK1 but not the 2.9, the other 2.15 and the 1.3 kb bands of λCFK1. From the known HindIII restriction map of λCFK1 (Figure 4.3) these data positioned the left end of λCFK9 within the 2.9 kb HindIII fragment of λCFK1 which contains gene B (Figure 4.3). The two remaining HindIII fragments of λCFK9 (6.8 and 3.4 kb) had no counterparts in λCFK1 which suggested that these DNA fragments, at least in part, were not contained in λCFK1. The single HindIII site in the λ Charon 4A vector will produce a HindIII vector fragment of 5.8 kb and a vector-insert fragment of more than 5.2 kb (Blattner et al., 1977; de Wet et al., 1980). Thus, the 6.8 kb HindIII fragment of λCFK9 contains 5.2 kb of λ vector DNA and (by subtraction) 1.6 kb of insert DNA. This was confirmed by EcoRI/HindIII digestion of λCFK9 (Figure 4.8). The only HindIII fragment so far not positioned in λCFK9, the 3.4 kb fragment, must extend the end 3.0 kb EcoRI/HindIII fragment of λCFK1 rightwards by 0.4 kb. This fragment is positioned between the 1.6 kb HindIII fragment, which contains most of gene D, and the 1.6 kb EcoRI/HindIII fragment which is at the end of the λCFK9 insert (Figure 4.3).

The map of λCFK9 derived from restriction analyses was confirmed by Southern hybridization experiments presented in Figure 4.2. Figure 4.2a shows the pattern of hybridization obtained with a complete feather keratin gene (the 2 kb EcoRI/HindIII fragment containing gene B, Figure 4.3) as probe. As expected, the 2.15 and 1.6 kb HindIII fragments present in both

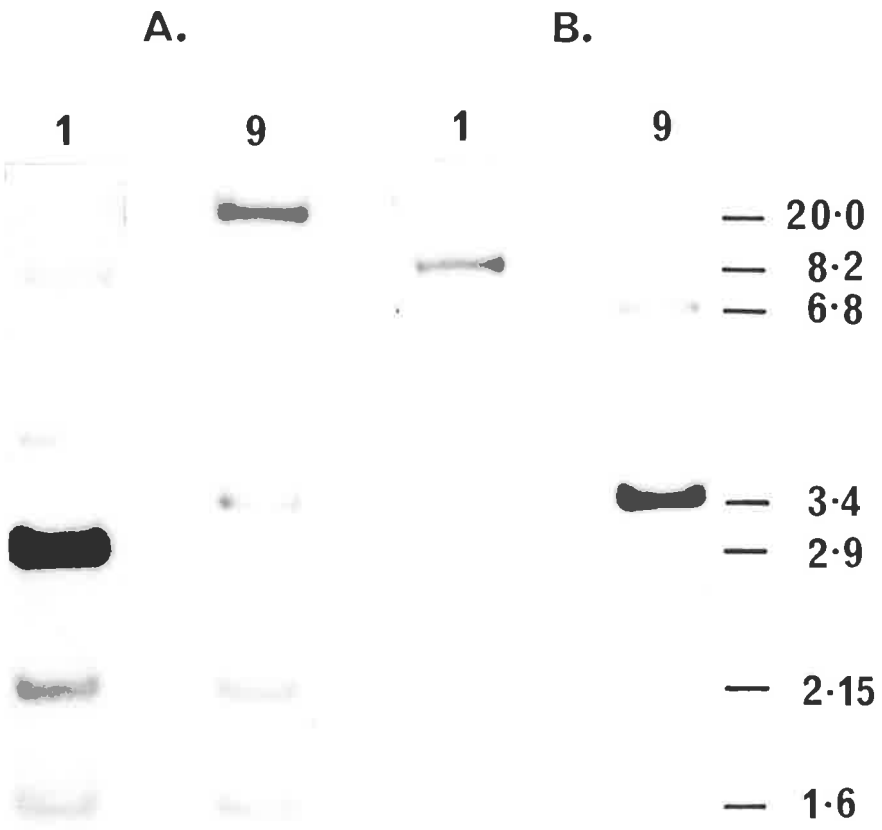
FIGURE 4.2 : SOUTHERN ANALYSIS OF λCFK1 AND λCFK9

λ CFK1 (1) and λCFK9 (9) DNAs which had been digested with HindIII were transferred to two nitrocellulose filters (Section 2.2.9) and hybridized with either

A - the feather keratin gene probe (Section 2.2.10)

B - the nick-translated (Section 2.2.11.2) insert of pλCFK1-2 (Section 4.3.1, Figure 4.3).

The filters were washed in 0.5 x SSC, 0.1% SDS at 65°C and bands detected by autoradiography at -80°C. The sizes (kb) of the hybridizing fragments are indicated.


A.
B.
1
9
1
9
20·0
8·2
6·8
3·4
2·9
2·15
1·6

clones, and the 2.9 and 8.2 kb HindIII fragments of λCFK1, hybridized with the probe. (The 8.2 kb HindIII fragment contained the end 3.0 kb of the λCFK1 insert and 5.2 kb of λ DNA, Figure 4.3). The remaining band hybridizing in λCFK1 did not correspond to any DNA fragment and was presumed to be a partial restriction fragment.

In λCFK9, the large 20 kb fragment hybridized strongly with the feather keratin gene B probe; this fragment contains the left λ vector arm and about 0.4 kb of the 2.9 kb HindIII fragment from which the gene B probe was prepared (see Figure 4.3). The 3.4 kb Hind III fragment of λCFK9 was also detected by the gene probe, consistent with it encompassing the 3.0 kb EcoRI/HindIII fragment of λCFK1 (Figure 4.3).

In order to confirm this, the 'walking' probe p λCFK1-2, which contains the end 3.0 kb EcoRI/HindIII fragment of λCFK1 and was used to isolate λCFK9 (Figure 3.2), was hybridized to HindIII digests of λCFK1 and λCFK9 (Figure 4.2b). As expected, the 'walking' probe hybridized strongly to the 8.2 kb HindIII fragment of λCFK1 (Figures 4.2b, 4.3) and, as predicted above, the 'walking' probe hybridized most strongly to the 3.4 kb HindIII fragment of λCFK9 (Figure 4.2b), suggesting that they overlapped each other (Figure 4.3). Since the p λCFK1-2 probe contains the 5' non-coding region of gene D and the 3' non-coding region of gene E (Figure 4.3), it also cross-hybridizes to the 2.15 kb HindIII fragments of λCFK1 and λCFK9 and to the 6.8 kb HindIII fragment of λCFK9 (Figure 4.2b). On a longer exposure, the p λCFK1-2 probe also weakly detected the 1.6 kb HindIII fragment, which contains the coding and 3' non-coding regions of gene D (Figure 4.3). The hybridization of the probe

FIGURE 4.3 : COMPARISON OF THE RESTRICTION MAPS OF λCFK1 AND λCFK9

The restriction mapping data for λCFK1 is from Molloy et al. (1982). The location of the five feather keratin sequences (A-E) in λCFK1 was determined by DNA sequence analysis (Gregg et al., manuscript in preparation). The map of λCFK9 was determined from the data presented in Section 4.3.1. The recombinant λCFK9 contains the whole of gene E which is only partially contained in λCFK1 (Molloy et al., 1982; Gregg et al., manuscript in preparation)

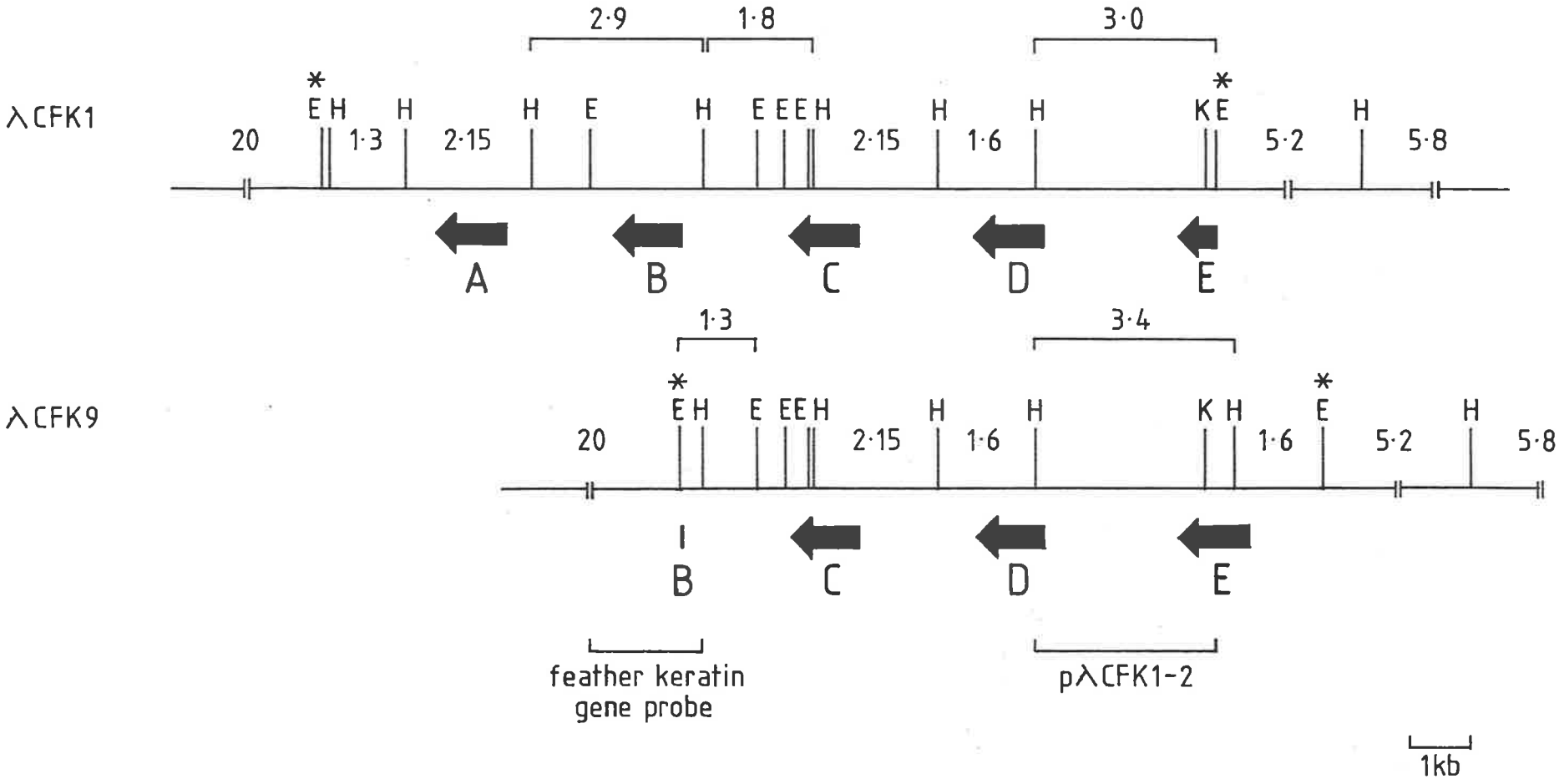
DNA segments contained in the two restriction fragments (the feather keratin probe, which comprises all of gene B (Section 2.2.10) and pλCFK1-2) used as hybridization probes in Figure 4.2 are shown in the figure. The sizes (kb) of some of the restriction fragments, including those which are contained in the λ Charon 4A vector (Blattner et al., 1977; de Wet et al., 1980), are indicated. The false EcoRI sites which are located at each end of the λ insert are starred (E*).

Restriction sites shown are:

E - EcoRI

H - HindIII

K - KpnI


λCFK1
2·9
1·8
3·0
E H
H
H
E
H
E E E H
H
H
K E
H
20
1·3
2·15
2·15
1·6
5·2
5·8
A
B
C
D
E
λCFK9
1·3
3·4
E H
E
E E H
H
H
K H
E
H
20
2·15
1·6
1·6
5·2
5·8
B
C
D
E
feather keratin
gene probe
pλCFK1-2
1kb

to the 6.8 kb fragment suggests that gene E spans the HindIII site in λCFK9 (Figure 4.3). This was later confirmed in studies on cosmid clones which span this region (see Section 4.3.4.2b).

The EcoRI restriction pattern obtained with λCFK9 served to confirm these results (Figure 4.1a). There are four EcoRI insert fragments in λCFK9, two easily visible of 9.0 and 1.3 kb and a doublet of 0.5 kb visible in EcoRI digests when more DNA was loaded on the gel. From the HindIII map of λCFK1 and λCFK9 (Figure 4.3), there are no EcoRI sites in the 1.6, 2.15 and 3.4 kb HindIII fragments. Therefore, the 9.0 and 1.3 kb EcoRI fragments of λCFK9 must span the regions to the right and left of the 0.5 kb EcoRI doublet, respectively (Figure 4.3). The size of the smaller EcoRI fragment (1.3 kb) enabled the location of the left of the clone insert to be determined precisely (Figure 4.3).

The final restriction maps of λCFK9 and λCFK1 are compared in Figure 4.3. The recombinant λCFK9 extends λCFK1 by a total of 2.0 kb and contains the whole of gene E. Although the clone λCFK9 provided little new data pertaining to the organization of feather keratin genes, it aided considerably in the mapping of cosmid 4 which is described below.

4.3.2 CHARACTERIZATION OF COSMIDS WHICH OVERLAP WITH λCFK1

Restriction mapping of the feather keratin gene locus was facilitated by the isolation from the chick cosmid library of a number of overlapping recombinants, some of which contained regions present in the previously characterized λ recombinants (see Figure 4.3). In the preliminary dot blot characterization of the cosmid clones described in Chapter 3, three cosmid

clones were strongly detected by the λCFK1-specific probe (a unique intergenic probe, see Section 3.3.3) and were designated 4, 12 and 25. This section describes the detailed characterization of these three clones. Cosmid 4, which was the first recombinant to be mapped, was subsequently used as the 'reference' clone to which all other keratin-containing cosmid recombinants were compared.

4.3.2.1 Restriction Mapping of Cosmid 4

A. The EcoRI and HindIII maps. Figure 4.4a shows the pattern of restriction fragments obtained after EcoRI digestion of cosmid 4. A total of 8 restriction fragments were generated, including the two small fragments of 0.45 and 0.46 kb both designated as 0.5 kb on stained agarose gels and restriction maps (Figures 4.3, 4.4, 4.13, etc.). Three EcoRI fragments (31, 8.4 and 2.9 kb) hybridized with the feather keratin coding probe.

Figure 4.4b compares the restriction patterns of cosmid 4 and λCFK1 digested with HindIII. All HindIII insert fragments of λCFK1 (2.9, 2.15 doublet, 1.8, 1.6 and 1.3 kb) are present in cosmid 4 (Figure 4.4b). Southern blot analysis using a feather keratin coding probe demonstrated that this cosmid contains several HindIII restriction fragments additional to those present in the λCFK1 region of the clone which hybridize to the coding probe (Figure 4.4b). In total, cosmid 4 contains 9 HindIII fragments which were detected by the feather probe, including the 8.1 and 7.8 kb HindIII fragments which migrate very close together on agarose gels and the 2.15 doublet (Figure 4.4b). The 7.8 kb HindIII fragment is cut by EcoRI to generate several EcoRI/HindIII fragments, including one of

FIGURE 4.4 : RESTRICTION AND SOUTHERN ANALYSIS OF COSMID 4 WITH EcoRI AND HindIII

A. Cosmid 4 DNA was digested with EcoRI and the DNA fragments electrophoresed on a 0.8% agarose gel (Section 2.2.7). DNA was detected under UV light after ethidium bromide staining. Following Southern transfer (Section 2.2.9), the filter bound DNA was hybridized with a 25-mer coding region oligonucleotide (Section 2.2.10) which is specific for feather keratin gene sequences. The filter was washed in 4 x SSC at 65°C and exposed to X-ray film overnight at -80°C. The sizes of all the EcoRI fragments are given in kilobases.

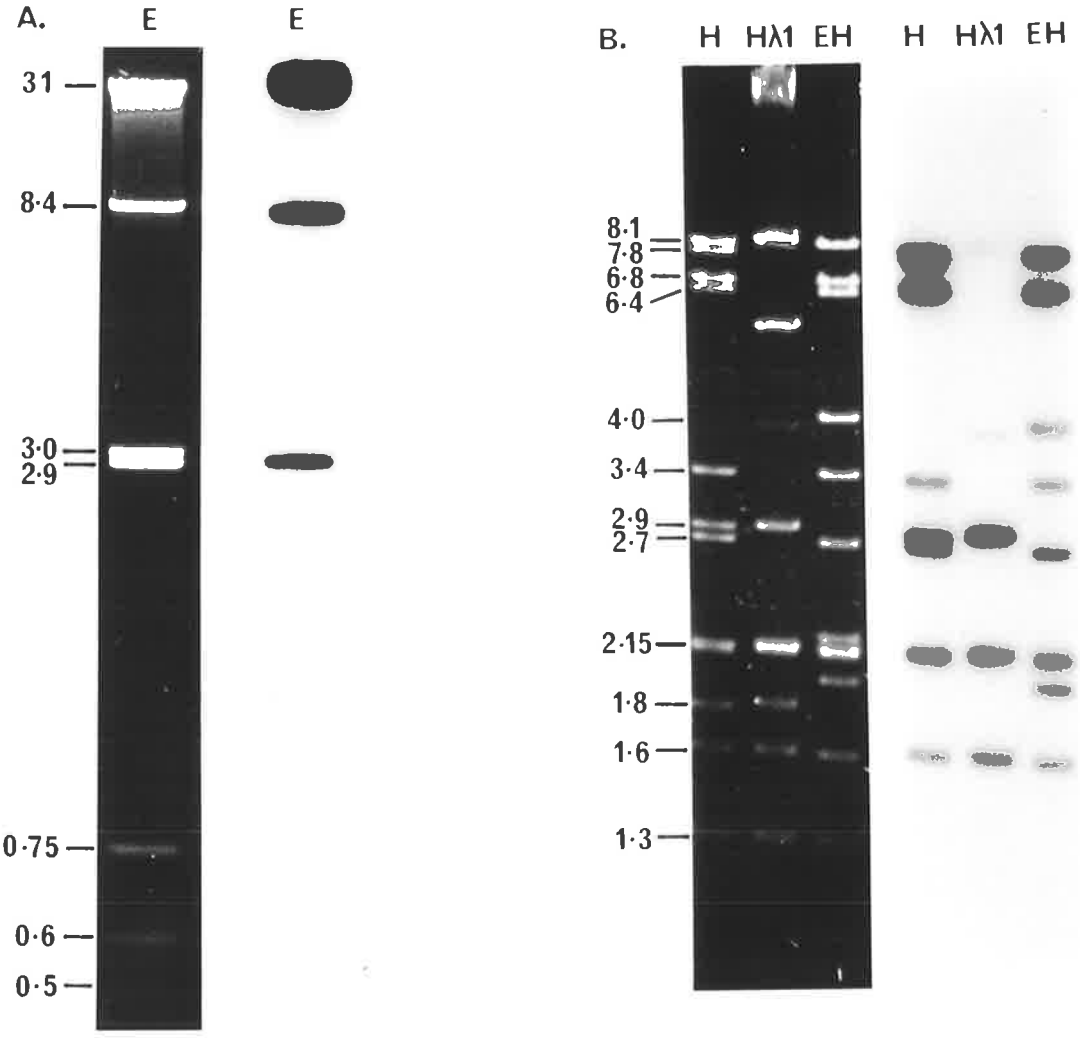
B. Cosmid 4 DNA was digested with HindIII and EcoRI/HindIII, and λCFK1 DNA was digested with HindIII, fractionated on a 0.9% agarose gel, stained with ethidium bromide and detected under UV light (Section 2.2.7). A Southern blot of this agarose gel was hybridized with probe, washed and exposed as described in Figure 4.4a above. The sizes (kb) of all the HindIII restriction fragments of cosmid 4, including those which were common with λCFK1, are shown.

The tracks contained

H - HindIII digest of cosmid 4

Hλ1 - HindIII digest of λCFK1

EH - EcoRI/HindIII digest of cosmid 4


A.
E
E
31
8·4
3·0
2·9
0·75
0·6
0·5
B.
H
HΛ1
EH
H
HΛ1
EH
8·1
7·8
6·8
6·4
4·0
3·4
2·9
2·7
2·15
1·8
1·6
1·3

4.0 kb which hybridized with the coding probe (Figure 4.4b). These results indicate that the feather keratin genes contained in the clone λCFK1 are part of a much larger gene cluster.

There are 6 HindIII fragments in cosmid 4 which have no counterparts in λCFK1 or λCFK9 of sizes 8.1, 7.8, 6.8, 6.4, 2.7 and 0.7 kb (Figure 4.4b, 4.6a). In an EcoRI digest of cosmid 4, there are 5 fragments (31, 8.4, 3.0, 0.75 and 0.6 kb) which have no counterparts in the λ clones (see Figure 4.7 below).

A first step in mapping this clone was to determine the location of the cosmid vector sequences. Since there is one EcoRI and one HindIII site in the cosmid vector pHC79 (Figure 4.5), two EcoRI and two HindIII cosmid DNA fragments should hybridize with a labelled pBR322 probe. This was indeed found to be the case (Figure 4.6); strong hybridization signals were produced by the 31 kb EcoRI fragment and by the 6.8/6.4 kb doublet, which was not resolved well enough on this gel to allow unambiguous identification, while weaker signals were produced by the 3.0 kb EcoRI and 0.7 kb HindIII fragments. To determine which of the two HindIII fragments in the 6.8/6.4 kb HindIII doublet contained most of the cosmid vector, a HindIII digest of cosmid 4 was restricted with SalI. The 6.8 kb HindIII fragment was cut by SalI (Figure 4.11a) and as the pHC79 vector sequence has a SalI site (Figure 4.5), this HindIII fragment must contain most of the vector sequence.

The order of EcoRI fragments in cosmid 4 was derived by comparison with the known maps of the two λ clones and from the Southern blot using pBR322 as probe and is detailed in the legend to Figure 4.7a.

FIGURE 4.5 : RESTRICTION MAP OF THE COSMID VECTOR

Restriction map of the pHC79 cosmid vector in the cosmid recombinants. The hatched region (▨) indicates λ-derived sequences which includes the cos site. The cosmid vector is reduced from 6.43 kb to ~ 6.1 kb during the preparation of the vector prior to ligation with partially-cut insert DNA and *in vitro* packaging (see Figure 3.3).

WWW Represents genomic (chromosomal) DNA sequences.

The restriction sites are indicated as follows:

B - BglII

E - EcoRI

H - HindIII

P - PstI

S - SalI

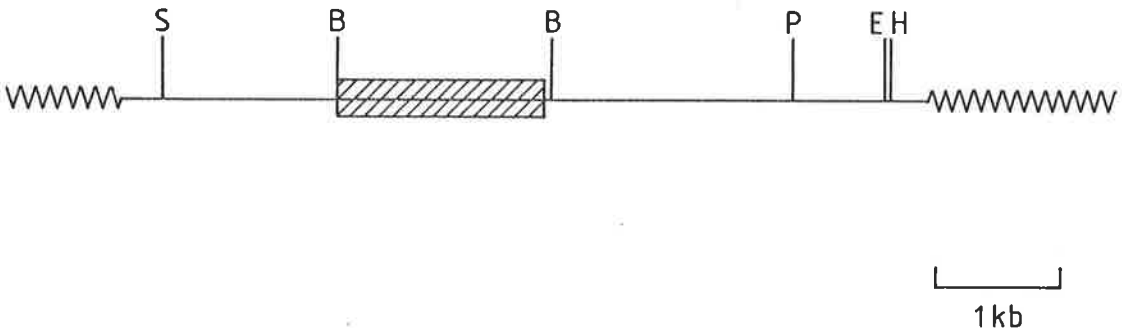

S
B
B
P
E
H
1kb

FIGURE 4.6 : SOUTHERN ANALYSIS OF COSMID 4 USING pBR322 AS PROBE

A. Cosmid 4 DNA was digested with HindIII, electrophoresed on a 0.9% agarose gel and the DNA fragments observed under UV light after ethidium bromide staining (Section 2.2.7). DNA was transferred to nitrocellulose (Section 2.2.9) and hybridized with pBR322 labelled by nick translation (Section 2.2.11.2). The filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and exposed to X-ray film for 1 day at -80°C.

B. Southern analysis of an EcoRI digest of cosmid 4 using pBR322 as probe. Details are as described in Figure 4.6a above.

The sizes of the fragments (kb) which hybridized with the pBR322 probe in each digest are shown.

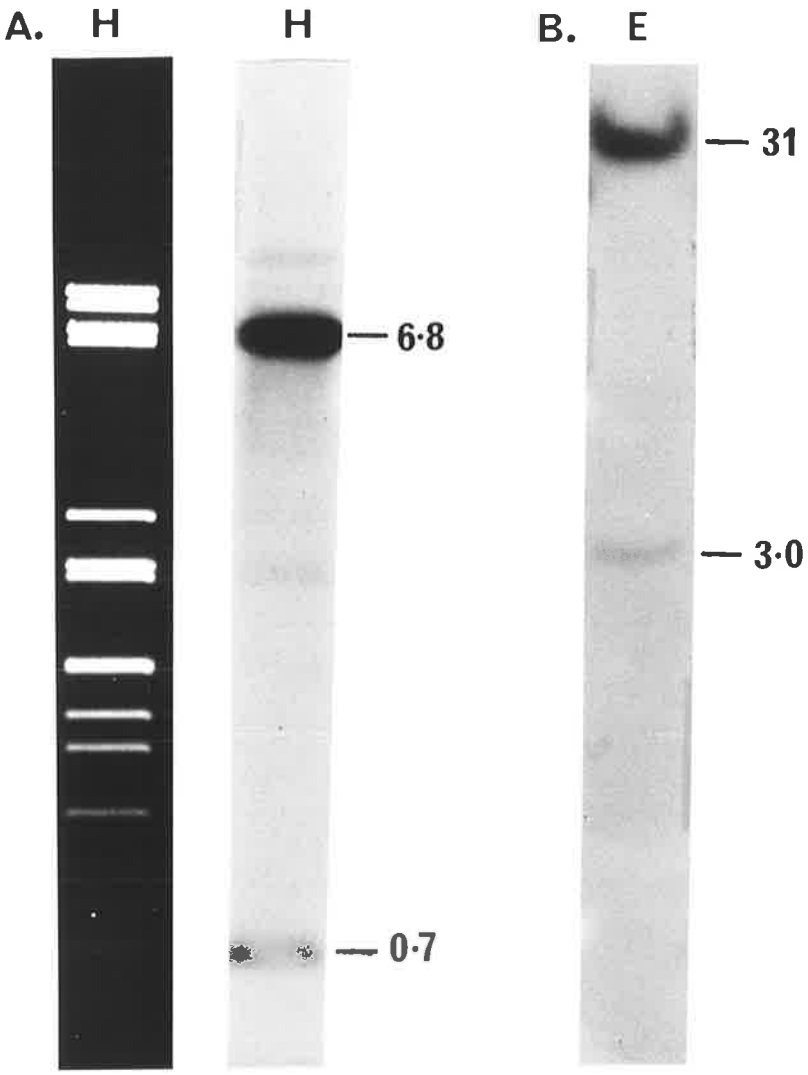
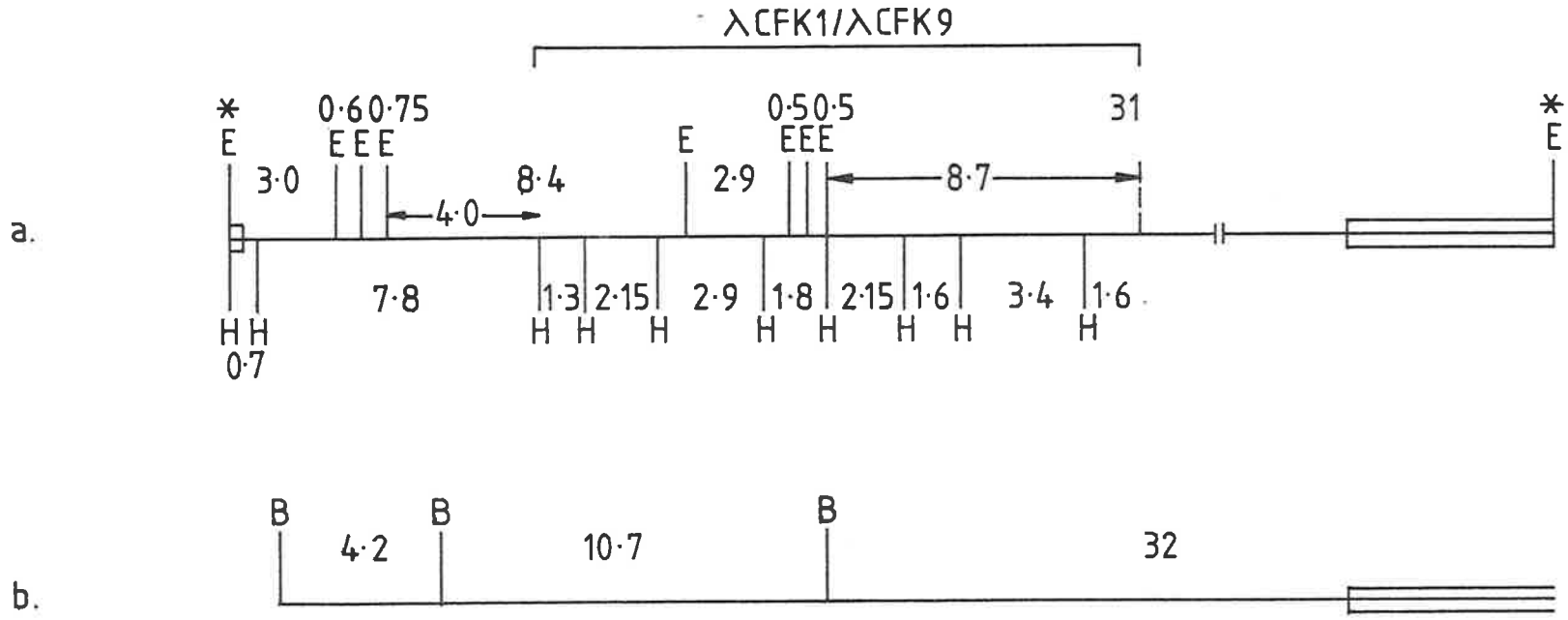

A.
H
H
6·8
0·7
B.
E
31
3·0

FIGURE 4.7 : DERIVATION OF THE EcoRI AND BamHI RESTRICTION MAPS OF COSMID 4

A. EcoRI Map. From the known maps of λCFK1 and λCFK9 (Figure 4.3), the 2.9 and 0.5 kb EcoRI fragments could be positioned in cosmid 4 (Figure 4.7a). From the restriction maps of the two λ clones, there are no EcoRI sites rightwards of the 0.5 kb EcoRI fragment for at least 8.7 kb. Therefore the 8.4 kb EcoRI fragment will lie leftwards of the 2.9 kb EcoRI fragment present in λCFK1 and cosmid 4. This was confirmed by subcloning the 8.4 kb EcoRI fragment into pBR322 (see Section 4.3.4.2d). The 31 kb EcoRI fragment, which hybridized strongly with the feather keratin gene and pBR322 probes (Figures 4.4a and 4.6b), must therefore span the region to the right of the 0.5 kb EcoRI fragment. This 31 kb fragment includes most of the cosmid vector at the right-hand end of the clone. The 3.0 kb EcoRI fragment hybridized weakly with the pBR322 probe (Figure 4.6b), and therefore contains a small region of the cosmid vector located at the left-hand end of the clone. The two remaining EcoRI fragments of 0.75 and 0.6 kb can only lie between the 3.0 and 8.4 kb EcoRI fragments in the left half of the insert. The EcoRI map of the left-hand end of cosmid 4 was confirmed in a later hybridization experiment using the 4.2 kb BamHI fragment as a probe (see Figure 1 of Appendix A).

B. BamHI Map. The order of BamHI fragments was determined as described in the text. The BamHI map is consistent with the following observations. (1) The 7.8 kb HindIII fragment, which is located at the left-hand end of the clone insert, is cut by BamHI to produce fragments of 4.2 kb, 2.5 and ~ 1.0 kb (see Figure 4.10a). (2) From the HindIII map and the mapping of the 10.7 kb BamHI fragment, there is a total of ~ 5.6 kb between the 10.7 kb fragment and the cosmid vector. (3) Hybridization of the 4.2 kb BamHI fragment to an EcoRI digest of cosmid 4 gave strong signals with the 8.4, 3.0, 0.75 and 0.6 kb EcoRI fragments (Figure 1 of Appendix A), confirming its position as shown in Figure 4.7b. (4) There are no BamHI sites in the HindIII fragments that lie to the right of the site located in the λ clones (see text), indicating that the 32 kb BamHI fragment spans this region.


λCFK1/λCFK9
a.
*
E
0·6 0·75
E E E
0·5 0·5
E E E
31
*
E
3·0
4·0
8·4
E
2·9
8·7
7·8
1·3
2·15
2·9
1·8
2·15
1·6
3·4
1·6
H H
0·7
H H H H H H H H
b.
B
B
B
4·2
10·7
32

There are four HindIII fragments in cosmid 4 (8.1, 7.8, 6.4 and 2.7 kb) which are not present in the two λ recombinants and do not hybridize with pBR322 (see Figures 4.4b, 4.6a, 4.7a). Of these, only the 7.8 kb HindIII fragment was restricted by EcoRI (Figure 4.4b). From the EcoRI map (Figure 4.7a), there are no EcoRI sites in the right half of cosmid 4 and it was calculated there was 7.7 kb of DNA between the 1.3 kb HindIII and the 0.7 kb HindIII fragments of cosmid 4 (Figure 4.7a). These data are consistent with the 7.8 kb HindIII fragment (which is cut by EcoRI) being located to the left of the 1.3 kb HindIII fragment as shown in Figure 4.7a.

The order of the remaining three HindIII fragments (of 8.1, 6.4 and 2.7 kb), which are contained in the right half of cosmid 4, was yet to be determined.

B. ORDER OF THE THREE HINDIII FRAGMENTS IN THE RIGHT HALF OF COSMID 4

It has been established from mapping data on the clone λCFK9 (Figure 4.3) that feather keratin gene E probably spans the HindIII site of the 3.4 kb fragment in that λ clone. Therefore, the HindIII fragment in cosmid 4 which is located next to the 3.4 kb HindIII fragment should have the 5' end of a feather keratin gene at one end. Results presented later indicate that the 8.1 kb HindIII fragment of cosmid 4 contained the 5' end of a feather gene at one end (see Section 4.3.4.2a, Figure 4.36) and therefore this HindIII fragment could be located next to the 3.4 kb HindIII fragment. More specifically, from the map of the 8.1 kb HindIII fragment reproduced in Figure 4.8c, the 5' non-coding exon and probably part of the

intron of gene E would appear to be in a 1.05 kb BglII/HindIII fragment. Since λCFK9 extends λCFK1 by 2 kb, the former λ clone should completely contain this 1.05 kb restriction fragment if pH 8.1 is located next to the 3.4 kb HindIII fragment (Figure 4.8c). To test this hypothesis, both λCFK9 and cosmid 4 were restricted with BglII and HindIII, fractionated on an agarose gel and probed with the 1.05 kb BglII/HindIII fragment from pH 8.1 (Figure 4.8a,b). Figure 4.8b shows that the probe strongly bound to a 1.05 kb BglII/HindIII fragment in λCFK9 and cosmid 4 (arrow in Figure 4.8b). Since the BglII/HindIII probe contains the 5' end of gene E, it also hybridizes with other feather keratin genes in cosmid 4 and λCFK9, although less strongly. In a parallel experiment the probe was hybridized to HindIII and EcoRI/HindIII digests of λCFK9 (Figure 4.8b) and, as expected, it bound strongly to the 6.8 kb HindIII and 1.6 kb EcoRI/HindIII fragments of that λ recombinant (see Section 4.3.1, Figure 4.3), demonstrating that the 5' end of gene E spans the HindIII site of the 3.4 and 8.1 kb fragments.

The order of the other two HindIII fragments (6.4 and 2.7 kb) remained to be determined. Mapping data from cosmid 4 with KpnI, shown below, suggested that the 2.7 kb fragment was located next to the 6.8 kb cosmid vector fragment (see Section 4.3.2.1d). If the gene in the 2.7 kb HindIII fragment of cosmid 4 is orientated (5'-3') in the same direction as those in λCFK1 (right to left, Figure 4.3), the 1.0 kb KpnI/HindIII fragment which contains the 5' half of this gene (see Figure 4.9d) should therefore be next to the cosmid vector fragment. From the known map of the cosmid vector (Figure 4.5)

FIGURE 4.8 : SOUTHERN BLOT ANALYSIS OF COSMID 4 AND λCFK9 WITH THE 1.05 KB BglII/HindIII PROBE

A. Cosmid 4 DNA was digested with BglII and HindIII; λCFK9 DNA was digested with BglII/HindIII, EcoRI/HindIII and HindIII. The resultant digestions were fractionated on a 1% agarose gel, stained with ethidium bromide and visualized under UV light (Section 2.2.7).

B. Following Southern transfer (Section 2.2.9), the filter-bound DNA was hybridized with the 1.05 kb BglII/HindIII fragment from pH 8.1 (Figure 4.8c), washed in 0.1 x SSC, 0.1% SDS at 65°C and autoradiographed overnight at -80°C. The arrow indicates the position of the 1.05 kb BglII/HindIII fragment present in both cosmid 4 and λCFK9. The size of the fragments in λCFK9 which are detected by the probe are also shown.

C. Restriction map of pH 8.1, showing the position and orientation of the feather keratin genes and the region of overlap with the recombinant λCFK9 (see text and Section 4.3.4.2a for further details).

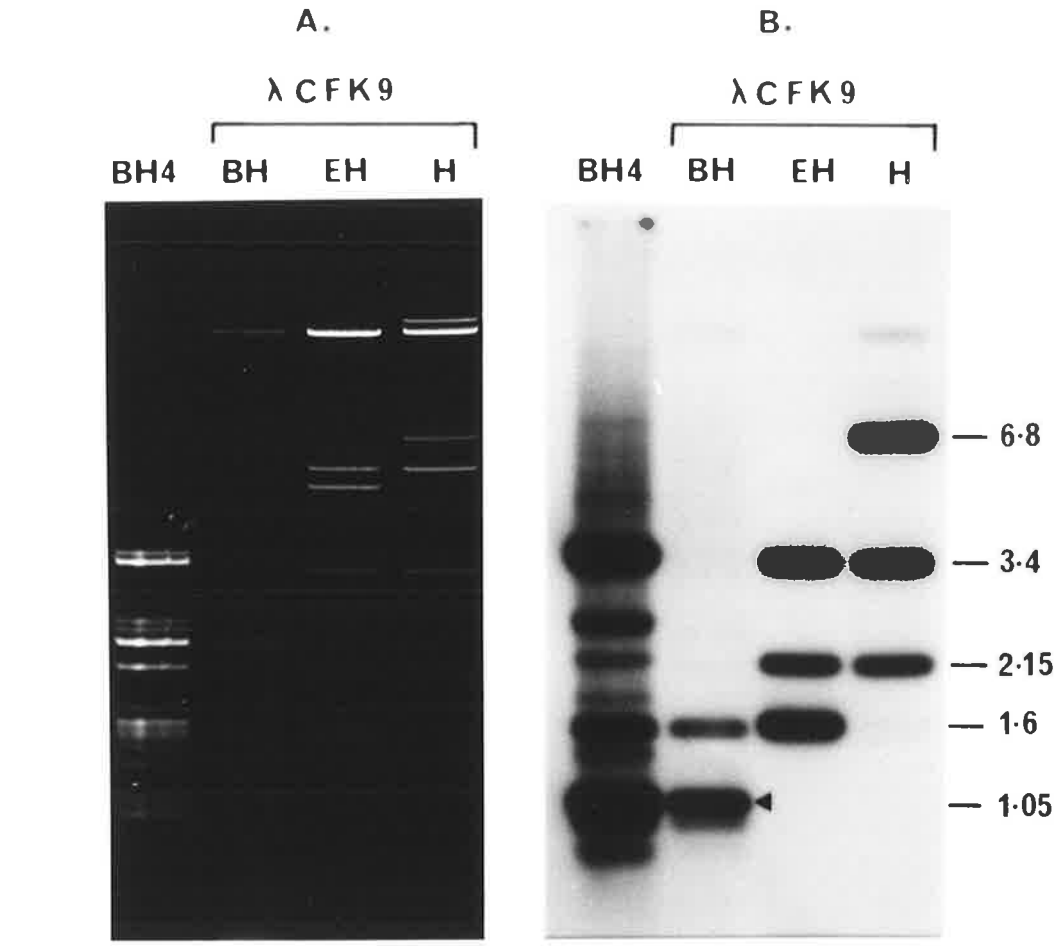
Restriction sites are indicated as follows:

B - BglII

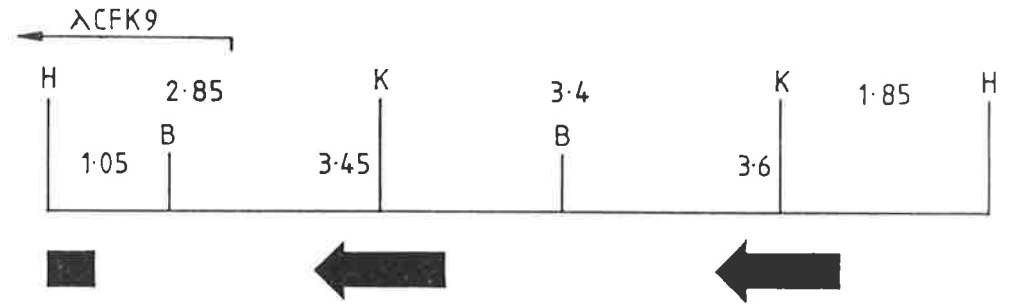
E - EcoRI

H - HindIII

K - KpnI


A.
B.
λCFK9
λCFK9
BH4
BH
EH
H
BH4
BH
EH
H
6·8
3·4
2·15
1·6
1·05

C.


λCFK9
H
2·85
K
3·4
K
1·85
H
1·05
B
3·45
B
3·6

and knowing the orientation of the cosmid vector in this recombinant (see EcoRI map of cosmid 4, Figure 4.7a), a 'hypothetical' map of this region could be drawn (Figure 4.9d). From this 'hypothetical' map, the sizes of certain restriction fragments could be predicted and, if correct, these restriction fragments should hybridize with both the 1.0 kb KpnI/HindIII fragment from pH 2.7 and pBR322 (Figure 4.9d).

To test this 'hypothetical map', cosmid 4 was restricted with the combinations of enzymes shown in Figure 4.9a and the gel blotted and hybridized with the two probes. The fragments which anneal with both probes are shown by arrows and their sizes are listed underneath the respective tracks (Figure 4.9b,c). There is good correlation between the sizes of the fragments which were predicted to hybridize with both probes (Figure 4.9d,e) and the observed sizes (Figure 4.9b,c). It was not expected that the two patterns of hybridization be identical, since the 1.0 kb KpnI/HindIII probe would be expected to hybridize to other bands in cosmid 4 which contain feather genes and the pBR322 probe would also bind other fragment(s) in cosmid 4 which contain vector sequences complementary to pBR322. From these results, it is apparent that the 2.7 kb fragment is located next to the cosmid vector-containing fragment as shown in Figure 4.9d. The 6.4 kb HindIII fragment of cosmid 4 must therefore lie between the 8.1 and 2.7 kb HindIII fragments.

The HindIII restriction map of cosmid 4 determined from the studies described in this section is shown in Figure 4.13. Characterization of another recombinant, cosmid 31, which contains the 2.7 kb and part of the 6.4 kb HindIII fragments of

FIGURE 4.9 : ORDERING OF HindIII FRAGMENTS AT THE RIGHTHAND END OF COSMID 4

A. Cosmid 4 DNA was digested with a number of restriction enzymes (see below), electrophoresed on 1% agarose and the restriction fragments visualized under UV light after ethidium bromide staining (Section 2.2.7). The DNA fragments were transferred to nitrocellulose using the bidirectional transfer procedure (Section 2.2.9) and the filters probed with:

B. The 1.0 kb KpnI/HindIII fragment of pH 2.7 (a subclone which contains the 2.7 kb HindIII fragment of cosmid 4, Section 4.3.4.2.b). The 1.0 kb fragment contains the 5' half of the gene in this fragment.

C. Nick translated pBR322 (Section 2.2.11.2).

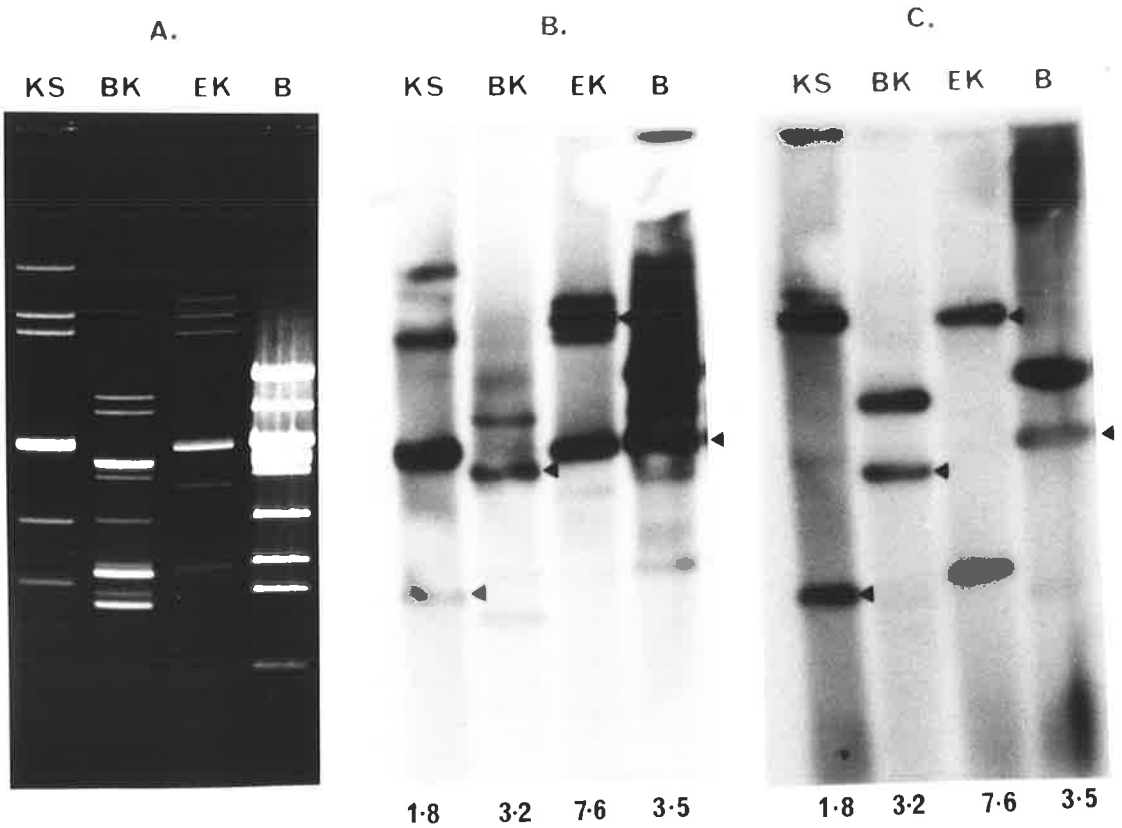
The filters were washed in 0.1 x SSC, 0.1% SDS at 65°C and autoradiographed for 1-2 days at -80°C. The arrows indicate the single fragment in each track which hybridized with both of the probes in B and C. Their sizes (kb) are listed underneath the respective tracks.

D. Restriction map of the right-hand end of cosmid 4, showing the position and orientation of the rightmost gene in cosmid 4 (arrow) and the cosmid vector (box). The regions spanned by the two probes used in the Southern hybridizations in B and C are depicted. The pBR322 probe does not hybridize with the BglII fragment in the middle of the cosmid vector, as this segment is derived from λ (see Figure 4.5). The restriction maps of the 2.7 kb HindIII fragment and the cosmid vector are from Sections 4.3.4.2b and Figure 4.5, respectively.

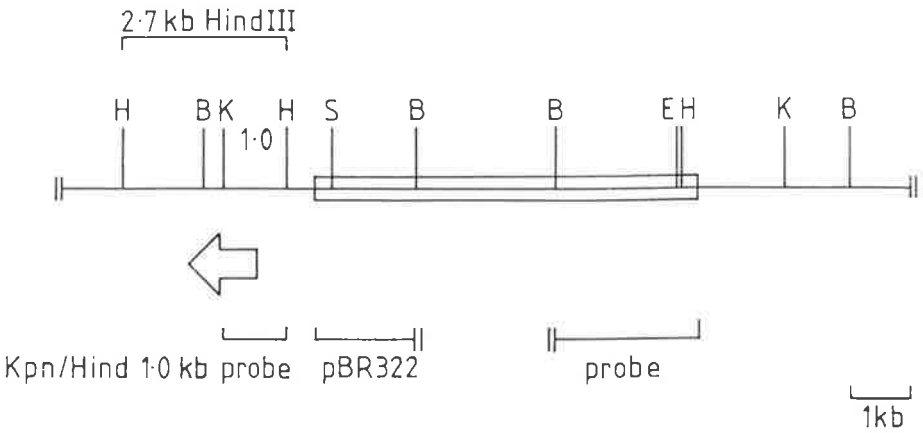
E. The sizes of the restriction fragments which hybridize with both probes (Figure 4.9b,c) closely match the predicted sizes based on the hypothetical map (Figure 4.9d,e) and the map of pHC79 (Figure 4.5). These results demonstrated that the map shown in Figure 4.9d is correct (see text for details).

Restriction sites are indicated as follows:

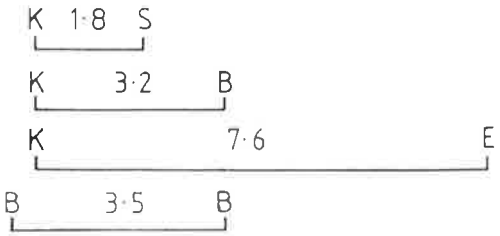
B - BglII
E - EcoRI
H - HindIII
K - KpnI
S - SalI


A.
KS BK EK B
B.
KS BK EK B
C.
KS BK EK B
1·8 3·2 7·6 3·5
1·8 3·2 7·6 3·5

D.


2·7 kb HindIII
H BK H S B B EH K B
1·0
Kpn/Hind 1·0 kb probe
pBR322
probe
1kb

E.


K 1·8 S
K 3·2 B
K 7·6 E
B 3·5 B

cosmid 4 (see Section 4.3.3.1) confirmed the positioning of the 6.4 and 2.7 kb HindIII fragment in cosmid 4.

C. BamHl map. In Figure 4.10a, the digestion pattens obtained with BamHI, HindIII and BamHI/HindIII are shown. Three BamHI restriction fragments were produced, of which two (32 and 10.7 kb) were detected by a feather keratin gene fragment used as a probe. There is one BamHI site in λCFKl, which is located in the 3' non-coding region of gene C downstream of the HindIII site which this gene spans (Figure 4.7b; Molloy et al., 1982). Therefore, the 32 and 10.7 kb BamHI fragments must encompass and extend the genomic region contained in λCFKl.

Since only the 7.8 kb HindIII fragment (which is located to the left of the 1.3 kb HindIII fragment of cosmid 4, Figure 4.7a) is cut by BamHI, the 10.7 kb BamHI fragment must encompass the region to the left of the BamHI site of λCFKl (Figure 4.7b). Since there are no BamHI sites in the 8.1, 6.8, 6.4 and 2.7 kb HindIII fragments which are located to the right of the genomic region contained in λCFKl (Figure 4.10a), the large 32 kb BamHI fragment therefore spans this region comprising part of the λCFKl genomic region and the 8.1, 6.4, 2.7, 6.8 and 0.7 kb fragments of cosmid 4 (Figure 4.7b, 4.13). This is consistent with the size of this fragment (32 kb).

The only BamHI fragment so far not mapped, the 4.2 kb fragment, must lie at the left-hand end of the clone insert next to the 10.7 kb BamHI fragment (for further details, see legend to Figure 4.7b).

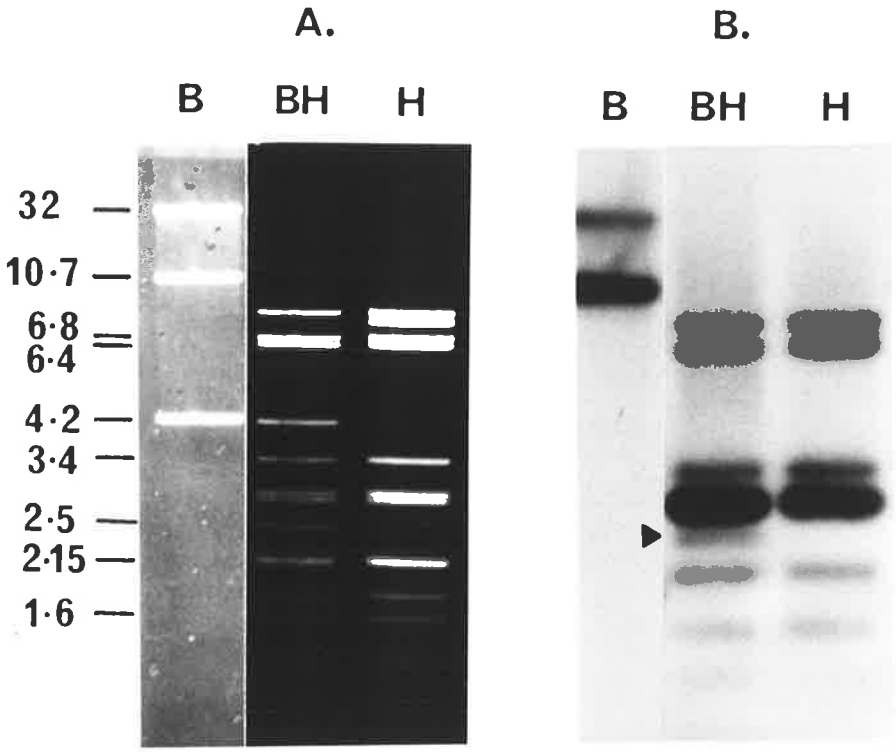
FIGURE 4.10 : RESTRICTION AND SOUTHERN ANALYSIS OF THE BamHI DIGEST OF COSMID 4

A. Cosmid 4 DNA was digested with BamHI, BamHI/HindIII and HindIII. The products were separated on a 0.9% agarose gel, stained with ethidium bromide and viewed under UV light (Section 2.2.7).

B. The DNA fragments were transferred to nitrocellulose (Section 2.2.9) and the immobilized DNA probed with a 0.33 kb PstI/SalI fragment which contains the intron of feather keratin gene B (Section 6.3.3). Following hybridization, the filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and exposed to X-ray film at -80°C for 1 day. The arrow indicates the hybridizing band at 2.5 kb which is one of several fragments produced from the 7.8 kb HindIII fragment (see legend to Figure 4.7b). The sizes (kb) of the BamHI and some of the BamHI/HindIII fragments are shown.

B - BamHI

H - HindIII


A.
B
BH
H
32
10·7
6·8
6·4
4·2
3·4
2·5
2·15
1·6
B.
B
BH
H

D. The KpnI map of cosmid 4: Presence of a repeated KpnI fragment. A KpnI digest of cosmid 4 produces three restriction fragments of sizes 21.5, 9.1 and 3.4 kb (Figure 4.11a). Since there is only a single KpnI site located at one end of the 15 kb insert of the λCFK1 clone (Figure 4.3), the 21.5 kb fragment must span the genomic region contained in the λ recombinant and the next KpnI site must lie close to the end of cosmid 4 (Figure 4.12a). It remained to position the 9.1 and 3.4 kb fragments, which must account for the remaining 25.5 kb of cosmid 4 (see legend to Figure 4.12). As there are no KpnI sites in the cosmid vector (Hohn and Collins, 1980), the vector, which is 6.1 kb, must be entirely contained in the 9.1 kb KpnI fragment (Figure 4.12).

Inspection of the KpnI gel pattern (Figure 4.11a) indicated that several co-migrating fragments were present at 3.4 kb. The calculations described in the legend to Figure 4.12 indicate that there are 5 tandem repeats of the 3.4 kb fragment between the 21.5 and 9.1 kb KpnI fragments, totalling 17 kb in length (Figure 4.12b), which begin at the right-hand end of the 3.4 kb HindIII fragment (Figure 4.3), encompassing the 8.1 and 6.4 kb fragments and terminating in the 2.7 kb HindIII fragment.

In order to confirm this, the 3.4 kb KpnI band was isolated, labelled and hybridized to a blot of the gel shown in Figure 4.11a. It can be seen from the Southern blot shown in Figure 4.11b that the 3.4 kb KpnI fragments indeed hybridized strongly to the 8.1, 6.4 and 3.4 kb HindIII fragments of cosmid 4. This result is consistent with the 3.4 kb KpnI repeat spanning the righthand segment of the cosmid 4 insert

FIGURE 4.11 : SOUTHERN ANALYSIS OF COSMID 4 WITH THE 3.4 KB KpnI FRAGMENTS

A. Cosmid 4 DNA was digested with SalI/HindIII, BamHI/HindIII and KpnI, electrophoresed on a 0.9% agarose gel, stained with ethidium bromide and visualized under UV light (Section 2.2.7).

B. The restriction fragments were blotted on to a nitrocellulose filter and the filter bound DNA hybridized with the 3.4 kb KpnI band of cosmid 4, which had been isolated from a KpnI digest of cosmid 4 (Section 2.2.8). The 3.4 kb band consists of 5 similar sized DNA fragments (see Figure 4.12). The filter was washed in 0.1 x SSC, 0.1% SDS at 65°C and autoradiographed for 8 hours at -80°C.

The arrows indicate the three fragments which were strongly bound by the probe - the 8.1, 6.4 and 3.4 kb HindIII fragments of cosmid 4 (see text).

The sizes of the KpnI and some of the other fragments are shown.

Restriction sites are indicated as follows:

B - BamHI

H - HindIII

K - KpnI

S - SalI

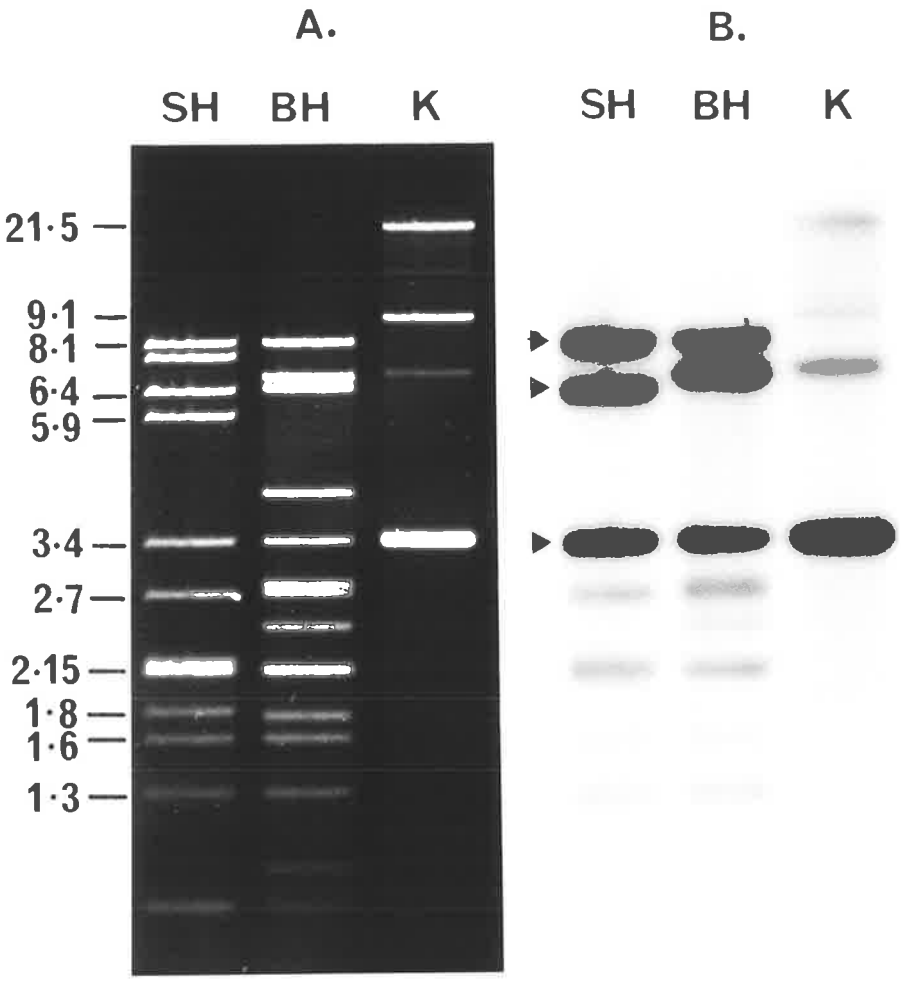

A.
B.
SH
BH
K
SH
BH
K
21·5
9·1
8·1
6·4
5·9
3·4
2·7
2·15
1·8
1·6
1·3

FIGURE 4.12 : DERIVATION OF THE KpnI RESTRICTION MAP OF COSMID 4

A KpnI digest of cosmid 4 produces three restriction fragments of 21.5, 9.1 and 3.4 kb (Figure 4.11a). Since there is only a single KpnI site located at one end of the 15 kb insert of the λCFK1 clone (Figure 4.3), the 21.5 kb fragment must span the genomic region contained in the λ recombinant and the next KpnI site must lie close to the end of cosmid 4. Summation of the sizes of the HindIII fragments gives a total size of cosmid 4 of 47 kb, of which 21.5 kb has been accounted for in the large KpnI fragment which encompasses almost the whole genomic region contained in λCFK1. It remained to position the 9.1 and 3.4 kb fragments, which must account for the remaining 25.5 kb of cosmid 4. As there are no KpnI sites in the cosmid vector, the vector (6.1 kb) must be entirely contained in the 9.1 kb KpnI fragment (Figure 4.12b).

Inspection of the KpnI gel pattern (Figure 4.11a) indicated that several co-migrating fragments were present at 3.4 kb. Given the location of the KpnI site in λCFK1 and from the HindIII map, it could be calculated that there is 18-19 kb between the KpnI site of λCFK1 and the cosmid vector sequence. It has been established previously from other mapping data on cosmid 4 that there is a KpnI site close to the left end of the insert (see Figure 4.9d). This implies that the 9.1 kb KpnI fragment extends at least 2 kb to the left from the vector, leaving 16-17 kb between the 21.5 and 9.1 kb KpnI fragments (Figure 4.12a).

Therefore there are 16-17/3.4 or 5 tandem repeats of the 3.4 kb KpnI fragment between the 21.5 and 9.1 kb KpnI fragments, which begin at the right-hand end of the 3.4 kb HindIII fragment (Figure 4.3), encompassing the 8.1 and 6.4 kb fragments and terminating in the 2.7 kb HindIII fragment (Figure 4.12b). This is consistent with the strong hybridization of the 3.4 kb KpnI band to the 8.1, 6.4 and 3.4 kb HindIII fragments of cosmid 4 (Figure 4.11b) and studies on subclones containing the 8.1 and 2.7 kb HindIII fragments (Section 4.3.4.2a,b).

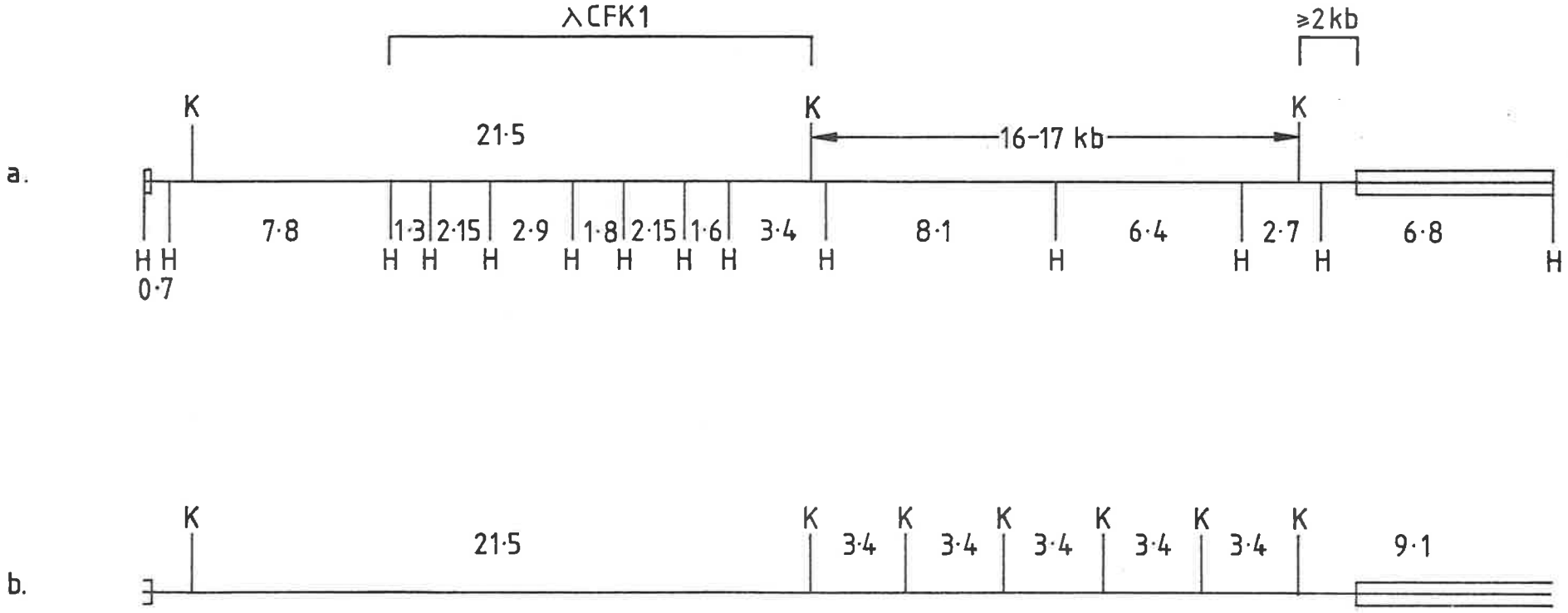
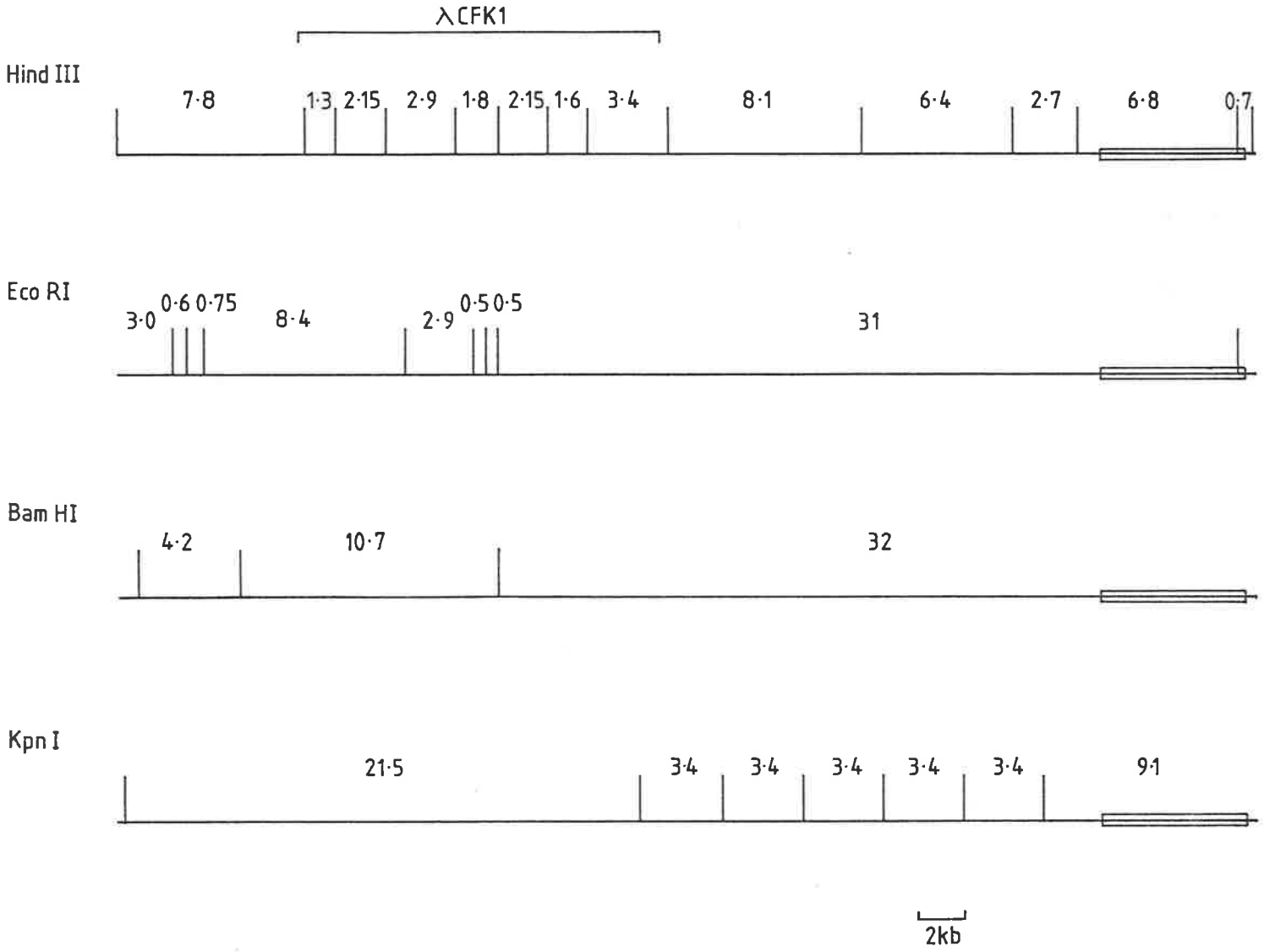

λCFK1
≥2kb
a.
K
K
K
21·5
16–17 kb
7·8
1·3
2·15
2·9
1·8
2·15
1·6
3·4
8·1
6·4
2·7
6·8
H H
0·7
H
b.
K
21·5
3·4
9·1

FIGURE 4.13 : RESTRICTION MAPS OF COSMID 4

The restriction maps of cosmid 4 for the enzymes HindIII, EcoRI, BamHI and KpnI, derived from the data presented in Section 4.3.4.1, are illustrated. The sizes of the restriction fragments are shown in kilobases. The portion of cosmid 4 which is contained in the recombinant λCFK1 (Figure 4.3) is indicated. The boxed region at the right-hand end of each map contains the cosmid vector.


λCFK1
Hind III
7·8
1·3
2·15
2·9
1·8
2·15
1·6
3·4
8·1
6·4
2·7
6·8
0·7
Eco RI
3·0
0·6
0·75
8·4
2·9
0·5
0·5
31
Bam HI
4·2
10·7
32
Kpn I
21·5
3·4
3·4
3·4
3·4
3·4
9·1
2kb

(Figure 4.12b). This map is in agreement with the location of KpnI sites in the 8.1 and 2.7 kb HindIII fragments mapped after subcloning into pBR322 (see Section 4.3.4.2a,b). Although the 6.4 kb HindIII fragment was not mapped extensively, the very strong hybridization signal of this fragment with the KpnI 3.4 kb probe is consistent with it spanning one or more of these KpnI fragments.

The restriction maps of cosmid 4 for the enzymes HindIII, EcoRI, BamHI and KpnI are summarized in Figure 4.13. Another recombinant, cosmid 3, was found to be identical to cosmid 4 (data not shown).

4.3.2.2 Restriction Mapping of Cosmid 12

A. HindIII map. Cosmid 12 also hybridized to the λCFK1-specific probe (Section 3.3.3, Figure 3.9c), indicating that cosmids 4 and 12 probably overlapped. This was confirmed by comparison of the HindIII digestion and Southern blot patterns of the two cosmids (Figure 4.14) which showed a number of similar-sized bands were present in both clones, that is the 7.8, 3.4, 2.9, 2.15 (doublet), 1.8, 1.6 and 1.3 HindIII fragments. The 6.9 kb HindIII fragment of cosmid 12 was detected strongly using a pBR322 probe, indicating that the cosmid vector was mostly contained in this fragment (data not shown) and, on a longer exposure, the probe also hybridized to a small (0.45 kb) HindIII fragment visible in a digest of cosmid 12 (see Figure A.2.b), locating the remaining 350 bp of the cosmid vector in this fragment (data not shown; see Figure 4.16).

Given the HindIII restriction map of cosmid 4 (Figure 4.13), all of the HindIII restriction fragments of cosmid 4 which had counterparts in cosmid 12 could be posit-

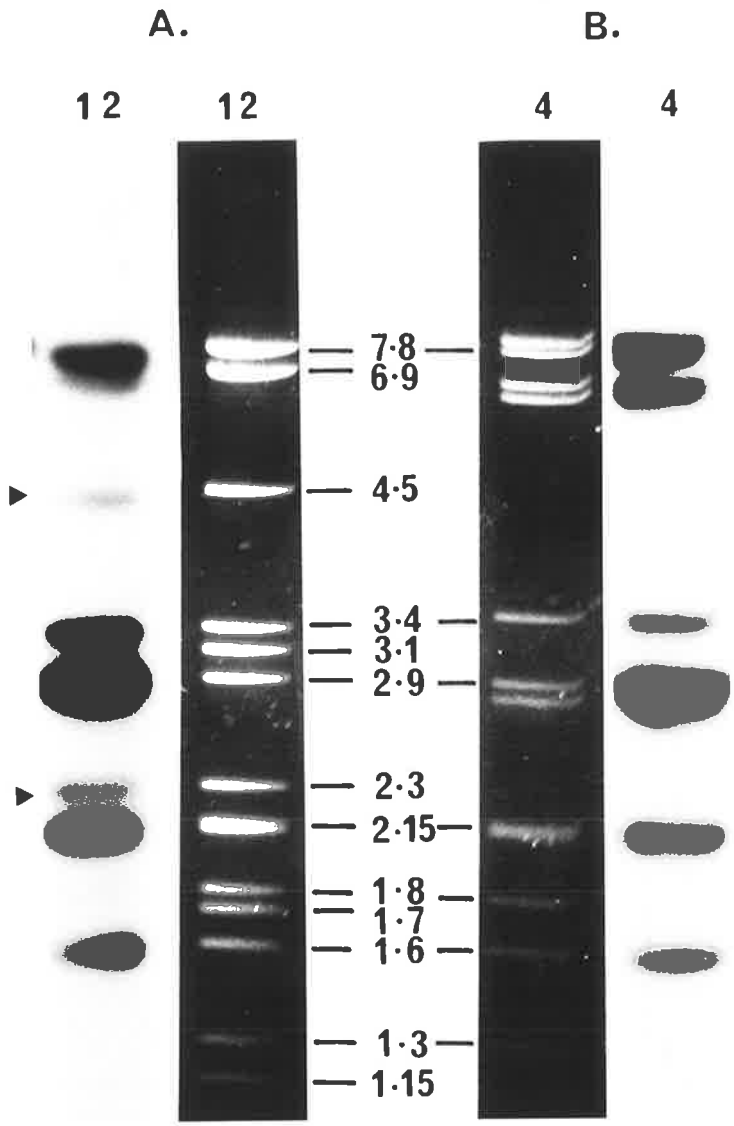
FIGURE 4.14 : COMPARISON OF THE HindIII PATTERNS OF COSMIDS 4 AND 12

Cosmid 4 and 12 DNAs were digested with HindIII, fractionated on 0.9% agarose, stained with ethidium bromide and visualized under UV light (Section 2.2.7). The DNA fragments were blotted on to nitrocellulose (Section 2.2.9), hybridized with the feather keratin gene probe (Section 2.2.10), washed in 1 x SSC, 0.1% SDS at 65°C and autoradiographed overnight at -80°C.

A. HindIII digest of cosmid 12 probed with the feather keratin gene probe. The two HindIII fragments which were detected weakly by the feather gene probe are indicated by arrows.

B. HindIII digest of cosmid 4 probed with the feather keratin gene probe.

The sizes, in kilobases, of all the restriction fragments in cosmid 12, are shown. The HindIII fragments of cosmid 12 which have counterparts in cosmid 4 are also indicated.


A.
B.
1 2
1 2
4
4
7·8
6·9
4·5
3·4
3·1
2·9
2·3
2·15
1·8
1·7
1·6
1·3
1·15

ioned (Figures 4.17 and A.3.a). The remaining HindIII fragments, of which there were six (4.5, 3.1, 2.3, 1.7, 1.15 and 0.65, Figures 4.14), had no counterparts in cosmid 4 and therefore must be located in the region of cosmid 12 which does not overlap with cosmid 4. Two of these 6 fragments, the 4.5 and 2.3 kb HindIII fragments, hybridized weakly with the feather keratin gene probe (arrows in Figure 4.14a) indicating that there were keratin sequences with some homology to feather keratin genes downstream of the genomic region contained in cosmid 4.

B. EcoRI map. In order to locate these genes, restriction maps for BamHI and EcoRI were derived, enabling the order of HindIII fragments also to be determined. Digestion of cosmid 12 with EcoRI generated a total of 12 fragments, including two doublets at 8.4 and 0.5 kb (Figure 4.15a). Of these, one of the 8.4 kb fragments, the 2.9, 0.75, 0.6 and 0.5 kb fragments of cosmid 12 appeared to have counterparts of a similar size in cosmid 4 and the 8.4 and 2.9 kb EcoRI fragments of both clones hybridized with the feather keratin gene probe. All of these common EcoRI fragments could be positioned from the known map of cosmid 4 (Figure 4.15a). The 8.4 kb EcoRI fragment of cosmid 12 is a doublet (Figure 4.15a); one of these fragments is identical to the 8.4 kb EcoRI fragment of cosmid 4 while the other is weakly detected by the pBR322 probe (Figure 4.15d). The latter fragment therefore lies at the left-hand end of the insert and contains part of the cosmid vector (Figure 4.16 and Appendix A).

Three other EcoRI fragments unique to cosmid 12 hybridized with the feather keratin gene probe (strongly, 14 kb; moder-

FIGURE 4.15 : COMPARISON OF THE EcoRI PATTERNS OF COSMIDS 4 AND 12

A. Cosmids 4 and 12 DNAs were digested with EcoRI and the resultant fragments fractionated on a 0.8% agarose gel. DNA was viewed under UV light after ethidium bromide staining (Section 2.2.7). Fragment sizes are indicated in kilobases.

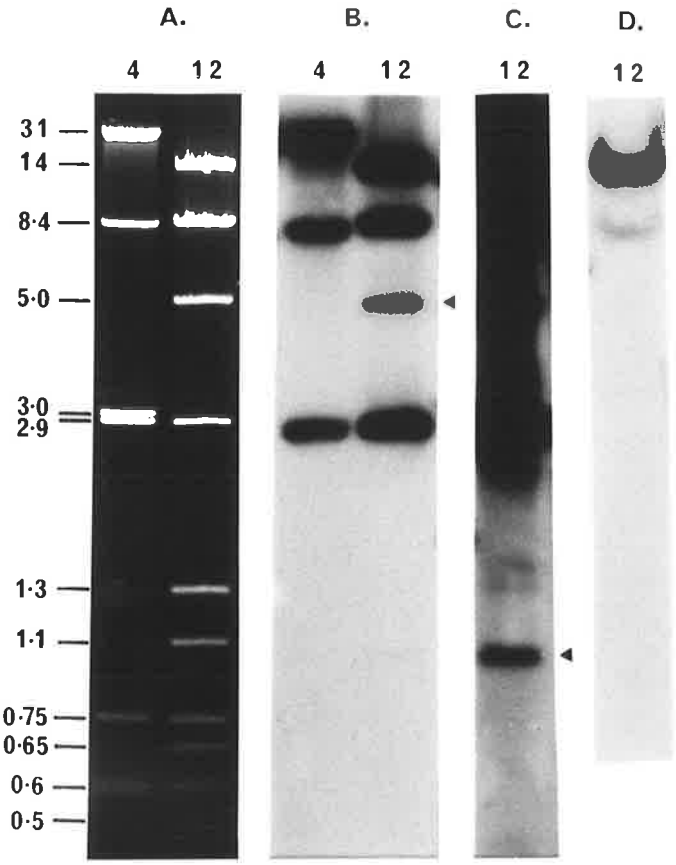
B. The restriction digests were transferred to a nitrocellulose filter (Section 2.2.9) and hybridized with the feather keratin gene probe (Section 2.2.10). Following hybridization, the blot was washed at 0.5 x SSC, 0.1% SDS at 65°C and autoradiographed at -80°C for 16 hours. The arrow indicates the 5.0 kb EcoRI fragment of cosmid 12 which was detected weakly by the probe.

C. 60 Hour exposure of the cosmid 12 track shown in B. The 1.1 kb EcoRI fragment (arrowed) is detected by the probe in this longer exposure, but not in the shorter exposure shown in B.

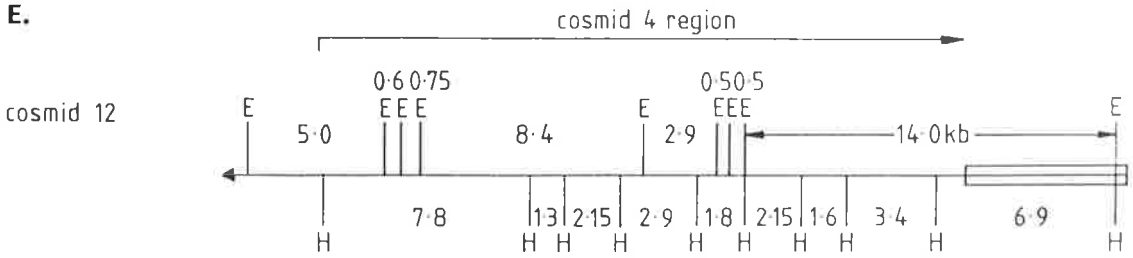
D. Southern blot of cosmid 12 hybridized with nick translated (Section 2.2.11.2) pBR322. The filter was washed in 0.1 x SSC, 0.1% SDS at 65°C and exposed overnight at -80°C.

The 14 kb and 8.4 kb EcoRI fragments hybridized strongly and weakly, respectively, with pBR322.

E. Partial map of cosmid 12, showing the region of overlap with cosmid 4. The 8.4, 2.9, 0.75 , 0.6 and 0.5 kb EcoRI fragments are common between both clones (Figure 4.15a) (see Section 4.3.2.2 and Appendix A for further details).


A.
B.
C.
D.
4
12
4
12
12
12
31
14
8·4
5·0
3·0
2·9
1·3
1·1
0·75
0·65
0·6
0·5

E.


cosmid 4 region
cosmid 12
E
5·0
0·6 0·75
E E E
8·4
E
2·9
0·5 0·5
E E E
14·0 kb
E
H
7·8
1·3
2·15
H H
H
2·9
H
1·8
H
2·15
H
1·6
H
3·4
H
6·9
H

ately, 5.0 kb; weakly, 1.1 kb, Figure 4.15b,c). The 14 kb EcoRI fragment of cosmid 12 hybridizes strongly to both pBR322 and the feather keratin gene probes (Figure 4.15b,d). As cosmid 12 contains the right 2.15, 1.6 and 3.4 kb HindIII fragments of cosmid 4 which contain feather keratin genes (Figure 4.14), the 14 kb fragment must span the region from the EcoRI site in the 1.8 kb HindIII fragment to the EcoRI site of the cosmid vector (Figure 4.15e). The keratin genes located in the 5.0 and 1.1 kb EcoRI fragments have been fully characterized and are discussed in Section 4.3.2.2c below.

It has previously been established that there is a 3.0 kb EcoRI fragment at the lefthand end of cosmid 4 which contains approximately 350 bp of cosmid vector (as it hybridizes weakly with pBR322, Figure 4.13). Therefore, the EcoRI fragment in cosmid 12 which contains the 'genomic' equivalent of the truncated 3.0 kb EcoRI fragment of cosmid 4 must be at least (3-0.35) kb. The only possible EcoRI fragment in cosmid 12 which is compatible with this is the 5.0 kb EcoRI fragment which must lie next to the small 0.75 and 0.6 kb EcoRI fragments in cosmid 12 (Figure 4.15e). This was confirmed using the 4.2 kb BamHI fragment of cosmid 4 as a hybridization probe (see Appendix A).

It remained to determine the location and order of the 1.3, 1.1 and 0.6 kb EcoRI fragments of cosmid 12. In order to resolve this, a BamHI restriction map of cosmid 12 was constructed. The construction of the BamHI map and the subsequent placement of the remaining EcoRI and HindIII fragments of cosmid 12 is described in Appendix A.

FIGURE 4.16 : RESTRICTION MAPS OF COSMID 12

The EcoRI, BamHI and HindIII restriction maps of cosmid 12 were derived as described in Section 4.3.2.2 and Appendix A. The sizes (kb) of the fragments, obtained after restriction enzyme cleavage, are indicated. The boxed region at the right-hand end of each map indicates the location of the cosmid vector.

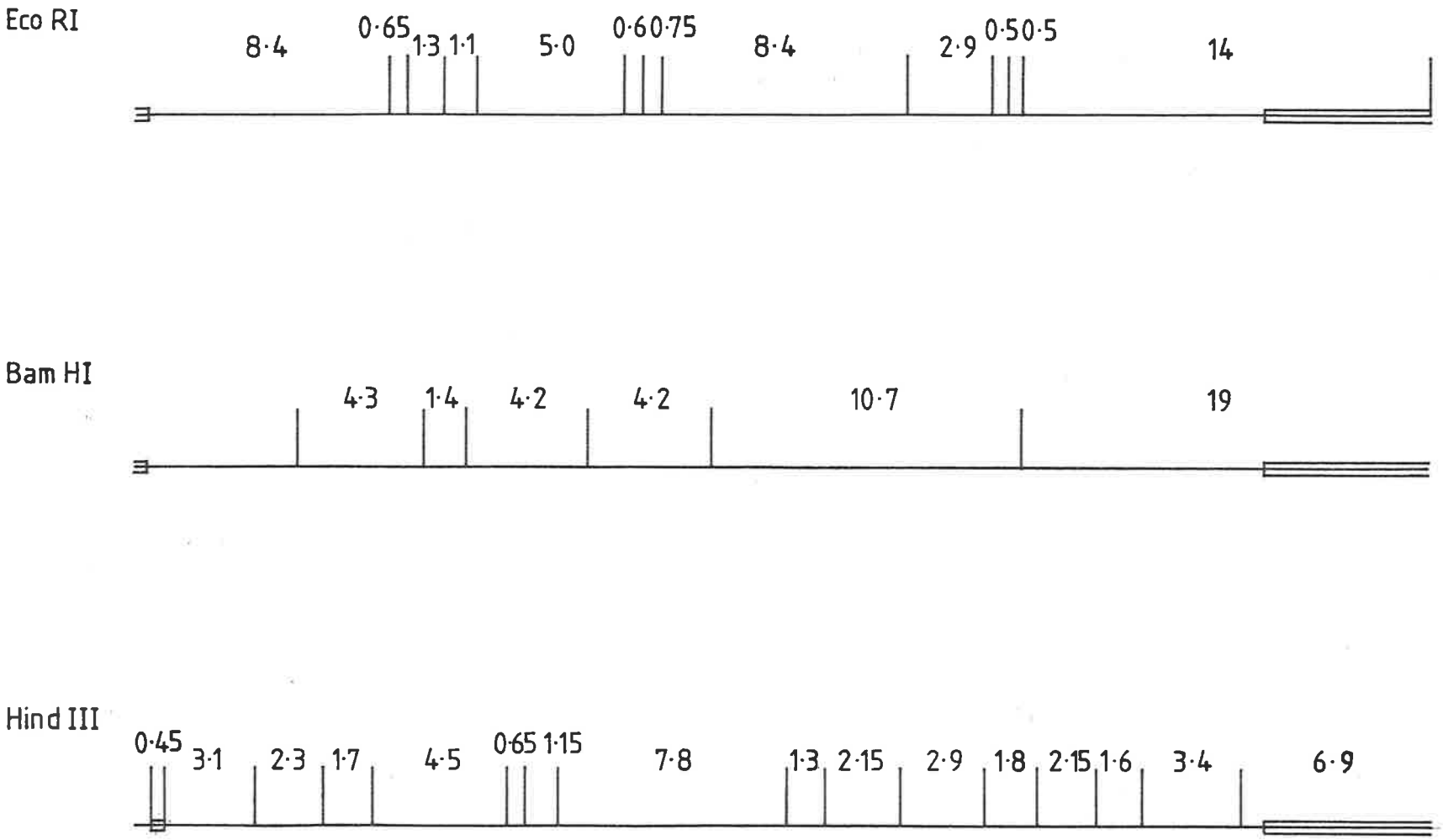
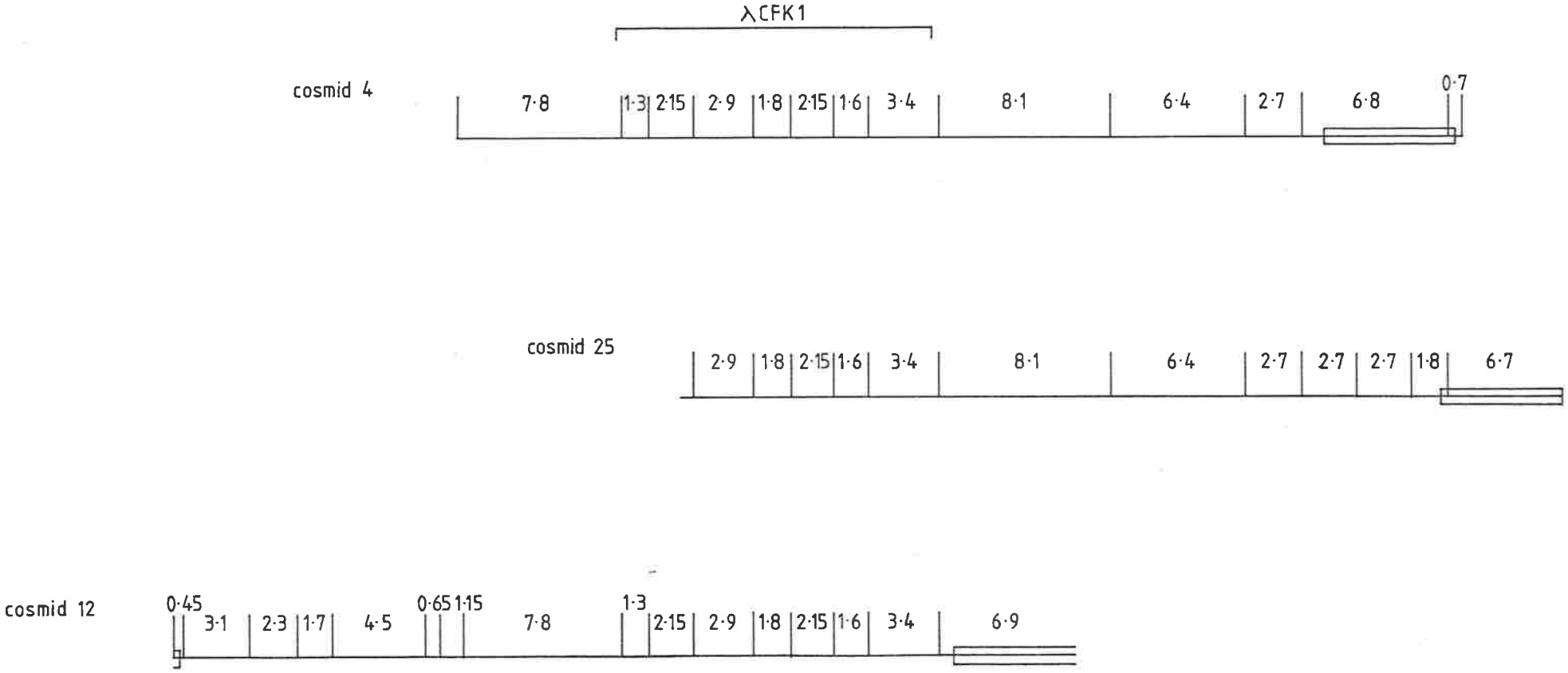

Eco RI
8·4
0·65
1·3
1·1
5·0
0·6
0·75
8·4
2·9
0·5
0·5
14
Bam HI
4·3
1·4
4·2
4·2
10·7
19
Hind III
0·45
3·1
2·3
1·7
4·5
0·65
1·15
7·8
1·3
2·15
2·9
1·8
2·15
1·6
3·4
6·9

FIGURE 4.17 : HindIII RESTRICTION MAPS OF THE OVERLAPPING COSMID RECOMBINANTS 4, 12 AND 25

The figure compares the HindIII restriction maps of the three overlapping recombinants, cosmids 4, 12 and 25, which hybridized strongly with the λCFK1-specific probe (Section 3.3.3., Figure 3.9c). The HindIII maps were taken from Figures 4.13, 4.16 and 4.22. The sizes, in kilobases, of the HindIII fragments are shown. The region contained in the previously isolated genomic clone λCFK1 (Molloy et al., 1982) is shown. The boxed region in each map illustrates the position of the cosmid vector.


λCFK1
cosmid 4
7·8
1·3
2·15
2·9
1·8
2·15
1·6
3·4
8·1
6·4
2·7
6·8
0·7
cosmid 25
2·9
1·8
2·15
1·6
3·4
8·1
6·4
2·7
2·7
2·7
1·8
6·7
cosmid 12
0·45
3·1
2·3
1·7
4·5
0·65
1·15
7·8
1·3
2·15
2·9
1·8
2·15
1·6
3·4
6·9

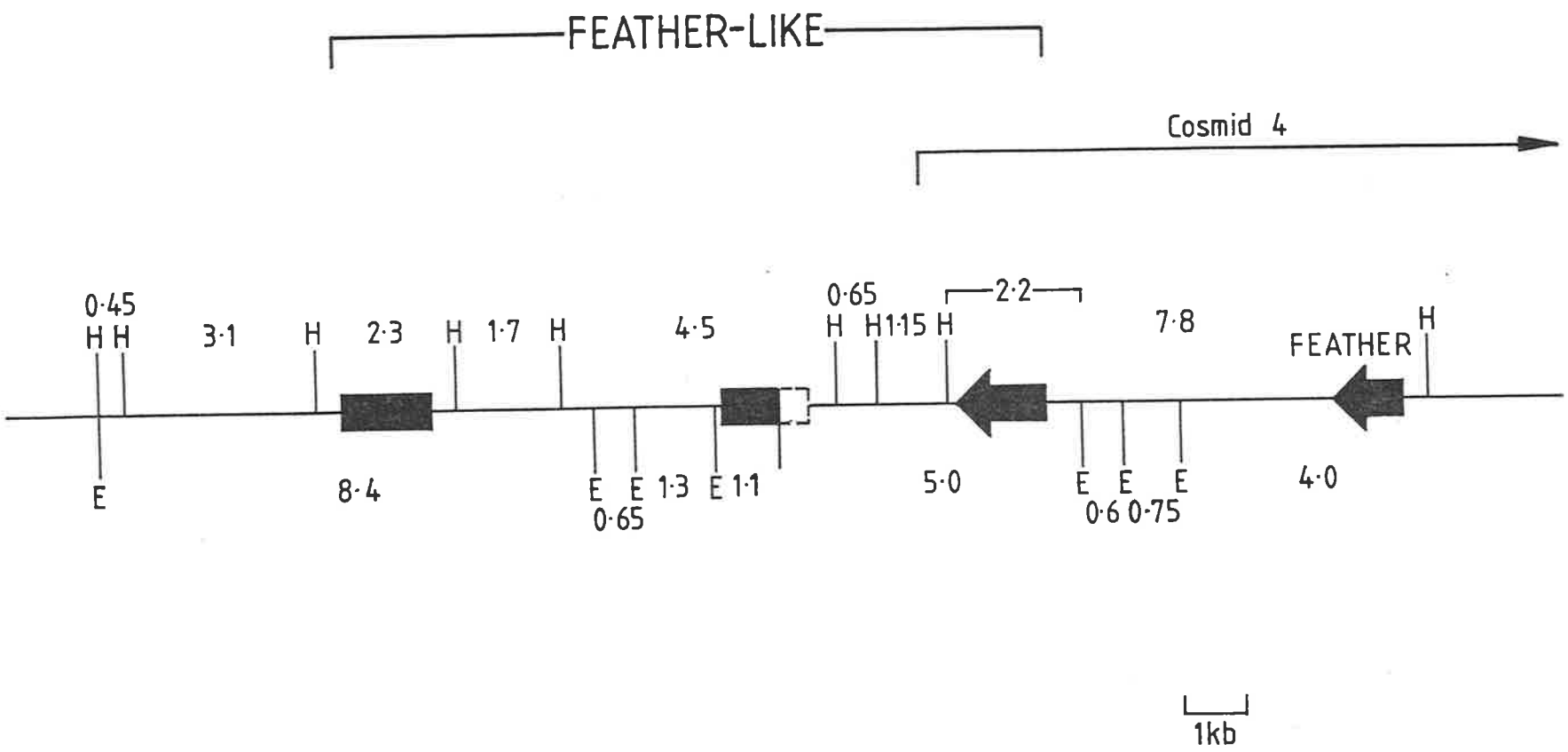
The complete EcoRI, BamHI and HindIII restriction maps of cosmid 12 are summarized in Figure 4.16. The HindIII maps of cosmid 12 and cosmid 4 are compared in Figure 4.17. Cosmid 12 extends cosmid 4 leftwards by 14 kb (Figure 4.17).

C. Characterization of the feather-like genes. During the mapping of cosmid 12, it was determined that in addition to several feather keratin genes, this cosmid contained three regions to the left of the feather gene locus which hybridized weakly with the feather keratin gene probe (see Figures 4.14, 4.15). In an EcoRI digest of cosmid 12, two restriction fragments of 5.0 and 1.1. kb were identified as containing these feather-like keratin sequences (Figure 4.15b,c). A third gene belonging to this 'family' of keratin sequences was identified from the HindIII blot of cosmid 12, in which the 2.3 kb HindIII fragment, which maps in the 8.4 kb EcoRI fragment at the left-hand end of the insert, hybridized weakly with the feather gene probe (see Figure 4.14). This third gene was not apparent from the EcoRI blot as this 8.4 kb EcoRI fragment, which maps at the left-hand end of the insert (Figure 4.16), co-migrates with the 8.4 EcoRI fragment which contains two feather keratin genes (Figure 4.15b).

Figure 4.18 shows the location of the three feather-like genes in the left half of cosmid 12 relative to the left-hand end gene of the feather cluster. The three genes are spaced 3-5 kb apart and each gene contains a BamHI site. In at least two of these genes (those located in the 5.0 and 1.1 kb EcoRI fragments), the BamHI site is present in an identical position in the protein coding region (see Chapter 5).

FIGURE 4.18 : ARRANGEMENT OF THE FEATHER-LIKE GENES IN COSMID 12

The figures shows the restriction map of the left half of the cosmid 12 insert, indicating the position of the three feather-like genes and the last (leftmost) feather keratin gene. The genes were mapped as described in Section 4.3.2.2 and Appendix A. The orientation of the feather-like gene closest to the feather gene cluster, which is present in both cosmids 4 and 12, was determined by DNA sequence analysis (Section 5.3.1). The two remaining feather-like genes are present only in cosmid 12 and map in the 4.5 kb HindIII (1.1 kb EcoRI) and the 2.3 kb HindIII (8.4 kb EcoRI) fragments of this recombinant (Section 4.3.2.2).


FEATHER-LIKE
Cosmid 4
0·45
H H
3·1
H
2·3
H
1·7
H
4·5
0·65
H
H1·15
H
2·2
7·8
FEATHER
H
E
8·4
E
0·65
E
1·3
E
1·1
5·0
E
0·6
E
0·75
E
4·0
1kb

Neither the 5.0 or 1.1 kb EcoRI fragments, which contain feather-like genes, hybridized with either the 5' or 3' non-coding oligonucleotide probes derived from feather keratin sequences (data not shown) suggesting somewhat unexpected variation in both 5' and 3' regions of these 'feather' keratin genes or the presence of genes from another keratin family.

The scale keratins are a related family of keratin genes which share considerable homology with feather keratins (Section 1.4.6; Gregg et al., 1984). Hybridization of a complete scale gene probe (Section 2.2.10), to cosmid 12 DNA suggested that these feather-like genes were not of the scale type, since the scale probe hybridized no more strongly to them than to the feather keratin gene-containing fragments (Figure 4.19). The relative signal intensities of the feather-like and feather keratin gene-containing fragments with the two probes suggest that the feather-like genes are an intermediate form between feather and scale keratin genes (see Sections 4.4.2.2 and 5.4.2 for further discussion).

Partial restriction mapping and Southern blot data of another cosmid, cosmid 2, which extends cosmid 12 leftwards by about 15 kb, indicated that there were no other keratin-hybridizing sequences in the genomic DNA downstream of the 2.3 kb HindIII fragment of cosmid 12 for at least 15 kb (data not shown).

D. Subcloning of a feather-like gene. The 5.0 kb EcoRI fragment was subcloned into pBR322 for further characterization. As in other subcloning experiments, recombinants were analyzed by restriction digestion to check that they contained a correctly sized, resectable insert (Section 4.3.4.2,

FIGURE 4.19 : SOUTHERN ANALYSIS OF COSMIDS 4 AND 12 WITH A SCALE KERATIN GENE

A. Cosmid 4 and 12 DNAs were digested with EcoRI, fractionated on 0.8% agarose, stained with ethidium bromide and visualized under UV light (Section 2.2.7).

B. The DNA fragments were transferred to a nitrocellulose filter (Section 2.2.9) and the immobilized DNA hybridized with the scale keratin gene probe (Section 2.2.10).

Following hybridization, the filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and autoradiographed for 36 hours at -80°C. The sizes (kb) of the hybridizing fragments are displayed.

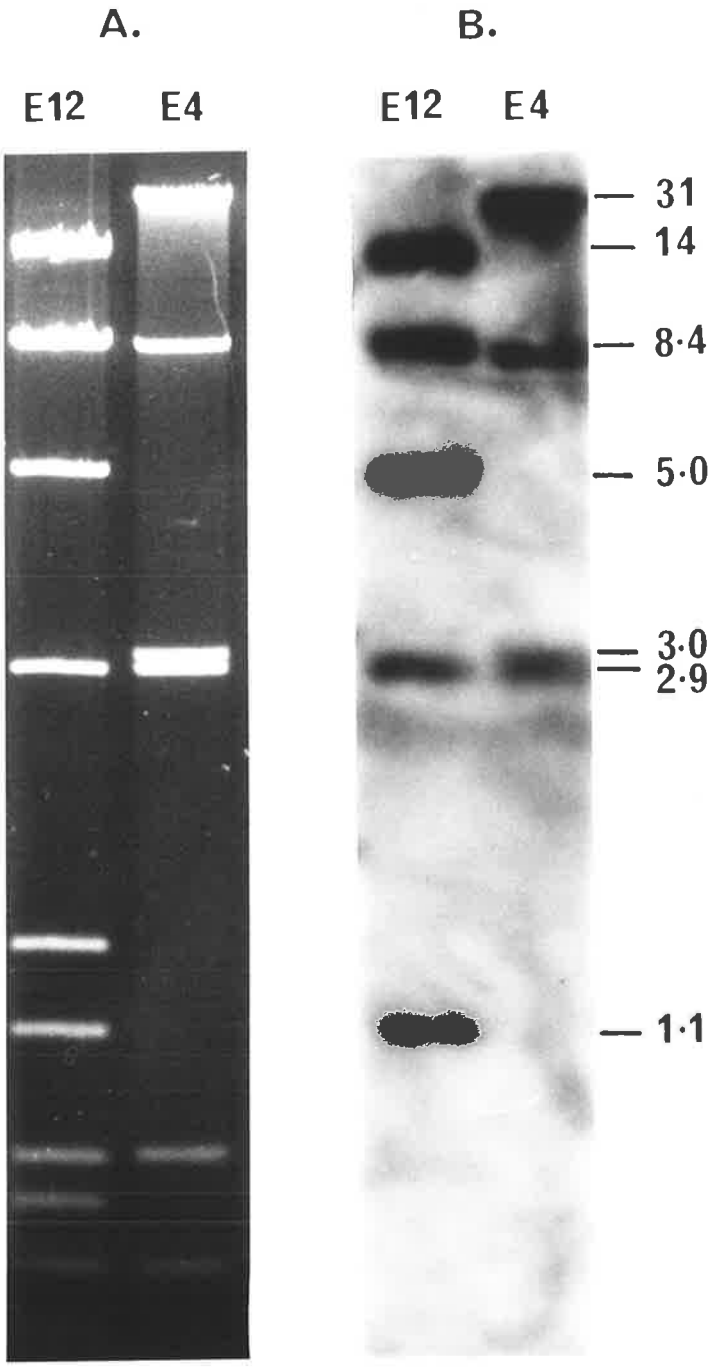

A.
B.
E12
E4
E12
E4
31
14
8·4
5·0
3·0
2·9
1·1

FIGURE 4.20 : RESTRICTION AND SOUTHERN ANALYSIS OF pE5.0

A. pE5.0 DNA was digested with EcoRI, HindIII and EcoRI/HindIII, electrophoresed in 1.6% agarose and visualized under UV light after ethidium bromide staining (Section 2.2.7). Restriction fragment sizes are indicated in kilobases.

The DNA fragments were transferred to nitrocellulose filters (Section 2.2.9) and hybridized with:

B. The feather keratin gene probe.

C. The scale keratin gene probe (Section 2.2.10).

Following hybridization, filters were washed in 1 x SSC, 0.1% SDS at 65°C and exposed at -80°C for the following periods:

B (feather probe) - 3 ½ days

C (scale probe) - 2 days

D. Restriction map of pE5.0. The map was derived as follows. Both the 1.3 kb BamHI/EcoRI and 2.2 kb EcoRI/HindIII fragments were detected by the two keratin probes (Figure 4.20b,c and other data not shown). Therefore, the 1.3 kb BamHI/EcoRI fragment and the 2.2 kb EcoRI/HindIII fragment overlap and lie at one end of the insert. The 1.0 kb EcoRI/HindIII fragment is not present in a HindIII digest (Figure 4.20a) and therefore must be produced from the largest (5.3 kb) HindIII fragment; the 1.0 kb EcoRI/HindIII fragment is located at the other end of the fragment. The 1.15 and 0.65 kb HindIII fragments are located in the middle.

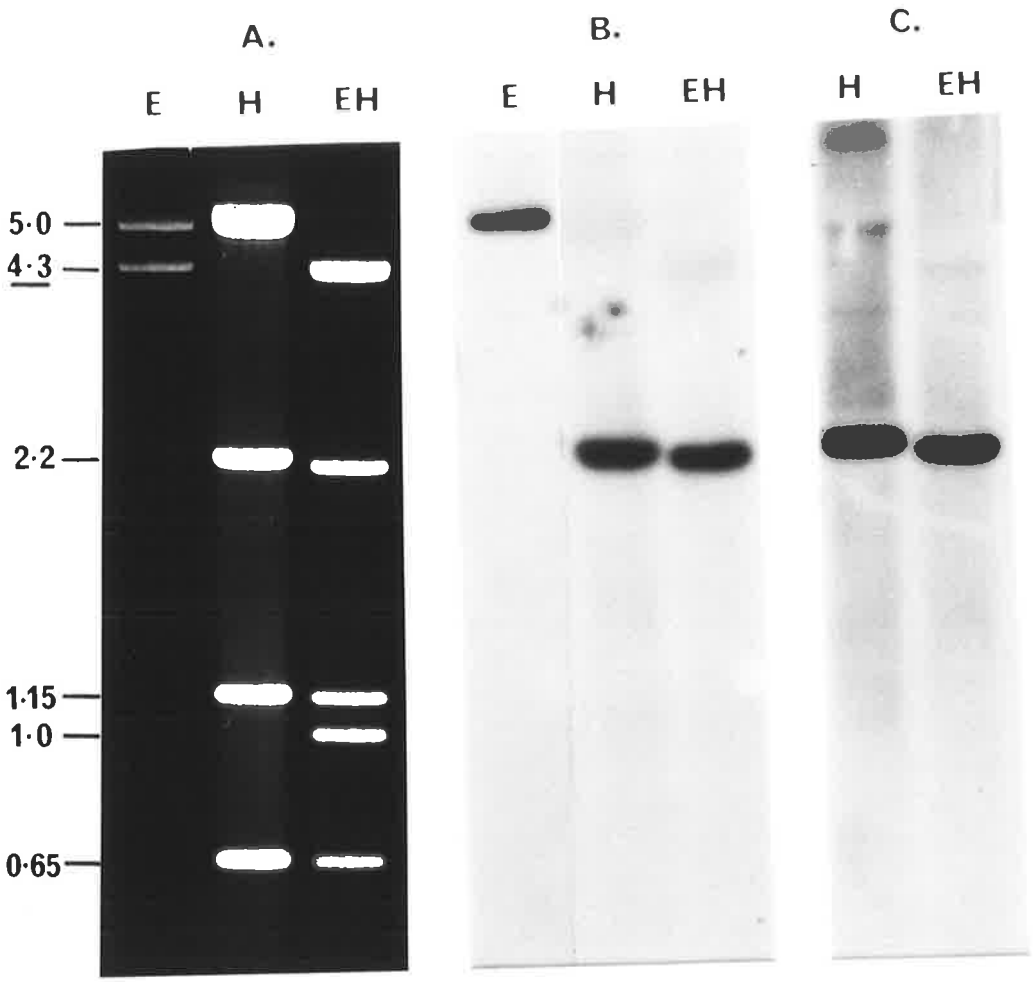
The solid box represents the approximate position of the feather-like gene, as determined from restriction and hybridization analysis.

∿∿ Represents pBR322 vector sequences (not to scale). Restriction sites shown are:

B - BamHI

E - EcoRI

H - HindIII


A.
B.
C.
E
H
EH
E
H
EH
H
EH
5·0
4·3
2·2
1·15
1·0
0·65

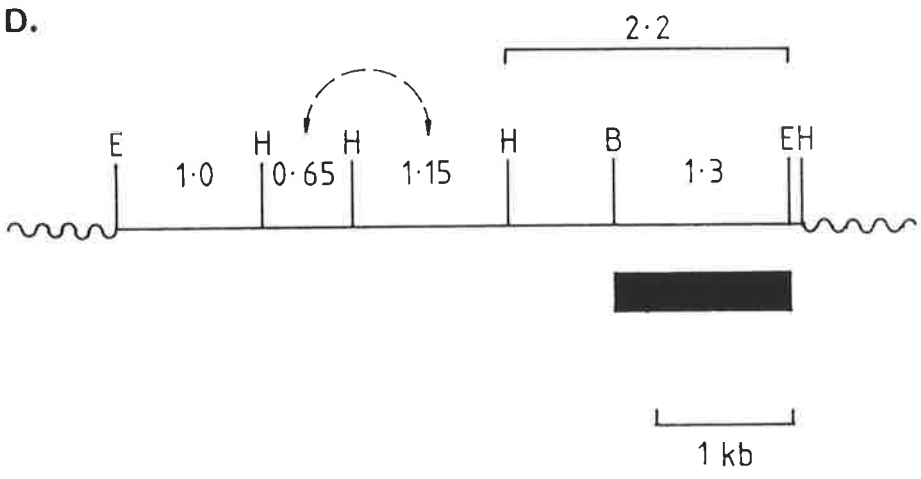

D.
2·2
E
H
H
H
B
EH
1·0
0·65
1·15
1·3
1 kb

Figure 4.35). To locate the genes in the recombinant carrying the 5.0 kb EcoRI insert, pE5.0 DNA was restricted, the resultant fragments transferred to nitrocellulose and probed with feather and scale keratin genes (Figure 4.20). Both feather and scale probes detected a 2.2 kb EcoRI/HindIII fragment (Figure 4.20b,c). In other experiments, not shown, the probes hybridized to a 1.3 kb BamHI/EcoRI fragment and a small 0.16 kb PstI fragment. The restriction map of pE5.0 and the location of the feather-like gene are displayed in Figure 4.20d. The 2.2 kb EcoRI/HindIII fragment is located in cosmid 4 and the insert of cosmid 4 terminates about 0.4 kb to the left of the HindIII site of this fragment (see Figure 4.18).

4.3.2.3 Restriction Mapping of Cosmid 25

Cosmid 25, like cosmid 4, also bound the intergenic λ CFK1-specific probe (Section 3.3.3, Figure 3.9c) and thus must at least partially span the region contained in λCFK1. By comparing the HindIII digestion and blot hybridization patterns of cosmid 25 with cosmid 4, there were a number of fragments which appeared to be present in both clones with sizes of 8.1, 6.7/6.8, 6.4, 3.4, 2.9, 2.7, 2.15 (one of them), 1.8 and 1.6 kb (Figure 4.21a,b). Cosmid 25 differs from cosmid 4 in that it does not contain the 7.8, one of the 2.15 or the 1.3 kb HindIII bands of cosmid 4 and that it also contains probably three extra fragments not present in cosmid 4 - two of 2.7 kb and one of 1.8 kb (Figure 4.21a).

From (i) the staining intensity of the 2.7 kb HindIII fragment in cosmid 25, (ii) the size of the EcoRI fragment encompassing this region of cosmid 25 (data not shown), (iii) other hybridization data described in Section 4.3.4.1

FIGURE 4.21 : COMPARISON OF THE HindIII PATTERNS OF COSMIDS 25 AND 4

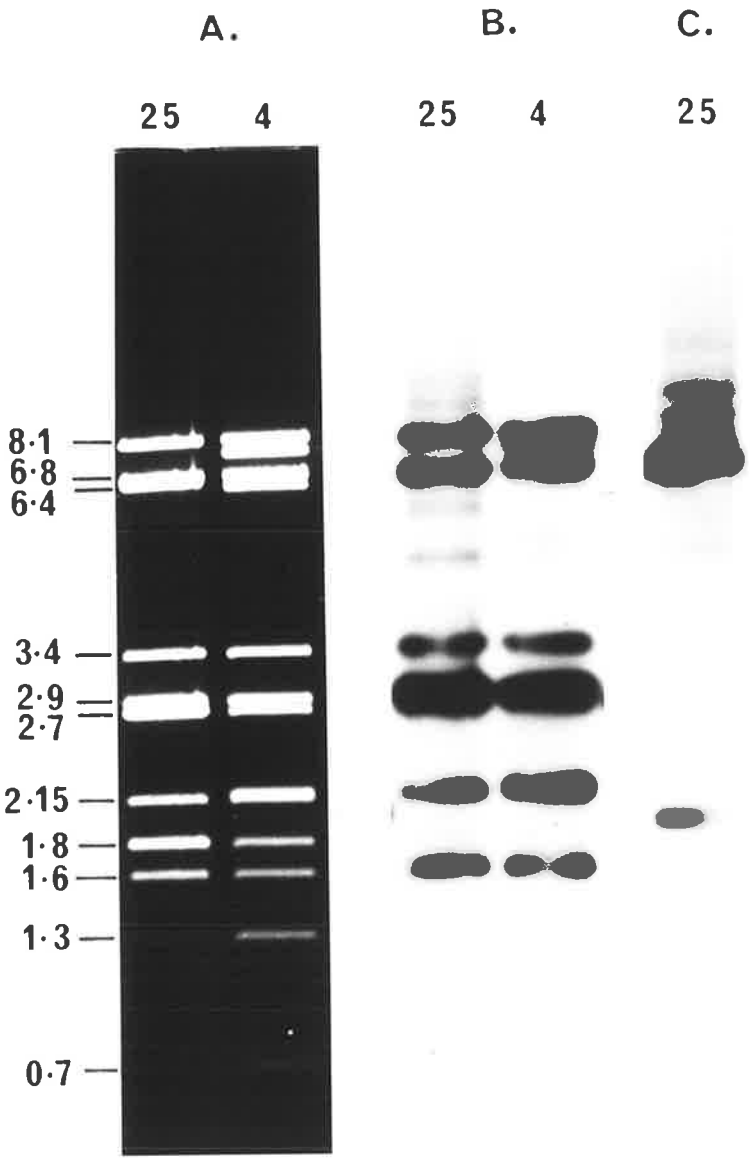
A. Cosmid 25 and 4 DNAs were digested with HindIII, electrophoresed on 0.9% agarose, stained with ethidium bromide and visualized under UV light (Section 2.2.7).

B. The DNA fragments were transferred bidirectionally to nitrocellulose and hybridized with the feather keratin gene probe (see Section 2.2.10). Following hybridization, the filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and autoradiographed for 8 hours at -80°C.

C. A HindIII blot of cosmid 25 probed with labelled pBR322.

The hybridized filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and autoradiographed for 12 hours at -80°C.

The sizes of the HindIII fragments are indicated.


A.
B.
C.
25
4
25
4
25
8·1
6·8
6·4
3·4
2·9
2·7
2·15
1·8
1·6
1·3
0·7

(Figure 4.32), it was estimated that there are three co-migrating 2.7 kb HindIII fragments in cosmid 25, one of which has a counterpart in cosmid 4 (Figure 4.17).

Southern blot analysis of HindIII-restricted cosmid 25 with a pBR322 probe showed a strong hybridization signal with the 6.7 kb fragment and a weak signal with one of the 1.8 kb fragments (Figure 4.21c), indicating that most of the cosmid vector is located in the 6.7 kb fragment while the remainder is contained in the 1.8 kb fragment. Similar experiments enabled the location of the cosmid vector in both BamHI and EcoRI digests of cosmid 25 to be determined (data not shown).

From the HindIII restriction map of cosmid 4 and the data shown in Figure 4.21, the left-hand end of cosmid 25 was positioned in the 2.15 kb HindIII fragment which contains gene A of λCFK1 (Figure 4.17). Therefore the three fragments unique to cosmid 25 (two of 2.7 kb and one of 1.8 kb) must be located rightward of the region spanned by cosmid 4. Since the 1.8 kb HindIII fragment contains part of the cosmid vector, the two 2.7 kb HindIII fragments must be located next to each other (Figure 4.22).

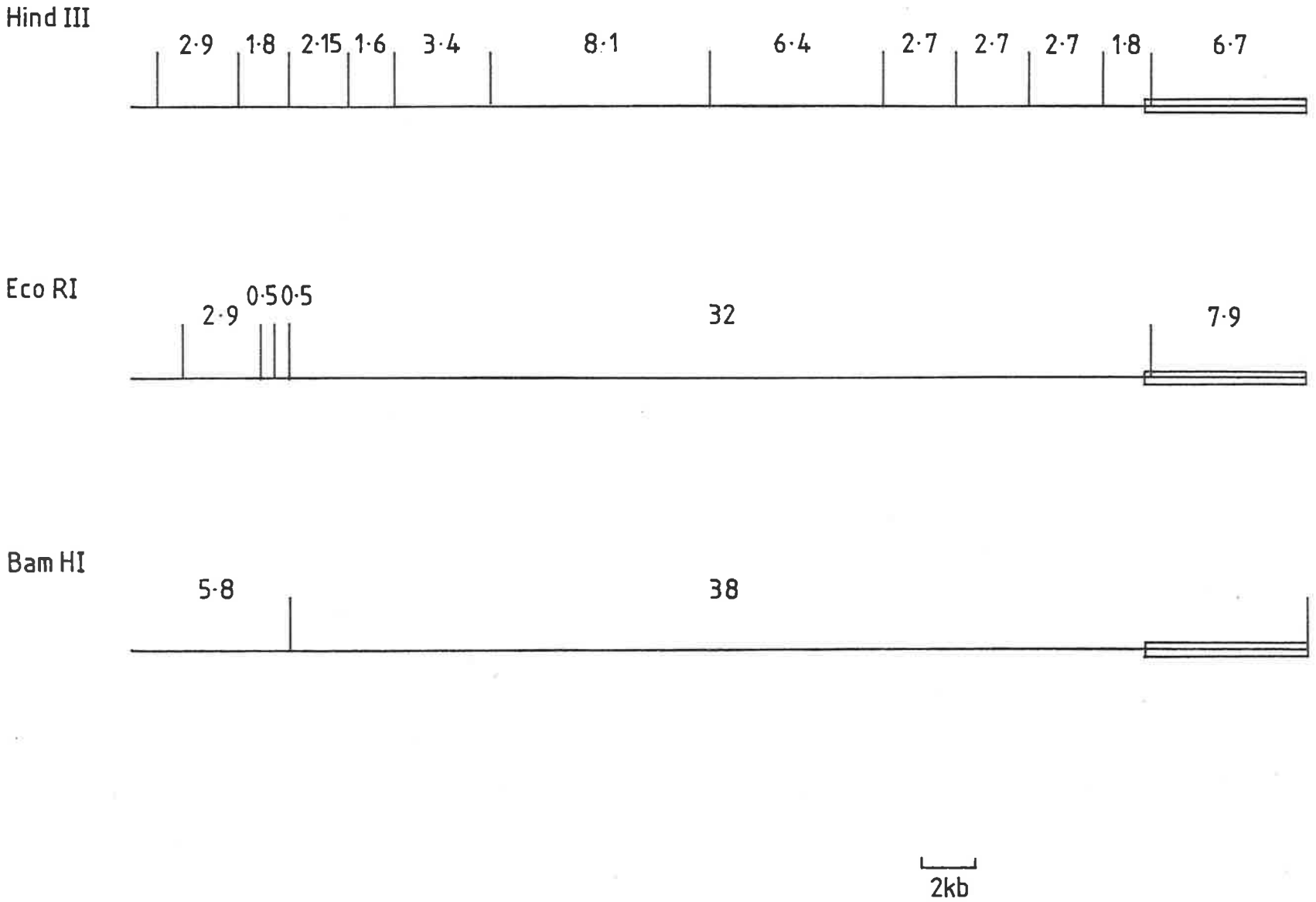
The HindIII map of cosmid 25 derived from the experimental data shown in Figure 4.21 is displayed in Figure 4.22. A comparison of the HindIII restriction maps of cosmids 4 and 25 is also shown (Figure 4.17). The EcoRI and BamHI maps of cosmid 25 presented in Figure 4.22 were derived in a similar manner. The BamHI site present at the extreme right-hand end of cosmid 25 may have been regenerated during the cloning step. Similar results have been found by other workers when using partially cut Sau3A chromosomal DNA as the insert DNA for cosmid cloning (B. Powell, personal communication).

FIGURE 4.22 : RESTRICTION MAPS OF COSMID 25

The HindIII, EcoRI and BamHI restriction enzyme maps were determined as described in Section 4.3.2.3 and from other data not shown.

The sizes, in kilobases, of the restriction fragments are shown. The boxed region indicates the location of the cosmid vector.

Note that one of the two BamHI sites in cosmid 25 abuts the right-hand end of the cosmid vector (see text).


Hind III
2·9
1·8
2·15
1·6
3·4
8·1
6·4
2·7
2·7
2·7
1·8
6·7
Eco RI
2·9
0·5
0·5
32
7·9
Bam HI
5·8
38
2kb

4.3.3 CHARACTERIZATION OF COSMIDS 31 AND 33: LINKAGE TO COSMIDS 4 AND 25

Restriction digestion and Southern analysis of cosmid 25, which extends cosmid 4 rightwards by about 6 kb (Figure 4.17), indicated that there are feather keratin genes in all three HindIII fragments which lie to the right of cosmid 4, including the 1.8 kb HindIII fragment which contains part of the cosmid vector (see Section 4.3.4.1, Figure 4.32 and 4.33) This suggested that the feather gene cluster extended beyond the genomic region contained in cosmid 25. During this time studies were initiated on two other recombinants, cosmids 31 and 33. These clones hybridized to both the complete feather gene probe and the feather-specific oligonucleotide but not the λCFK1-specific probe (Section 3.3.3, Figure 3.9). The detailed analysis of these two cosmids described in this section demonstrated that these cosmids were closely linked to the recombinants already described and enabled the feather keratin gene locus to be extended further rightwards.

4.3.3.1 Restriction Mapping of Cosmid 31

Figure 4.23 shows the pattern of restriction fragments obtained on restriction of cosmid 31 with EcoRI and HindIII and subsequent blot analysis with feather keratin gene and pBR322 probes. A HindIII digest of this clone produced DNA fragments of 7.1, 6.2, 4.7, 2.9 kb and several bands co-migrating at 2.7 kb (Figure 4.23a). Of these, the 6.2, 4.7, 2.9 and 2.7 kb bands hybridized to the feather gene probe (Figure 4.23b). The 7.1 and 4.7 kb HindIII fragments were detected strongly and weakly by pBR322 respectively (Figure 4.23c), indicating that most of the cosmid vector was located in the 7.1 kb fragment with the remainder being in the 4.7 kb fragment.

FIGURE 4.23 : RESTRICTION AND SOUTHERN ANALYSIS OF COSMID 31

A. Cosmid 31 DNA was digested with EcoRI and HindIII and the resulting DNA fragments were fractionated on a 0.9% agarose gel. The gel was stained with ethidium bromide and the DNA detected under UV light (Section 2.2.7). Restriction fragments sizes (kb) are indicated.

The DNA fragments were transferred bidirectionally to nitrocellulose (Section 2.2.9) and the two filters probed with:

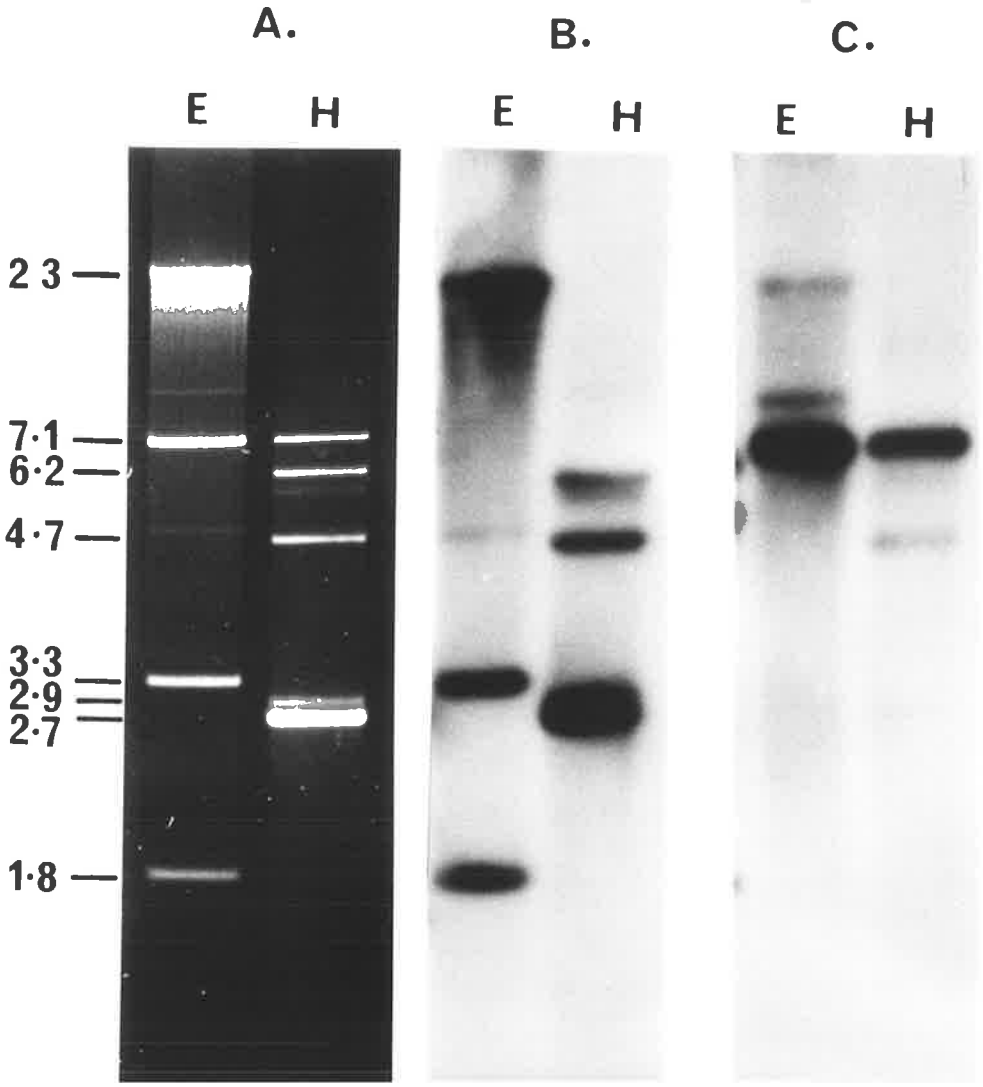
B. The feather keratin gene probe (Section 2.2.10).

C. pBR322 which had been labelled by nick translation (Section 2.2.11.2).

Filters were washed in 0.5 x SSC, 0.1% SDS at 65°C and exposed to X-ray film at -80°C for 8 hours.

E - EcoRI

H - HindIII


A.
B.
C.
E
H
E
H
E
H
23
7·1
6·2
4·7
3·3
2·9
2·7
1·8

An EcoRI digest of cosmid 31 generated four fragments of 23, 7.1, 3.3 and 1.8 kb, of which the 23, 3.3 and 1.8 kb bands hybridized with the feather keratin gene probe (Figure 4.23b). The 7.1 and 23 kb EcoRI fragments hybridized strongly and weakly with the pBR322 probe, respectively, indicating that the cosmid vector was contained in these two DNA fragments (Figure 4.23c).

There were a number of similarities in the restriction and blot hybridization patterns between cosmids 31, 25 and 33. The 3.3 and 1.8 kb EcoRI fragments of cosmid 31, which were detected by the feather gene probe, have counterparts of a similar size in an EcoRI digest of cosmid 33 (see Figure 4.26). Also, the 6.2, 2.9 and 2.7 kb HindIII fragments of cosmid 31 which hybridized with the feather gene probe (Figure 4.23b) have counterparts of a similar size in cosmids 33 (6.2, 2.9 and 2.7 kb), 25 and 4 (2.9 and 2.7 kb) (Figures 4.4b, 4.21, 4.26). This suggested that cosmid 31 may overlap both cosmids 25, 4 and cosmid 33 (Figure 4.25a).

To determine whether the cosmids 25 and 31 are linked as suggested, the 4.7 kb HindIII fragment of cosmid 31 was isolated from a gel and used to probe cosmid 4 under stringent conditions. The resulting Southern blot showed that the 4.7 kb DNA probe hybridized strongly to the 6.4 kb HindIII fragment (arrow in Figure 4.24b). However, the probe also detected the 8.1 and 3.4 kb HindIII fragments and, albeit very weakly, the other HindIII fragments in cosmid 4 which contain feather keratin genes (see Figure 4.4b). The strong hybridization signals observed with this probe to the 8.1, 6.4 and the 3.4 kb HindIII fragments of cosmid 4 (Figure 4.24b) are presumably due

FIGURE 4.24 : SOUTHERN ANALYSIS OF COSMID 4 USING THE 4.7 KB HINDIII FRAGMENT OF COSMID 31 AS PROBE

A. Cosmid 4 DNA was digested with various combinations of restriction enzymes (see below), electrophoresed on 0.9% agarose, stained with ethidium bromide and visualized under UV light (Section 2.2.7).

B. The DNA fragments were transferred on to a nitrocellulose filter (Section 2.2.9) and the immobilized DNA hybridized at 50°C (rather than the normal 42°C) with the 4.7 kb HindIII fragment of cosmid 31 (see Figure 4.25a). The filter was washed in 0.1 x SSC, 0.1% SDS at 65°C and autoradiographed for 6 hours at -80°C.

The arrows indicate the three HindIII fragments of 8.1, 6.4 and 3.4 kb, which are strongly detected by the 4.7 kb probe from cosmid 31.

Restriction enzymes used were:

B - BamHI

H - HindIII

K - KpnI

S - SalI

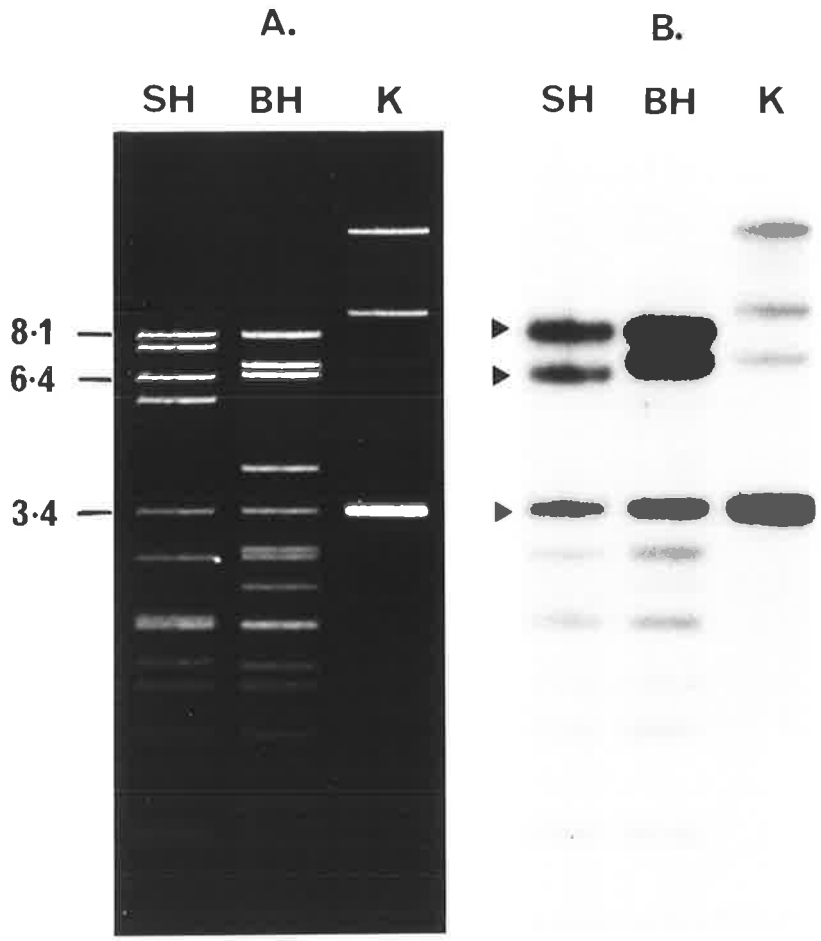
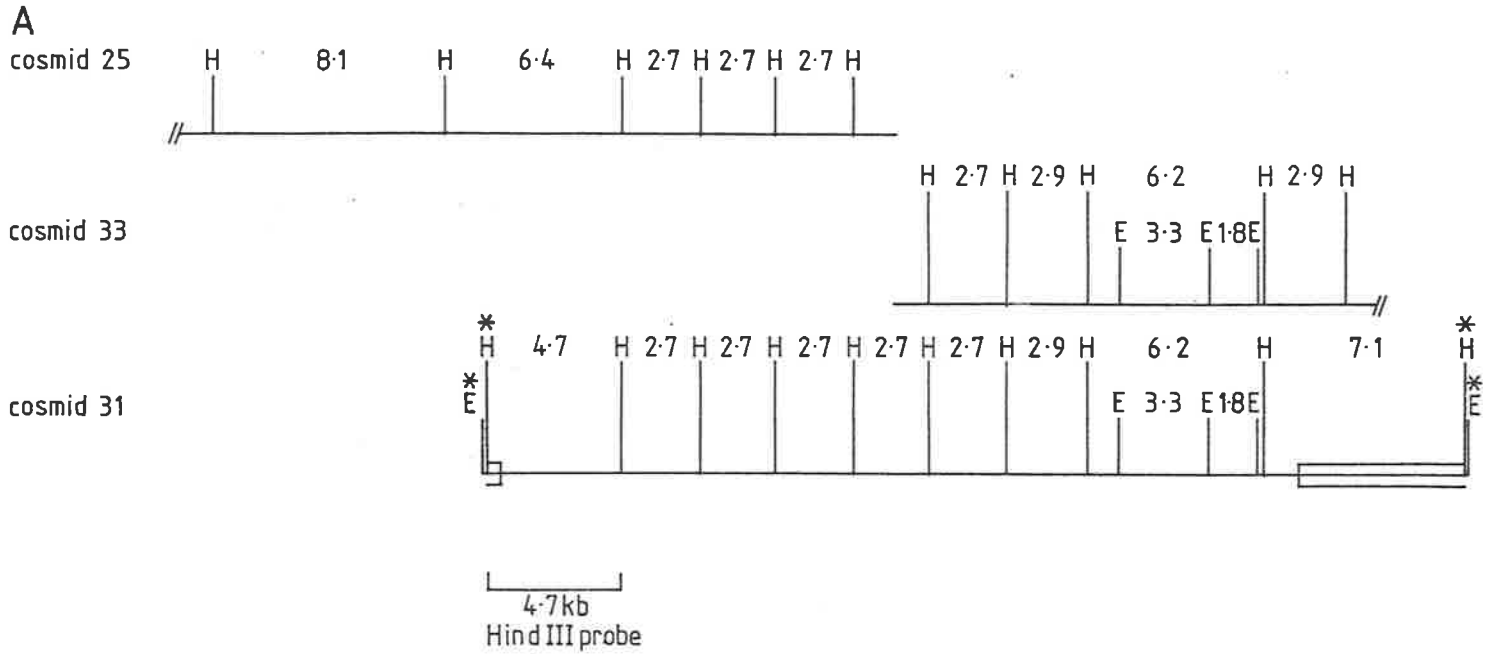
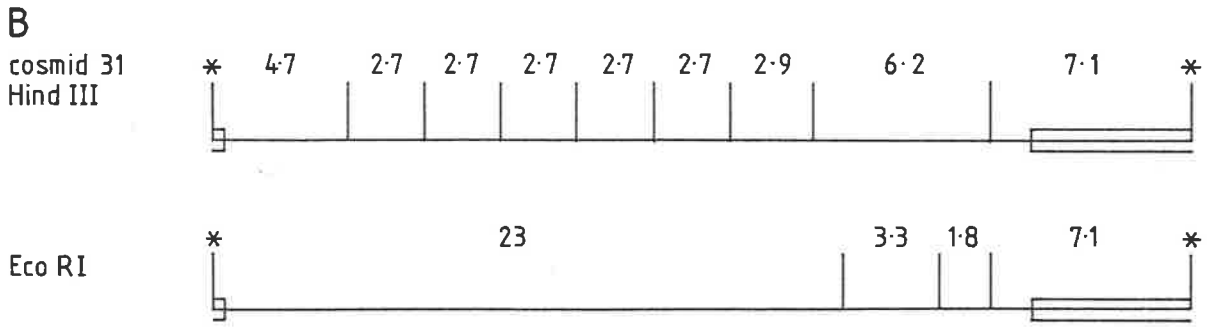

A.
B.
SH
BH
K
SH
BH
K
8·1
6·4
3·4

FIGURE 4.25 : RESTRICTION MAPPING OF COSMID 31

A. Relationship of cosmid 31 to cosmid 25 and 33. This result was derived from (1) the comparison of the restriction patterns of cosmid 31 (Figure 4.23) with cosmids 25 and 33 (see text), (2) the Southern blot shown in Figure 4.24. (3) The Southern blot of EcoRI-cut chick genomic DNA (see Section 4.3.7b). Only part of the cosmids 25 and 33 inserts are shown. The stars at either end of the cosmid 31 map denotes that these restriction sites are contained in the cosmid vector (boxed region) and are not part of the insert DNA.

B. Restriction maps of cosmid 31 for the enzymes EcoRI and HindIII.


A
cosmid 25
H 8·1 H 6·4 H 2·7 H 2·7 H 2·7 H
cosmid 33
H 2·7 H 2·9 H 6·2 H 2·9 H
E 3·3 E 1·8 E
cosmid 31
H 4·7 H 2·7 H 2·7 H 2·7 H 2·7 H 2·7 H 2·9 H 6·2 H 7·1 H
E E 3·3 E 1·8 E
4·7 kb
Hind III probe


B
cosmid 31
Hind III
4·7 2·7 2·7 2·7 2·7 2·7 2·9 6·2 7·1
Eco RI
23 3·3 1·8 7·1

to the close homology of the feather keratin genes located in the right half of cosmid 4 which all contain a KpnI site (see Section 4.3.2.1d). While this result was inconclusive, it did indicate that the 4.7 kb HindIII fragment contained feather keratin genes belonging to the 'Kpn family'. This result, the comparative restriction enzyme data of cosmids 31, 4 and 33 (Figure 4.25a) and the results obtained from probing EcoRI and HindIII-digested chick DNA with a feather keratin gene (Section 4.3.7), collectively demonstrate that these three cosmids do overlap as depicted in Figure 4.25a (see Discussion, Section 4.4.1).

Figure 4.25b displays the restriction map of cosmid 31 which was derived from the data shown in Figure 4.23 and the known restriction maps of cosmids 25 and 33 (Figure 4.25a). From the size of the large EcoRI fragment of cosmid 31 (23 kb, Figure 4.23), it was determined that there are 5 2.7 kb HindIII fragments in cosmid 31. This is compatible with the intensity of the 2.7 kb fragment of cosmid 31 relative to the other HindIII fragments (Figure 4.23a).

4.3.3.2 Restriction Mapping of Cosmid 33. Cosmid 33 was restricted with HindIII and EcoRI/HindIII, and after separation of the resulting fragments by agarose gel electrophoresis, the DNA was transferred to nitrocellulose and probed with the feather keratin gene probe (Figure 4.26). It was not possible to determine whether cosmids 4 and 33 overlapped from a comparison of the HindIII patterns.

Cosmid 33 contained four HindIII fragments which were detected by the feather keratin gene probe of sizes 6.7, 6.2, 2.9 and 2.7 kb (Figure 4.26). The 6.2 kb HindIII fragment was

FIGURE 4.26 : COMPARATIVE RESTRICTION AND SOUTHERN ANALYSES OF COSMIDS 4 AND 33

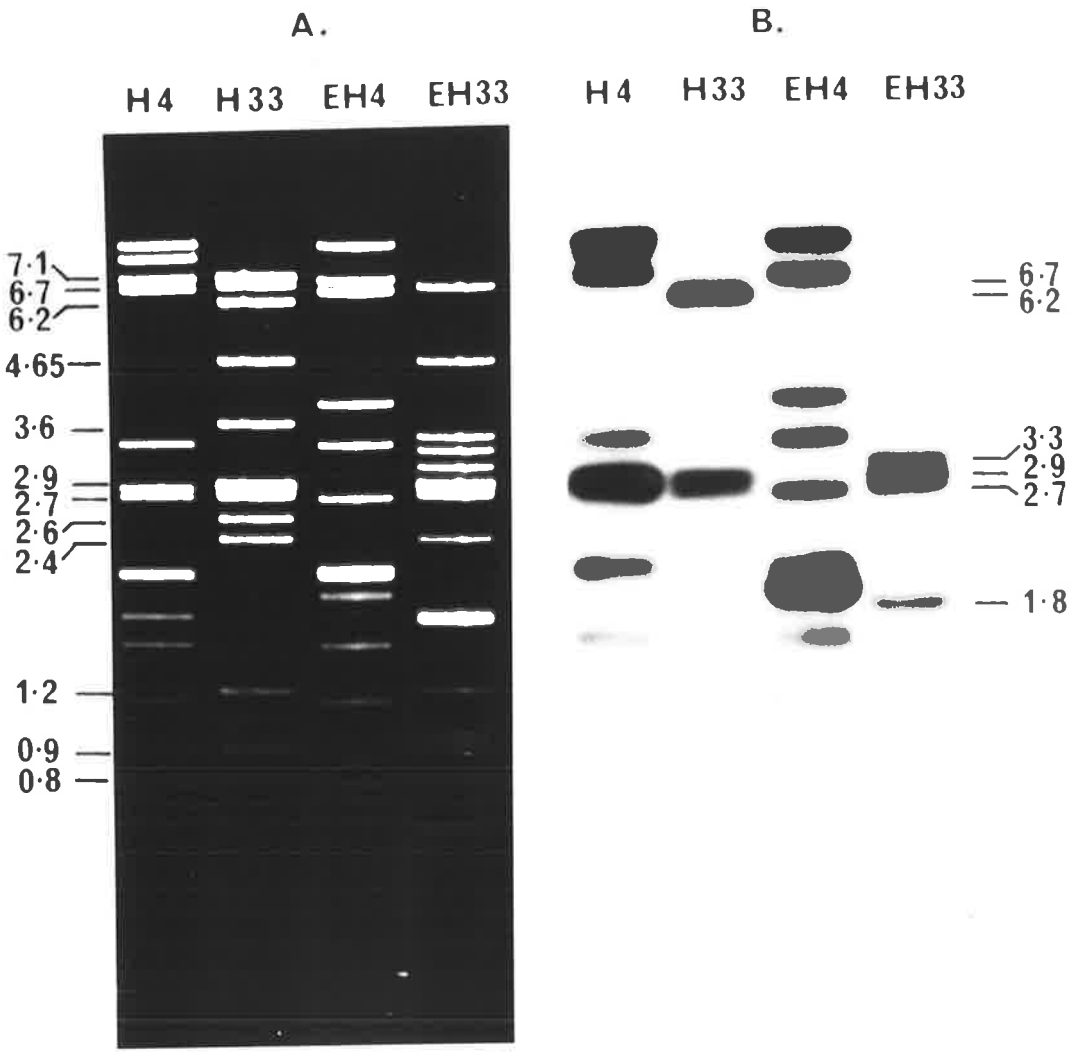
A. Cosmids 4 and 33 DNAs were restricted with HindIII and EcoRI/HindIII, the digestion products fractionated on a 0.9% agarose gel and visualized after ethidium bromide staining by UV illumination (Section 2.2.7). The sizes of the HindIII restriction fragments of cosmid 33 are indicated.

B. The DNA on this gel was transferred and immobilized on to a nitrocellulose filter (Section 2.2.9) and was then hybridized with the feather keratin gene probe (see Section 2.2.10). The filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and exposed to X-ray film for 16 hours at -80°C. The sizes of the restriction fragments which were detected by the probe are shown.

The restriction enzymes used were:

E - EcoRI

H - HindIII


A.
B.
H4
H33
EH4
EH33
H4
H33
EH4
EH33
7·1
6·7
6·2
4·65
3·6
2·9
2·7
2·6
2·4
1·2
0·9
0·8
6·7
6·2
3·3
2·9
2·7
1·8

cut by EcoRI (Figure 4.26a) and subsequent analysis of a sub-clone containing the 6.2 kb HindIII fragment showed that it generated the EcoRI fragments of 3.3 and 1.8 kb which hybridize with the feather keratin gene probe (see Appendix B). These results suggested that cosmid 33 contains up to four feather keratin genes. This was later confirmed using oligonucleotide probes representing various conserved sequences derived from feather keratin genes (see Section 4.3.4.1). Southern hybridization experiments described below demonstrated that three of the HindIII fragments in cosmid 33 (4.65, 2.7 and 2.4 kb) contain claw keratin genes. The cross-hybridization of these fragments in the mapping studies described here and in Appendix B is due to the homology amongst the claw genes contained in these fragments.

In order to map the location of the genes, the complete map of cosmid 33 was determined using the restriction enzymes HindIII, BamHI and KpnI. Examples of different experiments used to derive the map are shown here and in Appendix B and all the data are summarized in the table of Appendix B.

Figure 4.27 shows the restriction patterns of cosmid 33 obtained with HindIII, BamHI and HindIII/BamHI followed by blot hybridization using feather and claw keratin genes, and pBR322, as probes. The feather keratin gene probe detected two BamHI fragments of 11.0 and 6.7 kb and, as shown in the previous figure, four HindIII fragments of 6.7, 6.2, 2.9 and 2.7 kb (Figure 4.27b). The weakly hybridizing 6.7 kb HindIII fragment (arrow in Figure 4.27b) also gave a strong hybridization signal with the pBR322 probe (Figure 4.27c), indicating that this fragment contains most of the cosmid vector and at least part

of a feather keratin gene. Results obtained with a subclone containing the 6.2 kb HindIII fragment established that the 5.9 kb BamHI/HindIII fragment was produced from the 6.2 kb HindIII fragment (see Figure B.3). The small 0.4 kb BamHI/HindIII fragment, which also hybridized with the feather keratin gene probe (Figure 4.27b), therefore was produced from the 6.7 kb HindIII fragment.

The pBR322 probe hybridized strongly to the 6.7 kb HindIII fragment and weakly to the 1.2 kb HindIII fragment (Figure 4.27c), indicating that the cosmid vector is mostly contained in the 6.7 kb HindIII fragment but spans the HindIII site of the 1.2 kb fragment. The pBR322 probe hybridized strongly to the 6.4 kb BamHI fragment and a 6.0 kb BamHI/HindIII fragment (Figure 4.27a,c). These results show that the 6.7 kb HindIII fragment, which contains most of the cosmid vector, has a BamHI site and since there are no BamHI sites in the packaged form of the cosmid vector (Ish-Horowicz and Burke, 1981), the BamHI site must be located at the extreme left-hand end of the insert (Figure 4.28) and was probably regenerated during the construction of the library (see cosmid 25, Section 4.3.2.3, for another example). The pBR322 probe only detects the 9.0 kb KpnI fragment of cosmid 33 (Figure B.1), locating the cosmid vector in this KpnI fragment (Figure 4.28).

The 2.4 kb HindIII fragment of cosmid 33, which contains claw keratin sequences (see Section 4.3.3.3), only hybridized to the 8.5 kb BamHI fragment (Figure 4.27d). From the hybridization results shown in Figure 4.27d, all of the claw keratin genes appear to be closely related. Furthermore, the 2.4 kb HindIII fragment detected by cross-hybridization all three

FIGURE 4.27 : SOUTHERN ANALYSIS OF COSMID 33

A. Cosmid 33 DNA was restricted with BamHI, HindIII or BamHI/HindIII and the products fractionated on a 0.9% agarose gel. The DNA fragments were stained with ethidium bromide and visualized under UV light (Section 2.2.7). The sizes of all of the BamHI fragments, and some of the HindIII fragments, are indicated.

The DNA fragments were transferred bidirectionally to nitrocellulose (Section 2.2.9) and the filter bound DNA hybridized with the following probes:

B. The feather keratin gene probe (Section 2.2.10).

C. pBR322.

D. The 2.4 kb HindIII fragment of cosmid 33 which contains one of the claw keratin genes (see Section 4.3.3.3). The fragment was isolated from LGT agarose (Section 2.2.8) and labelled by nick translation (Section 2.2.11.2).

The filters were washed and autoradiographed as described previously (Figure 4.26). The arrow in B indicates the 6.7 kb HindIII fragment which weakly hybridizes with the feather keratin gene probe. The 6.7 kb fragment is strongly detected using pBR322 as probe (Figure 4.27c). The size of the restriction fragments hybridizing with each of the probes are shown.

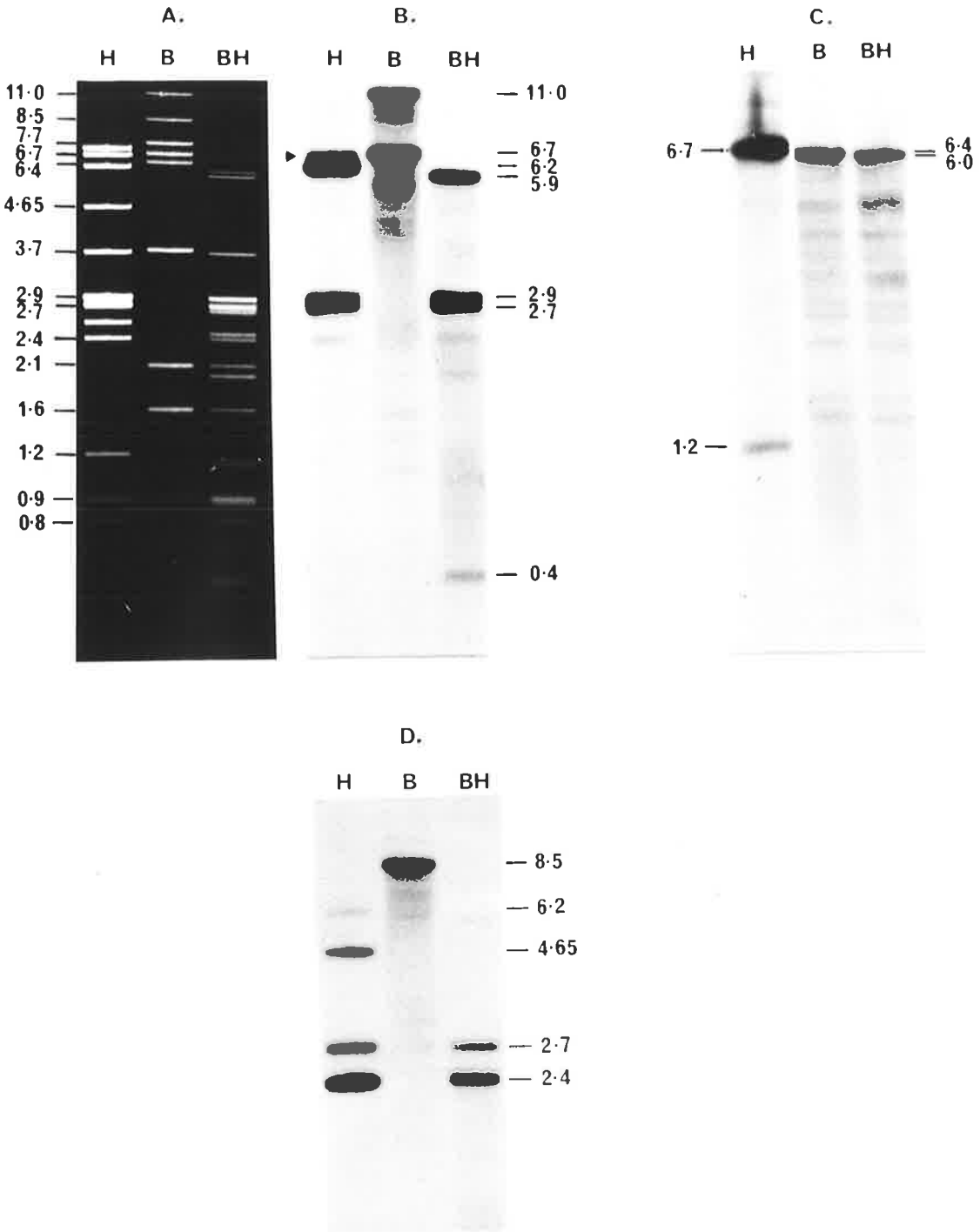
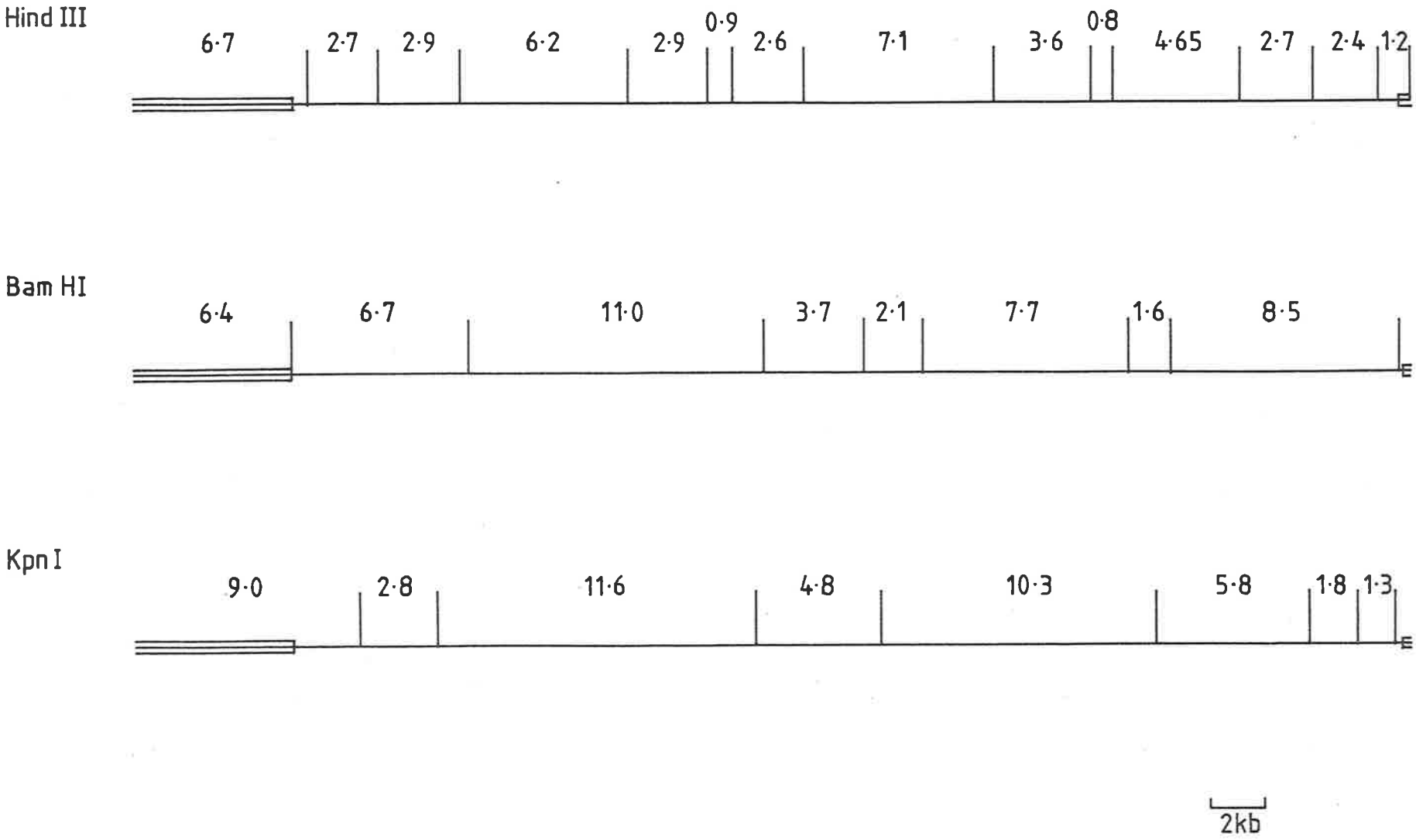

A.
H
B
BH
11·0
8·5
7·7
6·7
6·4
4·65
3·7
2·9
2·7
2·4
2·1
1·6
1·2
0·9
0·8
B.
H
B
BH
11·0
6·7
6·2
5·9
2·9
2·7
0·4
C.
H
B
BH
6·7
1·2
6·4
6·0
D.
H
B
BH
8·5
6·2
4·65
2·7
2·4

FIGURE 4.28 : RESTRICTION MAPS OF COSMID 33

The HindIII, BamHI and KpnI restriction maps of cosmid 33 are displayed. The sizes, in kilobases, of the restriction fragments are indicated. The maps for each restriction enzyme were derived from data presented in Section 4.3.3.2 and Appendix B. The boxed region at the left-hand end of each map indicates the position of the cosmid vector.


Hind III
6·7
2·7
2·9
6·2
2·9
0·9
2·6
7·1
3·6
0·8
4·65
2·7
2·4
1·2
Bam HI
6·4
6·7
11·0
3·7
2·1
7·7
1·6
8·5
Kpn I
9·0
2·8
11·6
4·8
10·3
5·8
1·8
1·3
2kb

HindIII fragments in cosmid 33 thought to contain claw keratin genomic sequences (Figure 4.29). Therefore, it is thought that the claw gene probe used here was suitable to study the genomic location of all the claw keratin genes in cosmid 33.

Similar blot hybridization experiments were performed using all of the HindIII fragments of cosmid 33 as probes to single and double digests of cosmid DNA restricted with BamHI, HindIII and KpnI (see Appendix B). These data (summarized in the table of Appendix B), and additional information from a number of subclones (Appendix B and data not shown) and cosmid 31 (Figure 4.25), enabled the restriction map for cosmid 33 to be derived (Figure 4.28). From hybridization studies described here and Appendix B, cosmid 33 appears to contain four complete feather keratin genes and the part of a fifth gene at its left-hand end and three or possibly four claw keratin genes at the right-hand end. The exact feather keratin gene content of this clone was determined using feather-specific oligonucleotide probes (see below, Section 4.3.4.1). The characterization of cosmids 31 and 33 allowed the right-hand end of the feather keratin gene cluster to be extended further rightwards and enabled the number of feather keratin genes in this locus to be determined.

4.3.3.3 Identification of the Claw Keratin Genes in Cosmid 33

Restriction mapping and blot hybridization data have established that cosmid 33 contains four feather keratin genes and part of a fifth gene in four HindIII fragments of 6.7, 6.2, 2.9 and 2.7 kb (see Section 4.3.3.2 and Figure 4.29, Track 2) which comprise the right-hand end of the feather keratin locus. A long autoradiographic exposure of a Southern blot of HindIII-

FIGURE 4.29 : IDENTIFICATION OF THE CLAW KERATIN GENES IN COSMID 33

TRACK 1: DNA fragments generated from a HindIII digest of cosmid 33 following agarose gel electrophoresis (see Figure 4.26).

TRACK 2: After bidirectional Southern transfer of track 1, one of the filters was hybridized with the feather keratin gene probe as described in the legend to Figure 4.26. Autoradiography was performed for 16 hours at -80°C.

TRACK 3: Five day (120 hour) exposure of Track 2. The hybridization of the feather gene probe to 4.65 and 2.4 kb HindIII fragments is apparent on this long exposure.

TRACK 4: A duplicate filter containing HindIII-digested cosmid 33 DNA was hybridized with a claw keratin cDNA clone, CCK22 (see Section 2.2.10). This autoradiograph was provided by L. Whitbread.

The sizes of the HindIII fragments detected with the feather gene probe (6.7, 6.2, 2.9 and 2.7 kb, Track 2) and claw cDNA clone (4.65, 2.7 and 2.4 kb, Track 4) are shown.

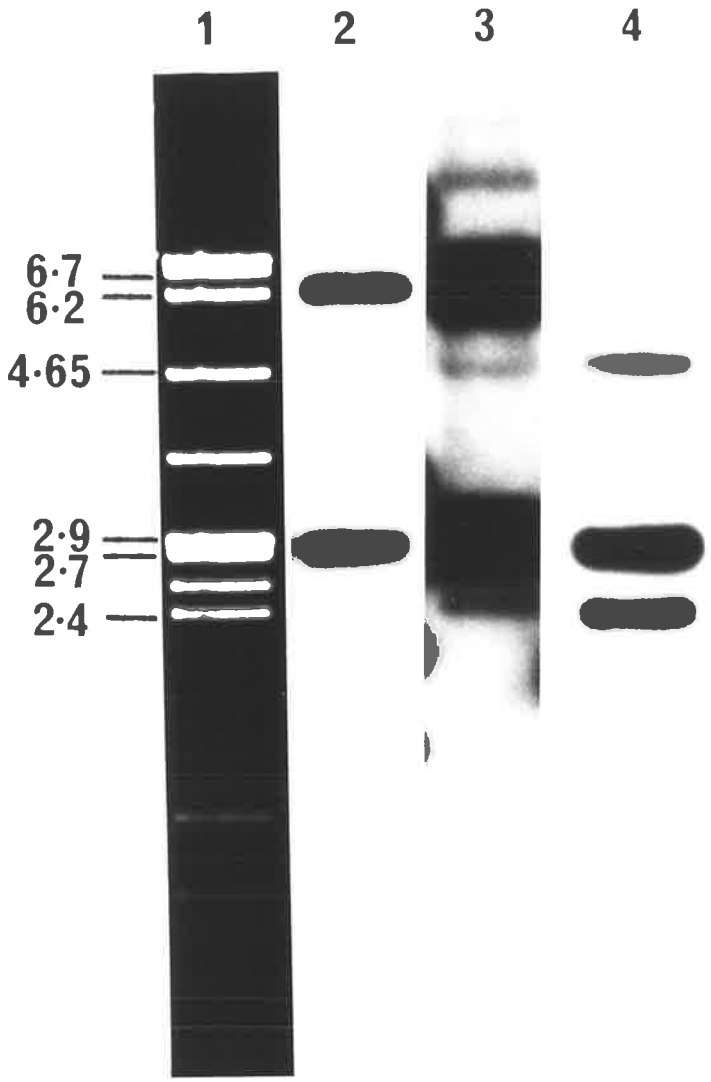

1
2
3
4
6·7
6·2
4·65
2·9
2·7
2·4

FIGURE 4.30 : ARRANGEMENT OF THE CLAW KERATIN GENES IN COSMID 33

The figure shows the location of the claw keratin genes in the 8.5 kb BamHI fragment (Figures 4.27 and 4.28) at the right-hand end of the cosmid 33 insert. The two claw genes which form a divergently orientated gene pair have been partially characterized by DNA sequence analysis (L. Whitbread, personal communication). The other two genes have been mapped by blot hybridization (Sections 4.3.3.2, 4.3.3.3) and DNA sequence analysis (L. Whitbread and K. Gregg, unpublished results). The HindIII and BamHI restriction sites are shown above the map, while the KpnI sites are shown below the map. The sizes (kb) of the restriction fragments are indicated.

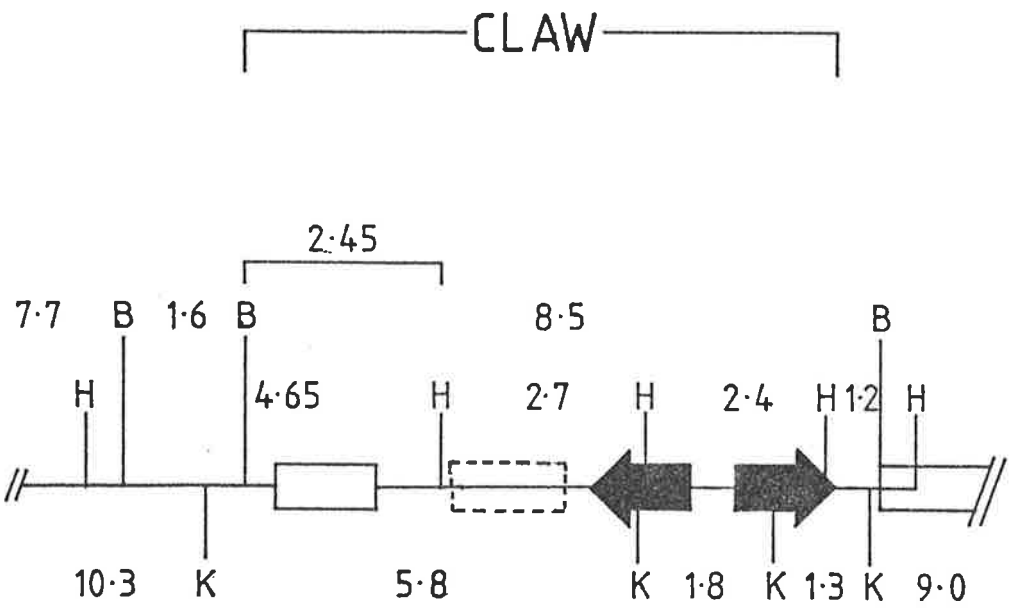
The boxed region partly shown at the right-hand end of the map depicts the cosmid vector.

Restriction enzymes sites are indicated as follows:

B - BamHI

H - HindIII

K - KpnI


CLAW
2·45
7·7
B
1·6
B
8·5
B
H
4·65
H
2·7
H
2·4
H
1·2
H
10·3
K
5·8
K
1·8
K
1·3
K
9·0

restricted cosmid 33 DNA probed with a feather keratin gene revealed two extra hybridizing bands which corresponded to the 4.65 and 2.4 kb fragments (Figure 4.29, Track 3), suggesting that keratin sequences with some homology to feather keratin genes were also present in this cosmid. At that time a cDNA clone, which was identified as encoding a claw keratin gene, had been isolated and partially sequenced (Whitbread, 1985). In dot blot hybridizations, the claw cDNA clone bound strongly to cosmid 33 and an uncharacterized clone, cosmid 35 (Whitbread, 1985). Figure 4.29 shows that this claw cDNA clone hybridizes strongly to the 4.65 and 2.4 kb fragments (which were weakly detected by the feather gene probe) and a third band of 2.7 kb. The restriction mapping of the claw keratin genes (Section 4.3.3.2) demonstrated that they are located in a 8.5 kb BamHI fragment at the right-hand end of the cosmid 33 insert (Figure 4.28). Detailed restriction and Southern blot analyses (Section 4.3.3.2 and Appendix B) and DNA sequencing data (L. Whitbread, personal communication) indicate that there are at least three and possibly four claw keratin genes in this segment which are spaced 1-2 kb apart (Figure 4.30). The gene pair located in the right half of the BamHI fragment appear to be tightly clustered and divergently orientated (Figure 4.30).

4.3.4. CHARACTERIZATION OF THE FEATHER KERATIN GENES IN THE COSMID RECOMBINANTS

The detailed restriction mapping of some of the cosmid recombinants which encompass the feather keratin gene locus was described in Sections 4.3.2 and 4.3.3 and is summarized for HindIII in Figures 4.17 and 4.25a. This section describes the mapping of the genes in these recombinants using Southern blot

hybridization with oligonucleotide probes derived from conserved regions of feather keratin genes and subcloning of some restriction fragments into pBR322 vectors.

4.3.4.1 Use of Oligonucleotide Probes to Determine the Number of Feather Keratin Genes and Their Sequence Conservation

Southern blot analysis using a whole feather keratin gene as probe demonstrated that cosmids 4 and 33 contained a total of 13 hybridizing HindIII restriction fragments (9 and 4, respectively) (Figures 4.26b). To obtain a more accurate estimate of the total number of feather keratin genes in these cosmids, three oligonucleotides prepared from segments of the 5' non-coding, intron and 3' non-coding regions of feather keratin genes (Section 2.2.10) were labelled and hybridized to blots of cosmids 4 and 33.

Figure 4.31 shows the result of hybridizing each of these oligonucleotide probes to HindIII and EcoRI/HindIII restriction digests of cosmids 4 and 33. Each of the probes detect almost all of the restriction fragments which were detected by the feather keratin gene probe (Figure 4.31 b-d). Careful examination of the blots indicated that, with each oligonucleotide probe, there were one or two feather gene-containing restriction fragments that did not hybridize to it. The feather gene-containing restriction fragments which did not bind the various oligonucleotides are arrowed in Figure 4.31b-d.

Figure 4.31 b shows the pattern of hybridization obtained with the 3' non-coding oligomer. Two feather gene-containing restriction fragments did not bind the probe, the 4.0 kb EcoRI/HindIII fragment of cosmid 4 and one of the two 2.9 kb HindIII fragments of cosmid 33 (arrows in Figure 4.31b). The feather

FIGURE 4.31 : SOUTHERN ANALYSIS OF COSMIDS 4 AND 33 WITH OLIGONUCLEOTIDE PROBES

A. Cosmid 4 and 33 DNAs were digested with HindIII and EcoRI/HindIII and the products analyzed by agarose gel electrophoresis (Figure 4.26). The sizes, in kilobases, of some of the restriction fragments are shown.

The cosmid DNA fragments were transferred to nitrocellulose filters using the bidirectional transfer procedure (Section 2.2.9). The filters were hybridized in turn with the following oligonucleotides, which are described in detail in Section 2.2.10.

B. 3' Non-coding probe (19-mer).

C. Intron probe (19-mer).

D. 5' Non-coding probe (20-mer) (see note below).

The filters were washed in 4 x SSC at 65°C and exposed to X-ray film at -80°C. The arrows indicate the feather keratin gene-containing fragments which were not detected by the various probes (see text).

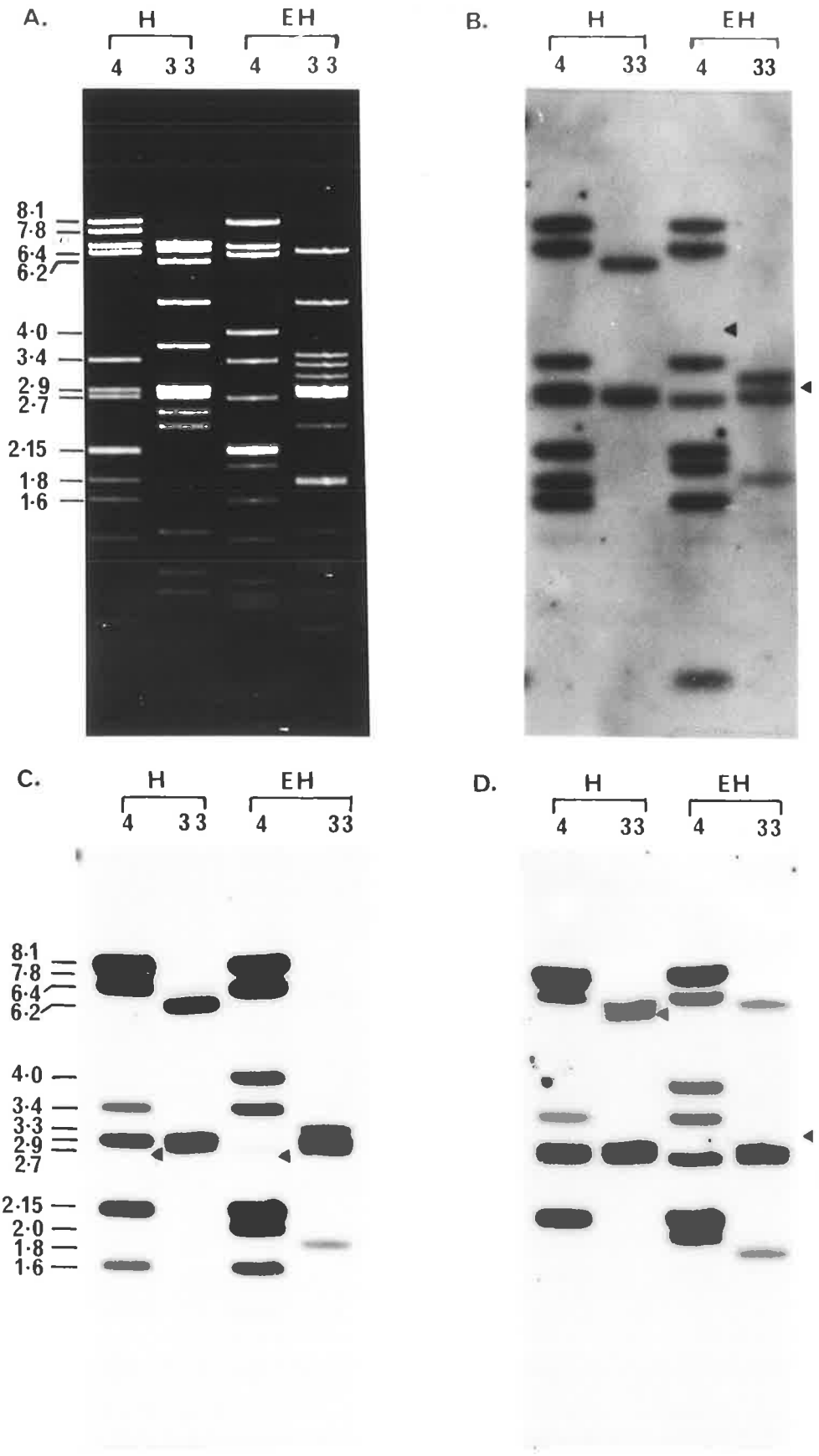
These restriction fragments were:

B - 4.0 kb EcoRI/HindIII fragment, cosmid 4.
2.9 kb EcoRI/HindIII or HindIII fragment, cosmid 33.

C - 2.7 kb HindIII fragment, cosmid 4.

D - 3.3 kb EcoRI/HindIII fragment or EcoRI fragment, cosmid 33.

N.B. After autoradiography of filter C, it was boiled in distilled water to remove annealed probe, prehybridized again and probed with the 5' non-coding oligomer (Figure 4.31d).


A.
H
EH
4
3 3
4
3 3
8·1
7·8
6·4
6·2
4·0
3·4
2·9
2·7
2·15
1·8
1·6
B.
H
EH
4
33
4
33
C.
H
EH
4
33
4
33
8·1
7·8
6·4
6·2
4·0
3·4
3·3
2·9
2·7
2·15
2·0
1·8
1·6
D.
H
EH
4
33
4
33

keratin gene located in one of the two 2.9 kb HindIII fragments of cosmid 33 did not hybridize with the probe, although this is much less apparent in the HindIII track than the EcoRI/HindIII track of cosmid 33 (Figure 4.31b). Hybridization data obtained with subclones containing these two genes (Section 4.3.4.2d and other data not shown) indicate that the genes in the 4.0 kb EcoRI/HindIII and the 2.9 kb HindIII fragments are entirely contained within these fragments and therefore, the absence of hybridization of this probe is not due to the 3' non-coding region of these genes being located in neighbouring restriction fragments. Similar results were obtained when the hybridizations with this probe were performed under reduced stringency conditions (e.g. hybridization at 30°C, washing at 5 x SSC, room temperature). The two complete genes located in the 8.1 kb HindIII fragment of cosmid 4 hybridized with the 3' non-coding oligomer (Section 4.3.4.2a, Figure 4.36b). As the intensity of the signal of the 6.4 kb HindIII fragment of cosmid 4 was similar to that of the 8.1 kb HindIII fragment (Figure 4.31b), it would appear that both genes in the 6.4 kb HindIII fragment (see Section 4.3.4.2c) contain this conserved sequence.

The intron oligonucleotide was found to hybridize to all the feather keratin gene-containing restriction fragments except the 2.7 kb HindIII fragment of cosmid 4 (arrows in Figure 4.31c, see EcoRI/HindIII track in particular). The two complete genes in the 8.1 kb HindIII fragment of cosmid 4 both annealed the intron oligomer (see Section 4.3.4.2a, Figure 4.36c). The two genes in the 6.4 kb HindIII fragment of cosmid 4 appear to have conserved the intron sequence, since

both of the PvuII/HindIII fragments produced from the 6.4 kb HindIII fragment (Section 4.3.4.2c) hybridized with the probe (data not shown). The gene contained in the 2.7 kb HindIII fragment of cosmid 4 has been further characterized by DNA sequencing and is discussed in Section 5.3.4.

Southern blot analysis of these two cosmids with the 5' non-coding probe detected all except one of the feather keratin gene-containing restriction fragments, i.e. the 3.3 kb EcoRI fragment of cosmid 33 which is contained in the 6.2 kb HindIII fragment (arrows in Figure 4.31d). However, it should be noted that the 3.3 kb EcoRI fragment hybridized with a M13 clone (Section 2.2.10) containing the whole of the 37 bp 5' non-coding exon and the 5' flanking sequences of feather keratin gene B (data not shown), suggesting that there is some conservation of the sequence between this gene and gene B. The 6.7 kb HindIII fragment of cosmid 33, which contains most of the cosmid vector, was only detected by the 5' non-coding probe (Figure 4.31d), demonstrating that the 5' end of a feather keratin gene abutted the cosmid vector.

The 8.1 kb HindIII fragment contains two complete genes and the 5' end of a third gene, all of which hybridized with the 5' non-coding oligomer (Section 4.3.4.2, Figure 4.36d). Comparison of the intensity of hybridization of the 6.4 kb HindIII fragment to the 8.1 kb HindIII fragment suggests that this 5' non-coding sequence is conserved in at least one of the two genes located in the 6.4 kb fragment.

In order to determine whether the two 2.7 kb HindIII fragments of cosmid 25 which lie rightwards of cosmid 4 contained feather keratin genes, cosmid 25 was probed with the intron

oligonucleotide. Figure 4.32 shows that the two 2.7 kb HindIII fragments of cosmid 25 which are not present in cosmid 4 both hybridize with this probe (arrow in Figure 4.32b). In the EcoRI/ HindIII track of cosmid 25, the 2.7 kb band is twice as intense as the 3.4 or 1.6 kb fragments, which both contain one feather keratin sequence. The third 2.7 kb HindIII fragment of cosmid 25, which has an identical counterpart in cosmid 4 (see Figure 4.17), is not detected by the intron probe (Figure 4.31c, 4.32b). This indicates that the two 2.7 kb HindIII fragments of cosmid 25 which lie to the right of the genomic region contained in cosmid 4 both contain feather keratin genes.

The right-hand end fragment of the cosmid 25 insert is a 1.8 kb HindIII fragment which contains about 1.4 kb of genomic DNA plus part of the cosmid vector (see Figure 4.22). Figure 4.33 shows the result of hybridizing the 1.8 kb HindIII fragment from cosmid 25 to a HindIII digest of cosmid 33. This recombinant contains four complete feather keratin genes located in HindIII fragments of 6.2, 2.9 and 2.7 kb (see Section 4.3.3.2). The 1.8 kb HindIII probe detected these three HindIII fragments (Figure 4.33) suggesting that the right-hand end fragment of cosmid 25 also contains part of a feather keratin gene. This 1.8 kb probe from cosmid 25 also detected the 1.2 kb HindIII fragment of cosmid 33 which, in that clone, contains part of the cosmid vector (see Figure 4.28).

The length of the exposure that was required to give a signal (3 days) suggests that the 1.8 kb fragment contains part of the 3' non-coding region of a feather keratin gene, a part

FIGURE 4.32 : SOUTHERN ANALYSIS OF COSMIDS 4 AND 25 WITH THE INTRON OLIGONUCLEOTIDE PROBE

A. Cosmid 4 and 25 DNAs were digested with HindIII and EcoRI/HindIII, the products fractionated on an agarose gel and visualized under UV light after ethidium bromide staining (Section 2.2.7). The sizes of the HindIII fragments of cosmid 4, many of which are also present in cosmid 25 (see text), are shown in kilobases.

B. Following Southern transfer (Section 2.2.9), the filter-bound DNA was hybridized to the intron oligonucleotide (Section 2.2.10).

The filter was washed in 4 x SSC at 65°C and autoradiographed overnight at -80°C. The arrow indicates the position of the 2.7 kb band of cosmid 25 which hybridized strongly with the probe.

E - EcoRI

H - HindIII

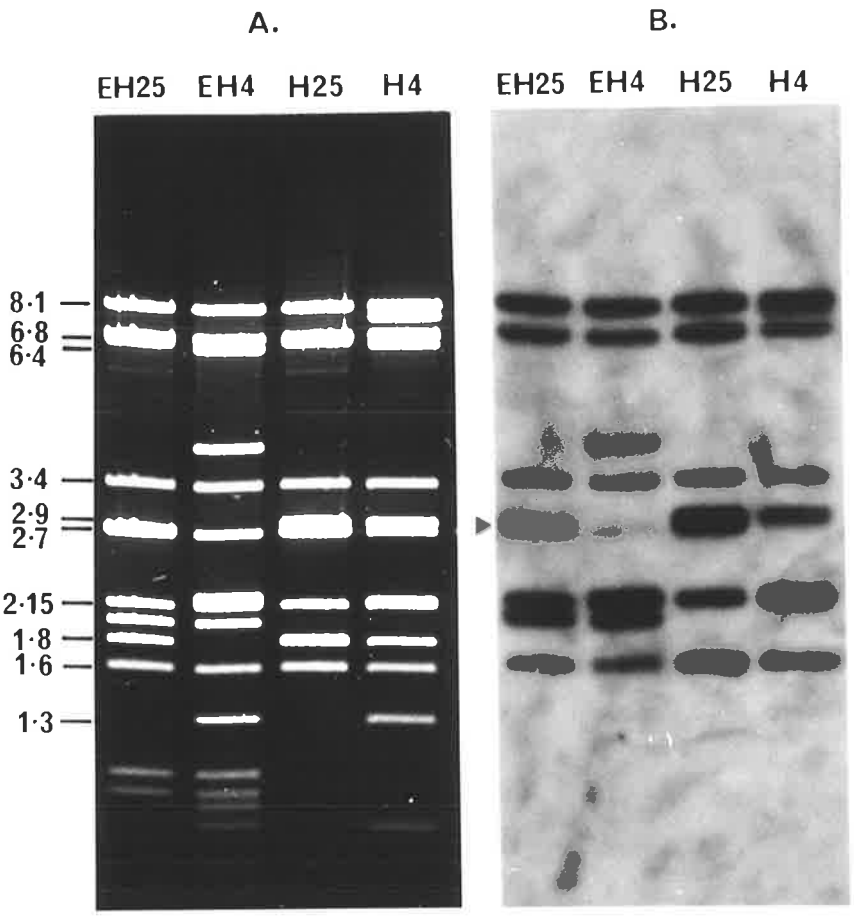
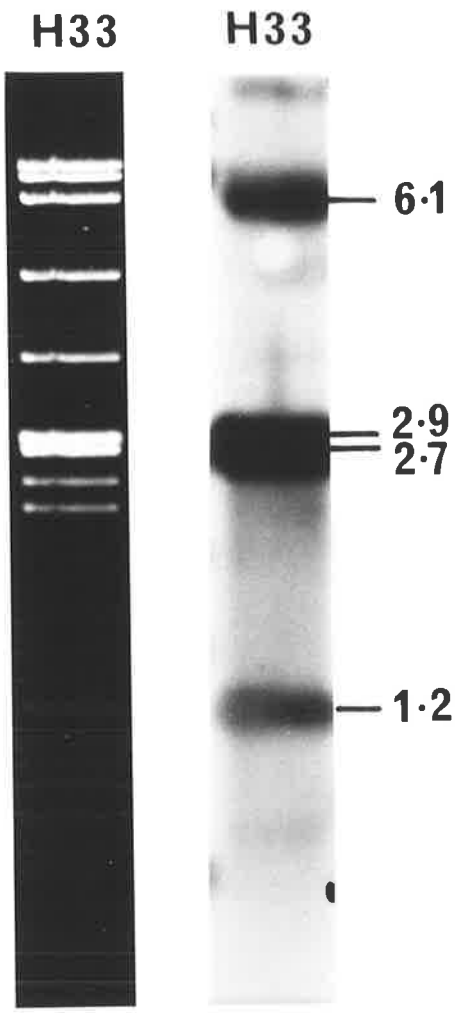

A.
B.
EH25
EH4
H25
H4
EH25
EH4
H25
H4
8·1
6·8
6·4
3·4
2·9
2·7
2·15
1·8
1·6
1·3

FIGURE 4.33 : HYBRIDIZATION OF THE RIGHT-HAND END FRAGMENT OF COSMID 25 TO COSMID 33

Cosmid 33 DNA was digested with HindIII and the products electrophoresed on a 0.9% agarose gel (see Figure 4.26). After blotting on to nitrocellulose (Section 2.2.9), the filter was hybridized with the 1.8 kb HindIII fragment of cosmid 25 which contains the right-hand end of the cosmid 25 insert plus 0.4 kb of cosmid vector (Figure 4.22). Following hybridization, the filter was washed in 0.1 x SSC, 0.1% SDS at 65°C and exposed to X-ray film at -80°C for 3 days.

The three HindIII fragments of cosmid 33 which contain feather keratin sequences and the 1.2 kb fragment, which contains part of the cosmid vector, were detected by the probe (see Figure 4.27).


H33
H33
6·1
2·9
2·7
1·2

which is generally not well conserved amongst feather keratin sequences (Section 1.4.4, Figure 1.7). This may explain why the 1.8 kb fragment was not detected in the comparatively short exposure of the Southern blot probed with the feather keratin gene (Figure 4.21b).

4.3.4.2 Characterization of Feather Keratin Gene-Containing Subclones from Cosmid 4

The three HindIII fragments of cosmid 4 located to the right of the genomic region present in λCFK1 (8.1, 6.4 and 2.7 kb, Figure 4.13) all hybridized strongly with feather keratin probes (see Figures 4.4.b, 4.31). In order to determine the number and arrangement of the feather keratin genes within these fragments, the three HindIII fragments were subcloned into pBR322 as described in Section 4.2. Transformants were screened by colony hybridization and miniscreen analysis of small scale cultures. A typical result of a colony hybridization experiment is shown in Figure 4.34. Some, but not all, of the colonies hybridized with the feather keratin gene probe.

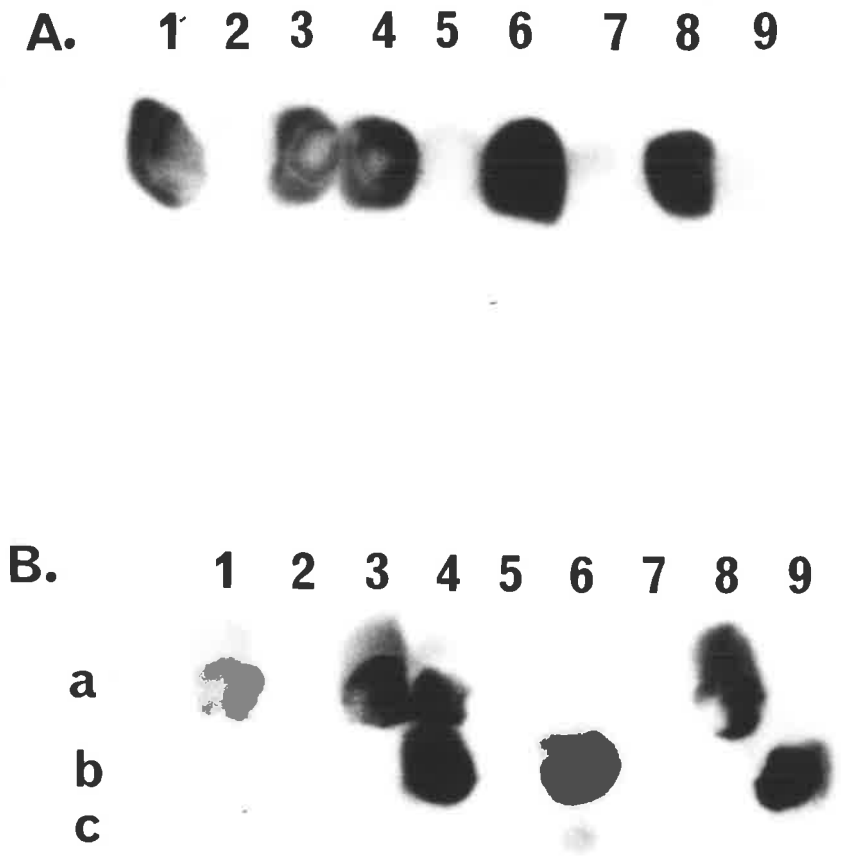
Recombinants that contained feather keratin sequences were further tested by miniscreen analysis to check that the plasmid was carrying the correct size insert. Figure 4.35 shows the result of a number of miniscreen analyses of clones from four different cloning events. In all cases, except one, the clone insert was the correct size and was resectable with the appropriate enzyme. Figure 4.35a shows recombinants generated from the cloning of the 8.1 and 2.7 kb HindIII fragments and Figure 4.35b shows recombinants generated from the cloning of the 6.4 kb HindIII and 5.0 kb EcoRI fragments. The latter contains one of the feather-like genes located to the left of

FIGURE 4.34 : DETECTION OF RECOMBINANTS CONTAINING FEATHER KERATIN GENE SEQUENCES BY COLONY HYBRIDIZATION

Recombinants that were Amp^r Tet^s were grown on a nitrocellulose filter layered on to a L-agar plate containing ampicillin (40 µg/ml). The plasmid DNA was lysed and fixed to the filters using the colony lysis procedure described by Maniatis et al. (1982). The cloned DNA was hybridized with the feather gene probe, washed in 0.5 x SSC, 0.1% SDS at 65°C and exposed to X-ray film at -80°C for 1 day.

PANEL A shows 9 Amp^R Tet^s colonies obtained from the cloning of the 2.7 kb HindIII fragment hybridized with the feather gene probe.

PANEL B shows 27 Amp^r Tet^s colonies from the cloning of the 8.1 kb HindIII fragment screened with the feather gene probe (see text for further details).


A.
1 2 3 4 5 6 7 8 9
B.
1 2 3 4 5 6 7 8 9
a
b
c

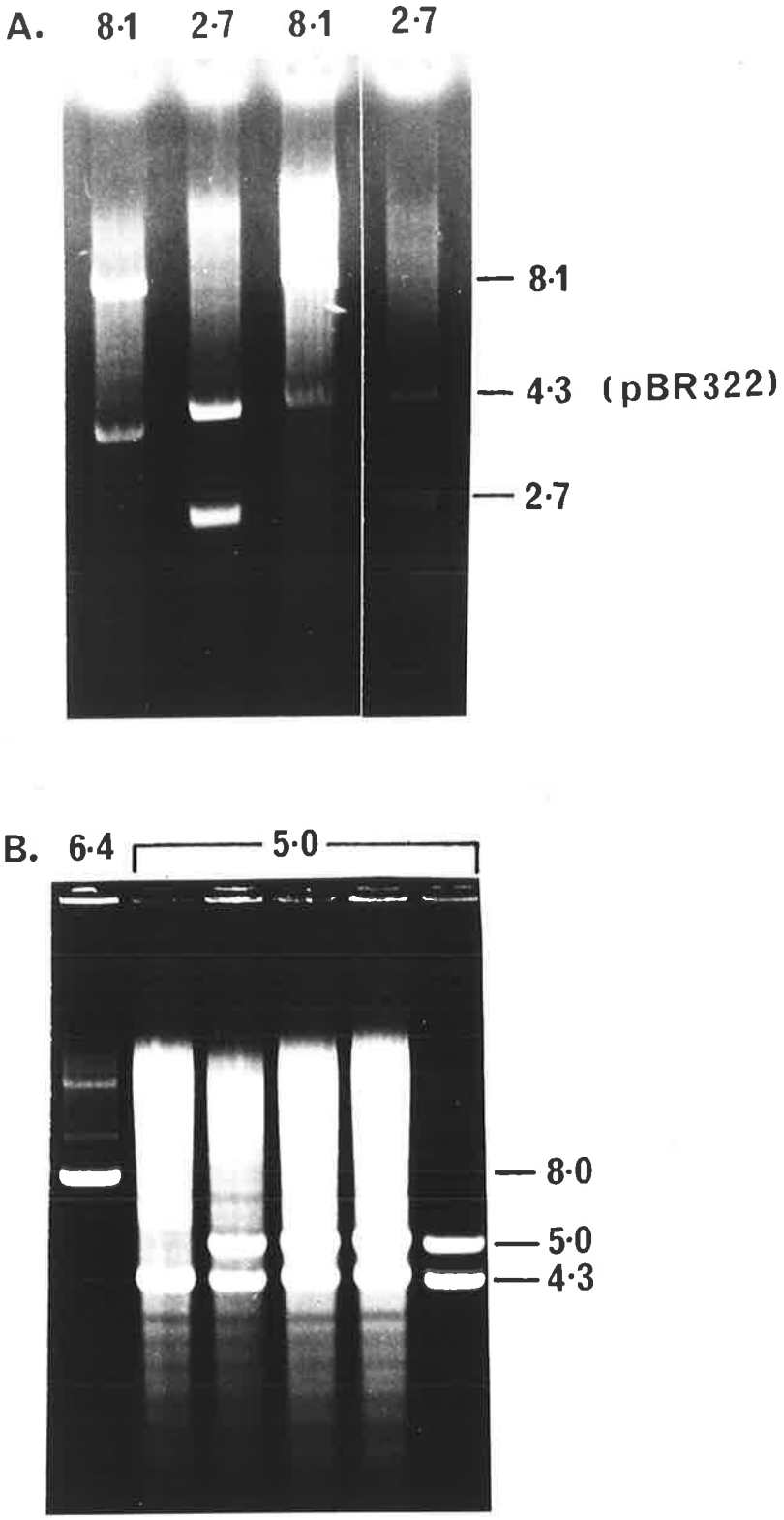
FIGURES 4.35 : MINISCREEN ANALYSIS OF PLASMID RECOMBINANTS

DNA was prepared from small scale cultures of the plasmid recombinants as described in Section 4.2.6. A 0.5 µg sample of each DNA preparation was digested with the appropriate enzyme (see below), the products fractionated on a 1% agarose gel and visualized under UV light after staining with ethidium bromide (Section 2.2.7). The gels show:

A. Digestion pattern of pH 8.1 and pH 2.7 subclones restricted with HindIII.

B. Digestion patterns of pE5.0 subclones and a single recombinant generated from the cloning of the 6.4 kb HindIII fragment. The 6.4 clone has lost part of the 6.4 kb insert and pBR322 sequences and is only 8.0 kb in size (see text for further details).

The sizes of the resected inserts (8.1, 2.7 and 5.0 kb), the 6.4 deletion clone (8.0 kb) and the pBR322 vector fragment (4.3 kb) are indicated.


A.
8·1
2·7
8·1
2·7
— 8·1
— 4·3 (pBR322)
— 2·7
B.
6·4
5·0
— 8·0
— 5·0
— 4·3

the feather cluster, see Section 4.3.2.2c). As discussed below, the 6.4 kb HindIII fragment could not be successfully cloned. Only recombinants which had deleted a portion of the insert and one HindIII site were ever isolated (Figure 4.35).

A. Characterization of a recombinant containing the 8.1 kb HindIII fragment (pH 8.1). In order to map the gene(s) in the 8.1 kb HindIII fragment, pH 8.1 DNA was digested with HindIII, KpnI and BglII and combinations thereof and the products fractionated on an agarose gel (Figure 4.36a). The pH 8.1 clone contained two KpnI and BglII sites in the insert (there are no KpnI or BglII sites in pBR322 (Sutcliffe, 1978). The digestion patterns obtained with these enzymes (Figure 4.36a) enabled the restriction map shown in Figure 4.36e to be constructed (for a detailed description of the derivation of the map, see Appendix C).

Preliminary Southern blot analysis with a whole feather keratin gene as probe detected all the insert restriction fragments, suggesting the presence of more than one gene. In order to precisely determine the number of genes and their orientation, the blots were hybridized with three different oligonucleotide probes from the 3' non-coding region, intron and 5' non-coding regions (see Section 2.2.10 for a description). The results of these hybridization experiments to pH 8.1 are shown in Figure 4.36 b-d. Comparison of the KpnI/HindIII tracks of these three blots indicated differences in the pattern of hybridizing bands. The 1.85 kb KpnI/HindIII fragment, for example, hybridized with the 5' non-coding and intron probes but not the 3' non-coding probe, indicating that the gene contained in this fragment spans the KpnI site

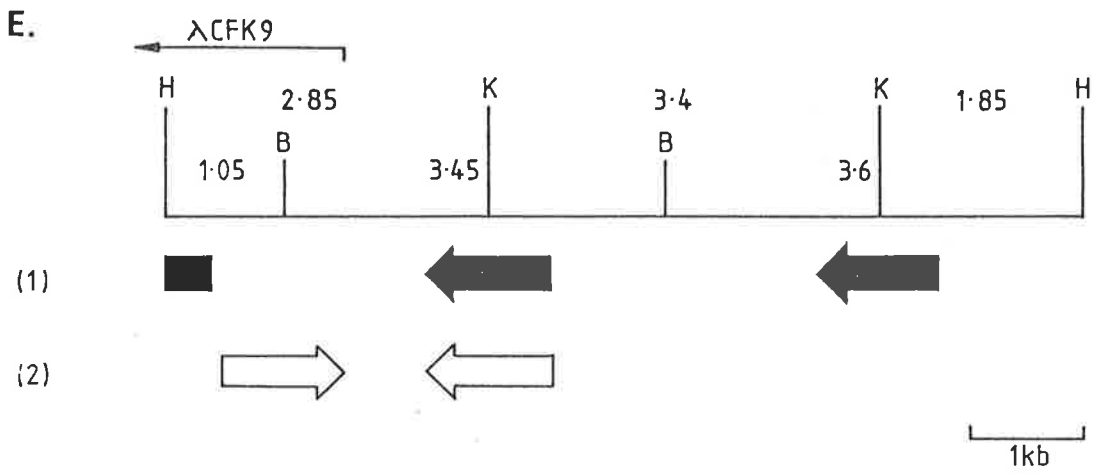


FIGURE 4.36 : RESTRICTION AND SOUTHERN ANALYSIS OF pH 8.1

A. pH 8.1 DNA was digested with a number of restriction enzymes (see below), fractionated on a 1.1% agarose gel, stained with ethidium bromide and visualized under UV light (Section 2.2.7).

The DNA fragments were transferred bidirectionally to nitrocellulose filters (Section 2.2.9), immobilized and hybridized with the following oligonucleotide probes (see Section 2.2.10 for details):

B. The 3' non-coding probe (19-mer)

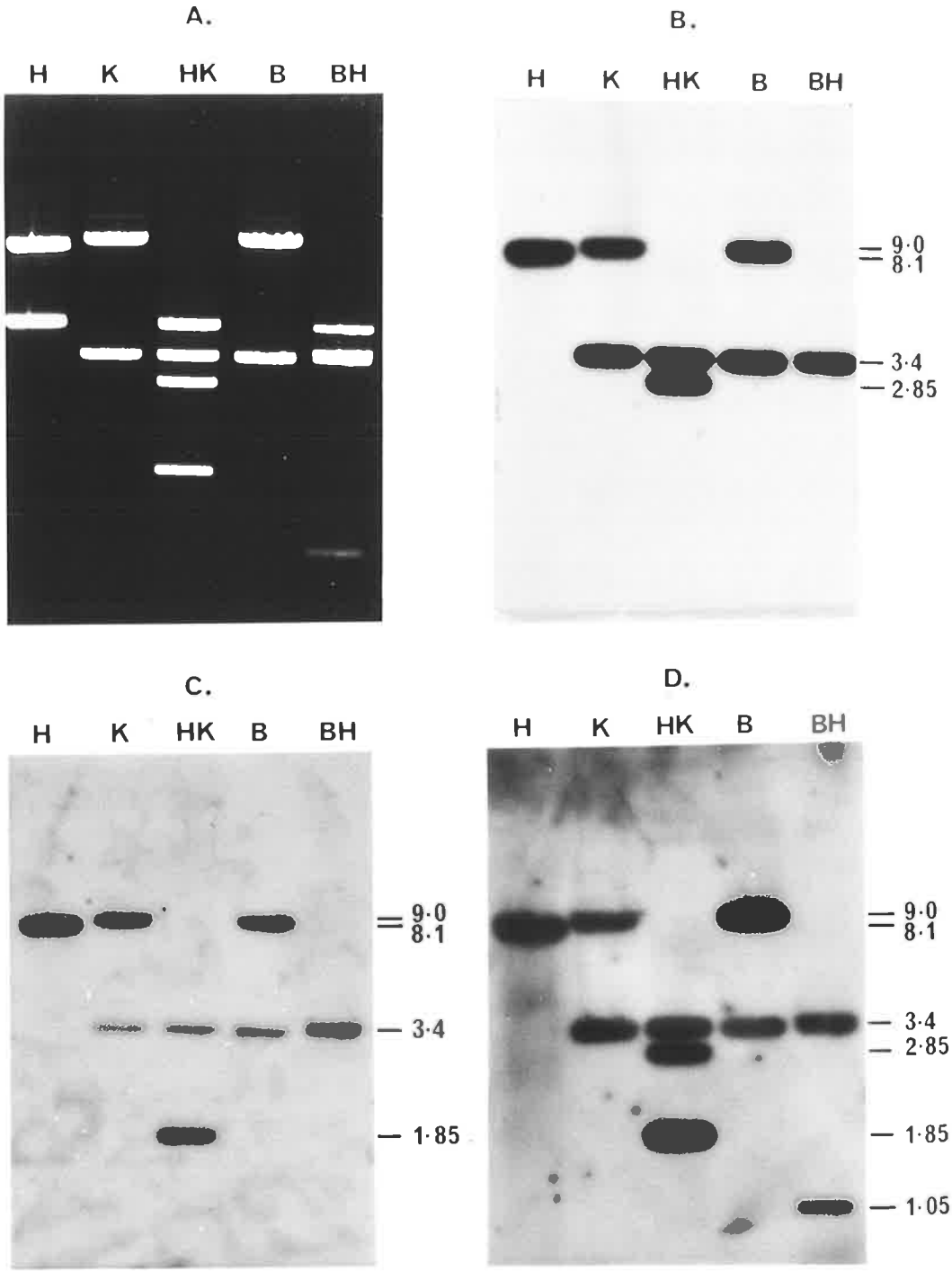
C. The intron probe (19-mer)

D. The 5' non-coding probe (20-mer).

The filters were washed in 4 x SSC at 65°C and exposed to X-ray film at -80°C. The sizes of the fragments which were detected by each of the probes are shown.

E. Location of the feather keratin genes in pH 8.1. The restriction mapping of pH 8.1 is described in Appendix C. The sizes, in kilobases, of the restriction fragments are shown with BglII and BglII/HindIII fragment sizes at the lower level and KpnI and KpnI/HindIII sizes at the upper level. The arrangement of the genes are shown in option (1) (filled-in arrows). An alternative possible arrangement (option 2, open arrows) was ruled out (see text for details). The part of the pH 8.1 insert which is also present in the clone λCFK9 is shown. Restriction enzymes used were:

B - BglII
H - HindIII
K - KpnI


A.
H
K
HK
B
BH
B.
H
K
HK
B
BH
9·0
8·1
3·4
2·85
C.
H
K
HK
B
BH
9·0
8·1
3·4
1·85
D.
H
K
HK
B
BH
9·0
8·1
3·4
2·85
1·85
1·05

(Figure 4.36e). This was later confirmed by DNA sequencing (see Section 5.3.3).

The 2.85 kb KpnI/HindIII fragment hybridized with the 5' and 3' non-coding probes but not the intron probe. This hybridization data suggested that either **(1)** the 2.85 kb fragment contained the 5' end of one feather keratin gene and the 3' end of another gene or **(2)** it contained the 3' end of one gene and an additional feather keratin gene whose intron differed substantially from the intron probe. [Options (1) and (2) are shown in Figure 4.36e]. **[N.B.** The possibility that the 2.85 kb KpnI/HindIII fragment contains just <u>one</u> gene (option (2) in Figure 4.36e) with a divergent intron was ruled out by DNA sequence analysis from the KpnI site which showed that a feather keratin gene spanned the KpnI site at a position just upstream of the translation 'stop' codon (data not shown)]. If option (2) is correct, one would predict that the signal intensity of the 2.85 kb KpnI/HindIII fragment when probed with the 3' non-coding probe ought to be twice that of the 3.4 kb KpnI fragment (see Figure 4.36e). Figure 4.36b shows that this is not the case; the intensity of the hybridizing bands at 3.4 and 2.85 kb is similar. This indicates that the gene arrangement in option (a) (shaded gene 'boxes') is probably correct.

The demonstration that gene E of λCFK9 spans the HindIII site of the 3.4 kb fragment (see Section 4.3.1, Figure 4.3) and that the 8.1 kb HindIII fragment maps next to the 3.4 kb HindIII fragment in cosmid 4 (Section 4.3.2.1b) is further proof that the gene map shown in option (1) of Figure 4.36e is correct. Thus, the 1.05 kb BglII/HindIII fragment contains the

5' non-coding region and part of the intron of a feather keratin gene (that is, gene E of λCFK9 (Figure 4.3, 4.36e)).

The final restriction map of pH8.1 is shown in Figure 4.36e (option 1) indicating the position and orientation of the two complete genes. These genes, and the 5' end of the third gene located at the left-hand end of the insert are orientated in the same direction as the genes in λCFK1 (right to left), since the 1.05 kb BglII/HindIII fragment is present in both cosmid 4 and λCFK9 (see Figure 4.8).

B. Characterization of a recombinant containing the 2.7 kb HindIII fragment (pH 2.7). This subclone was mapped in a similar fashion as pH 8.1 with BglII and KpnI. The 2.7 kb HindIII insert contained one BglII and one KpnI site (Figure 4.37a). By hybridization analysis with radiolabelled oligonucleotides representing 3' and 5' non-coding regions (Figure 4.37b,c), it was established that this fragment contained only one gene, the 5' end of which is contained in the 1.0 kb KpnI/HindIII fragment (Figure 4.36b,d) and the 3' end of which is contained in the 1.7 kb KpnI/HindIII fragment (Figure 4.36 c,d). Since only the 1.27 kb BglII/HindIII fragment was detected by both probes, the gene must be almost entirely contained in this fragment. Therefore, the 1.27 kb BglII/HindIII fragment must be located at the right-hand end of the insert, as shown in Figure 4.37d.

C. Gene content of the 6.4 kb HindIII fragment. Several attempts to subclone the 6.4 kb HindIII fragment into pBR322 were unsuccessful. All three of the feather gene-containing HindIII fragments (of 8.1, 6.4 and 2.7 kb) were subcloned at the same time. In the case of the 6.4 kb fragment a large

FIGURE 4.37 : RESTRICTION AND SOUTHERN ANALYSIS OF pH 2.7

A. pH 2.7 DNA was digested with either BglII/HindIII or KpnI/HindIII, electrophoresed on 1% agarose and visualized under UV light after ethidium bromide staining (Section 2.2.7). Restriction fragment sizes are shown in kilobases.

After performing a bidirectional Southern transfer on the gel (Section 2.2.9), the filter-bound DNA was annealed with the following oligonucleotide probes:

B. The 5' non-coding oligomer.

C. The 3' non-coding oligomer (Section 2.2.10). Filters were processed as described in Figure 4.36.

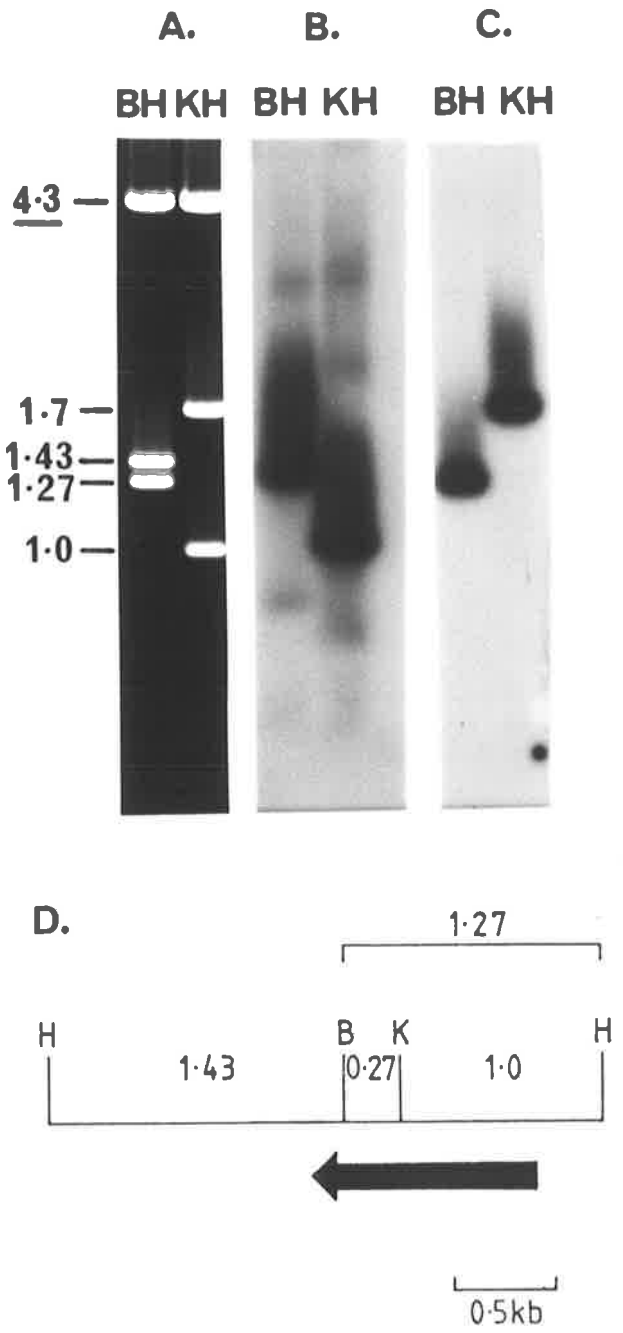
D. Restriction map of pH 2.7 showing the location of the single feather keratin gene. The position of the KpnI site in the gene was determined by DNA sequence analysis (Section 5.3.3.). The size of the restriction fragments are given.

Restriction enzymes are indicated as follows:

B - BglII

H - HindIII

K - KpnI

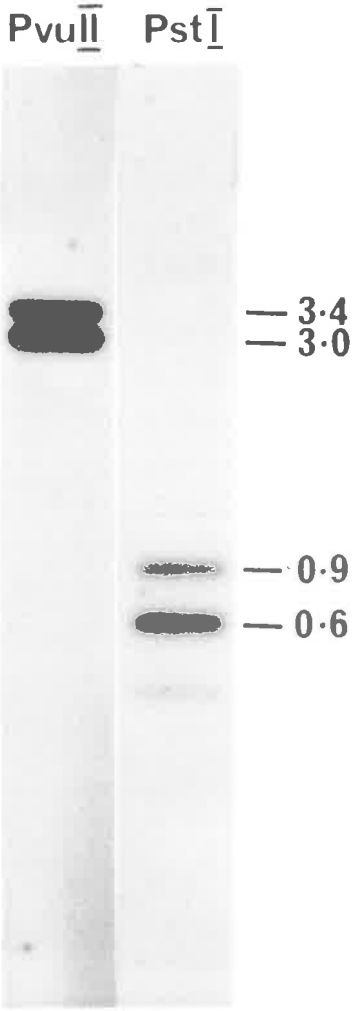

A.
B.
C.
BH KH
BH KH
BH KH
4·3
1·7
1·43
1·27
1·0
D.
1·27
H
B K
H
1·43
0·27
1·0
0·5kb

number of non-resectable recombinants were generated, an example of which is shown in Figure 4.35b. It can be seen that although the clone is harbouring an insert, it is not resectable with HindIII. These non-resectable subclones gave a positive signal when analyzed by Grunstein colony hybridization with the feather keratin gene probe (data not shown). Therefore they appeared to contain part of the 6.4 kb fragment but were not the correct size. Similar results were obtained when the E. coli strain HB101 (recA) was used as the host bacterium (Section 2.1.2). Therefore, an alternative approach was used to determine the number of genes in this fragment.

The 6.4 kb HindIII fragment was isolated from a HindIII digest of cosmid 4 and the fragment cut with PvuII or PstI. The resulting products were transferred to nitrocellulose and hybridized to the 25 base feather coding oligonucleotide (see Section 2.2.10). Digestion of the 6.4 kb HindIII fragment with PvuII generated two fragments of 3.4 and 3.0 kb, both of which hybridized with the 25-mer (Figure 4.38). Digestion of the 6.4 kb HindIII fragment with PstI generated about 10 fragments (data not shown), two of which hybridized with the oligonucleotide probe (Figure 4.38). The two PvuII/HindIII fragments also hybridized to the intron oligonucleotide (see Section 2.2.10, data not shown). Taken together, these results indicate that this fragment contains two feather keratin genes. The orientation and exact location of these genes has not been determined.

FIGURE 4.38 : DETERMINATION OF THE GENE CONTENT OF THE 6.4 KB HindIII FRAGMENT OF COSMID 4

The 6.4 kb HindIII fragment was isolated from a HindIII digest of cosmid 4 (Section 2.2.8), the DNA fragment restricted with either PvuII or PstI and electrophoresed on a 1.4% agarose gel (Section 2.2.7). The products of the double digests were transferred to nitrocellulose (Section 2.2.9) and the filter bound DNA hybridized with a 25-mer designed from the coding region of the feather keratin genes (see Section 2.2.10). The sizes of the restriction fragments in the HindIII/PvuII and HindIII/PstI digests of the purified fragment which hybridize with the probe are shown.


PvuII
PstI
3·4
3·0
0·9
0·6

D. Characterization of recombinants containing the 8.4 kb EcoRI and 2.5 kb BamHI/HindIII fragments (pE8.4 and pBH2.5).

Previous restriction mapping and Southern blot data from cosmid 4 indicated that the 8.4 kb EcoRI fragment contained part of λCFK1 and extended it leftwards (see Section 4.3.2.1a, Figure 4.7a). From these results, the 8.4 kb EcoRI fragment contains the 0.95 kb EcoRI/HindIII, the 2.15 and 1.3 kb HindIII fragments present in both λCFK1 and cosmid 4, and a 4.0 kb EcoRI/HindIII subfragment of the 7.8 kb HindIII fragment of cosmid 4 (Figures 4.7, 4.13) Of these fragments, the 4.0 kb EcoRI/HindIII and the 2.15 kb HindIII fragments contain feather keratin genes (see Figures 4.4b and 4.31), the latter being the same fragment which is present in λCFK1 and contains gene A (Figure 4.3). In order to confirm these results, the 8.4 kb EcoRI fragment was subcloned into pBR322.

Subsequent restriction and blot analysis of pE8.4 confirmed the previous results (Figure 4.39). As predicted above, an EcoRI/HindIII digest of pE8.4 produced insert fragments of sizes 4.0, 2.15, 1.3 and 0.95 kb, of which the 4.0 and 2.15 kb fragments were detected by the feather gene probe (Figure 4.38b). The hybridization of the 4.0 kb EcoRI/HindIII fragment with the feather gene probe confirmed that in cosmid 4, there is at least one gene beyond gene A of λCFK1. The pE8.4 clone also contains a BamHI site 1.5 kb from the left-hand end of the fragment (Figure 4.39), consistent with the map of this region of cosmid 4 (Figure 4.13). Since the 1.5 kb BamHI/EcoRI fragment does not hybridize with the probe, the gene is completely contained in the remaining 6.9 kb BamHI/EcoRI fragment. The restriction map of pE8.4 and organization of the two genes within it is shown in Figure 4.39c.

FIGURE 4.39 : RESTRICTION AND SOUTHERN ANALYSIS OF pE8.4

A. pE8.4 DNA was digested with a number of restriction enzymes, fractionated on 1.2% agarose, stained with ethidium bromide and visualized under UV light (Section 2.2.7).

B. The DNA fragments were transferred on to a nitrocellulose filter (Section 2.2.9) and hybridized with the feather keratin gene probe (Section 2.2.10). The filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and autoradiographed overnight at -80°C. The sizes (kb) of the restriction fragments generated from the pE8.4 insert, as well as the 4.3 and 0.4 kb fragments from pBR322 (underlined), are shown.

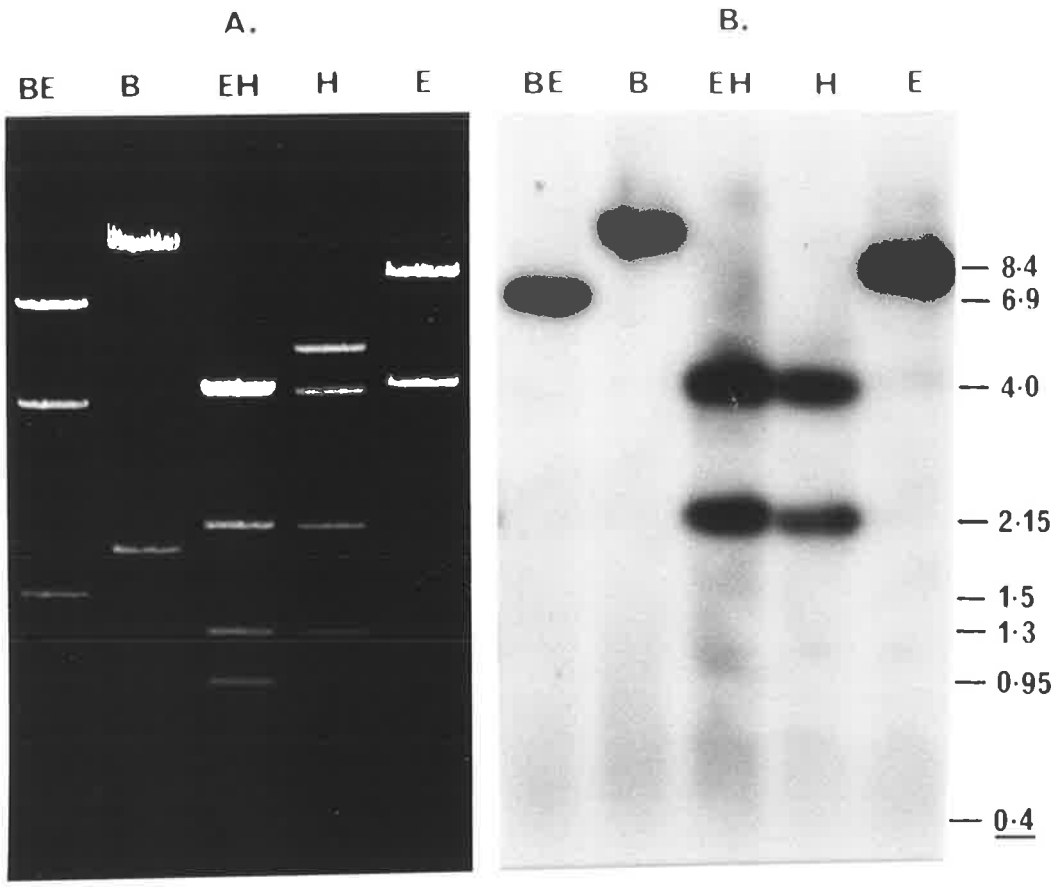
C. Restriction map of pE8.4 showing arrangement of the two feather keratin genes. The sizes of the restriction fragments are included. The position and orientation of the gene in the 2.15 kb HindIII fragment was determined by DNA sequencing (Gregg et al., manuscript in preparation). Similarly, the gene in the 2.5 kb BamHI/HindIII fragment was characterized by subcloning and sequence analysis (see text).

The restriction enzymes used were:

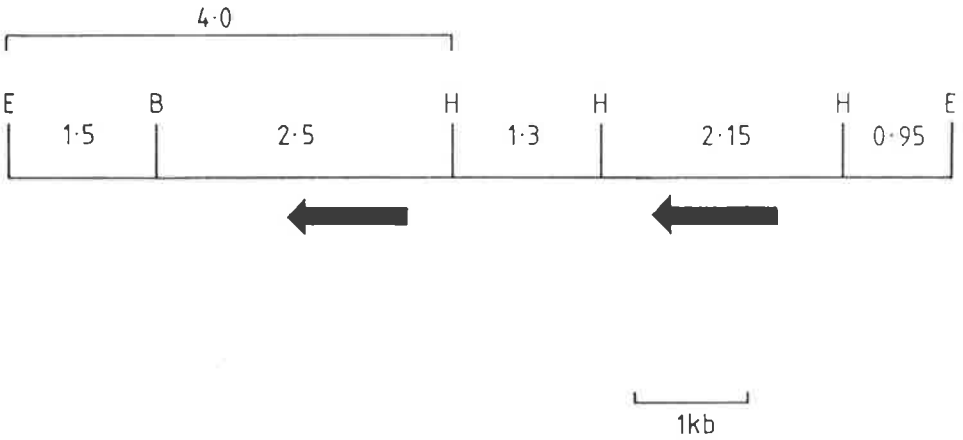
B - BamHI

E - EcoRI

H - HindIII


A.
B.
BE
B
EH
H
E
BE
B
EH
H
E
8·4
6·9
4·0
2·15
1·5
1·3
0·95
0·4

C.


4·0
E
B
H
H
H
E
1·5
2·5
1·3
2·15
0·95
1kb

The location and orientation of the gene in the 4.0 kb EcoRI/HindIII fragment was determined by subcloning the 2.5 kb BamHI/HindIII fragment into a pBR322 vector. The cloned 2.5 kb BamHI/HindIII fragment, designated pBH2.5, was digested with BamHI, HindIII and PstI or PvuII and the products analyzed by agarose gel electrophoresis (Figure 4.40a). The insert of the subclone contained one PvuII and two PstI sites and there are single PvuII and PstI sites in the pBR322 vector (Sutcliffe, 1978). The orientation of the PstI sites with respect to the BamHI and HindIII sites was determined by restricting the subclone with PstI, PstI/HindIII and all three enzymes and analyzing the products on an agarose gel (Figure 4.40a and other data not shown). This enabled the restriction map shown in Figure 4.40d to be constructed, the details of which are described in the figure legend.

Blot analysis with a feather keratin gene probe (Figure 4.40b) indicated that the gene was contained in the 0.45 kb PstI/HindIII and 1.1 kb PstI fragments. The 0.95 kb PstI/BamHI fragment did not contain any keratin sequences. Hybridization of an M13 clone, which contains the 5' non-coding exon and flanking regions of feather keratin gene B (Section 2.2.10), to digests of pBH2.5 indicated that the 5' end of the pBH2.5 gene is contained in the 0.45 kb PstI/ HindIII and 0.8 kb PvuII/HindIII fragments (Figure 4.40c). These studies enabled the location and orientation of the gene in this subclone to be determined. The exact position of the gene in the pBH2.5 subclone was determined by DNA sequencing which is described in Section 5.3.5.

FIGURE 4.40 : RESTRICTION AND SOUTHERN ANALYSIS OF pBH2.5

A. pBH2.5 DNA was digested with a number of restriction enzymes, electrophoresed in 1.1% agarose, stained with ethidium bromide and the DNA viewed under UV light (Section 2.2.7).

The digested DNA was transferred bidirectionally to nitrocellulose (Section 2.2.9) and the two filters probed with:

B. The feather keratin gene probe (Section 2.2.10).

C. An M13 clone containing the 5' non-coding exon and 5' flanking regions of gene B (Section 2.2.10).

The sizes of the insert DNA fragments and the pBR322 vector fragment (underlined) are shown.

Filters were washed in 0.5 x SSC, 0.1% SDS at 65°C and autoradiographed at -80°C.

The following restriction enzymes were used :

Track 1 - BamHI, HindIII

Track 2 - BamHI, HindIII, PvuII

Track 3 - BamHI, HindIII, PstI

D. Restriction map of pBH2.5 showing the location of the feather keratin gene. The map was derived as follows. From the blot data (Figure 4.40b,c), the 5' end of the gene was located in the 0.45 kb PstI/HindIII and 0.8 kb PvuII/HindIII fragments.

The 0.45 kb PstI/HindIII fragment was present in a PstI/HindIII digest but not a BamHI/PstI digest, locating it at the right end of the insert (Figure 4.40d). The 1.1 kb PstI fragment was present in all digests containing PstI and either HindIII or BamHI. The 0.95 kb PstI/BamHI fragment was not present in a PstI or PstI/HindIII digest, indicating that this fragment was located at the left of the insert (Figure 4.40d).

The PvuII site must be located 0.8 kb from the HindIII site, as the 0.8 kb PvuII/HindIII fragment was detected by the 5' end probe (Figure 4.40c).

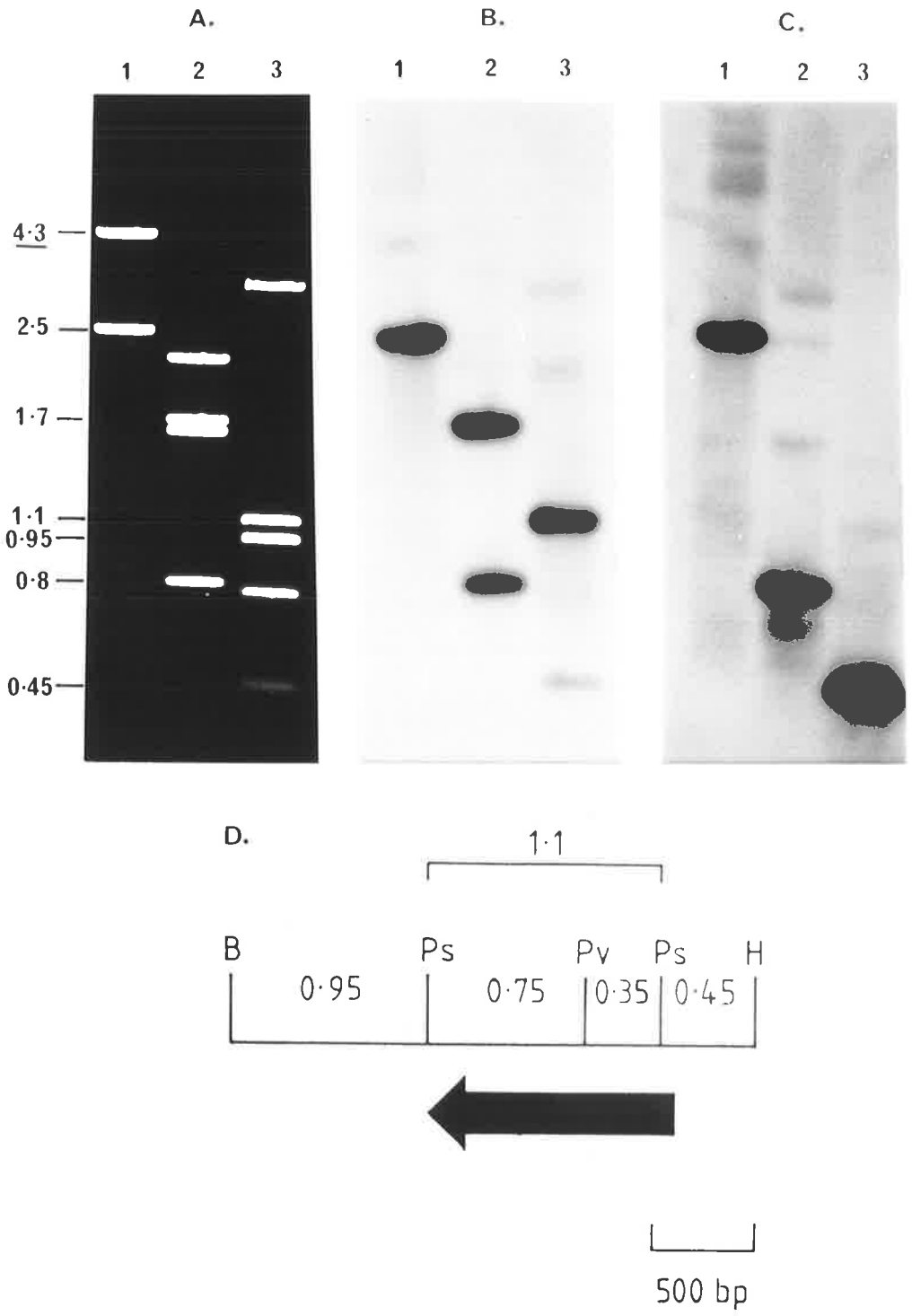
The sizes (kb) of all the relevant restriction fragments are shown. The symbols refer to the following restriction enzymes:

B - BamHI

H - HindIII

Ps - PstI

Pv - PvuII


A.
B.
C.
1
2
3
4·3
2·5
1·7
1·1
0·95
0·8
0·45
D.
1·1
B
Ps
Pv
Ps
H
0·95
0·75
0·35
0·45
500 bp

As the 5' end of the pBH2.5 gene is adjacent to the 1.3 kb HindIII fragment of cosmid 4 (Figure 4.39 and 4.40), the gene in the 2.5 kb BamHI/HindIII fragment, which comprises the left-hand end gene of the feather keratin cluster, is orientated in the same direction as the genes in λCFK1.

4.3.5. PRELIMINARY CHARACTERIZATION OF THE COSMID CLONES WHICH DO NOT OVERLAP WITH THIS KERATIN GENE LOCUS

By the criteria of restriction digestion and Southern hybridization, only three of the 15 cosmid recombinants isolated from the chick library did not appear to be derived from the genomic region defined by the cosmid clones 12, 4, 31 and 33 (see Section 4.4.1, Table 4.2). Two of these cosmids (1 and 5) were examined briefly and are described in this section.

4.3.5.1 Cosmid 5

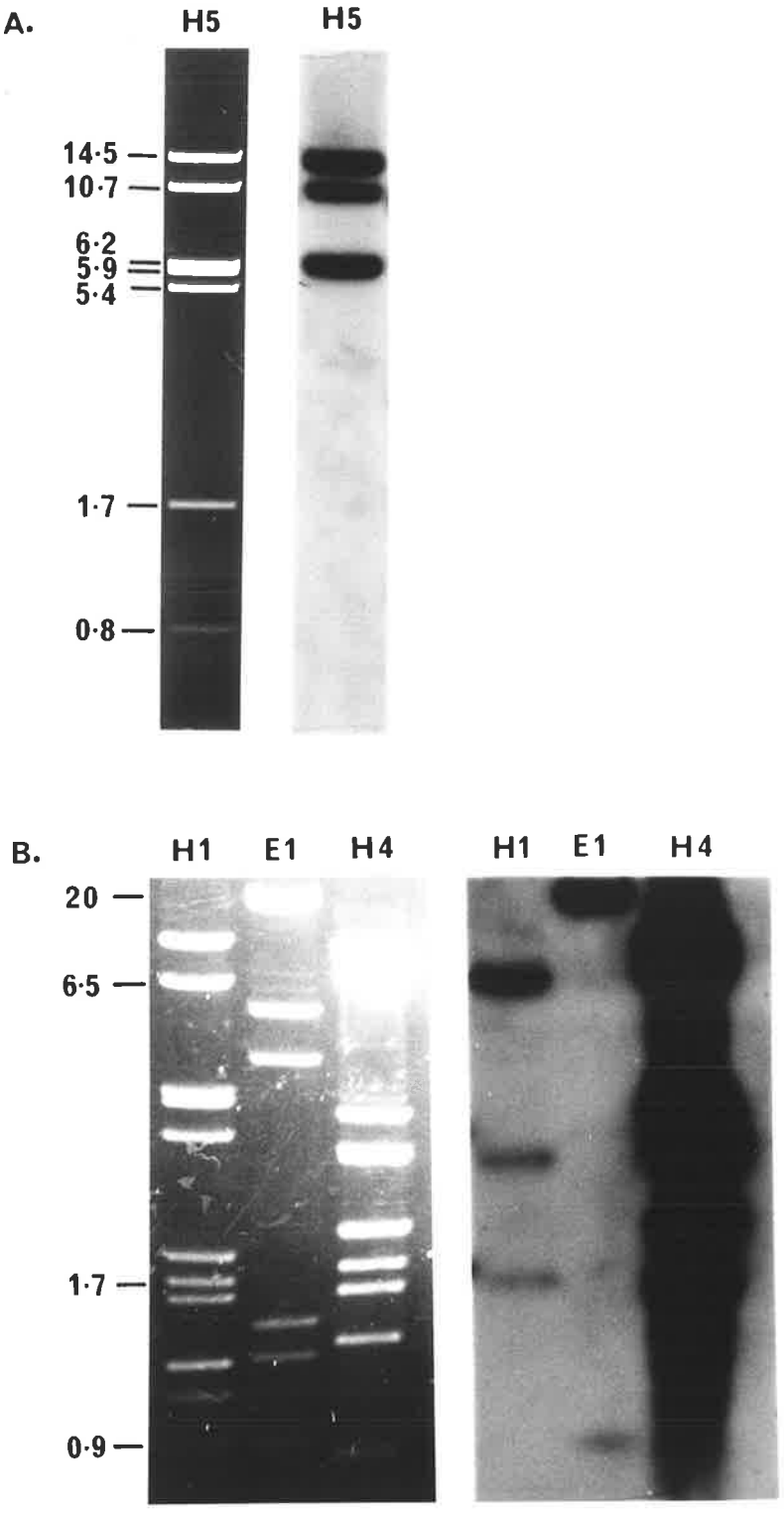
Previously, in dot blot hybridizations, cosmid 5 was found to bind the feather and scale keratin probes equally as well but failed to be detected at all with the feather-specific oligonucleotide or λCFK1-specific probes (see Section 3.3.3, Figure 3.9). A HindIII digest of this clone generate a relatively simple pattern of 7 restriction fragments and subsequent Southern blot hybridization with a scale keratin gene probe detected three bands (Figure 4.41a). The exact nature of the keratin genes in cosmid 5 is at present not known, although the observation that this cosmid did not hybridize with the feather-specific probe (Figure 3.9c) suggests that the keratin genes contained in this cosmid may not be of the feather type.

FIGURE 4.41 : RESTRICTION AND SOUTHERN ANALYSES OF COSMID 5 AND 1

Cosmid 5 and 1 DNAs were cleaved with restriction enzymes, fractionated on a 1% agarose gel, stained with ethidium bromide and visualized under UV light (Section 2.2.7). The cosmid DNA fragments were transferred to nitrocellulose (Section 2.2.9) and hybridized with keratin probes as described below.

A. Restriction digestion of cosmid 5 with HindIII and Southern analysis with a scale keratin gene probe (Section 2.2.10). The filter was washed in 1 x SSC, 0.1% SDS at 65°C and exposed to X-ray film for 6 hours at -80°C. The sizes, in kilobases, of all the HindIII fragments are shown.

B. Restriction digest of cosmid 1 with EcoRI and HindIII and Southern analysis with the feather keratin gene probe (Section 2.2.10). Following hybridization, the filter was washed in 1 x SSC, 0.1% SDS at 65°C and exposed to X-ray film for 3 days at -80°C. The sizes (kb) of the weakly hybridizing fragments in cosmid 1 are given. A HindIII digest of cosmid 4 was included for comparison.


A.
H5
H5
14·5
10·7
6·2
5·9
5·4
1·7
0·8
B.
H1
E1
H4
H1
E1
H4
20
6·5
1·7
0·9

4.3.5.2 Cosmid 1

Preliminary characterization of cosmid 1 (Section 3.3.3, Figure 3.9) suggested that it contains keratin genes with only poor homology to feather keratins, since it hybridized only weakly with the feather keratin gene probe and did not anneal the feather-specific or λCFK1-specific probes. These previous results were confirmed by hybridization analysis with a feather keratin gene probe. Cosmid 1 hybridizes weakly to the feather keratin probe as can be seen in Figure 4.41b. Two EcoRI (~ 20 and 0.9 kb) and two HindIII (6.5 and 1.7 kb) fragments hybridized with the feather gene probe (Figure 4.41b). The third band visible in the HindIII track did not correspond to a visible DNA fragment but may represent a partial digestion product. No further hybridization or mapping studies were carried out on this recombinant. The nature of the presumptive keratin genes contained in this recombinant are not known (see Discussion, Section 4.4.3).

4.3.6 ANALYSIS OF COSMID CLONES WITH A cDNA PROBE

In order to determine if the keratin gene locus described in this study contained any other gene(s) such as fast protein genes which are expressed in the developing feather, ^{32}P-labelled cDNA was prepared from 14-day feather poly A^+ RNA and used to probe recombinants representative of most of the chromosomal segment described in this study (Figure 4.45), that is cosmids 12, 4 and 33. As the feather keratin genes (either by sequence cross-hybridization and/or because they are expressed in the developing feather) will be detected by the cDNA probe, it was only possible to determine if any expressed genes are present in restriction fragments which did not contain feather (or perhaps other) keratin genes.

FIGURE 4.42 : SOUTHERN ANALYSIS OF COSMIDS 4 AND 33 WITH A cDNA PROBE

A Southern filter containing HindIII digests of cosmids 4 and 33 (see Figure 4.26) was hybridized with a cDNA probe, prepared from embryonic feather mRNA as described in Section 2.2.11.1. Following hybridization, the filter was washed in 0.5 x SSC , 0.1% SDS at 65°C and autoradiographed for 1 day at -80°C.

The sizes of all the fragments detected by the probe are shown (see Figure 4.26).

4 3 3

The results of some of these Southern blots are presented in Figure 4.42. No additional restriction fragments other than those which contain feather keratin genes were observed, although two of the claw gene-containing fragments (4.65 and 2.4 kb) were visible on a longer exposure. In an EcoRI blot of cosmid 12, no fragments other than those which contain feather keratin genes were detected except for the 5.0 kb EcoRI fragment of cosmid 12, which gave a weakly positive hybridization signal (data not shown). The hybridization of the cDNA probe to the feather-like gene in the 5.0 kb EcoRI fragment may be a result of cross-hybridization to keratin mRNA sequences in the probe or because the feather-like gene is expressed in feather tissue.

4.3.7 GENOMIC SOUTHERN ANALYSIS OF CHICKEN DNA WITH THE FEATHER KERATIN GENE PROBE

A. HindIII blot. Figure 4.43 shows the pattern of restriction fragments containing feather keratin genes in chicken DNA digested with HindIII compared to cosmid 25, included for comparison. The sizes of all of the hybridizing bands are indicated. A total of about 21-22 hybridizing restriction fragments were estimated by densitometric scanning to be present in chicken DNA (Table 4.1). Table 4.1 compares the sizes expected for the fragments (predicted from the mapping data) to the observed sizes obtained from Southern analysis of chicken genomic DNA (Figure 4.43). It can be seen that the feather keratin and feather-like genes contained in cosmids 4, 12, 31 and 33 probably account for 17 of the 21-22 HindIII fragments detected in restricted chick DNA (Table 4.1). Of the remainder, the 14 and 5.9 kb HindIII fragments have been tenta-

FIGURE 4.43 : GENOMIC SOUTHERN ANALYSIS OF HindIII-DIGESTED CHICKEN DNA WITH THE FEATHER GENE PROBE

A. A 20 µg sample of two chicken genomic DNA preparations (Section 2.1.10) were digested to completion with HindIII and electrophoresed on a 1% agarose gel. The genomic DNA was visualized under UV light after ethidium bromide staining.

B. The restricted genomic DNA was transferred to a nitrocellulose filter (Section 2.2.9) and was hybridized with the feather keratin gene probe (Section 2.2.10). The filter was washed in 1 x SSC, 0.1% SDS at 65°C and autoradiographed for 7 days at -80°C. Cosmid 25, which contains part of the feather gene cluster (see Section 4.3.2.3) was run alongside the chick DNA for comparison and to provide molecular weight markers. The sizes of the restriction fragments detected by the probe are shown. The autoradiograph was scanned using a Zeineh Soft Laser Scanning densitometer. These results are summarized in Table 4.1 alongside their likely counterparts in the mapped cosmid clones.

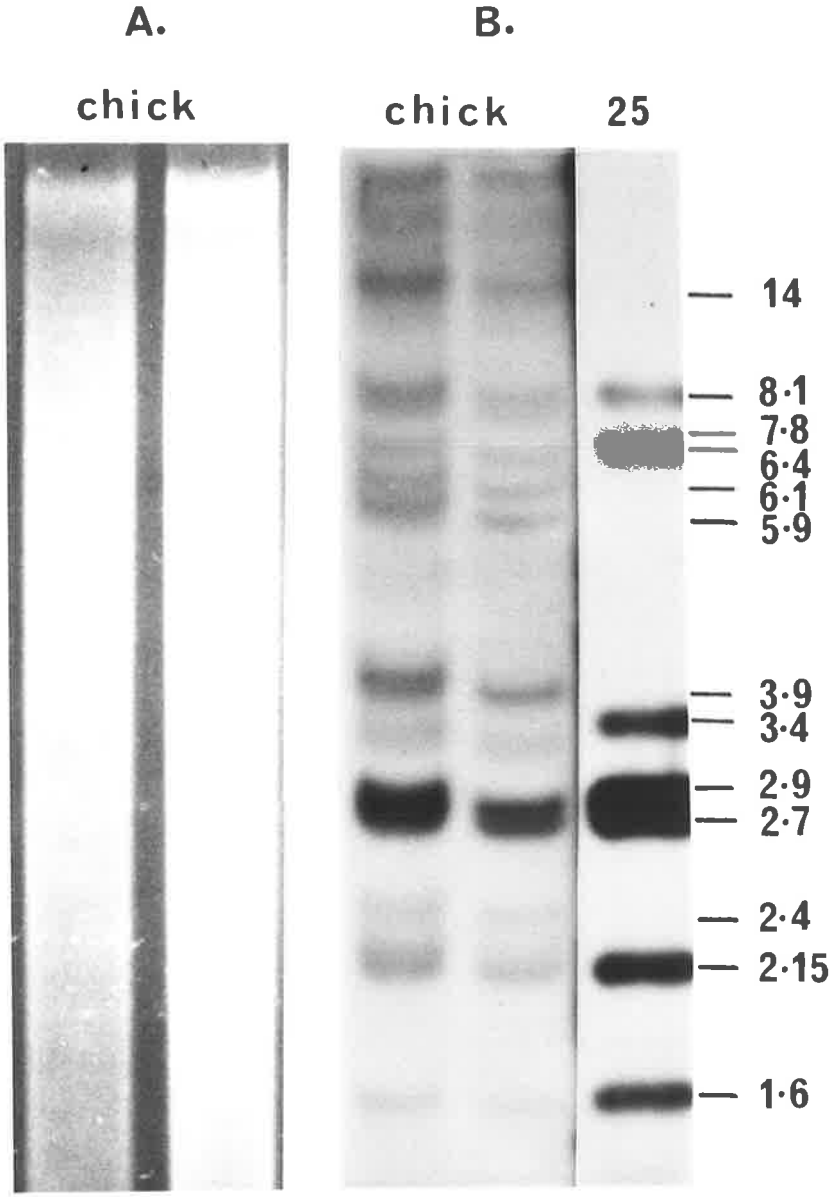

A.
chick
B.
chick
25
14
8·1
7·8
6·4
6·1
5·9
3·9
3·4
2·9
2·7
2·4
2·15
1·6

TABLE 4.1 : COMPARISON OF THE SIZES OF THE FEATHER KERATIN GENE-CONTAINING HINDIII RESTRICTION FRAGMENTS IN CHICKEN DNA AND THE COSMID CLONES

The table compares the sizes of the hybridizing fragments detected in chick DNA (Figure 4.43) with those obtained in the cosmid clones characterized in this chapter. The total number of restriction fragments in the genomic Southern blot was determined by densitometric scanning of the autoradiograph using a Zeineh soft laser densitometer. The 5 cosmid recombinants (4, 12, 25, 31 and 33) characterized in detail probably account for 17 of the 21-22 hybridizing fragments detected in chicken DNA. Two of the remaining four or five fragments detected in chick DNA Have counterparts of a similar size in cosmid 5. Cosmid 5 does not overlap with the characterized keratin locus (Figure 4.45) and may not contain keratin genes of the feather type (see Section 4.3.5.1).

FRAGMENT SIZE (kb) FROM		
CHICKEN GENOMIC SOUTHERN	MAPPING OF CLONES	COSMID(S)
14.0	14.6	5
8.1	8.1	4, 25
7.8	7.8	4, 12
6.4	6.4	4, 25
6.1	6.2	31, 33
5.9	5.9	5
3.9	4.5	12
3.9	-	-
3.4	3.4	4, 12, 25
2.9	2.9	4, 12, 25
2.9	2.9	31, 33
2.9[1]	-	-
2.7 x 5	2.7	31 (all); 4, 25, 33 (one or more)
2.4	2.3	12
2.15	2.15	4, 12, 25
2.15	2.15	4, 12
1.6	1.6	4, 12, 25

[1] May be two fragments of 2.9 kb which cannot be accounted for from the cosmid clones.

tively assigned to cosmid 5 (Table 4.1). The remaining 2-3 hybridizing bands of 2.9 and 3.9 kb detected in chicken DNA which had no counterparts in the cosmid clones may be contained in cosmid 35, which is currently being mapped in this laboratory (see Section 4.4.3).

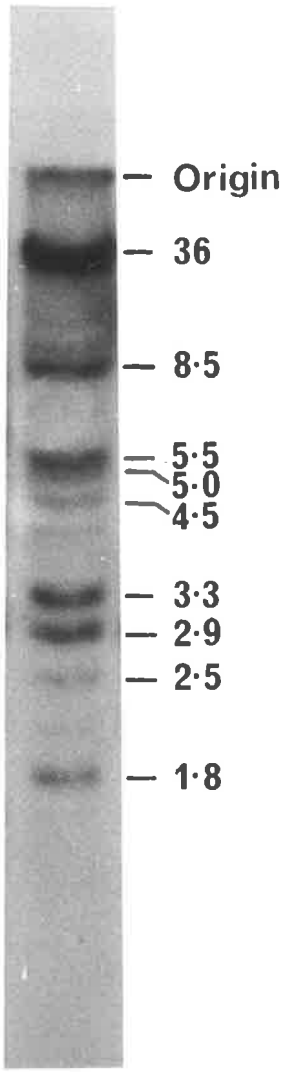
B. EcoRI blot. The genomic Southern analysis of EcoRI digested-chick DNA using a feather keratin gene as probe gave a much simpler pattern than that obtained above with HindIII (Figure 4.44). A total of 6 strong and 5 weak bands were detected, some of which are indicated on the figure. Five of the 6 strongly hybridizing fragments (36, 8.4, 3.3, 2.9 and 1.8 kb) and one of the fragments detected weakly (5.0 kb) could be accounted for from the mapped locus containing feather keratin and a feather-like gene. The 1.1 kb EcoRI fragment which contains one of the feather-like genes is apparently not detected in this genomic blot.

The size of the large (36 kb) EcoRI restriction fragment, which hybridized strongly to the feather gene probe, is consistent with the calculated size (39 kb) of this EcoRI fragment estimated from the mapping data. (A 10-15% error is not unexpected given the large size of this hybridizing fragment). This large EcoRI fragment spans the right half of cosmid 4 and most of cosmid 31 and confirms the linkage of these clones (see Sections 4.3.3.1 and 4.4.1).

The weakly hybridizing EcoRI fragments detected in chick DNA presumably represent 'variant' feather keratin genes, or cross-hybridizing keratin genes from a related family, such as scale (Section 1.4.6, Figure 1.9a).

FIGURE 4.44 : GENOMIC SOUTHERN ANALYSIS OF EcoRI-DIGESTED CHICKEN DNA WITH THE FEATHER GENE PROBE

Chicken DNA (20 µg) was digested with EcoRI, blotted on to nitrocellulose and hybridized with the feather keratin gene probe as described in Figure 4.43. The sizes of all of the strong and some of the weakly hybridizing fragments detected in the EcoRI-digested chicken DNA are shown.


Origin
36
8·5
5·5
5·0
4·5
3·3
2·9
2·5
1·8

4.4 DISCUSSION

4.4.1 RESTRICTION MAPPING OF THE λ AND COSMID RECOMBINANTS

This chapter described the characterization of one λ clone and a number of cosmid recombinants containing chicken keratin genes. Of the 15 cosmid recombinants isolated using a feather keratin gene as probe, 12 of them (and the χCFK9 clone) were found to be derived from a single chromosomal region spanning 115 kb (Table 4.2 and Figure 4.45). The genomic insert contained in the previously described clone χCFK1 (Molloy et al., 1982) is also derived from this region.

Of the four λ recombinants isolated, two were found to be identical to χCFK1 and one was judged to be a rearrangement of the χCFK1 segment (Section 4.3.1). Internal recombination events have been observed previously in chicken keratin genomic clones (Molloy et al., 1982) as well as in studies of other multigene families where there are repeated regions of homology (Adams et al., 1980; Lauer et al., 1980). The fourth clone χCFK9 extended χCFK1 rightwards by 2 kb and was useful in the subsequent mapping of cosmid clones extending this region.

The three cosmids which hybridized with the χCFK1-specific probe were mapped with several restriction enzymes (Section 4.3.2). One of these, cosmid 4, was found to encompass the genomic region contained in the two λ recombinants and extended them in both directions (Figure 4.13). Not unexpectedly, the two other recombinants which hybridized with the χCFK1-specific probe, cosmids 12 and 25, overlapped with cosmid 4 and extended it leftwards and rightwards, respectively (Figure 4.17 and Table 4.2). Cosmids 4 and 12 contained the left portion of the feather keratin gene cluster

and the subsequent characterization of cosmids 31 and 33 (Section 4.3.3) enabled the right-hand end of the feather gene cluster to be defined. The following observations lead to the conclusion that cosmid 31 overlapped the right end of cosmid 4 and the left end of cosmid 33 as shown in Figure 4.45.
(1) The similarities in the restriction patterns of cosmid 31 and cosmids 4 and 33 (Section 4.3.3.1). **(2)** The genomic Southern blots of chick DNA using a feather keratin gene as probe. The size of the large, strongly hybridizing EcoRI fragment detected in chicken genomic DNA (36 kb) was consistent with the predicted size of the EcoRI fragment which spans the right half of cosmid 4 and the left half of cosmid 31 (39 kb, Section 4.3.7b). Furthermore, from the densitometric scanning of a Southern blot of HindIII-digested chick DNA, it was clear that there were only about 5 2.7 kb HindIII fragments in the genome which contained feather keratin genes (Section 4.3.7a), all of which are present in cosmid 31 (Figure 4.25).
(3) Southern analysis of cosmid 4 using the left-hand end fragment of cosmid 31 as a probe (Section 4.3.3.1, Figures 4.24 and 4.25). Both the restriction enzyme and Southern blot data indicate that the cosmid 31 insert terminates in the 6.4 kb HindIII fragment of cosmid 4. The strong hybridization of the probe from cosmid 31 to the 8.1, 6.4 and 3.4 kb HindIII fragments, and the 3.4 kb KpnI fragments, of cosmid 4 is indicative of the close similarity of the genes and intergenic regions in this DNA segment.

Table 4.2 summarizes the data obtained from all of the cosmid recombinants. All of the recombinants except 1, 5 and 35 were derived from the chromosomal segment shown in

TABLE 4.2 : NATURE AND CODING POTENTIAL OF THE CHICK KERATIN COSMID RECOMBINANTS

The table summarizes the data obtained from the 15 cosmid recombinants isolated using a feather keratin gene as probe (Sections 3.3.2, 3.3.3).

Some of the cosmid clones were characterized in detail (Chapter 4), while others were only partially examined and the data was not presented (indicated as 'data not shown').

Twelve of the 15 recombinants were derived from the 100 kb region depicted in Figure 4.45. The nature and number of the genes contained in each recombinant is indicated.

COSMID	GENERAL DESCRIPTION	KERATIN GENE CONTENT	REFERENCE
1	Contains keratin genes of unknown identity (not feather). Does not overlap with the mapped keratin locus[A].	Beak keratins??	Section 4.3.5.2
2	Extends cosmid 12 leftwards by 15 kb. Contains no keratin genes additional to those present in the cosmid 12 insert.	Two feather, three feather-like	Data not shown
3	Identical to cosmid 4, but grows very poorly in liquid culture.	Eleven feather, one feather-like	Section 3.3.3 and 4.3.2.1
4	Completely encompasses the two λ recombinants and extends them in both directions.	Eleven feather, one feather-like	Section 4.3.2.1
5	Contains keratin genes of unknown identity (probably not feather). Does not overlap with the mapped keratin locus[A].	Scale keratins??	Section 4.3.5.1
11	A rearrangement of part of the mapped keratin locus[A]. Contains part of cosmids 4 and 31.	Feather keratins	Data not shown
12	Contains the whole of the ΧCFK1 region. Extends cosmid 4 leftwards by 14 kb.	Six feather, three feather-like	Section 4.3.2.2
16	Similar to cosmid 25[B].	Feather keratins (see Note B below)	Data not shown
25	Contains most of the cosmid 4 region and extends it rightwards by 6 kb.	Twelve Feather	Section 4.3.2.3
28	Identical to cosmid 25 except that it contains two tandem copies of the cosmid vector.	Feather keratins	Data not shown
29	Identical to cosmid 11.	Feather keratins	Data not shown
31	Extends cosmid 4 rightwards by 20 kb.	Nine Feather	Section 4.3.3.1
33	Extends cosmid 31 rightwards by 26 kb. Does not overlap with cosmid 4.	Four feather (complete) three or four claw	Section 4.3.3.2
34	Similar to cosmid 31, but contains a smaller insert.	Feather	Data not shown
35	Does not overlap with the mapped keratin locus[A].	Feather, scale and claw	Section 4.4.3 L. Whitbread, personal communication

[A] Refers to the characterized 100 kb region shown in Figure 4.45.

[B] Cosmid 16 contains all of the HindIII fragments present in the cosmid 25 insert, except for the 2.9 and 1.8 kb HindIII fragments which are located at the left end of cosmid 25 (see Section 4.3.2.3).

Figure 4.45. Cosmid 2 overlaps with cosmid 12 and extends it leftwards by about 15 kb (data not shown). Two interesting aberrations of the normal cosmid clones were isolated (Table 4.2). Cosmids 11 and 29 contained a rearrangement of part of the feather keratin gene locus, which consisted of two non-contiguous segments (data not shown). Similar rearrangements of cosmid clones have been reported by other workers (Meyerowitz et al., 1980; Grosveld et al., 1981). Cosmid 28 was similar to cosmid 25, but differed from cosmid 25 in that it contained two copies of the cosmid vector. The presence of more than one copy of the cosmid vector has also been observed in this laboratory during studies of cosmid clones containing sheep keratin genes (Fietz, 1985).

4.4.2 ORGANIZATION OF THE KERATIN GENE LOCUS

The structure and organization of the keratin gene locus is presented in Figure 4.45, showing the position of BamHI and HindIII restriction sites. In total, the mapped region covers a continuous segment of DNA 100 kb in length containing 18 feather keratin genes, three feather-like genes and three, or possibly four, claw keratin genes. The feather keratin genes span a total of about 53 kb and are spaced evenly apart, with a centre-to-centre distance of about 3 kb. Of the feather keratin genes, 10 have been shown to be transcribed in the same direction. The tandem arrangement of at least some of these feather keratin genes and their even spacing suggests that, as proposed by Molloy et al. (1982) in studies on the recombinant λCFK1, they arose by a series of tandem duplications (see Section 7.3). To the left of the feather keratin gene cluster, 5 kb away from the last feather gene, lie three feather-like

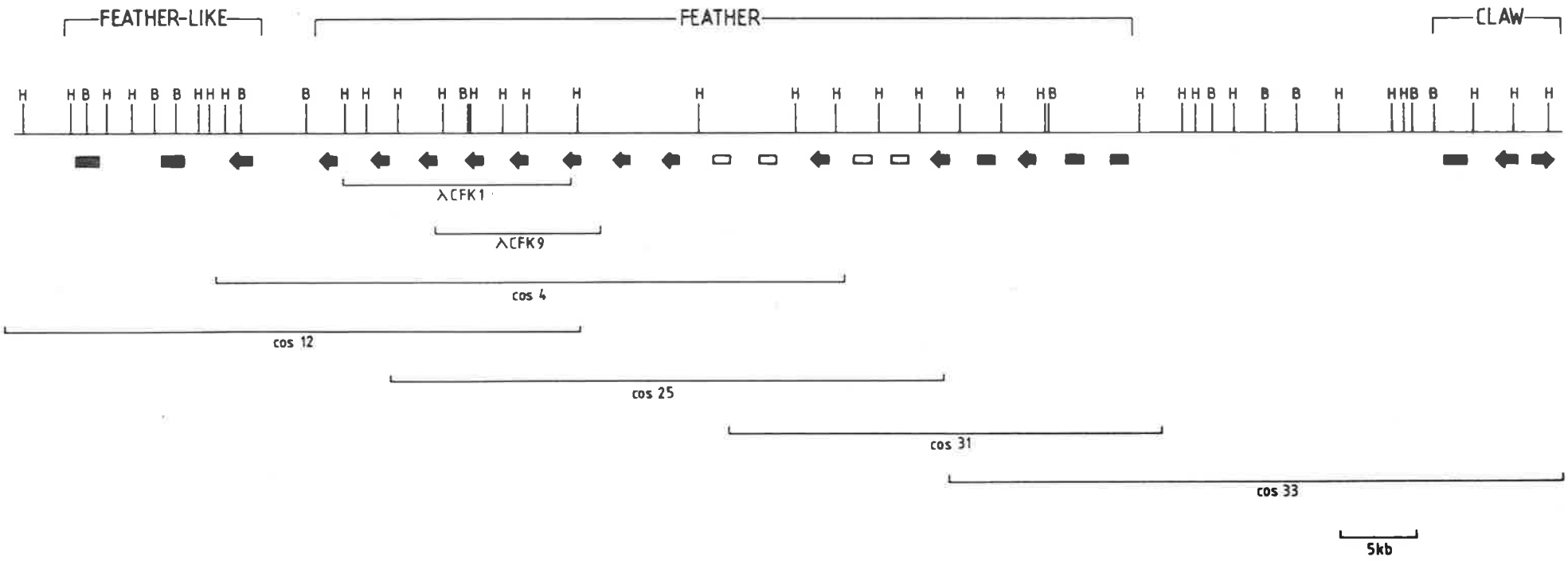
FIGURE 4.45 : CHROMOSOMAL ORGANIZATION OF THE AVIAN KERATIN GENES

The region contained in the 5 overlapping cosmid clones spans a continuous segment of 100 kb which includes the genomic region contained in the previously isolated recombinant λCFK1 (Molloy *et al.*, 1982). The arrangement of the keratin genes was determined by hybridization studies on the cosmid clones and derived plasmid subclones and, in some cases, DNA sequence analysis (Chapters 4 and 5). The transcriptional orientation of the genes, where known, are indicated by arrows. The four genes not mapped in detail are shown as open boxes. The location of these unshaded genes is only approximate.

Restriction sites shown are:

B - BamHI

H - HindIII


FEATHER-LIKE
FEATHER
CLAW
λCFK1
λCFK9
cos 4
cos 12
cos 25
cos 31
cos 33
5kb

genes which are themselves spaced about 4 kb apart. At least one of these genes has the same transcriptional polarity as most of the feather keratin genes (right to left, Figure 4.45). To the right of the feather gene cluster, 21 kb away from the last feather gene, lie a cluster of three or four claw keratin genes (Figure 4.30). These claw genes are spaced 1-2 kb apart and recent results have shown that two of these genes form a tightly linked, divergently transcribed pair (Figure 4.45; L. Whitbread, personal communication).

4.4.2.1 The Feather Keratin Genes

One of the central aims of this study was to determine the size of the feather keratin locus and the number of feather genes within it. By Southern analysis of the overlapping cosmid recombinants with three oligonucleotides specific for feather keratin genomic sequences (Section 4.3.4.1) and similar analyses on subclones generated from the cosmid recombinants (Section 4.3.4.2), a detailed picture of the number of feather keratin genes in this locus, and their arrangement, was constructed.

A number of important findings arising out of these studies are worthy of mention. The three oligonucleotides which were used as hybridization probes were constructed from highly conserved segments of the four complete genes from λCFK1 (Gregg and Rogers, 1986; see Section 2.2.10). Not unexpectedly, each of these oligonucleotides detected most of the restriction fragments of cosmids 4 and 33 which contained feather keratin genes, with one or two exceptions in each case (Section 4.3.4.1). These results showed that the conservation of these sequences was not peculiar to the genes located in

the λCFK1 clone, but is a general feature of the feather keratin gene family. Analysis of subclones generated from the cosmids demonstrated that the failure of one or two genes to hybridize to each of the oligonucleotides is not due to that part of the gene(s) being present in neighbouring restriction fragments. Rather it must be due to alterations in, or removal of, these sequences from these genes so that the probe no longer hybridizes.

The function of the sequences defined by these oligonucleotides is at present unknown. The conservation of the 5' non-coding exon, part of which is represented in the 5' non-coding oligonucleotide, implies that this part of the gene is of important functional significance (see Sections 7.3 and 7.4). That this 37 bp sequence might have some role in transcriptional regulation is suggested by the finding (Morris, 1984) that a gene encoding one of the fast proteins which are unrelated to, but co-ordinately expressed with, the feather keratins (Section 1.3.3), has a 37 bp 5' non-coding exon with a sequence similar to that of feather keratins.

The intron oligonucleotide, which comprises the conserved 3' end of the intron plus the 3' splice site, detected all of the feather keratin gene-containing restriction fragments of cosmids 4 and 33 except one (Figure 4.31c). DNA sequence analysis (Section 5.3.4) demonstrated that the failure of the intron probe to hybridize to that gene was due to several base changes in the sequence upstream from the 3' splice site and not to the loss of the intron or intron boundary sequences. It is unclear what the functional significance of the changes in this gene would be.

The 3' non-coding oligonucleotide detected all except two of the 15 genes in cosmids 4 and 33 (Figure 4.31b). The function of this sequence is unknown. Its conservation implies that it serves some function, particularly in view of the comparatively high degree of divergence observed in the 3' non-coding region compared to the 5' non-coding or protein coding regions (see Section 1.4.4, Figure 1.7). Presumably these conserved sequences in the feather keratin genes have been maintained by an active gene correction mechanism such as gene conversion (see Section 7.3.1).

During the restriction mapping of the feather keratin gene locus two sets of symmetrical or repeated restriction sites, which were each repeated six times, were observed. In cosmid 4, a KpnI site is repeated to generate 5 KpnI fragments of 3.4 kb (Figure 4.13) and in cosmid 31, a HindIII site is repeated to generate 5 2.7 kb HindIII fragments (Figure 4.25b). DNA sequencing studies (Section 5.3.3) indicate that at least some of the KpnI sites are located in the protein coding region 9 bp from the stop codon. These genes were designated the 'Kpn family' of feather keratin genes since they all contain a KpnI site. Similar studies have shown that most, and probably all, of the HindIII sites in the repeated 2.7 kb region lie in the intergenic regions between the feather genes. The significance of these repeated restriction sites are discussed further in Chapter 7 (Section 7.3.1).

4.4.2.2 The Feather-Like Genes

Hybridization studies on cosmid 12 demonstrated that there were three regions located leftwards of the feather gene

cluster which contained sequences with some homology to feather and scale keratin gene sequences. The three genes, named feather-like genes because of their similarity with feather keratin genes (Section 5.3.1), are spaced about 4 kb apart with the gene closest to the feather cluster being 5 kb away from the last feather gene (Section 4.3.2.2).

The hybridization results obtained using feather and scale genes as probes suggested that the feather-like genes had some similarities between feather and scale genes, and that the genes were of an 'intermediate' type (see Section 4.3.2.2c). DNA sequence analysis (Section 5.3.1) of two of the feather-like genes demonstrated that they do not contain the gly-gly-x repeat in their coding regions. This repeat is characteristic of scale (Wilton, 1983; Gregg et al., 1984) and claw (Whitbread, 1985; L. Whitbread and K. Gregg, unpublished results) keratin genes. The feather-like genes appear to be more closely related to feather keratin genes than any other known keratin type (Section 5.4.1).

The nature and possible evolutionary origins of the feather-like genes are discussed further in Section 5.4.2.

4.4.2.3 The Claw Keratin Genes

Recently, a clone was isolated from a cDNA library prepared from chicken claw tissue. This mRNA-derived clone has been shown by DNA sequencing and hybridization to claw RNA to encode a claw keratin (Whitbread, 1985). The partial sequencing of one of the claw genes from cosmid 33 demonstrated that it encodes a protein of 126 amino acids (L. Whitbread and K. Gregg, unpublished results) which is intermediate in size between feather (97 amino acids) and scale (154 amino acids)

(Figure 1.9a; Gregg et al., 1984). Like scale keratins, the protein coding region of the claw gene was found to contain a repeated gly-gly-x region. However, in the claw gene sequenced, the gly-gly-x region consists of a 9 amino acid (27 bp) segment which is repeated twice compared to scale, which has a 13 amino acid segment which is repeated four times. These results demonstrated that the claw keratins are probably a separate gene family from the scale keratins.

Cosmid 33, which contains the four feather keratin genes which comprise the right-hand end of the feather cluster, was shown by hybridization with the claw cDNA clone to contain three or possibly four claw keratin genes. These genes are located 21 kb from the last feather gene at the right-hand end of the cosmid insert (Figure 4.45 and Section 4.3.3.3). The claw keratin genes are located on a 8.5 kb BamHI fragment and appear to be more closely linked than the feather genes, being spaced 1-2 kb apart (Figure 4.30). The claw keratin genes contained in cosmid 33 account for three of the five hybridizing bands detected in HindIII-restricted chick genomic DNA, using the claw keratin cDNA clone as a probe (Whitbread, 1985). The remaining few claw keratin genes may lie just beyond the genomic region contained in cosmid 33 or alternatively, these genes might be present at a different chromosomal locus. Further studies on cosmid 35, which contains claw keratin genes on restriction fragments different in size from those of cosmid 33, should enable this to be resolved.

4.4.3 THE NATURE OF THE KERATIN GENES CONTAINED IN THE THREE COSMIDS NOT DERIVED FROM THE CHARACTERIZED LOCUS

Only three of the 15 cosmid recombinants isolated using the feather keratin gene probe did not overlap with any part of the keratin locus shown in Figure 4.45, that is, cosmids 1, 5 and 35. Cosmid 1 hybridized very poorly with the feather gene probe compared to cosmid 4 (Section 4.3.5.2, Figure 4.41b) and was not detected at all with the feather-specific, scale or claw probes (Section 3.3.3, Figure 3.9; Whitbread, 1985). Cosmid 1 may contain genes encoding beak keratin proteins, which comprise the other major keratinizing epidermal appendage of birds (Frenkel, 1975; Frenkel and Gillespie, 1976). Little is known about the beak keratins except that they have a similar molecular weight to scale keratins (14,500 daltons), and like scale and claw keratins, they apparently also contain an insoluble glycine-rich (gly-gly-x) peptide (Gibbs, P.E.M., Walker, I.D. and Rogers, G.E., unpublished data).

Cosmid 5 gave strong hybridization signals with feather and scale gene probes but failed to hybridize with the feather-specific oligonucleotide (Section 3.3.3., Figure 3.9 and Section 4.3.5.1, Figure 4.41a), suggesting that this recombinant contained either 'variant' feather keratin genes or keratin genes of another type, such as scale. DNA sequencing would distinguish between these two possibilities.

The third recombinant, cosmid 35, which hybridized with the feather and scale keratin probes but not the λCFK1-specific probe (Section 3.3.3, Figure 3.9), is currently being characterized in detail in this laboratory. Recent Southern blot hybridization studies using DNA probes specific for feather,

scale and claw keratin gene types suggests that this clone contains all three types of keratin genes (L. Whitbread, personal communication). Cosmid 35 does not appear to overlap with cosmid 33 (which also contains claw keratin genes) or any of the other cosmid clones, but may extend cosmid 33 rightwards.

4.4.4 THE CHARACTERIZED FEATHER KERATIN LOCUS COMPRISES MOST OF THE FEATHER KERATIN GENES IN THE CHICK GENOME

Genomic Southern analysis using a feather keratin gene as probe demonstrated that the feather-like and feather keratin genes isolated in λ and cosmid recombinants accounts for most of the feather genes in the chick genome (see Section 4.3.7). Southern analysis of EcoRI-digested chick genomic DNA using a feather keratin gene as probe demonstrated that 5 out of the 6 strongly hybridizing bands could be accounted for from the EcoRI map of the feather keratin gene locus. Similarly, these genes account for 17 of the 21-22 bands detected in a Southern blot of HindIII-restricted chick DNA (see Section 4.3.7., Table 4.1). The location and arrangement of the remaining feather keratin genes, or at least fragments detected on a genomic blot, are not known. Some or all of the remaining genes may be present in cosmids 5 and 35 (see Section 4.4.3).

The interpretation of these results is complicated by the fact that, since scale and feather keratin genes show a considerable degree of homology (~ 70%) in their coding regions (Gregg et al., 1984), the feather probe may detect keratin genes other than those of the feather type. However, this does not appear to be the case here, since most of the bands detected by genomic Southern analysis of chick DNA have counterparts

of a similar size in the characterized feather keratin genes contained in the cosmid clones. One possible exception is cosmid 5, which did not hybridize with the feather-specific oligonucleotide probe, yet the restriction fragments which hybridized with the scale gene used as a probe had counterparts of a similar size in chick genomic Southerns (see Section 4.3.7). The nature of the genes in this recombinant are unknown.

CHAPTER 5.

DNA SEQUENCE ANALYSIS OF

THE KERATIN GENES.

5.1 INTRODUCTION

The preceding chapter described the detailed analysis of a bank of cosmid recombinants containing chicken keratin genes. Most of the clones were found to be derived from a single chromosomal locus which contains 18 feather keratin genes, three, or possibly four, claw keratin genes and three keratin genes of an unknown type (see Figure 4.45). These genes were named feather-like genes because of their similarity with feather keratin genes (see Sections 5.3.1 and 5.4.1). This chapter describes a more detailed analysis by DNA sequencing of some features of this keratin gene locus.

Firstly, it was important to establish the nature of the keratin (feather-like) genes located to the left of the feather cluster, that is, whether they are feather, scale or some other keratin gene type, or pseudogenes. In order to do this, the feather-like gene most proximal to the feather gene cluster was partially sequenced.

Secondly, restriction mapping and blot hybridization studies indicated that a number of feather keratin genes in cosmids 4 and 33 contain a conserved Kpn I site which were subsequently named 'the Kpn family' (Section 4.4.2.1). In one of these genes (gene E of the λCFK1 clone, see Figure 4.3), the Kpn I site is located in the protein coding region close to the 'stop' codon (Gregg et al., manuscript in preparation). Southern blot analyses of subclones containing other gene members of 'the Kpn family' suggested that the position of the KpnI site in at least some of the other genes was similar (see Section 4.3.4.2.a,b). To investigate this further, two of these genes were cloned into KpnI M13 vectors and partially sequenced.

Thirdly, hybridization studies using three synthetic oligonucleotide probes, which contain the sequences of short, highly conserved segments found outside the protein coding region of the feather keratin genes in λCFK1 (Gregg and Rogers, 1986; Gregg et al., manuscript in preparation), indicated that these regions are highly conserved between nearly all of the feather keratin genes in the cluster. For example, in the case of the intron oligonucleotide, only one out of the 15 genes in cosmids 4 and 33 was not detected by the probe (Section 4.3.4.1). The 3' end of its intron was sequenced to determine why it did not hybridize with the oligomer.

Lastly, the 5' non-coding and 5' flanking regions of the feather keratin gene located at the left-hand end of the cluster were sequenced. The 5' flanking sequence of this gene is compared to the 5' flanking sequences of two other feather keratin genes and a fast protein gene (Morris, 1984).

5.2 METHODS

5.2.1 PREPARATION OF M13 VECTORS

The replicative form (RF) of M13 DNAs (M13mp8, M13mp9, M13mp18, M13mp19, Section 2.1.3) used in these cloning experiments were digested with the appropriate enzyme(s) to generate the desired termini. The linearized RF DNA was fractionated on a 1% LGT agarose gel (Section 2.2.7) to remove any intact molecules, extracted as described in Section 2.2.8 and resuspended in water to a final concentration of 20 ng/µl.

5.2.2 LIGATION CONDITIONS

The DNA fragment and appropriate M13 vector were combined in a molar ratio of approximately 3:1 in a 20 µl reaction mix containing 20 ng vector, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$,

1 mM rATP, 1 mM DTT, 0.5-1 U ligase and the DNA fragment to be cloned. All ligations, regardless of the termini (blunt or sticky ends), were incubated for 4-8 hours at 15°C or overnight at 4°C.

5.2.3 M13 TRANSFORMATION OF **E. COLI** JM101

Competent cells of E. coli strain JM101 were prepared as described by Messing (1983). The cells were finally resuspended in 50 mM $CaCl_2$ and left for at least one hour before use.

Transformation of the competent JM101 cells with the ligated M13 DNA was performed as follows (see Messing, 1983). An aliquot of the cells (0.2 ml) was added to 5-10 µl of the ligation mix and incubated on ice for 40-60 minutes. After heat shock treatment at 42°C for 2 minutes, the cells were mixed with 3 ml of molten YT soft agar (Section 2.1.8) containing 20 µl IPTG (24 mg/ml in water) and 20 µl BCIG (20 mg/ml in dimethyl formamide) (Section 2.1.9). The density of the competent cells was such that no additional bacteria were needed to act as a feeder lawn. This mixture was poured on to minimal (plus glucose) plates (Section 2.1.8) and the plates incubated overnight at 37°C.

5.2.4 PREPARATION OF SINGLE-STRANDED M13 DNA FOR SEQUENCING

The white plaques were grown up and the DNA harvested essentially as described by Winter and Fields (1980). An overnight culture of JM101 was diluted 1/100 into 2 x YT broth and 1.5 ml aliquots dispensed into screw-cap tubes. The cultures were infected with a white plaque using a toothpick and incubated at 37°C with aeration for 6-7 hours. The cells were sedimented at 12,000 rpm for 10 minutes and phage

particles precipitated from the supernatant (1 ml) by adding 250 µl of 20% PEG, 2.5 M NaCl and leaving the tubes at room temperature for 15 minutes. The precipitated phage were collected by centrifugation at room temperature for 10 minutes and then resuspended in 100 µl of TE pH 8.0. The phage DNA were extracted once with buffer-saturated phenol (50 µl), once with chloroform (50 µl) and once with ether (0.5 ml) before ethanol precipitation overnight at -20°C (Section 2.2.2). The DNA was collected by centrifugation, washed with cold 70% ethanol and resuspended in 30 µl TE pH 7.5 (Section 2.1.7).

5.2.5 HYBRIDIZATION

The universal sequencing primer (17-mer) (Section 2.1.6) was annealed to the template M13 DNA by mixing 1 µl of primer (2.5 ng), 1 µl of 10 x TM buffer (100 mM Tris-HCl pH 8.0, 100 mM $MgCl_2$) and 5-6 µl of M13 DNA in a final volume of 10 µl and heating at 100°C for 2 minutes in a boiling water bath. The samples were allowed to cool slowly to room temperature over a period of about 1 hour.

5.2.6 DIDEOXY CHAIN TERMINATION PROCEDURE FOR DNA SEQUENCING

Most of the DNA sequencing performed during the course of this work was done using the BRESA Dideoxy Sequencing Kit which provides all the necessary components, including those required for the hybridization step (Section 5.2.5), except the α-[^{32}P]dCTP.

The chain termination reactions were performed essentially as described previously (Sanger et al., 1977; Sanger et al., 1980; Messing et al., 1981). The incubations were carried out at 37°C for 15 minutes and the 'chase' performed for 10 minutes after which formamide loading mix was added to stop the reaction.

5.2.7 DNA SEQUENCING GELS

Just prior to loading on to denaturing 6% polyacrylamide/ 7 M urea gels (acrylamide to bisacrylamide ratio of 19:1), the reaction mixtures (plus formamide loading mix) were heated at 100°C for 3-5 minutes and then chilled on ice. A 0.5-1 µl sample of each sequencing reaction was loaded on to the gels which were then electrophoresed at 15-20 mA for 2-4 hours until the tracker dyes had migrated the required distance.

Following electrophoresis the gel, supported by a glass plate, was fixed with 12% acetic acid for 10-15 minutes (or until all the urea had been removed). The gel was then washed with 20% (v/v) ethanol and baked in an oven at 110°C for at least 30 minutes. The gel was covered with thin plastic (Vitafilm) and exposed to X-ray film at room temperature.

5.2.8 DNA SEQUENCE ANALYSIS

DNA sequence data was analyzed on a VAX 11-785 computer using programs generously donated by R. Staden and M. Kanehisa. The programs obtained from M. Kanehisa (National Cancer Institute, National Institutes of Health, Washington) were collectively known as IDEAS (for Integrated Database and Extended Analysis System for nucleic acids and proteins). The programs commonly used in this study were:

ANALYSEQ (R. Staden) - for general nucleotide sequence analysis (translation, codon usage, base content, restriction sites, etc.).

SEQA (M. Kanehisa) - for global homology alignment of two nucleotide sequences.

SEQH (M. Kanehisa) - determines local homology between two nucleotide sequences.

SQRVCM (R. Staden) - produces a file containing the reverse-complement of a given nucleotide sequence file.

5.3 RESULTS

5.3.1 THE FEATHER-LIKE GENE

To characterize the feather-like gene, fragments from the subclone containing the gene (pE5.0, Section 4.3.2.2.d) which hybridized with the feather keratin gene probe were cloned into M13 and sequenced (Section 5.2). Figure 5.1 shows a restriction map of the pE5.0 clone, reproduced from Figure 4.20d, indicating the position and orientation of the feather-like gene derived from DNA sequencing. The sequencing strategy is shown below the pE5.0 map in exploded view (Figure 5.1). Initially, the 1.16 BamHI/EcoRI and 0.9 kb Bam HI/HindIII fragments, containing the right and left halves of the 2.1 kb EcoRI/HindIII fragment, respectively, were subcloned into M13 and the sequence determined from the BamHI site in each case (Figure 5.1). The small 0.16 kb and 0.17 kb PstI fragments were later cloned into M13 and sequenced, enabling the sequence of both DNA strands in this region to be determined. Finally, in order to obtain the N-terminus of the coding region as well as upstream sequences, the 1.0 kb PstI/EcoRI fragment was cloned and partially sequenced (Figure 5.1).

Figure 5.2 shows examples of DNA sequencing of M13 clones containing the 0.16 kb PstI, 1.0 kb PstI/EcoRI and 0.9 kb BamHI/HindIII fragments. Routinely, at least 200 bases could be read from each sequenced clone.

The partial nucleotide sequence of the feather-like gene and its conceptual translation are presented in Figure 5.3. The restriction sites used in the M13 cloning experiments are

FIGURE 5.1 : MAP AND DNA SEQUENCING STRATEGY OF THE FEATHER-LIKE GENE CONTAINED IN pE5.0

The restriction map of pE5.0 is taken from Figure 4.20d. The position and transcriptional orientation of the feather-like gene was determined by DNA sequence analysis (see Section 5.3.1 and 5.3.2). The exact location of the 5' end of the gene is unknown (open, dotted box). The arrows below the gene indicate the position of the following:

1 - initiation codon

2 - termination codon

3 - polyadenylation site (putative)

The strategy used to determine the partial DNA sequence of the feather-like gene is shown below the map in exploded view. The arrows indicate the direction and length of sequence obtained from a particular restriction site. The 0.1 kb PstI/HindIII fragment sequenced in both directions contains the 3' end of the gene.

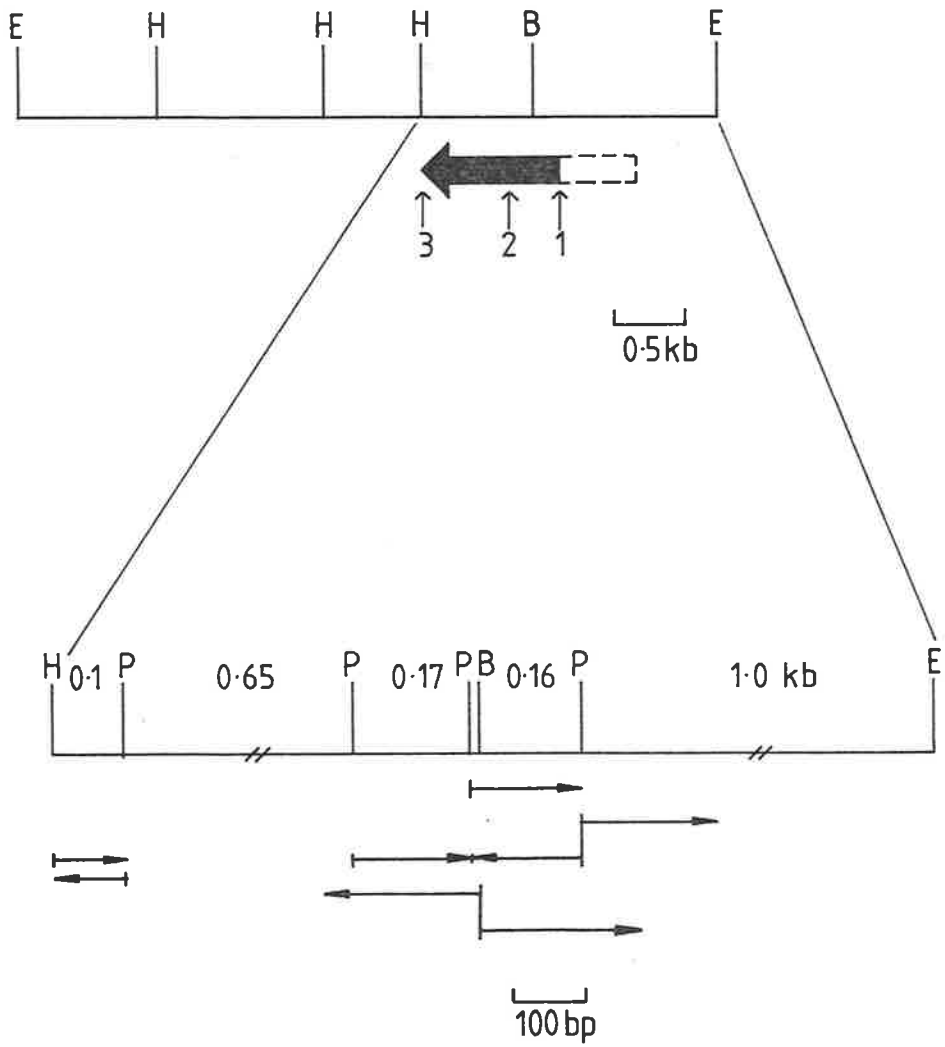

E
H
H
H
B
E
3
2
1
0·5kb
H
0·1
P
0·65
P
0·17
P
B
0·16
P
1·0 kb
E
100 bp

FIGURE 5.2 : DNA SEQUENCE DETERMINATION

The DNA fragments (see below) were subcloned into M13 vectors (Section 5.2) and the single stranded DNA isolated from M13 clones as described in Section 5.2.4. The cloned DNA was used as a template for dideoxy sequencing (Section 5.2.5 and 5.2.6) and the radiolabelled products electrophoresed on a 6% polyacrylamide DNA sequencing gel (Section 5.2.7).

The sequencing ladders are of M13 clones of the following DNA fragments (see Figures 5.1 and 5.3):

A. 0.16 kb PstI fragment

B. 1.0 kb PstI/EcoRI fragment

C. 0.9 kb BamHI/HindIII fragment

The position of the cloning sites (PstI, A, B; BamHI, C) are indicated by bars.

The four base-specific termination reactions were:

A dideoxyadenosine

G dideoxyguanosine

C dideoxycytosine

T dideoxythymidine

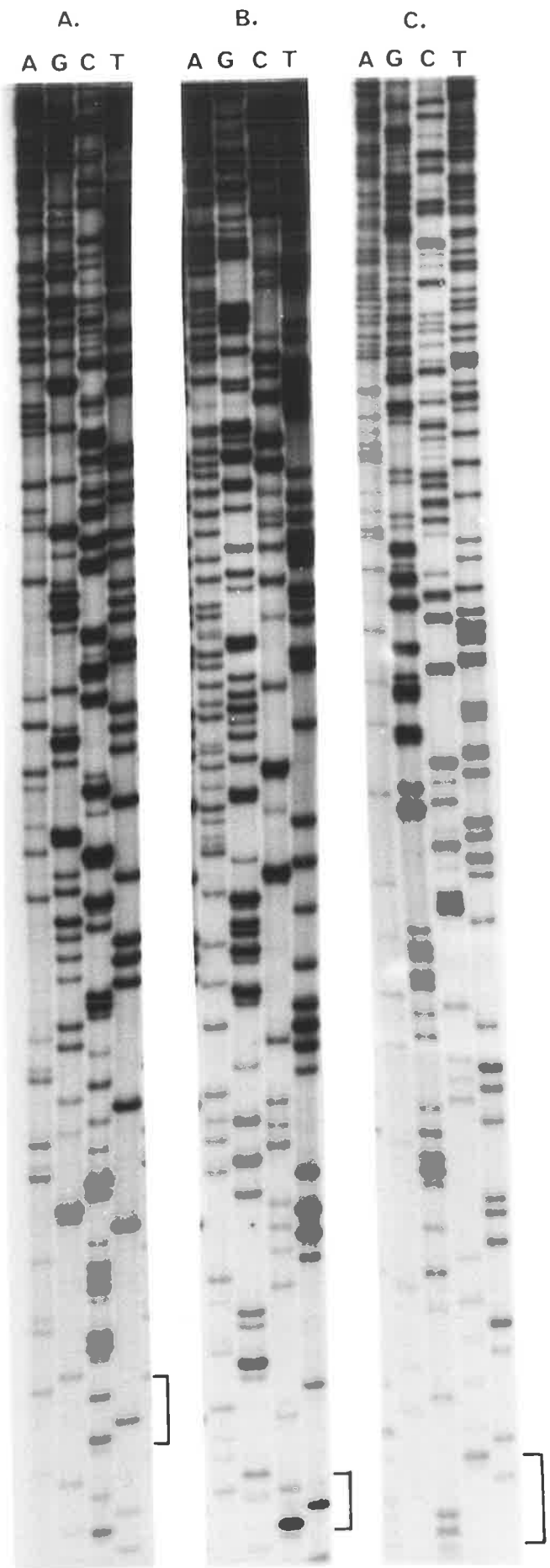
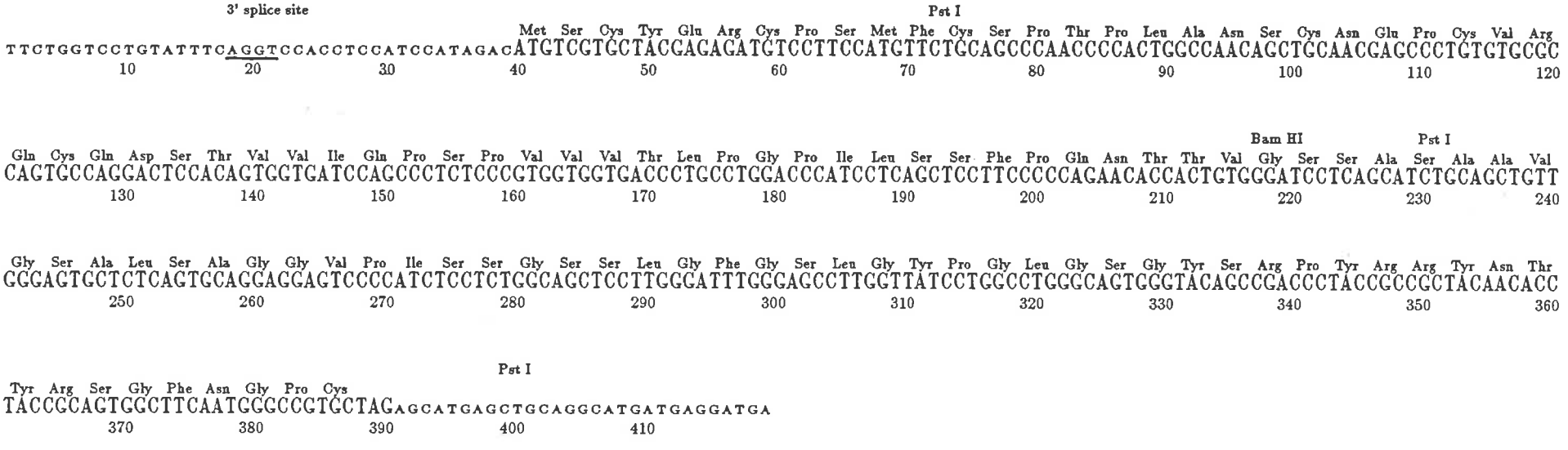

A.
A G C T
B.
A G C T
C.
A G C T

FIGURE 5.3 : THE PARTIAL NUCLEOTIDE SEQUENCE OF THE FEATHER-LIKE GENE

The nucleotide sequence obtained from the feather-like gene located in pE5.0 and its conceptual amino acid sequence are shown (see also Figures 5.1, 5.2). The restriction sites used for subcloning and sequencing experiments are indicated. The 3' splice site is underlined.


3' splice site
Pst I
Met Ser Cys Tyr Glu Arg Cys Pro Ser Met Phe Cys Ser Pro Thr Pro Leu Ala Asn Ser Cys Asn Glu Pro Cys Val Arg
TTCTGGTCCTGTATTTCAGGTCCACCTCCATCCATAGACATGTCGTGCTACGAGAGATGTCCTTCCATGTTCTGCAGCCCAACCCCACTGCCAACAGCTGCAACGAGCCCTGTGTGCGC
10 20 30 40 50 60 70 80 90 100 110 120
Bam HI
Pst I
Gln Cys Gln Asp Ser Thr Val Val Ile Gln Pro Ser Pro Val Val Val Thr Leu Pro Gly Pro Ile Leu Ser Ser Phe Pro Gln Asn Thr Thr Val Gly Ser Ser Ala Ser Ala Ala Val
CAGTGCCAGGACTCCACAGTGGTGATCCAGCCCTCTCCCGTGGTGGTGACCCTGCCTGGACCCATCCTCAGCTCCTTCCCCCAGAACACCACTGTGGGATCCTCAGCATCTGCAGCTGTT
130 140 150 160 170 180 190 200 210 220 230 240
Gly Ser Ala Leu Ser Ala Gly Gly Val Pro Ile Ser Ser Gly Ser Ser Leu Gly Phe Gly Ser Leu Gly Tyr Pro Gly Leu Gly Ser Gly Tyr Ser Arg Pro Tyr Arg Arg Tyr Asn Thr
GGGAGTGCTCTCAGTGCAGCAGGAGTCCCCATCTCCTCTGGCAGCTCCTTGGGATTTGGGAGCCTTGGTTATCCTGGCCTGGGCAGTGGGTACAGCCGACCCTACCGCCGCTACAACACC
250 260 270 280 290 300 310 320 330 340 350 360
Pst I
Tyr Arg Ser Gly Phe Asn Gly Pro Cys
TACCGCAGTGGCTTCAATGGGCCGTGCTAGAGCATGAGCTGCAGGCATGATGAGGATGA
370 380 390 400 410

shown. The opening reading frame of 348 bp encoded a protein which, excluding the initiating methionine, contained 115 amino acids, compared to 97 amino acids for the feather keratins (Gregg and Rogers, 1986). The conceptual protein product of the feather-like gene was calculated to have a molecular weight of 11,900 daltons. In the 5' non-coding region, 20 bp upstream from the translation initiation codon, is a splice site with the consensus sequence 5' AGGT 3' (nucleotides 18-21, Figure 5.3), indicating that the 5' non-coding region of this gene is probably interrupted by an intron.

Figure 5.4 compares the DNA sequences of the feather-like gene and feather keratin gene B. The two sequences show a high degree of homology (85%) between the 3' splice site (nucleotides 18-21) and position 281 of the feather-like DNA sequence. In particular, the region contained in the 0.16 kb PstI fragment (nucleotides 76-234) shows 92% homology with the feather keratin sequence. The DNA sequence of the feather-like gene downstream of base 281 showed little homology with the feather sequence, with two exceptions; a region of 12 bp near the C-terminus (nucleotides 315-326, Figure 5.3) and at the termination codon (nucleotides 382-395, Figure 5.3) (data not shown). Overall, the coding region of the feather keratin gene was calculated to be about 75% homologous with the feather-like sequence. The portion of the 5' non-coding region of the feather-like gene between the 3' splice site and the translation initiation codon was 80% homologous to that portion of the feather keratin gene (Figure 5.4).

Figure 5.5 illustrates the comparison of the conceptual translation product of the feather-like gene with feather

FIGURE 5.4 : COMPARISON OF THE DNA SEQUENCES OF THE FEATHER-LIKE GENE AND FEATHER KERATIN GENE B

The nucleotide sequences from the protein coding, and part of the 5' non-coding and intron of the feather-like gene (see Figure 5.3), is compared with feather keratin gene B of λCFK1 (Molloy et al., 1982; Gregg et al., 1984). The comparison spans from 19 bp upstream of the 3' splice site to about two-thirds the way through the protein coding region of the feather-like gene. Matches between the two sequences are indicated by a dot. Downstream of nucleotide 281, there was very little homology between the two sequences (see text). Hyphens (--) indicate gaps in the sequence which were introduced by the computer program to assist in their alignment (see Section 5.2.8).

The 3' splice site (bases 18-21) and the initiation codon (bases 40-42) are highlighted by boxes and the relevant restriction sites are underlined.

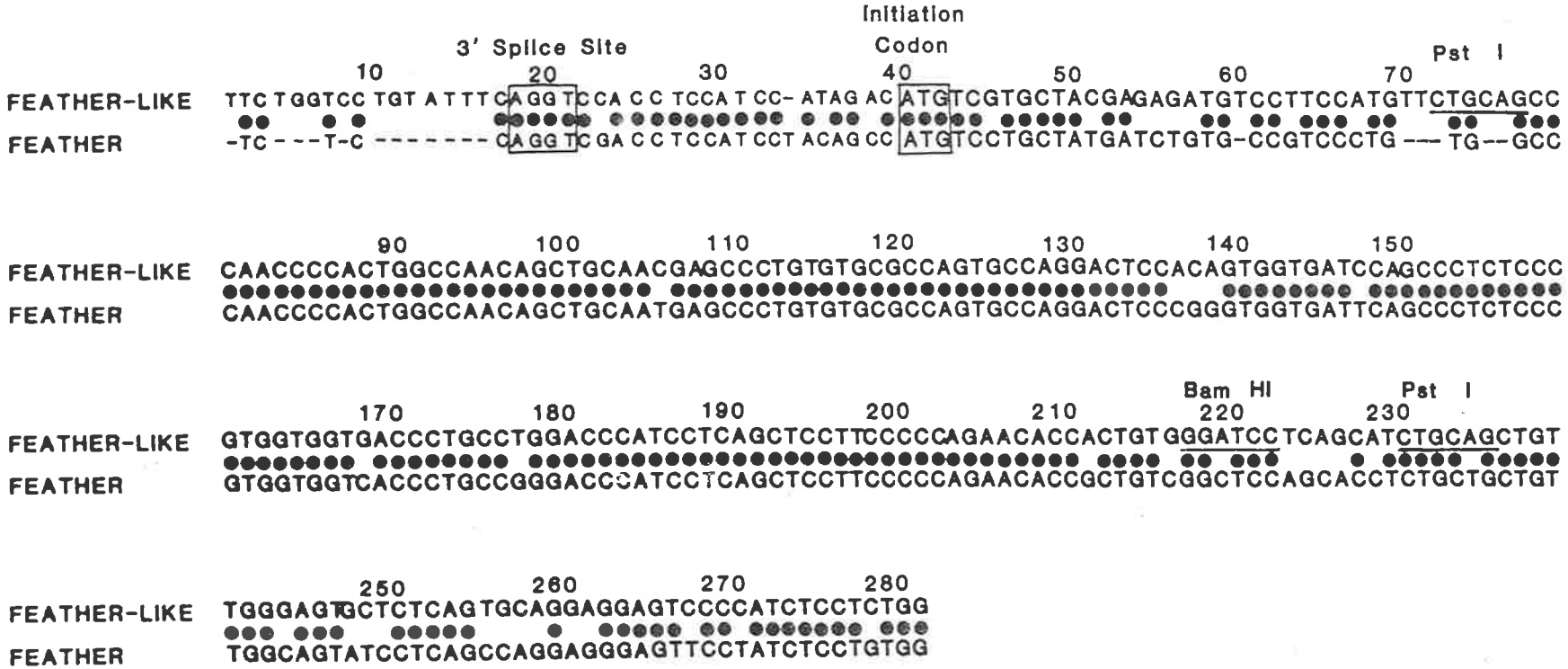
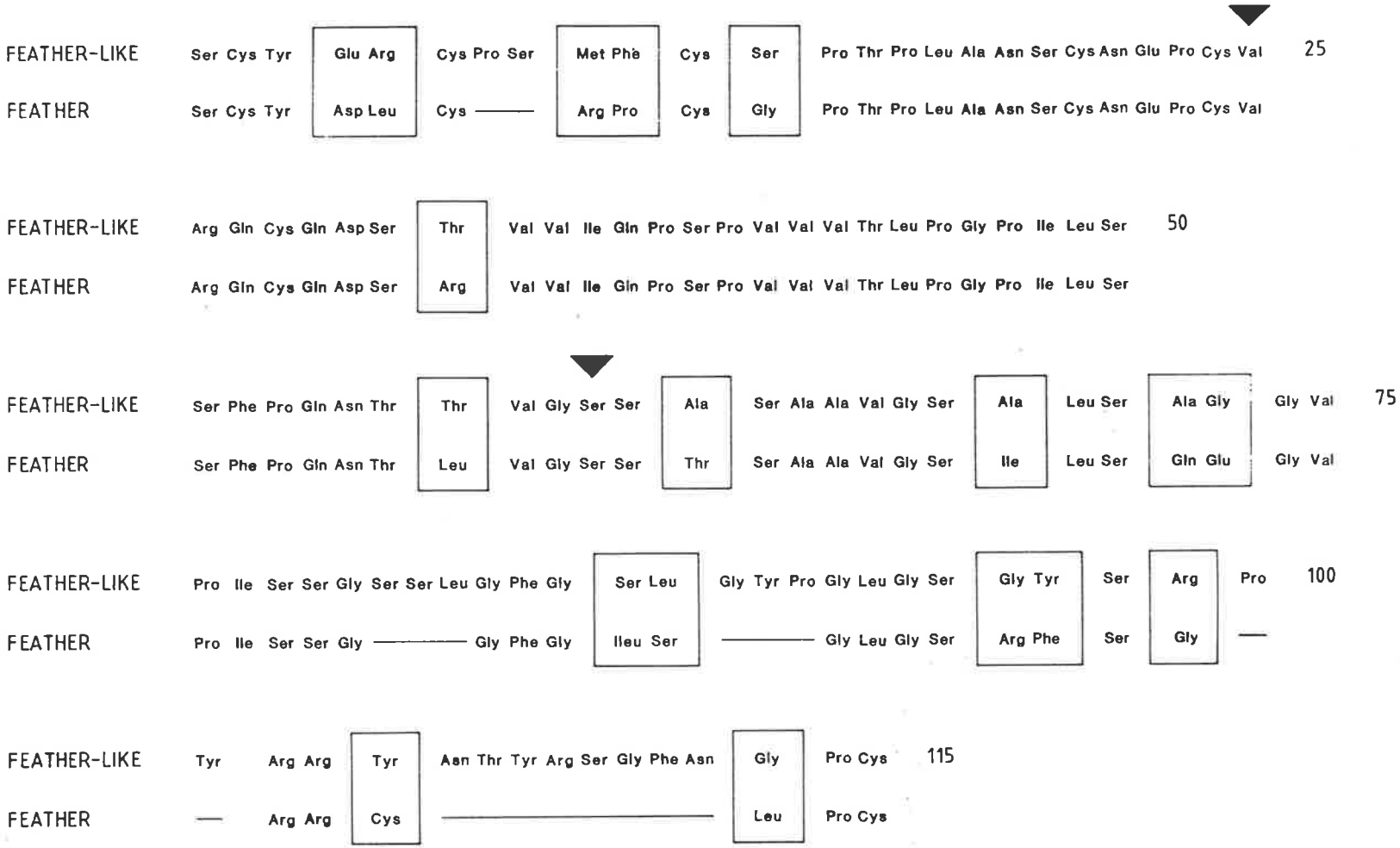

Initiation
Codon
3' Splice Site
Pst I
10
20
30
40
50
60
70
FEATHER-LIKE
TTC T GGTCC TGT A TTT CA GG TC CA C C TCCA T CC- ATAG AC ATG TCGTGCTACGAGAGATGTCCTTCCATGTTCTGCAGCC
FEATHER
-TC - - - T-C - - - - - - - CA GG TC GA C C TCCA T CCT A CAG CC ATG TCCTGCTATGATCTGTG-CCGTCCCTG --- TG--GCC
90
100
110
120
130
140
150
FEATHER-LIKE
CAACCCCACTGGCCAACAGCTGCAACGAGCCCTGTGTGCGCCAGTGCCAGGACTCCACAGTGGTGATCCAGCCCTCTCCC
FEATHER
CAACCCCACTGGCCAACAGCTGCAATGAGCCCTGTGTGCGCCAGTGCCAGGACTCCCGGGTGGTGATTCAGCCCTCTCCC
Bam HI
Pst I
170
180
190
200
210
220
230
FEATHER-LIKE
GTGGTGGTGACCCTGCCTGGACCCATCCTCAGCTCCTTCCCCCAGAACACCACTGTGGGATCCTCAGCATCTGCAGCTGT
FEATHER
GTGGTGGTCACCCTGCCGGGACCCATCCTCAGCTCCTTCCCCCAGAACACCGCTGTCGGCTCCAGCACCTCTGCTGCTGT
250
260
270
280
FEATHER-LIKE
TGGGAGTGCTCTCAGTGCAGGAGGAGTCCCCATCTCCTCTGG
FEATHER
TGGCAGTATCCTCAGCCAGGAGGGAGTTCCTATCTCCTGTGG

FIGURE 5.5 : COMPARISON OF THE PROTEINS ENCODED BY THE FEATHER-LIKE GENE AND FEATHER KERATIN GENE B

The predicted amino acid sequences of the feather-like gene and feather keratin gene B have been aligned to maximize homology between the two protein sequences. The feather-like gene codes for a protein of 115 amino acids, compared to 97 amino acids for the feather keratins.

Amino acid substitutions are boxed and the fine line indicates apparent deletion. The region between amino acid residues 25-60 (arrowed in figure) depicts the segment which, in the feather keratin sequence, could form a regular β-sheet conformation (Gregg et al., 1984). The portion between the arrows is highly conserved between the two sequences, with a total of only two amino acid differences.


FEATHER-LIKE
Ser Cys Tyr Glu Arg Cys Pro Ser Met Phe Cys Ser Pro Thr Pro Leu Ala Asn Ser Cys Asn Glu Pro Cys Val 25
FEATHER
Ser Cys Tyr Asp Leu Cys —— Arg Pro Cys Gly Pro Thr Pro Leu Ala Asn Ser Cys Asn Glu Pro Cys Val
FEATHER-LIKE
Arg Gln Cys Gln Asp Ser Thr Val Val Ile Gln Pro Ser Pro Val Val Val Thr Leu Pro Gly Pro Ile Leu Ser 50
FEATHER
Arg Gln Cys Gln Asp Ser Arg Val Val Ile Gln Pro Ser Pro Val Val Val Thr Leu Pro Gly Pro Ile Leu Ser
FEATHER-LIKE
Ser Phe Pro Gln Asn Thr Thr Val Gly Ser Ser Ala Ser Ala Ala Val Gly Ser Ala Leu Ser Ala Gly Gly Val 75
FEATHER
Ser Phe Pro Gln Asn Thr Leu Val Gly Ser Ser Thr Ser Ala Ala Val Gly Ser Ile Leu Ser Gln Glu Gly Val
FEATHER-LIKE
Pro Ile Ser Ser Gly Ser Ser Leu Gly Phe Gly Ser Leu Gly Tyr Pro Gly Leu Gly Ser Gly Tyr Ser Arg Pro 100
FEATHER
Pro Ile Ser Ser Gly —— Gly Phe Gly Ileu Ser —— Gly Leu Gly Ser Arg Phe Ser Gly ——
FEATHER-LIKE
Tyr Arg Arg Tyr Asn Thr Tyr Arg Ser Gly Phe Asn Gly Pro Cys 115
FEATHER
—— Arg Arg Cys —— Leu Pro Cys

keratin gene B aligned to maximize amino acid sequence homology. As expected from analysis of the DNA sequences, there is a close similarity between the sequences for the first 80 amino acids. In this region, 67 of the 80 amino acids are identical between the two sequences (Figure 5.5). The remainder of the feather-like sequence beyond the glycine residue at position 80 shares only short segments (2-4 amino acids) of homology and contains a number of insertions when compared to the feather keratin sequence. The three amino acids at the N-terminus and the two amino acids at the C-terminus are conserved between the two sequences (Figure 5.5). In all, 79 amino acids of the feather-like sequence were identical with the protein encoded by feather keratin gene B, equivalent to an amino acid sequence homology of 81%. Out of the remaining 36 amino acids of the feather-like sequence which differed from feather, 18 could be accounted for by amino acid substitutions and 18 by insertions (Figure 5.5).

5.3.2 THE 3' END OF THE FEATHER-LIKE GENE

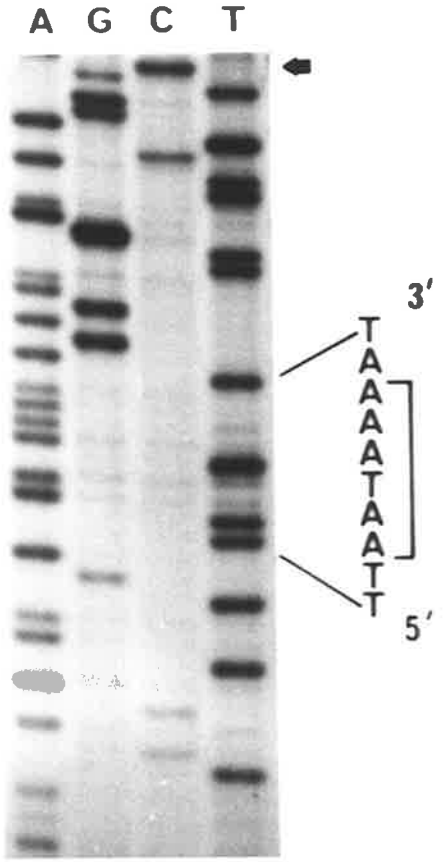
The sequencing of a small 0.1 kb PstI/HindIII fragment located downstream of the stop codon enabled the polyadenylation signal and putative poly(A) addition site of this gene to be identified. Figure 5.6 illustrates part of the sequence between the PstI and HindIII sites (5' → 3'), showing the sequence of the putative polyadenylation signal and immediately downstream. The sequence around the poly(A) signal of the feather-like gene is compared to that of feather keratin gene B, which have been aligned to maximize homology. Most of the invariant bases found in the four feather keratin genes between the polyadenylation signal and the site of poly(A)

FIGURE 5.6 : DNA SEQUENCE ANALYSIS OF THE 3' END OF THE FEATHER-LIKE GENE

To determine the DNA sequence of the 3' end of the feather-like gene, a 0.1 kb PstI/HindIII fragment from pE 5.0 was subcloned into M13mp8 and M13mp9 and sequenced in both directions (see Section 5.2). The gel illustrates part of the DNA sequence obtained, reading from the PstI to the HindIII sites (5' → 3'). The polyadenylation signal is bracketed and the arrow indicates the proposed poly(A) addition site (see text).

Below, the DNA sequence of the 3' end of the feather-like gene is compared to that of feather keratin gene B of λCFK1. The polyadenylation signal is bracketed and the point of polyadenylation of the mRNA is indicated by an arrow. The bases common between the two sequences are depicted by a dot. The bases which are invariant between the feather keratin genes sequenced (see Figure 1.8d) but are not conserved in the feather-like gene are boxed.

By comparison with the feather keratin genes, the point of polyadenylation in the feather-like gene is proposed to be 29 bp from the 5' adenine of the polyadenylation signal.



Poly(A) Signal

FEATHER-LIKE 5′ AATGATTAATAAAATAGAGAATTGAATTACTAGGTGCAACTTCTA 3′

FEATHER KERATIN TGGACTCAATAAAATGTATA-CTGCATTG-TAATCTCAGTCTCCT

addition are also conserved in the feather-like gene (Figure 5.6). By comparison with the feather keratin genes, it is postulated that the point of poly(A) addition in the feather-like gene is 29 bp downstream from the first base of the polyadenylation signal (arrows in Figure 5.6).

5.3.3 IDENTICAL LOCATION OF THE KpnI SITE IN THREE MEMBERS OF THE 'Kpn FAMILY'

In Chapter 4, it was observed that a number of feather keratin genes in the cluster contained a conserved KpnI site and it was plausible that the location of this site was similar or identical in each of these genes (see Section 4.4.2.1). In order to test this hypothesis, two genes belonging to the 'Kpn family' were partially sequenced from their KpnI sites after subcloning into KpnI/HindIII vectors. The genes chosen were the gene contained in the 2.7 kb HindIII fragment of cosmid 4 (pH 2.7, Section 4.3.4.2b) and the 1.85 kb KpnI/HindIII fragment of pH 8.1 which contains the 5' end of a feather keratin gene (Section 4.3.4.2a).

Figure 5.7a presents part of the sequence obtained from the KpnI site of the 1.0 kb KpnI/HindIII fragment of the pH 2.7 gene compared to gene E of λCFK1. Gene E has been previously shown to contain a KpnI site in the protein coding region 9 bp from the stop codon (Gregg et al., manuscript in preparation). Figure 5.7a shows that the position of the KpnI site is identical in these two genes, that is, 9 bp upstream from the stop codon. In the case of the feather keratin gene fragment contained in the 1.85 kb KpnI/HindIII fragment, the KpnI site was also in an identical location to these other two genes, but the DNA sequence obtained from the KpnI site corresponded to

FIGURE 5.7 : PARTIAL DNA SEQUENCE ANALYSIS OF GENES BELONGING TO THE 'KPN FAMILY'

A. DNA sequence around the KpnI site of the gene located in the 2.7 kb HindIII fragment of cosmid 4 (Section 4.3.4.2b) compared to gene E of λCFK1 (Gregg et al., manuscript in preparation). In each gene the KpnI site is located in the protein coding region, 9 bp upstream from the stop codon.

B. DNA sequence upstream of, and including, the KpnI site of the gene in the 1.85 kb KpnI/HindIII fragment of pH 8.1 (Section 4.3.4.2a) compared to gene E of λCFK1. The sequence obtained from the KpnI site is from the 3' end of the coding region of this gene.

C. Comparison of the DNA and conceptual amino acid sequences of the 'Kpn family' genes (see Figure 5.7a) and genes B-D of λCFK1 (Gregg et al., manuscript in preparation). The two bases underlined in genes B-D are substituted in the 'Kpn genes' to generate the KpnI site. This has resulted in a single amino acid change from a cysteine to a tyrosine residue.

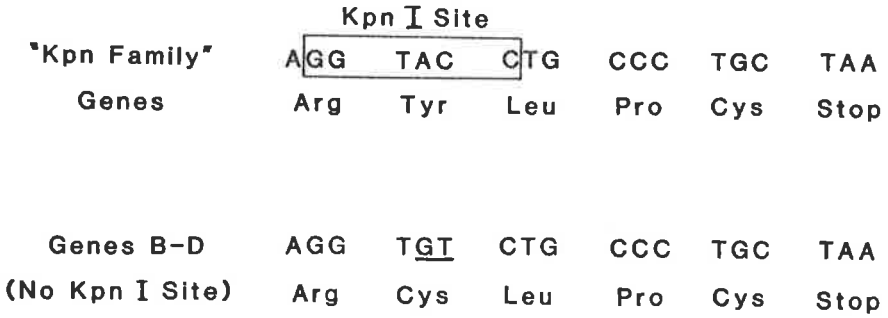
A


Kpn I Site
SEQUENCE FROM 2.7 KB HIND III FRAGMENT
AGGTACCTGCCCTGC TAA
GENE E
AGGTACCTGCCCTGC TAA
Stop

B


Coding
Kpn I Site
SEQUENCE FROM 1.85 KB KPN I / HIND III FRAGMENT
GGTACCTCCTGCCAGAGAA
GENE E
GGTACCTACTGCCAGAGAA

C


Kpn I Site
"Kpn Family" AGG TAC CTG CCC TGC TAA
Genes Arg Tyr Leu Pro Cys Stop
Genes B-D AGG TGT CTG CCC TGC TAA
(No Kpn I Site) Arg Cys Leu Pro Cys Stop

the DNA sequence upstream of the stop codon, reading into the coding region (Figure 5.7b).

In Figure 5.7c, the DNA and conceptual amino acid sequence of the region downstream of, and including, the KpnI site is compared to the same regions of genes B-D of λCFK1 which do not have the KpnI site (see Figure 4.3 and Gregg et al., manuscript in preparation). This demonstrated that a two base change in genes B-D is sufficient to generate the KpnI site present in the genes of 'the Kpn family'. At the amino acid sequence level this has resulted in a change from a cysteine to a tyrosine residue (Figure 5.7c).

5.3.4 SEQUENCE ANALYSIS OF THE 3' SPLICE JUNCTION OF THE pH2.7 GENE

The DNA sequence upstream of the initiation codon of this gene, including part of the 5' non-coding region and the 3' splice junction, was analyzed to determine the molecular basis for the failure of the intron oligonucleotide probe to hybridize to this gene (Section 4.3.4.1, Figure 4.31c). M13 clones containing the 5' half of the pH 2.7 gene (the 1.0 kb KpnI/HindIII fragment, Section 4.3.4.2b) were sequenced using the 25-mer made from the 5' end of the coding region as a sequencing primer. Figure 5.8a shows part of an autoradiograph of the resulting DNA sequencing gel, indicating the position of the initiation codon and the 3' splice site. Figure 5.8b compares the sequence obtained from this gene upstream of, and including, the 3' splice site to the sequence of the intron oligonucleotide probe designed from feather keratin genomic sequences (Section 2.2.10). Unfortunately, the DNA sequence in the region from which the intron oligonucleotide was designed was

FIGURE 5.8 : PARTIAL DNA SEQUENCE ANALYSIS OF THE pH 2.7 GENE

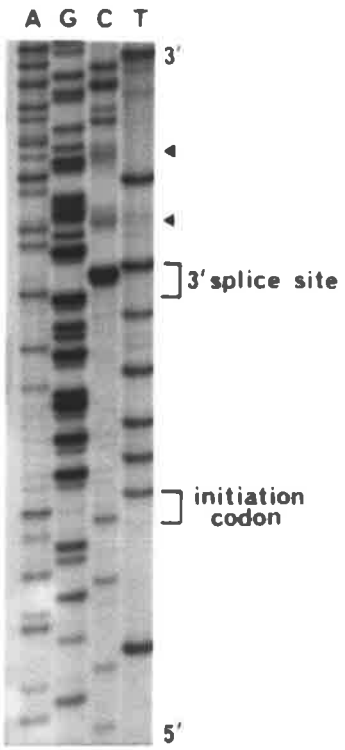
A. Autoradiograph of DNA sequencing gel showing the DNA sequence of the pH 2.7 gene (the gene located in the 2.7 kb HindIII fragment of cosmid 4, Section 4.3.4.2b) between the 3' end of the intron and the 5' end of the coding region. The sequencing primer used was a 25-mer from the 5' end of the coding region (see Section 2.2.10, Figure 2.2). The arrows indicate compressions in the sequencing ladder, which introduce ambiguities in the sequence obtained (see below).

B. The sequence of the 3' end of the intron of the pH 2.7 gene, obtained in Figure 5.8a, is compared to that of the intron oligonucleotide probe. This oligonucleotide did not hybridize with the pH 2.7 gene (Section 4.3.4.1). There is a total of 5 possible base changes in the gene sequence compared to the sequence of the oligomer. The arrowed bases indicate ambiguities in the gene sequence due to compressions (see Figure 5.8a).

C. Complementary sequence to the DNA sequence shown in Figure 5.8a. The 3' splice site and initiation codon are boxed. The DNA sequence downstream of the 3' splice site is very similar to that of other feather keratin genes.

The arrows highlight the bases where there are compressions in the DNA sequence.

A



B

INTRON SEQUENCE (FROM 2.7 KB HIND III FRAGMENT)

INTRON OLIGONUCLEOTIDE PROBE SEQUENCE

C

difficult to read unequivocally due to compressions in the sequence (arrows in Figure 5.8a and b). However, it is apparent from this sequencing data and the complementary DNA sequence of this gene shown in Figure 5.8c that (1) the gene probably contains an intron 21 bp upstream from the initiation codon (Figure 5.8c) and shows a high degree of sequence homology to feather keratin sequences downstream of the 3' splice site (data not shown), and (2) there are a number of base changes in the intron sequence of this gene immediately 5' to the splice site when compared to the sequence of the probe (Figure 5.8b). This would account for the absence of hybridization of this gene with the oligonucleotide probe.

5.3.5 SEQUENCING OF THE 5' NON-CODING AND PROMOTER REGIONS OF THE pBH2.5 GENE

The DNA sequence of the 5' end of the feather keratin gene contained in the pBH2.5 subclone (Section 4.3.4.2d) was determined by subcloning the 0.45 kb PstI/HindIII fragment into M13 and sequencing from the PstI site. Figure 5.9a presents the DNA sequence of part of this fragment. The PstI site is located in the intron 11 bp downstream from the 5' splice site (nucleotides 266-269 of Figure 5.9a). The gene contains a 5' non-coding exon which is identical in length, i.e. 37 bp, and sequence to those of genes A, C and D of λCFK1 (Figure 5.9b). The cap site (nucleotide 231) was located by homology with other feather keratin genes (Molloy et al., 1982; Gregg and Rogers,1986).

FIGURE 5.9 : DNA SEQUENCE ANALYSIS OF THE 5' END OF THE pBH2.5 GENE

A. Nucleotide sequence of the 5' flanking region and 5' non-coding exon of the pBH2.5 gene (a subclone which contains the feather keratin gene from the lefthand end of the cluster, Section 4.3.4.2d).

The following features are underlined:

Nucleotides 78-85	'-150 box'
Nucleotides 132-136, 157-160	two potential CAAT boxes.
Nucleotides 200-203	TATA box
Nucleotide 231	Cap site
Nucleotides 266-269	5' splice site
Nucleotides 274-279	PstI site

B. Comparison of the 37 bp 5' non-coding sequence of the pBH2.5 gene with gene C of λCFK1. The first base shown is the cap site. The 5' splice site and the PstI site, present in the pBH2.5 gene, are underlined.

The sequence of the 5' non-coding exon in genes A and D is identical to that of gene C. Gene B of λCFK1 has a single base changed in the 5' non-coding exon relative to the other four sequences (Gregg and Rogers, 1986).

A.

```
AGGTCTCTAATTCTGTCACTTGAAAGGGGCTGTGTGGCAAAATGGACTTGTCATGAACGG
        10         20         30         40         50         60

GGAAATGAAAAGGGAGCATTACAGGAAAGAGGAATAAGGCTACAATTTCTGTAATTGTGT
        70         80         90        100        110        120

CGTTTTGCATACAAATGGCTTGCCCTACTATAGCTCCATTCGCTAGGGTTGCCTCTGGGC
       130        140        150        160        170        180

                  TATA box                              cap site
CTGGGGTGGTCCAGATCAGTATAAAAGGCAGTCCAGCACCAGGCTCTCTCATCCACTTCT
       190        200        210        220        230        240

                          5' splice
                             site     PstI
CTTGCCTTCTCCTCCTTGGTGAACAAGGTGAGTCTGCAG
       250        260        270
```

B.

```
                                                5' splice
          5'                                       site          3'
End gene  ATCCACTTCTCTTGCCTTCTCCTCCTTGGTGAACAAGGTGAGTCTGCAG

          ::::::::::::::::::::::::::::::::::::::::::::::::::::

Gene C    ATCCACTTCTCTTGCCTTCTCCTCCTTGGTGAACAAGGTGAG-CTGC
```

5.4 DISCUSSION

5.4.1 THE FEATHER-LIKE GENES

As described in Chapter 4, three regions that mapped leftwards of the feather keratin gene cluster hybridized to varying degrees with feather and scale gene probes (see Sections 4.3.2.2). Using the sequencing strategy shown in Figure 5.1, the DNA sequences of the protein coding region and part of the 5' non-coding, intron and 3' non-coding regions of one of these genes (the gene located in the 5.0 kb EcoRI fragment of cosmid 12, Figure 4.18) were determined (Figures 5.3, 5.6). The open reading frame of this gene encoded a protein of 115 amino acids (Figure 5.3), 18 amino acids longer than the conceptual protein products of the characterized feather keratin genes (see Figures 1.6, 5.5).

A number of features of this feather-like gene bore a resemblance to the feather keratin genes - hence their name feather-like genes. Firstly, the amino acid composition was very similar. Like feather keratins, the protein product encoded by the feather-like gene had a high content of glycine (12.2%), proline (11.3%) and serine (18.3%) with no histidine, lysine or tryptophan. This compares to glycine (11.2%), proline (11.2%) and serine (15.3%) for feather keratin gene B (see Gregg et al., 1984; Gregg et al., manuscript in preparation).

Secondly, as is the case in feather keratin genes, there appears to be an intron in the 5' non-coding region which, in the feather-like gene, is 20 bp upstream of the initiation codon (Figure 5.3). Moreover, in all keratin genes sequenced to date (i.e. scale, feather, claw and this feather-like gene)

there appears to be an intron in the 5' non-coding region 20-23 bp upstream of the initiation codon (Gregg et al., 1983; Wilton, 1983; Gregg and Rogers, 1986; L. Whitbread and K. Gregg, unpublished observations).

Thirdly, comparison of the feather-like sequence to feather keratin gene B demonstrated that the portion of the 5' non-coding region between the 3' splice site and the initiation codon and the first 240 bp of the feather-like coding sequence show a high level of homology with the feather sequence (85%, Figure 5.4). Beyond this point, there is little homology except for a 12 bp stretch near the end of the coding region and around the stop codon itself (data not shown).

Comparison of the coding regions of the feather-like sequence with scale and claw keratin gene sequences showed a much lower degree of homology than that found with feather sequences (data not shown). Notwithstanding the similarity of the feather-like and feather keratin genes in the coding and 5' non-coding regions, the feather-like genes were not detected by any of the three feather keratin oligonucleotides which were designed from highly conserved sequences outside the protein coding region (Section 2.2.10). These results suggest that, outside the protein coding region and the portion of the 5' non-coding region downstream of the 3' splice site, the feather-like gene has little homology with feather keratin genomic sequences. In the case of the intron oligonucleotide, this was directly demonstrated by comparison of the feather-like and feather keratin gene sequences upstream of the 3' splice site (Figure 5.4).

The sequence of the feather-like gene immediately upstream of the initiation codon (5' GACATG 3') fits the consensus sequence for efficient translational initiation (5' $\begin{matrix}A\\G\end{matrix}$CCATG; Kozak, 1984, 1986) reasonably well. Of the bases immediately preceding the ATG initiation codon, a purine three nucleotides upstream is the most crucial for efficient initiation of translation (Kozak, 1986). All four genomic feather keratin sequences conform to this consensus sequence (Gregg et al., manuscript in preparation).

The feather-like gene shows a codon bias at the third base which is similar to the feather keratin genes (Table 5.1; Gregg et al., manuscript in preparation). The third base preference is strongly biased towards C with G and T at approximately random frequencies and with A residues severely limited (Table 5.1). Codon usage in various mRNAs has been observed to differ widely and is apparently specific to genome or cell type (Grantham et al., 1980). These differences possibly reflect tissue or species specific differences in tRNA populations or requirements for mRNA secondary structure and may be important in the regulation of translation.

Alignment of the amino acid sequences encoded by the feather-like gene and a feather keratin gene showed that, as expected from the homology observed at the DNA sequence level, the conceptual translation products are highly homologous at least for the first 80 amino acids of the feather-like sequence (Figure 5.5). If allowance is made for a two amino acid insertion near the N-terminus (Pro-Ser, Figure 5.5), then 67 of the first 80 amino acids are identical to the feather sequence. This 80 amino acid region includes the region between Val_{25} and

TABLE 5.1 : THIRD BASE PREFERENCE IN THE CODONS UTILIZED BY THE FEATHER-LIKE AND FEATHER KERATIN GENES

BASE	% FEATHER KERATIN[1]	% FEATHER-LIKE[2]
C	50.8	48.7
G	25.5	20.9
T	20.1	18.2
A	3.6	12.2

[1] Data compiled from the sequences of the four complete genes in the clone λCFK1 (Gregg et al., manuscript in preparation).

[2] Calculated from the data in Figure 5.3.

Ser_{60} (arrows in Figure 5.5) which, in the feather keratin sequence, has been predicted by Chou-Fasman rules to form a highly regular β-conformation (Gregg et al., 1984). This region also comprises the segment of greatest similarity between scale and feather keratin sequences (Gregg et al., 1984). However, it is more highly conserved between the feather-like and feather sequences (two amino acid substitutions, Figure 5.5) than between scale and feather sequences (8 amino acid substitutions, see Figure 1.9a).

It is interesting to note that all of the cysteines in the feather sequence, except one, are also present in homologous positions in the feather-like sequence (Figure 5.5). The cysteine residue present in the feather sequence four amino acids from the C-terminus is changed to a tyrosine in the feather-like sequence. This compares with three cysteines which differ between feather and scale keratin sequences (Gregg et al., 1984). The disulphide bonds formed between the cysteine residues of adjacent keratin monomers are thought to be important for the stabilization of the keratin filaments (Fraser et al., 1972; Fraser and MacRae, 1980) suggesting that the feather keratin and feather-like gene products could fulfil similar roles in keratinized feather epithelia.

Between amino acid 80 and the stop codon, the two conceptual protein sequences diverge considerably. The point at which these two protein sequences diverge, that is, Gly_{80} of the feather-like sequence (Figure 5.5) corresponds exactly to the position at which the two DNA sequences diverge, i.e. base 281 of the feather-like gene, (Figure 5.3, and 5.4). In this C-terminal portion, the feather-like sequence contains a total

of four insertions each of 2-8 amino acids and well as 7 amino acid substitutions relative to the feather keratin sequence (Figure 5.5). Two of the three short segments of significant amino acid sequence homology between the sequences in the C-terminus correspond to the Gly-Leu-Gly-Ser (amino acids 92-95) and Pro-Cys (amino acids 114-115) regions which are conserved between the two DNA sequences (data not shown).

DNA sequence analysis of the 3' half of the coding region of the feather-like gene contained in the 1.1 kb EcoRI fragment of cosmid 12 (Section 4.3.2.2.c) indicated that this gene was very similar to the gene in pE5.0 (data not shown). The BamHI site was located in an identical position as in the pE5.0 gene (see Figure 5.3) and nucleotide sequence differences between the genes were generally restricted to the third base of codons (data not shown).

Sequence analysis of a 100 bp PstI/HindIII fragment enabled the putative polyadenylation signal and site of poly(A) addition in the feather-like gene to be identified. From restriction mapping of the feather-like gene and the location of the translation 'stop' codon within it (Figure 5.1), it was estimated that the length of the 3' non-coding region was 700 bp. This compares with 440-460 bp for the feather keratin genes (Gregg and Rogers, 1986; Gregg et al., manuscript in preparation) and 350 and 362 bp for two scale keratin mRNA-derived clones (Wilton, 1983). The feather-like gene has a canonical AATAAA polyadenylation signal and shows considerable homology downstream of this point with feather keratin genes (Figure 5.6). It is postulated that, by comparison with feather keratin sequences, the poly(A) addition site is 29 bp from the

first base of the polyadenylation signal (Figure 5.6). However, the point of addition of the poly(A) tail is somewhat equivocal as there is an adenine immediately downstream of the putative poly(A) addition site in feather keratin genes and the feather-like sequence. The CAYTG motif, which is thought to be important for correct 3'-end formation (see Section 1.4.4; Berget, 1984, Gil and Proudfoot, 1984), is conserved to some extent in the feather-like gene (AATTA, Figure 5.6). Finally, it is interesting to note that the polyadenylation signal of the feather-like gene is feather-like (AATAAA, Gregg and Rogers, 1986) rather than scale-like (ATTAAA, Wilton, 1983; Wilton et al., 1985).

5.4.2 NATURE AND POSSIBLE ORIGINS OF THE FEATHER-LIKE GENES

The hybridization data obtained with feather keratin probes (Section 4.3.2.2) suggested that the feather-like genes were either variant feather keratin genes, scale or claw keratins, pseudogenes or an uncharacterized type of keratin gene, such as beak.

Partial DNA sequence analysis allowed at least some of these possibilities to be ruled out. Since this gene does not encode a gly-gly-x amino acid repeat characteristic of scale and claw keratins (Wilton, 1983; Wilton et al., 1985; Whitbread, 1985; L. Whitbread and K. Gregg, unpublished data), it would appear that it is not a scale or claw keratin gene. Although no beak keratin genes have been isolated, preliminary protein chemical studies on beak keratins suggested that, like the scale and claw keratins, they contain a gly-gly-x repeating unit (Gibbs, P.E.M., Walker, I.D. and Rogers, G.E., unpublished results). Therefore, the feather-like genes are probably not beak keratins either.

Another possibility is that the feather-like genes are feather keratin pseudogenes, 'evolutionary debris' of the duplication events of these genes. From the sequencing data described in this thesis, it appears that at least one of the feather-like genes is not a pseudogene as it does not have any in-phase 'stop' codons or deletions in its coding region which are characteristic of many pseudogenes (Proudfoot and Maniatis, 1980; Harris et al., 1984; Townes et al., 1984; Frenkel, 1985).

From the hybridization and sequencing data, two alternative schemes are proposed to account for the origin of the three feather-like genes (Figure 5.10).

MODEL A. Gregg et al. (1984) have proposed that the ancestral scale keratin gene gave rise to the primordial feather keratin gene by the deletion of the gly-gly-x region and the other insertions present in the scale sequence (see Section 1.4.6).

It is proposed here that the feather-like gene(s) represent an 'intermediate' form between scale and feather keratin genes which have subsequently evolved into the feather-like genes. This hypothesis is supported by DNA sequencing which indicates that, although the feather-like gene does not contain a gly-gly-x region, it does contain a number of amino acid insertions relative to the feather sequence reminiscent of scale keratins. These insertions, as in the scale genes, predominate in the C-terminal half of the gene and their composition is similar, e.g. rich in tyrosine and proline. The feather keratin genes lack these amino acid insertions characteristic of the feather-like gene and thus the feather keratins represent the final evolutionary stage.

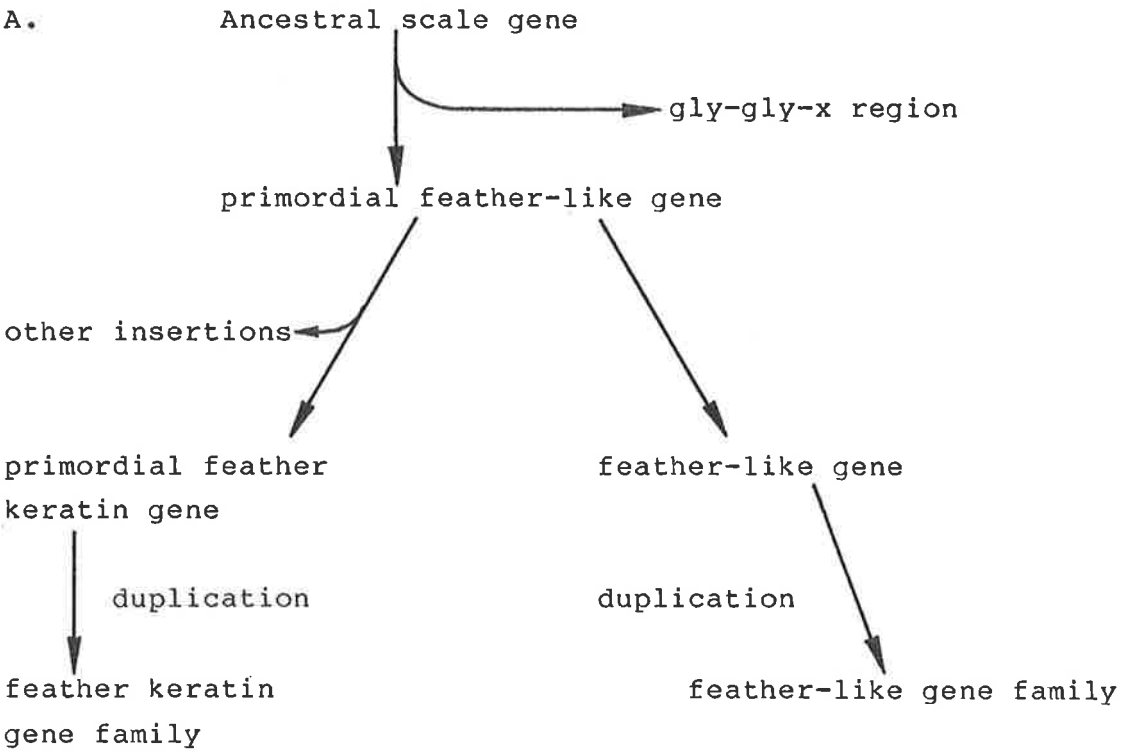
FIGURE 5.10 : POSSIBLE ORIGINS OF THE FEATHER-LIKE GENES

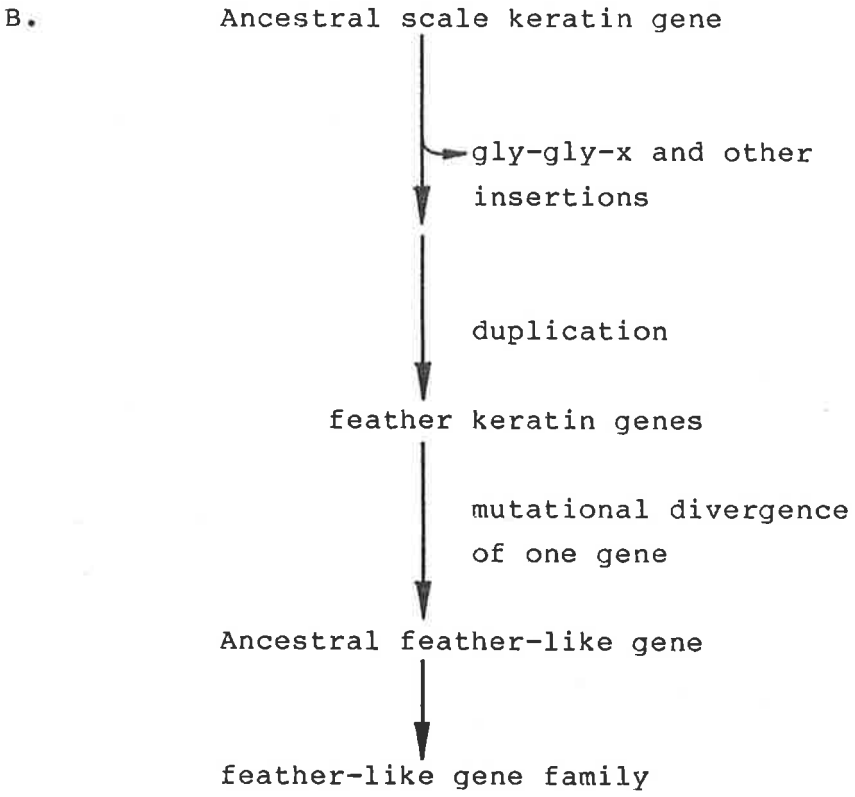
The figure illustrates the two models proposed for the evolution of the feather-like genes.

MODEL A: The first feather-like gene represented an intermediate stage in the transition between scale and feather keratin genes (see text). The feather-like gene was fixed and the gene subsequently duplicated to generate the three feather-like genes. At about the same time, the remaining insertions in the protein coding region were deleted to produce the primordial feather keratin gene.

MODEL B: The first feather-like gene arose from one of the feather keratin genes. This would imply that the feather-like gene(s) evolved much more recently than in Model A.

One of the feather keratin genes (probably at the end of the cluster where the opportunities for gene conversion may be fewer) may have mutated by base substitution/insertion and subsequently diverged, such that in subsequent gene conversion events (Section 7.3.1), poor alignment of sequences restricted homogenization. This primordial feather-like gene continued to diverge, particularly outside the protein coding region, where the gene shows little apparent homology with feather keratin genes. The feather-like gene subsequently duplicated to generate the three feather-like genes. Further correction events occurred between these three genes, perhaps restricting the evolutionary drift.


A.
Ancestral scale gene
gly-gly-x region
primordial feather-like gene
other insertions
primordial feather keratin gene
feather-like gene
duplication
duplication
feather keratin gene family
feather-like gene family


B.
Ancestral scale keratin gene
gly-gly-x and other insertions
duplication
feather keratin genes
mutational divergence of one gene
Ancestral feather-like gene
feather-like gene family

MODEL B. The second model to account for the origin of the feather-like genes is displayed in Figure 5.10b. In this hypothesis, it is proposed that the feather-like genes arose from the feather keratin genes after the appearance of the primordial feather keratin gene.

Of the proposed models to account for the origin of the feather-like genes, Model B appears to be the more likely. The evidence for this is as follows. Firstly, the coding region of feather-like gene shows a higher degree of sequence homology with feather keratins than scale or claw keratins (Section 5.4.1). Allowing for a number of insertions in the feather-like sequence relative to feather keratins, the coding regions of the two genes are 75-80% homologous at the DNA and protein sequence levels (Section 5.3.1). Secondly, the feather-like gene has a feather-type polyadenylation signal (AATAAA, Figure 5.6) rather than that which is characteristic of the scale genes sequenced to date (ATTAAA, Wilton, 1983; Wilton et al., 1985). Thirdly, the feather-like genes are located in close proximity to the feather keratin gene cluster (Section 4.4.2, Figure 4.45), favouring the idea that they evolved from the feather keratin genes. Taken together, these results favour the idea (Figure 5.10b) that the feather-like genes evolved from the feather keratins by mutational divergence and subsequent gene duplication.

5.4.3 'THE Kpn FAMILY': APPARENT CONSERVATION OF A KpnI SITE

A number of feather keratin genes in cosmids 4 and 33 contain a KpnI site (e.g. Figure 4.13) and gene mapping studies indicated that, at least in some of these genes, the KpnI site is located between the initiation codon and the conserved

segment in the middle of the 3' non-coding region (see Section 4.3.4.2a,b). Restriction mapping and DNA sequence analysis of gene E of λCFK1, which is partially contained in this λ recombinant (Figure 4.3; Gregg et al., manuscript in preparation), had demonstrated that this gene contains a KpnI site which is located in the protein coding region 9 bp from the stop codon. The results presented here show that, in at least two other genes belonging to the 'Kpn family', the KpnI site is in an identical location to that of gene E (Figure 5.7a,b). Blot hybridization studies indicate that in at least some of the other members of 'the Kpn family', the KpnI site is probably located in an identical position.

This data adds weight to the proposal (Section 4.4.2.1) that the 'Kpn family' genes are the result of a recent duplication and further that the KpnI site is perhaps being maintained by some recombination mechanism such as gene conversion (Baltimore, 1981; Kourilsky, 1986). Comparison of the amino acid sequences of the genes which do not contain a KpnI site with those that do show that they differ by a single amino acid (Figure 5.7c). The significance of this change (from a cysteine to a tyrosine residue) with respect to protein function, if any, is not known. It would be interesting to examine the intergenic regions between members of the 'Kpn family' by either DNA sequencing and/or Southern hybridization to see whether these regions contained highly conserved segments or short direct repeats which are often associated with gene duplication events (see Shen et al., 1981; Wang et al., 1985).

5.4.4 THE 3' SPLICE SITE OF THE GENE LOCATED IN THE 2.7 kb HindIII FRAGMENT OF COSMID 4

Of the 13 HindIII fragments in cosmids 4 and 33 that contained feather keratin genes, only one failed to hybridize with the oligonucleotide prepared from the conserved 3' end of the intron (see Section 4.3.4.1, Figure 4.31c). Hybridization of these cosmids, and subclones prepared from them indicated that the only gene which did not bind the probe was that contained in the subclone pH 2.7. The data presented in Figure 5.8 shows that the reason why no hybridization was detected by this intron probe was that there were a number of base substitutions in the region of the gene spanned by this oligonucleotide. Because of the compressions in the sequencing ladder (arrows in Figure 5.8a), it was not possible to unequivocally determine the DNA sequence of this region. However, it was apparent that, as expected, the gene appears to contain an intron 21 bp 5' from the initiation codon (Figure 5.8c).

The significance of the base changes in the intron sequence of this gene is not known. It is of interest to note that although the actual DNA sequence in this gene has altered, the intron sequence is still very pyrimidine rich, a feature commonly observed around splice junctions (Breathnach and Chambon, 1981; Mount, 1982). It is interesting to note that this intron oligomer does not hybridize with the claw keratin genes in cosmid 33 (Figure 4.31c) and shows little DNA sequence homology with the feather-like gene (Figure 5.4.1).

5.4.5 SEQUENCE ANALYSIS OF THE 5' PROMOTER OF THE pBH2.5 GENE

DNA sequence analysis of part of the 0.45 kb PstI/HindIII fragment of the pBH2.5 subclone (Section 4.3.4.2d) enabled the exact location of this gene to be determined and confirmed its orientation in the subclone (Figure 4.40d). The PstI site was found to be located in the intron, 11 bp from the 5' splice site (Figure 5.9a). The gene contains a 5' non-coding exon which is identical in length and sequence to gene C (Figure 5.9b). This 37 bp 5' non-coding exon is the most highly conserved segment of feather keratin genes, with only one base change between the five known sequences including the sequence reported here (Section 1.4.4, Figure 1.8b and this work, Figure 5.9b).

The probable cap site could be located by comparison with the available genomic sequences from λCFK1 (Figure 5.9a,b; base 231). The gene contains a TATA box located 31 bp 5' to the cap site (base 200) and two potential CAAT boxes located 74 (base 157, CATT) and 99 bp (base 132, CAAAT) upstream from the cap site (Figure 5.9a). Comparison of the 5' flanking regions of the gene sequenced in this study to feather keratin gene B from λCFK1 gave a similar high level of homology as has been reported previously between genes A and B of λCFK1 (see Figure 1.8a, Gregg and Rogers, 1986). However, as pointed out in Section 1.4.4, because of the similarity of these genes it is not possible from such analyzes to determine whether the observed sequence homology is a result of tandem gene duplication or conservation of sequences which are important for tissue-specific expression of the keratin genes, as observed for other eukaryotic genes (e.g. Karin et al., 1984; Parker and

Topol, 1984). An alternative approach to identifying possible regulatory sequences was to compare the 5' flanking regions of feather keratins and a fast protein gene, which are co-ordinately expressed in the developing chick feather (Powell and Rogers, 1979). This comparison is described in the following section.

5.4.6 COMPARISON OF THE 5' FLANKING REGIONS OF FEATHER KERATIN AND FAST PROTEIN GENES

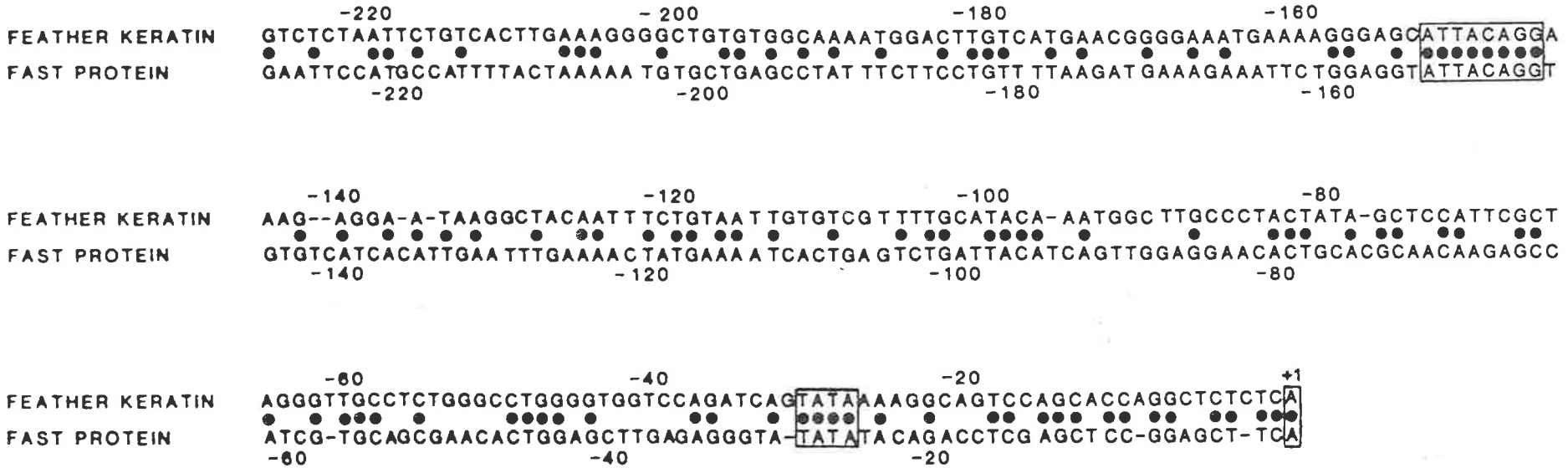
The feather keratins and fast protein genes are co-ordinately expressed in the developing chick feather (Powell and Rogers, 1979; Morris, 1984). Comparison of the 5' flanking regions of these two gene types may reveal DNA sequences which are conserved, and might therefore play a functional role in the tissue-specific regulation of gene transcription, as has been found for other genes.

Figure 5.10 presents a comparison of the 5' flanking regions of a feather keratin gene (Figure 5.9a) and the fast protein gene sequenced by Morris (1984). Inspection of the figure shows that there is one region of 8 bp which is conserved between the two sequences located between -146 and -153 of the feather keratin sequence relative to the cap site at +1 (Figure 5.10). There are some smaller regions of 3-4 bp which are similar, although apart from the TATA box at -30, these are unlikely to be of any significance. This 8 bp sequence is conserved in 7/8 positions in all four flanking sequences available, that is, three feather keratin genes and one fast protein gene. In gene B, there is a one base change

FIGURE 5.11 : COMPARISON OF THE 5' FLANKING REGIONS OF FEATHER KERATIN AND FAST PROTEIN GENES

The nucleotide sequence of the 5' flanking region of the feather gene sequenced in this work (the pBH2.5 gene, Figure 5.9a) is compared to that of a fast protein gene (Morris, 1984). Deletions have been introduced by the computer program SEQA (Section 5.2.8) to maximize the homology.

Negative numbers refer to the number of nucleotides upstream from the cap site at +1, which is boxed. The TATA box at about -30 and the 8 bp region at about -150 (the -150 box), which are conserved between the feather keratin and fast protein sequences, are highlighted by boxes (see text).


-220
-200
-180
-160
FEATHER KERATIN
GTCTCTAATTCTGTCACTTGAAAGGGGCTGTGTGGCAAAATGGACTTGTCATGAACGGGGAAATGAAAAGGGAGCATTACAGGA
FAST PROTEIN
GAATTCCATGCCATTTTACTAAAAA TGTGCTGAGCCTAT TTCTTCCTGTT TTAAGAT GAAAGAAATTC TGGAGGT ATTACAGGT
-220
-200
-180
-160
-140
-120
-100
-80
FEATHER KERATIN
AAG--AGGA-A-TAAGGCTACAATT TCTGTAA T TGTGTCG T TTTGCATACA- AATGGC TTGCCCTACTATA-GCTCCATTCGCT
FAST PROTEIN
GTGTCATCACATTGAA TTTGAAAAC TATGAAA A TCACTGA GTCTGATTACATCAGTTGGAGGAACACTGCACGCAACAAGAGCC
-140
-120
-100
-80
-60
-40
-20
+1
FEATHER KERATIN
AGGGTTGCCTCTGGGCCTGGGGTGGTCCAGATCAGTATAAAAGGCAGTCCAGCACCAGGCTCTCTCA
FAST PROTEIN
ATCG-TGCAGCGAACACTGGAGCTTGAGAGGGTA-TATATA CAGACCTCG AGCT CC-GGAGCT- TCA
-60
-40
-20

in the 8 bp sequence. The 8 bp consensus sequence for the '-150 box' is:

$$5' \text{ ATT}\begin{matrix}\text{A}\\\text{G}\end{matrix}\text{CAGG } 3'$$

The functional significance of this sequence, if any, is not known, although it may be important in the co-ordinate transcriptional regulation of the feather keratin and fast protein genes. It is interesting to note that gene B, which has a one bp change in the conserved 8 bp sequence also has a one bp change in the otherwise strictly conserved 37 bp 5' non-coding exon of the 5' non-coding region (Figure 1.8a; Gregg and Rogers, 1986).

CHAPTER 6.

SEARCH FOR A PRECURSOR TO

CHICKEN FEATHER KERATIN mRNA.

6.1 INTRODUCTION

Since the discovery that an adenovirus gene contained an intervening sequence or intron in its coding region (Sambrook, 1977), most eukaryotic genes have been found to contain them (Breathnach and Chambon, 1981) with a number of exceptions (e.g. most histone genes; Hentschel and Birnstiel, 1981). In general, introns have been found to occur in the coding region of eukaryotic genes (Breathnach and Chambon, 1981; Blake, 1985) and in a few cases the 5' non-coding region (Benoist et al., 1980; Nunberg et al., 1980; Young et al., 1981; Molloy et al., 1982; Bernard et al., 1983; Steiner et al., 1985). Gilbert (1978) proposed that the split gene organization found in eukaryotes may be a mechanism for allowing the rapid evolution of new proteins. He proposed that exons correspond to protein functions which can be recombined or reshuffled within the introns between the coding portions to reassort these functions and produce new kinds of proteins from parts of existing ones. Furthermore, the reading of the gene over a larger region of DNA would considerably enhance the probability of recombination. This hypothesis was extended by Blake (1978) who suggested that if exons also correspond to folded protein structures or domains, the problem of folding the new protein would be made easier. Accumulating evidence suggests that this hypothesis may be at least partly correct, although exceptions have been found. (For reviews see Craik et al., 1982; Blake, 1983, 1985; Cornish-Bowden, 1985; Rogers, 1985).

The function of introns which are present in the 5' non-coding regions of a number of eukaryotic genes (Benoist et al., 1980, Young et al., 1981; Molloy et al., 1982; Bernard et al.,

1983; Steiner et al., 1985) is unknown except in the case of the α-amylase gene (Young et al., 1981), where it was found that a different promoter and 5' non-coding region was used in the two tissues where the gene was expressed (the salivary gland and liver), giving rise to two different mRNAs. It is interesting to note that no eukaryotic genes sequenced to date have introns in the 3' non-coding regions, indicating the lesser importance in evolutionary terms of that part of eukaryotic genes compared to the protein coding and 5' non-coding regions.

The introns of probably all eukaryotic genes are bounded at both ends by the consensus sequence 5' AGGT 3' (Breathnach and Chambon, 1981) which is thought to be involved in the splicing of pre-mRNAs (Keller, 1984). Keller and Noon (1984) have noticed another consensus sequence CTPuAPy (where Pu is a purine and Py is a pyrimidine) present in the introns of a number of vertebrate genes near the 3' splice site which has been proposed to base-pair with a complementary sequence downstream of the 5' splice site, forming the RNA splicing intermediates known as lariats (Keller, 1984).

Experiments in the late 1970's demonstrated that eukaryotic genes are transcribed initially into mRNA precursors (pre-mRNAs) up to four times larger than the mature (cytoplasmic) mRNAs which were thought to contain the introns present in the chromosomal genes (Macnaughton et al., 1974; Kwan et al., 1977; Rabbitts, 1978; Roop et al., 1978; Ross and Knecht, 1978; Strair et al., 1978; Tilghman et al., 1978a, b). These introns were thought to be excised and the exons ligated together by a mechanism termed splicing. The approach taken by these workers

was to isolate whole cell or nuclear RNA, separate the RNA species by agarose gel electrophoresis or on denaturing sucrose gradients and attempt to identify high molecular weight pre-mRNAs by hybridization with specific cDNA probes. Using this approach, and with little or no knowledge of the fine gene structure, nuclear RNA transcripts significantly larger than the cytoplasmic mRNA were reported for the immunoglobulins (Schibler et al., 1978; Rabbitts, 1978), ovalbumin (Roop et al., 1978), albumin (Strair et al., 1978), globin (Kwan et al., 1977; Ross and Knecht, 1978) and milk proteins (Bathurst et al., 1979). Tilghman et al. (1978b) demonstrated directly the presence of introns in a cloned mouse β-globin gene by hybridization to mature globin mRNA and visualization of the R loops under the electron microscope. The absence of any R loops when purified precursor globin mRNA was hybridized to the genomic clone was consistent with the introns being present in this RNA species.

The genomic clone λCFK1 contains four complete feather keratin genes and part of a fifth one, all of which have been sequenced (Molloy et al., 1982; Gregg and Rogers, 1986; Gregg et al., manuscript in preparation). Perhaps the most interesting feature of these genes was the discovery that all of them contain an intron of between 320-340 bp in the 5' non-coding region, which interrupts the genes 21 or 22 bases prior to the initiation codon (see Figure 1.5b). By analogy with other studies, it was thought that a precursor for feather keratin mRNA would be present in detectable levels in either whole cell or nuclear RNA extracts. This chapter describes attempts to identify and partially characterize the putative precursor mRNA

for feather keratin. In an earlier study using total feather cDNA, Gibbs (1977) could not detect a nuclear precursor for feather keratin mRNA which was larger than the 12S cytoplasmic feather keratin mRNA. With the availability of more sensitive molecular techniques and the isolation of feather keratin genomic sequences which indicated that these genes contain an intron (Gregg and Rogers, 1986), another attempt to identify the pre-mRNA for feather keratin was considered to be justified.

6.2 METHODS

6.2.1. RNA PREPARATION

Total RNA was isolated as described by Brooker et al. (1980), but with some modifications.

Feather tissue plucked from about 20 14-day chick embryos (Section 2.1.1) was homogenized in 50 ml of ice cold 6 M guanidine-HCl, 0.2 M Na acetate pH 5.2, 10 mM β-mercaptoethanol using a motor-driven Potter-Elvehjem teflon homogenizer. Two volumes of pre-chilled ethanol were added to the homogenate and after standing for at least 2 hours at -20°C, the mixture was centrifuged at 12,000 rpm for 20 minutes at 4°C to recover the precipitate. The pellet was resuspended in about 30 ml of 6 M guanidine-HCl, 0.2 M Na acetate pH 5.2, 10 mM EDTA and again ethanol precipitated as described above. The pellet was resuspended in 30 ml of freshly made 7 M urea, 0.1 M Tris-HCl pH 8.5, 0.1 mM EDTA and 0.1% SDS and extracted twice with an equal volume of phenol/chloroform (1:1), once with chloroform and then ether and finally ethanol precipitated.

To remove high molecular weight DNA from the preparation, a selective high salt precipitation was performed according to

a procedure reported by Diaz-Ruis and Kaper (1978). An equal volume of 4 M LiCl was added to the nucleic acid solution (resuspended in water) and left overnight at 4°C. The high molecular weight RNA (greater than 4S) becomes insoluble and can be recovered by centrifugation (15,000 r.p.m., 15 minutes, 4°C). The pellet was then washed with 70% ethanol, before a final precipitation to remove excess salt. The RNA was stored in sterile H_2O at -20°C.

6.2.2 ISOLATION OF POLY A^+ RNA

Total RNA isolated from feather tissue was fractionated to purify the poly A^+ RNA by oligo (dT)-cellulose chromatography using a modification of the method described by Aviv and Leder (1972). Briefly, RNA in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1% SDS was dissociated by heating at 65°C for 3 minutes and then chilled on ice to prevent reaggregation of RNA chains. The RNA was adjusted to 0.5 M NaCl and layered on to a 3 x 0.5 cm column of oligo (dT)-cellulose which had previously been equilibrated with binding buffer (0.5 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA). The column was washed with binding buffer until the A_{254} (absorbance at 254 nm, 1 cm path length) had returned to the zero baseline. The bound poly A^+ RNA was then eluted from the column using 10 mM Tris-HCl pH 7.5, 1 mM EDTA and precipitated by the addition of Na acetate and ethanol (Section 2.2.2). RNA samples were resuspended in sterile double distilled water and stored at -20°C.

6.2.3 CELL-FREE TRANSLATION OF RNA

A commercial nuclease-treated rabbit reticulocyte lysate was used as specified by the manufacturer (Amersham). Poly A^+ RNA (0.25-2 μg) was incubated with the lysate and [^{3}H]leucine for 1 hour at 30°C.

The incorporation of [^{3}H]leucine into protein during cell-free translation was determined as instructed by the manufacturer. Briefly, duplicate aliquots of 1 µl were added to 1 M NaOH, 5% H_2O_2 (0.5 ml) and heated at 37°C for 15 minutes, the solution cooled and the proteins precipitated with 25% TCA, 2% casamino acids (3 ml). After cooling on ice, the proteins were collected on GF/A filters and washed with 5% TCA, ethanol and finally ether. The filters were immersed in Triton/toluene scintillant and counted in a Packard liquid scintillation counter.

6.2.4 S-CARBOXYMETHYLATION OF TRANSLATION PRODUCTS

The translation products were S-carboxymethylated as described by Frenkel (1985). The lysate sample (30 µl) was mixed with 24 mg urea and 5 µl of reduction buffer (0.5 M Tris-HCl, pH 9.5, 0.5 M DTT) and incubated for several minutes at room temperature. The proteins were then alkylated by adding 25 µl of a solution containing 30% iodoacetic acid and 3 M Tris-HCl pH 8.0 and incubating at room temperature for 20 minutes. The reaction was terminated by the addition of 2 µl β-mercaptoethanol. The samples were then ready for loading on to acrylamide gels.

6.2.5 POLYACRYLAMIDE GEL ELECTROPHORESIS AND FLUOROGRAPHY OF TRANSLATION PRODUCTS

The S-carboxymethylated translation products were electrophoresed on 1 mm thick 15% polyacrylamide slab gels (bis:acrylamide ratio of 1:100) as described by Laemmli (1970) except that the stacking gel was omitted. Electrophoresis was carried out at 15 mA until the bromophenol blue tracker dye had almost reached the bottom of the gel (usually 9-12 hours). The radio-

labelled proteins were detected in the gel using the fluorographic procedure of Laskey and Mills (1975).

6.2.6 ISOLATION OF FEATHER NUCLEI AND PREPARATION OF NUCLEAR RNA

Nuclei were prepared from the feather tissue of 14-day old chick embryos using the method of Marshall and Burgoyne (1976). The tissue was homogenized in the sucrose medium described by these workers using a Sorvall Omnimix (4 x 15 seconds at full speed) and was filtered through four layers of muslin cloth to remove large cell debris and unhomogenized material. The nuclei were purified by centrifugation as described by Marshall and Burgoyne (1976) and stored as a pellet at -20°C. The RNA was prepared from the feather nuclei as described by Bathurst et al. (1980a).

6.2.7 ELECTROPHORESIS OF RNA ON AGAROSE GELS

RNA samples were analyzed for their integrity on 1.2% agarose gels in TAE buffer (Section 2.2.7).

RNA samples (1-10 µg) destined for hybridization experiments were fractionated on 1.2% slab horizontal (20 x 20 cm x 2.5 mm) agarose gels after denaturation with glyoxal exactly as described by Thomas (1980) except that DMSO was omitted from the denaturation step. The electrode buffer (10 mM Na phosphate pH 6.5-7.0) was recirculated to maintain a near constant pH and thus prevent glyoxal dissociation from the RNA. Electrophoresis was normally performed for 4-5 hours at 40-60 mA.

6.2.8 NORTHERN TRANSFER PROCEDURE

RNA fractionated on 1.2% agarose gels (Section 6.2.7) was transferred to nitrocellulose using the procedure described by Thomas (1980, 1983).

Briefly, following electrophoresis, the gel on which duplicate samples were run was cut in half; one half was stained with ethidium bromide and visualized under U.V. light (see Section 2.2.7) while the other half was retained for transfer. This was done in view of the report (Thomas, 1983) that ethidium bromide reduces the efficiency of transfer. The other half of the gel, without prior treatment, was overlain with a single sheet of nitrocellulose and three layers of Whatman 3 MM paper which had been pre-soaked in 20 x SSC (Section 2.1.7) and the transfer was performed as described (Thomas, 1980, 1983). After baking the filters at 80°C for 2 hours *in vacuo*, they were washed in boiling 20 mM Tris-HCl pH 8.0 for 5 minutes to remove residual glyoxal (Thomas, 1983). The Northern filters were hybridized with radioactive probes and washed as described in Section 6.3.3.

6.3 RESULTS

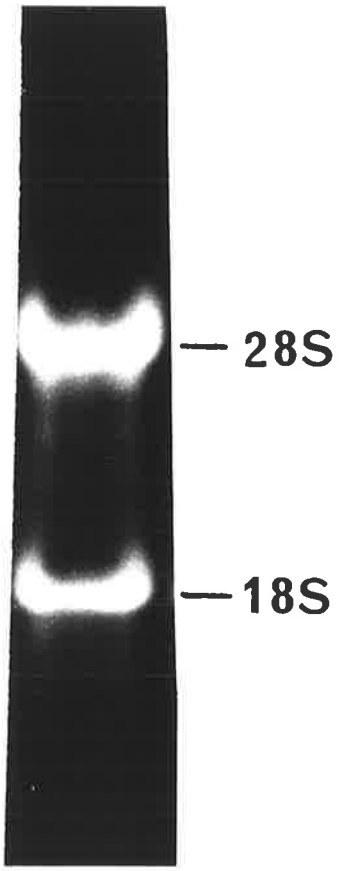
6.3.1 ISOLATION OF RNA

Two different methods of isolating whole cell RNA were attempted. Initially, a method involving homogenization of the feather tissue in a guanidine-HCl extraction buffer followed by centrifugation through a 5.7 M cushion of CsCl was tried (Glisin *et al.*, 1974; Kaplan *et al.*, 1979). However, poor yields of total RNA which were often largely degraded (as judged by the intensity of the ribosomal RNA bands on an agarose gel) were obtained (data not shown). The alternative method now commonly used in our laboratory is that of Brooker *et al.* (1980) described in the methods (Section 6.2.1). Routinely, yields of about 2 mg of total feather RNA from 12 14-day old chick embryos were obtained.

FIGURE 6.1 : ANALYSIS OF TOTAL FEATHER RNA BY AGAROSE GEL ELECTROPHORESIS

Total cellular RNA was isolated from 14-day embryonic feather tissue as described in Section 6.2.1 and a 4 µg sample electrophoresed on a 1.2% agarose gel in TAE buffer as described in Section 2.2.7.

The figure shows the profile obtained with the 28S and 18S rRNA bands being clearly visible.


28S
18S

The RNA prepared by this method appeared to be undegraded, as judged by agarose gel electrophoresis (Figure 6.1), since the larger 28S species was more intense than the 18S species. In order to enrich for polyadenylated RNA, some of the total RNA preparation was fractionated on oligo (dT)-cellulose and the bound poly A^+ RNA (about 1.5% of the total) was collected.

6.3.2 CELL-FREE TRANSLATION OF FEATHER RNA AND PRODUCT ANALYSIS

Whole cell poly A^+ RNA isolated from feather tissue as described (Sections 6.2.1, 6.2.2) was assayed for its translational activity using a reticulocyte lysate by determining total incorporation of [^{3}H]leucine into acid precipitable protein (Section 6.2.3). The RNA preparation stimulated protein synthesis up to about five fold above the lysate (-RNA) control with the maximum incorporation occurring with about 1.5 µg of added RNA. In order to demonstrate that the radiolabelled proteins were mostly feather keratins, the translation products were S-carboxymethylated and fractionated on a denaturing Laemmli gel after which the proteins were visualized by fluorography (Sections 6.2.4, 6.2.5). Figure 6.2 clearly shows that a prominent band is present in the 10 kD region of the gel which is the size of feather keratin proteins (Gregg and Rogers, 1986); this band accounts for more than 90% of the radioactivity incorporated.

6.3.3 ATTEMPTS TO DETECT FEATHER KERATIN PRE-mRNA IN RNA PREPARATIONS FROM EMBRYONIC FEATHER TISSUE

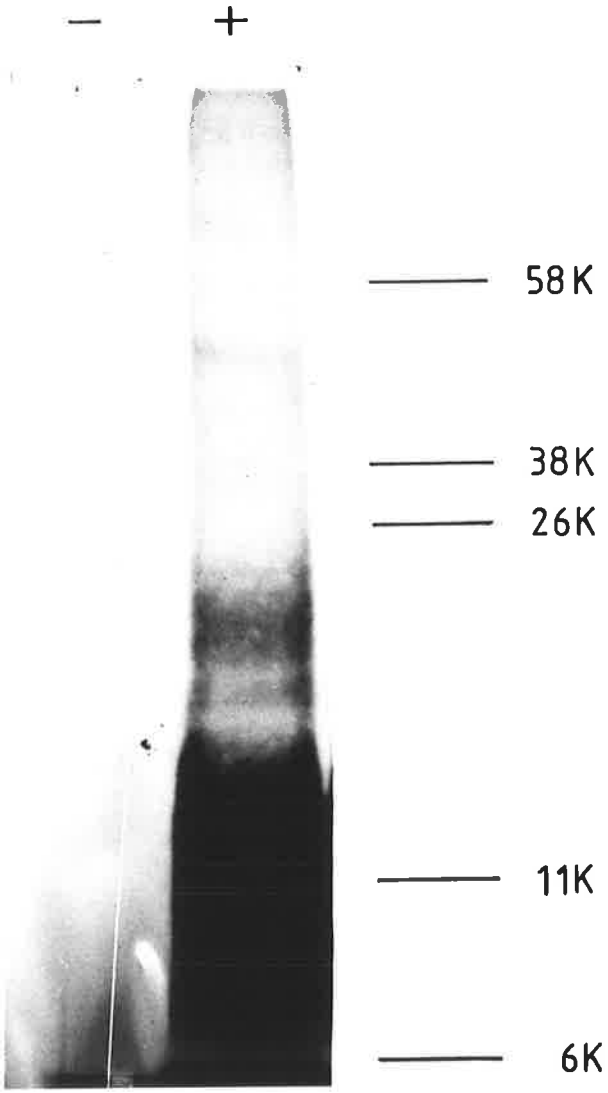
This section describes experiments designed to detect feather keratin pre-mRNA molecules (i.e. unspliced transcripts) in total RNA preparations from keratinizing feather tissue.

FIGURE 6.2. : ELECTROPHORETIC ANALYSIS OF THE IN VITRO TRANSLATION PRODUCTS DIRECTED BY FEATHER mRNA

Poly A^+ RNA (1.5 µg) prepared from 14-day embryonic feather tissue (Section 6.2.1, 6.2.2) was used to programme a rabbit reticulocyte lysate using [^{3}H]leucine as the radioactive amino acid (Section 6.2.3). The ^{3}H-labelled translation products were S-carboxymethylated, fractionated on a 15% SDS/polyacrylamide gel and the radiolabelled proteins visualized by fluorography as described in Sections 6.2.4 and 6.2.5. The molecular weight markers shown were ^{14}C-labelled wool keratins kindly provided by Dr. E. Kuczek.

The tracks show translation assays which contained:

- \- no added RNA
- \+ 1.5 µg feather poly A^+ RNA


−
+
58K
38K
26K
11K
6K

The approach involved using a whole feather keratin gene and later intron-specific DNA sequences as hybridization probes to Northern blots of total and poly A^+ RNAs isolated from whole feather tissue.

In the initial RNA transfer experiments it was important to determine that the RNA had transferred efficiently to nitrocellulose. This was determined in three ways. Firstly, the agarose gels were stained with ethidium bromide after transfer was complete, i.e. after removing the nitrocellulose filter for baking. Generally no RNA was visible on the gel. However, sometimes at the highest RNA concentration transferred (10 µg) a small amount of fluorescence was visible. Secondly, a Northern filter was probed with ribosomal cDNA made from poly A^- RNA from chick liver (Figure 6.3). It can be seen that this ribosomal probe detects complementary sequences in the feather RNA and it was concluded that the RNA samples had indeed transferred to nitrocellulose. As a final test, a Northern filter was probed with feather keratin gene B from λCFK1 (Section 2.2.10), the result of which is illustrated in Figure 6.4. An intense band is visible at approximately 12S, the size of feather keratin mRNA (Partington et al., 1973) in both total RNA (track A) and poly A^+ RNA (track B) preparations from embryonic feather tissue; no hybridization was detected to the ribosomal standards used as markers or to liver RNA (data not shown).

In Figure 6.4, the approximate position at which the pre-mRNA would be expected to run is indicated (~ 14S). It is apparent that no band corresponding to the expected size of the feather keratin pre-mRNA is visible. In order to confirm this,

FIGURE 6.3 : DETECTION OF SEQUENCES COMPLEMENTARY TO RIBOSOMAL cDNA IN TOTAL FEATHER RNA

Total feather RNA (~ 1 µg), isolated as described in Section 6.2.1, was glyoxylated, fractionated on a 1.2% agarose gel and transferred to nitrocellulose (Sections 6.2.7, 6.2.8). The Northern filter was hybridized with a ribosomal cDNA probe prepared from the poly A^- fraction of chicken liver RNA (a gift from Dr. Steve Wilton). Following hybridization, the filter was washed in 0.1 x SSC, 0.1% SDS at 65°C and autoradiographed at -80°C for 2 days (see Section 2.2.12).

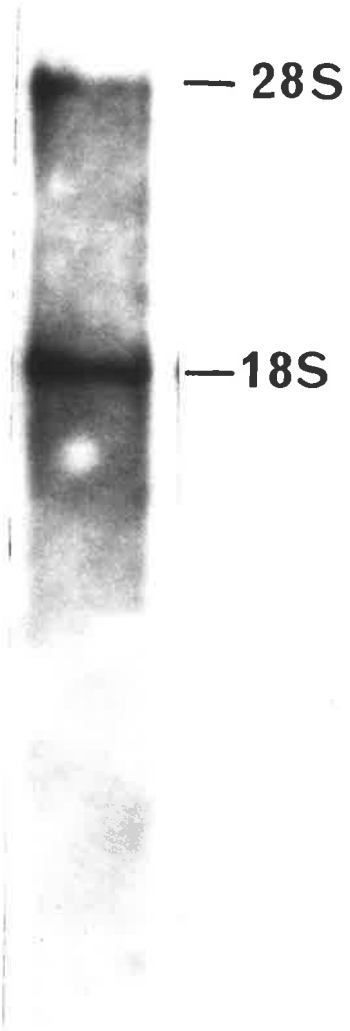

— 28S
—18S

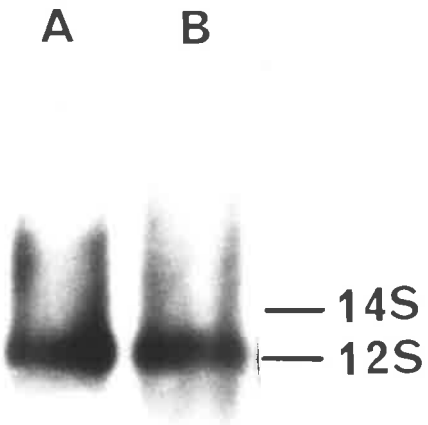
FIGURE 6.4 : DETECTION OF FEATHER KERATIN SEQUENCES IN FEATHER RNA BY NORTHERN BLOT ANALYSIS

Total and poly A^+ RNAs isolated from 14-day embryonic feather tissue as described (Sections 6.2.1, 6.2.2), was glyoxylated, fractionated on 1.2% agarose gels and blotted on to nitrocellulose filters (Sections 6.2.7 and 6.2.8). The filter was hybridized with a complete feather keratin gene probe (gene B from λCFK1, see Section 2.2.10), washed in 0.5 x SSC, 0.1% SDS at 65°C and exposed overnight at -80°C (Section 2.2.12). The 12S and 14S markers indicate the size of cytoplasmic feather keratin mRNA and the putative precursor feather keratin mRNA, respectively (see text).

The tracks contained:

A - 6 µg total feather RNA

B - 1.5 µg feather poly A^+ RNA


A
B
14S
12S

an intron-specific probe was prepared by digestion of the gene B-containing subclone p λCFK1-4 with a combination of Pst I and Sal I (see Figures 2.1 and 2.2). The small 0.33 kb Pst I/Sal I fragment, which contains the entire intron of gene B, was subsequently isolated by polyacrylamide gel electrophoresis (Figure 6.5), labelled by nick translation (Section 2.2.11.2) and used to probe a Northern filter containing varying amounts (0.5-10 μg) of whole cell and poly A^+ RNA from 14-day feathers. No hybridization was detected to either total cell RNA or poly A^+ RNA, even after long exposure times (3 weeks at -80°C). Similar results were obtained with several RNA preparations.

Later in the project a DNA synthesizer became available in the Department (see Section 2.1.6). Using the 'Gene Machine', a 19-mer was synthesized from the 3' terminal region of the feather keratin intron which includes the 3' splice site (see Section 2.2.10). This 19 base sequence is highly conserved amongst the four genes sequenced (Gregg and Rogers, 1986). The 19-mer was kinased using γ-ATP and polynucleotide kinase (Section 2.2.11.3), purified by electrophoresis on a 20% polyacrylamide gel to remove unincorporated radioactivity (Section 2.2.5) and hybridized to a filter containing whole cell RNA from feather tissue as well as poly A^+ RNA. No hybridization of this oligonucleotide probe was observed to the RNA, even though in a control experiment this probe strongly bound to the 0.33 kb Pst I/Sal I fragment from p λCFK1-4 which contains the intron of gene B (data not shown). It appears then that the level of unspliced feather keratin mRNA in whole cell RNA extracts is below the level of detection of this method of assay. In an attempt to obtain RNA preparations

FIGURE 6.5 : ISOLATION OF AN INTRON-SPECIFIC RESTRICTION FRAGMENT FROM A Pst I/Sal I DIGEST OF p λCFK1-4

p λCFK1-4 DNA (5 μg), a subclone from the genomic clone λCFK1 which contains gene B (Section 2.2.10), was restricted with Pst I and Sal I and the products electrophoresed on a 5% polyacrylamide gel (Section 2.2.5). The small 0.33 kb Pst I/Sal I fragment (arrow) which contains the intron of gene B, was excised and isolated from the acrylamide gel as described (Section 2.2.6).

which ought to be enriched for pre-mRNA molecules, nuclei were prepared from feather tissue and Northern blots of the nuclear RNA probed with the intron-specific probes. The results of these experiments are described in the next section.

6.3.4 ISOLATION OF NUCLEAR RNA AND ATTEMPTS TO DETECT UNPROCESSED TRANSCRIPTS

Considerable difficulty was experienced in obtaining nuclei preparations which were not contaminated with keratin or feathers. This was partially overcome by increasing the homogenization time and filtering the extract through the muslin cloth prior to the centrifugation steps. However, this problem was never completely overcome. Giemsa staining of the final nuclei pellet after resuspension indicated that the majority of the material consisted of intact nuclei (data not shown).

A low yield of RNA, prepared from the nuclei by phenol/chloroform extraction after lysis in SDS and proteinase K (Section 6.2.6), was obtained and it was somewhat degraded. A second preparation of nuclear RNA gave a similar poor yield and was largely degraded. Attempts to demonstrate the presence of a precursor for feather keratin in these nuclear RNA preparations using the two intron-specific gene probes described in Section 6.3.3 were unsuccessful. Insufficient time was available to further investigate the reasons for the poor quality and low yields of nuclear RNA obtained from the feather tissue. At this time the study was concluded.

6.4 DISCUSSION

This chapter describes an attempt to detect the pre-mRNA for feather keratin by Northern analysis of whole cell RNA,

poly A^+ RNA and nuclear RNA from embryonic feather tissue with intron-specific probes. The intron has been shown to be present in four of the genes of the genomic clone λCFK1 by DNA sequencing. In each of the genes the intron splits the 5' non-coding region 37 bp from the transcription start site (see Figure 1.5b; Molloy et al., 1982; Gregg and Rogers, 1986; Gregg et al., manuscript in preparation).

Initially, whole cell RNA was isolated from embryonic feather tissue and the poly A^+ RNA was purified by affinity chromatography (Section 6.3.1). Since feather keratin mRNA constitutes more than 50% of the transcripts present in 14-day embryonic feather tissue (Powell et al., 1976), and as whole cell RNA would include nuclear RNA, it was considered that the pre-mRNA ought to be present in detectable quantities in such a preparation. It is apparent from a number of studies of abundant mRNAs that a significant proportion of these pre-mRNAs present in the nucleus are polyadenylated (Macnaughton et al., 1974; Ross and Knecht, 1978; Schibler et al., 1978) implying that polyadenylation occurs prior to processing, although exceptions have been reported (Bathurst et al., 1980b). This is consistent with the observation (Bantle and Hahn, 1976; Hahn et al., 1978) that a considerable proportion of most heteronuclear RNAs are polyadenylated. Therefore, in order to enrich for mRNAs which would include pre-mRNAs, the whole cell RNA preparations were passaged through an oligo(dT)-cellulose column. Before carrying out this enrichment step, the RNA preparation was first analyzed on an agarose gel which showed that it was not degraded (Figure 6.1).

However, hybridization of two different intron-specific probes (a restriction fragment containing the entire intron of gene B and an oligonucleotide made from the 3' end of the feather keratin intron) to Northern filters containing up to 10 µg of both whole cell and poly A^+ RNA failed to detect the presence of any transcripts containing the intron. Several lines of evidence suggested that the whole intron probe from gene B ought to have been suitable for the detection of feather keratin pre-mRNA. Firstly, it is known that gene B is expressed in 14-day embryonic feather tissue (Gregg et al., manuscript in preparation) and therefore it was expected that detectable levels of pre-mRNA for gene B would be present in embryonic feather tissue. Secondly, this whole intron probe cross-hybridizes, albeit weakly, to all the Hind III restriction fragments in cosmid 4 which contain feather keratin genes (Section 4.3.2.1c, Figure 4.10). Thus, it was thought that this probe would cross-hybridize with other pre-mRNAs presumably transcribed from at least some of the 11 genes in cosmid 4. The second probe used to look for feather keratin pre-mRNA molecules was a 19-mer which was synthesized from a conserved region of the feather keratin intron adjacent to, and including, the 3' splice site. This oligonucleotide probe ought to have detected pre-mRNAs for a number of feather keratin genes, since under the same hybridization and washing conditions as used here (4 x SSC, 37°C-65°C), the probe bound to all except one and possibly two of the 18 feather keratin genes characterized in Chapter 4 (see Section 4.3.4.1). Furthermore, it is known that at least four and possibly most of these 18 genes are expressed in the developing feather (Gregg et al.,

unpublished data; see Section 7.4.1). Therefore, the reason why feather keratin pre-mRNAs were undetected in the RNA preparations was not because these genes are not expressed in embryonic feather tissue.

A second possibility is that the assay procedure may not be sensitive enough to detect the levels of pre-mRNA present. However, the improved Northern blot procedure (Thomas, 1983) can apparently detect as little as one or a few copies of specific mRNA per cell using high specific activity probes. In this study, the procedure appeared to be working at optimum sensitivity (Section 6.3.3) and the probes used were of a high specific activity which is important for optimum sensitivity (Thomas, 1980).

In order to perhaps improve the likelihood of detecting unspliced transcripts for feather keratin mRNA, attempts were made to prepare nuclear RNA from feather tissue. Because of the proteinaceous nature of keratinizing feather tissue (Matulionis, 1970) at least moderate contamination with protein is almost unavoidable. Two nuclear RNA preparations were attempted which gave poor yields of RNA and were considerably degraded. The reason why the RNA prepared from feather nuclei was partially degraded is not clear. Insufficient time was available to continue this work so it was eventually terminated. Nevertheless, hybridization of these nuclear RNA preparations with the two intron-specific probes did not reveal any detectable bands.

The results obtained here corroborate those of an earlier study by Gibbs (1977). In that previous study, nuclear RNA was prepared from 13-day feather tissue, subjected to sucrose

gradient centrifugation in the presence of formamide (Suzuki et al., 1972) and the fractions assayed for feather keratin sequences by Rot analysis with a total cDNA probe. Using these methods, no keratin sequences larger than the 12S cytoplasmic feather mRNA were detected. Similar results were obtained when total RNA was fractionated on denaturing polyacrylamide gels and subjected to the same analysis. However, since it is now known from DNA sequencing (Molloy et al., 1982; Gregg and Rogers, 1986) that the theoretical size of the feather keratin pre-mRNA (~14S) would only be marginally larger than the mature mRNA (12S), it was possible that the pre-mRNA might not have been resolved under the conditions used by Gibbs (1977). Alternatively, the detection method used by Gibbs (1977) may not have been sufficiently sensitive to detect the levels of feather keratin pre-mRNA present in the developing feather.

The failure to detect any precursor mRNA for feather keratin in embryonic feather tissue is not easily explained. In view of the sensitivity of the Northern blot procedure providing probes of high specific activity are used (Thomas, 1983), the presence of even a few molecules of unspliced RNA per cell in whole cell RNA should have been detected. There does not appear to be a significant amount of pre-mRNA in nuclear RNA either, since no detectable hybridization was observed to the two nuclear RNA preparations used in this study. However, since the nuclear RNA was considerably degraded as judged by agarose gel electrophoresis, it is possible that any feather keratin pre-mRNA molecules that were present were sufficiently broken down as to be unable to hybridize to the probes.

Why then should the levels of feather keratin pre-mRNAs be so low when the pre-mRNAs of other abundant mRNAs are present in significant amounts in their corresponding tissues? For some reason, the precursor mRNA may be very rapidly processed to the mature mRNA during transcription by RNA polymerase. This may be related to the fact that the intron in feather keratin genes is located in the 5' non-coding region close to the cap site, resulting in rapid maturation, ligation and rejoining of the two exons. The reason why these genes contain an intron in the 5' non-coding region at all is a mystery. It does not appear to be related to the use of alternative transcription initiation sites (Young et al., 1981) as was suggested by Molloy et al. (1982), since no evidence has been found for this hypothesis to date (Gregg and Rogers, 1986). Another possibility is that the half life of feather keratin mRNA is very long (24 hours or more) and that de novo keratin gene transcription is relatively low by day 14 of embryonic development. Very little information is available pertaining to this question. Bell (1964) found that incubation of 13-day embryonic feather tissue with actinomycin D for 24 hours, which inhibited RNA synthesis by more than 90%, resulted in only a 40% reduction in protein synthesizing activity whereas this same treatment abolished protein synthesis in cells which had not commenced the keratinization process. Neither the polysome profile (Bell, 1964) or keratin protein synthesis (Bell and Merill, 1967; Gibbs, 1977) were significantly affected by this treatment.

Since feather keratins constitute the bulk of the cellular mRNA in terminally differentiating feather cells (Powell

et al., 1976), these early findings suggest that feather keratin mRNA is a relatively stable mRNA and that it turns over relatively slowly in the keratinizing feather. If this is the case, it is possible that the rate of de novo keratin gene transcription (and hence the levels of unspliced feather keratin mRNA) is relatively low by day 14 of development. This possibility could be tested by determining the transcription rates in purified feather nuclei isolated at different times of development. It would also be of interest to measure the half life of feather keratin mRNA in vivo using the pulse chase methods used to determine the half life of other mRNAs (Lowenhaupt and Lingrel, 1978; Guyette et al., 1979).

CHAPTER 7.

CONCLUDING DISCUSSION.

7.1 INTRODUCTION

The keratinization of embryonic chick feather is a terminal differentiation process which leads to the death of the keratin-synthesizing cells. This process is characterized by the synthesis of large amounts of two families of structural proteins; the feather keratins, which constitute the major gene product, and the fast proteins (or histidine-rich proteins). The work described in this thesis involved a detailed analysis of the chromosomal organization of a feather keratin gene locus which comprises most of the feather keratin genes in the chick genome. The data obtained here also demonstrates the linkage of three different keratin gene subfamilies. This chapter summarizes these findings and compares the organization and probable evolutionary history of the avian keratin genes to that of other eukaryotic multigene families and considers how the co-ordinate expression of the feather keratin and fast protein gene families might be controlled.

7.2 THE ORGANIZATION OF AVIAN KERATIN GENES

7.2.1 THE NUMBER OF FEATHER KERATIN GENES

In this work, 18 feather keratin and three feather-like genes were mapped in a region of DNA covering approximately 70 kb of DNA (Figure 4.45).

These genes account for about 80% of the feather keratin genes in the chick genome, as judged by Southern analysis of EcoRI and HindIII-restricted chick genomic DNA (see Section 4.3.7 and 4.4.4). From these results, it was estimated that there are 21-24 feather keratin genes in the chick genome. This figure may be a slight overestimate because of the possibility of cross-hybridization of the feather and scale keratin sequences (Section 4.4.4).

It was originally estimated that there were 100-240 keratin genes in the chick genome (Kemp, 1975). These values were obtained by measuring the kinetics of hybridization and reannealing of feather cDNA to chick DNA. The available data on the reiteration frequency of the feather keratin genes (this work) and including other keratin gene subfamilies (Gregg et al., 1983; Wilton, 1983; Whitbread, 1985; Wilton et al., 1985) indicates that this previous figure of 100-240 genes is clearly an overestimate. It is noteworthy that the calculations of Kemp (1975) included a correction factor to account for mismatching which multiplied the original Cot values by a factor of four. This would appear now not to have been warranted.

The complex band pattern observed on the genomic Southern blots probed with feather keratin cDNA is not due to restriction site polymorphisms since the same pattern was found for DNA isolated from a number of individual chickens, including those of different strains (Saint, 1979).

7.2.2 THE ORGANIZATION OF AVIAN KERATIN GENES

Two extreme patterns of organization have been established in eukaryotic multigene families. They can be either clustered or dispersed, that is, scattered at different chromosomal locations. However, such patterns are not static and clustered gene families can give rise to individual dispersed members known as orphons (Childs et al., 1981; Leder et al., 1981; Engel et al., 1982). The keratin gene system, like many multigene families, is composed of distinct subfamilies, namely feather, scale, beak and claw that are expressed during terminal differentiation of the respective avian epidermal append-

ages. In cases where subfamilies can be distinguished two gene arrangements have been found.

1. Genes of a particular subfamily clustered together and separate from genes in another subfamily. An example of this type of arrangement are the mammalian globin genes, where the α- and β-globins are arranged in separate clusters located on different chromosomes (Efstratiadis et al., 1980; Karlsson and Nienhuis, 1985).

2. The genes for the different subfamilies are intermingled and arranged in clusters. The A and B chorion gene families of silkmoths (Jones and Kafatos, 1980; Eickbush and Kafatos, 1982) and the histone genes of most organisms (Hentschel and Birnstiel, 1981; Maxson et al., 1983) are examples of gene families in this category.

The avian keratin gene subfamilies are members of the first category, since individual gene members belonging to the same subfamily are grouped together (Figure 4.45). A total of 18 feather keratin genes, which represents most of the feather keratin genes in the chick genome (see Section 7.2.1), are located at one chromosomal locus and span a segment of DNA 53 kb in length. An intergenic distance of about 3 kb is maintained throughout the entire gene cluster and the genes are probably all orientated in the same direction. This indicates that, as suggested earlier by Molloy et al. (1982), the feather gene cluster probably arose by a series of tandem duplications of an ancestral feather keratin gene.

The organization of other avian keratin gene subfamilies is similar in that members of the same subfamily are closely linked. Thus, 5 kb leftwards of the feather gene cluster are

three feather-like genes which are themselves spaced about 4 kb apart. One of these genes was partially sequenced (see Section 5.3.1, 5.3.2). It encodes a protein which is 20% larger than the proteins encoded by feather keratin genes and is about 75% homologous to them. Located rightwards of the feather gene cluster, 21 kb distant from the last feather gene, is a cluster of three, or possibly four, claw keratin genes. These claw genes appear to be evenly spaced but, interestingly, two of them form a divergently orientated gene pair (L. Whitbread, personal communication). The scale keratin genes are also clustered to some extent in the chick genome (see Section 1.4.6, Gregg et al., 1983; Wilton, 1983).

An important difference between the organization of mammalian globin and avian keratin genes is that the two subfamilies of α- and β-globin genes are located on different chromosomes, except in Xenopus (see Flavell and Grosveld (1983) and references therein), whereas at least some of the keratin gene subfamilies are closely linked on a single chicken chromosome. It is possible that all of the avian keratin genes are located in a giant locus on a single chromosome, as appears to be the case for the chorion genes of B. mori (Eickbush and Kafatos, 1982) and the major histocompatability complex of the mouse (Steinmetz et al., 1986). In the case of the avian keratin genes, this could be determined by further chromosome walking experiments or possibly by separating isolated chick chromosomes by pulse field gradient gel electrophoresis followed by Southern blot analysis using keratin gene probes (Schwartz and Cantor, 1984).

7.2.3 THE POSSIBLE FUNCTION OF THE CLUSTERED ORGANIZATION OF MULTIGENE FAMILIES

One of the most intriguing questions regarding clustered multigene families is whether the close linkage often observed between homologous genes has any functional significance or merely reflects their evolutionary history.

Early work on the histone genes of sea urchin and Drosophila demonstrated that they are arranged into a repeating unit, each tandem unit containing genes for the five histone proteins (reviewed in Hentschel and Birnstiel, 1981). This led to the proposal that the tandem arrangement of these genes could solve the problem of co-ordinate expression for the five histone genes and facilitate rapid transcription at times of high histone demand (Kedes, 1975; Birnstiel et al., 1979). However, more recent studies indicate that these hypotheses are incorrect (Maxson et al., 1983). Furthermore, in the chicken (D'Andrea et al., 1985) and most mammals (Marzluff and Graves, 1984; Stein et al., 1984), the histone genes are organized into small variable clusters with no obvious repeat, demonstrating that a highly ordered tandem arrangement is not necessary for providing stoichiometric amounts of histone proteins throughout development. Similarly, in most animals, the α- and β-globin genes are co-ordinately expressed in erythroid tissues but are found on separate chromosomes, except in Xenopus where the two gene families are closely linked (Jeffreys et al., 1980). Therefore, neither a regular tandem arrangement or even close linkage are necessary for the co-ordinate expression of gene families.

The alternative and more tangible hypothesis, is that the clustering or chromosomal linkage of multigene families is a record of their evolutionary history and perhaps provides a positive selective advantage. Holt and Childs (1984) proposed that the homogenization of genes within but not between arrays is a consequence of gene conversion processes because these occur more readily between closely linked sequences (see Roeder et al., 1984). Thus, a mechanism for keeping the members of a tandem array of genes nearly identical and independent of another related family of genes is to have the two gene families on separate chromosomes (Coen and Dover, 1983). In this situation recombination events between genes of the separate families will be rare.

Orphons are single, isolated genes derived from tandem repeats but present at a different chromosomal location (Childs et al, 1981; Leder et al., 1981; Engel et al., 1982). Once removed from the constraints that retain homogeneity amongst the co-evolving, tandem repeats, orphons would be free to diverge in both sequence and function to some extent.

The proposal discussed above would explain the linkage of keratin genes belonging to a single subfamily. In the case of the feather keratin gene locus, for example, they have been maintained at a single location presumably to maximize gene correction by mechanisms such as gene conversion (see Section 7.3.1). Similar arguments would apply to the other keratin gene subfamilies, including the feather-like genes, which probably arose from a feather keratin gene by mutational divergence and subsequent gene duplication (see Section 5.4.2). The linkage of different keratin gene subfamilies in the

chicken (Figure 4.45) may enable recombination events to occur between the different subfamilies, generating new genetic variants. This is possibly how the different subfamilies arose. Nevertheless, it is still possible that the keratin genes are clustered to allow co-ordinate expression of the gene locus; for example, the feather keratin genes may form a single chromatin domain and become transcriptionally active as a single unit.

7.3 EVOLUTION OF THE AVIAN KERATIN GENE SUBFAMILIES

7.3.1 THE SEQUENCE HOMOGENEITY OF MULTIGENE FAMILIES

Hood et al. (1975) proposed that the multigene family represents a unit of chromosomal organization where the gene members are closely linked, homologous in sequence and exhibit related or overlapping phenotypic functions. However, given the inevitability of mutation and the difficulty of group selection, some gene correction mechanism(s) must exist to maintain their sequence homogeneity. From studies on many different gene families, two classes of recombination mechanisms can act on families of genes to maintain their sequence homogeneity. They are unequal crossing-over and gene conversion (Figure 7.1).

A. Unequal Crossing-Over

Smith (1976) proposed that multiple occurrences of unequal crossing-over within a tandem array could maintain homogeneity of the repeating units. Any variant that arises within a tandem array may be either fixed or lost by this process. Unequal crossing-over can only occur among tandemly arranged genes of a family and results in a reciprocal increase and decrease in gene number on different chromosomes (Figure 7.1).

Unequal crossing-over events have been demonstrated to occur in the ribosomal RNA repeat unit of yeast (Petes, 1980; Szostak and Wu, 1980) at a rate sufficient to maintain the sequence homogeneity of these genes. This mechanism may explain the striking differences in the structure and repetition frequency of histone gene clusters observed between closely related species (Maxson et al., 1983).

The tandem arrangement of the feather keratin genes and their almost equidistant spacing along the DNA indicates that the feather keratin gene subfamily arose by a recombination mechanism such as unequal-crossing over. Further evidence for this type of mechanism is the presence of two sets of repeated, or symmetrical, restriction enzyme sites in two different regions of the feather gene locus, producing 5 KpnI fragments of 3.4 kb (see cosmid 4 map, Figure 4.13) and 5 HindIII fragments of 2.7 kb (see cosmid 31 map, Figure 4.25b). At least some, and possibly all, of the KpnI sites in the right half of cosmid 4 lie in the protein coding region of a feather keratin gene, 9 bp from the 'stop' codon. Furthermore, hybridization studies indicate that the genes belonging to this 'Kpn family' are more homologous to each other than is the feather keratin gene subfamily as a whole (see Section 4.3.3.1). Taken together, these data provide strong evidence for the involvement of unequal crossing-over during the evolution of this gene subfamily.

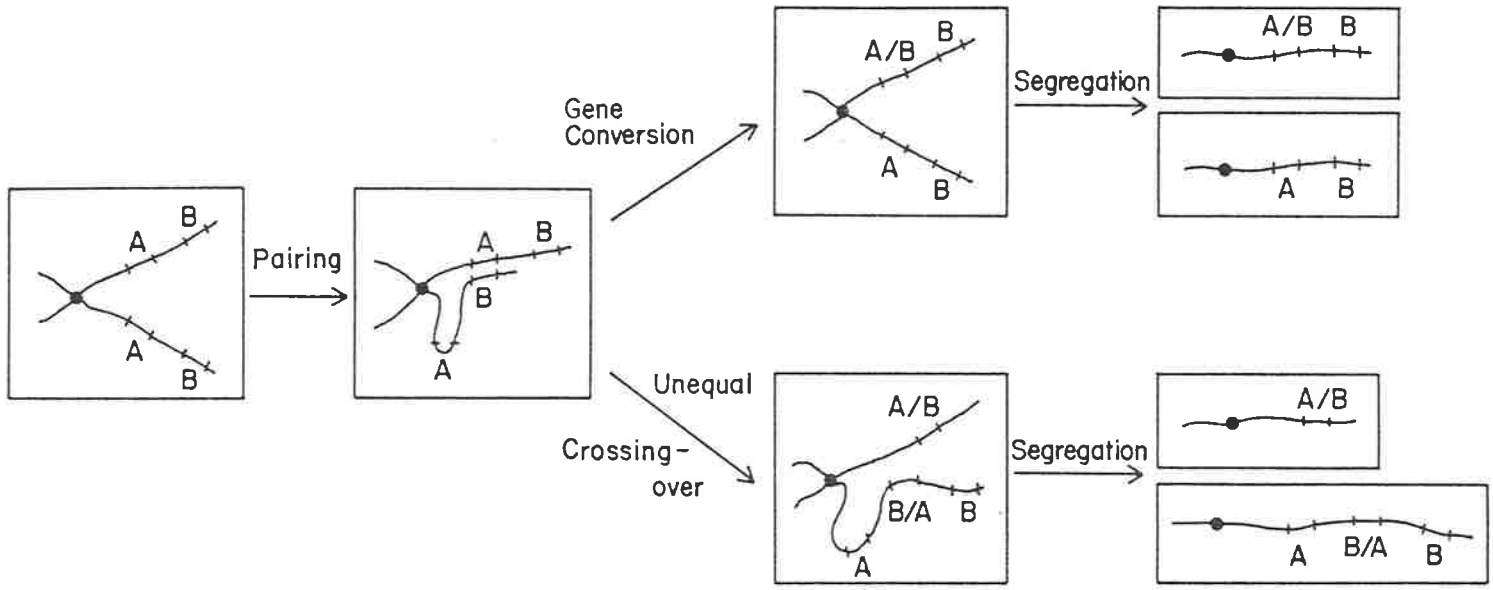
Duplication events which may have telltale symmetrical or repeated restriction enzyme sites are well documented in other reiterated gene families. For example, among the human foetal globin genes, there are two γ-globin genes which have arisen by

FIGURE 7.1 : SCHEMATIC DIAGRAM ILLUSTRATING GENE CONVERSION AND UNEQUAL CROSSING-OVER MECHANISMS

The figure compares gene conversion and unequal crossing-over for an idealized case of a two-gene family (A and B) on the long arm of a single chromosome. The events are shown taking place between sister chromatids and could occur at mitosis or meiosis.

In gene conversion, gene A interacts with gene B in such a way that the nucleotide sequence of part or all of gene A becomes identical to that of gene B. In the interaction, gene A and gene B retain their integrity and their physical locations but a non-reciprocal alteration in the structure of one partner occurs: this is aptly called conversion. Gene A and gene B could be alleles on homologous chromosomes; tandem copies of a gene on the same chromosome; tandem genes interacting from sister chromatids; or a pair of related genes located anywhere in the chromosomes of an organism.

Unequal crossing-over events can take place only among tandemly arranged genes of a family, i.e. between two or more genes of very similar sequence that arose by gene duplication. The two interacting genes could be on sister chromatids, as illustrated in the figure, or on homologs. The result of an unequal crossing-over event is either an increase or a decrease in the size of the gene family - should the family be decreased to one gene, unequal crossing-over events could no longer take place. Any single unequal crossing-over event is reciprocal; one segregant has an increased number of genes and the other has a decreased number (from Baltimore, 1981).


A
B
A
B
Pairing
A
B
B
A
Gene
Conversion
A/B
B
A
B
Segregation
A/B
B
A
B
Unequal
Crossing-
over
A/B
B/A
B
A
Segregation
A/B
A
B/A
B

the duplication of a 5 kb region (Shen et al., 1981). These workers found a short direct repeat at each end of the duplicated segment of DNA containing the γ-globin genes, and they and others (Jeffreys and Harris, 1982) proposed that this duplication occurred by mispairing and unequal crossing-over between short repeated elements on homologous chromosomes. Once a segment is duplicated in tandem by a process such as this, it provides an excellent substrate for unequal crossing-over and can rapidly amplify to form an array of closely related genes, such as those found in each of the keratin gene subfamilies. Similar duplications have been described in the chick histone genes, where a 2.1 kb region is almost exactly duplicated in the reverse orientation (Wang et al., 1985) and in the tRNA genes of Xenopus laevis, where a 3.18 kb repeating unit is tandemly reiterated some 300 times (Müller and Clarkson, 1980).

B. Gene Conversion

The occurrence of gene conversion, in which a DNA sequence is transferred between two members of a homologous gene family (see Figure 7.1), was first demonstrated in fungi between genes on homologous chromosomes (see Radding, 1978 for a review). More recently, intrachromosomal gene conversion has been demonstrated between a pair of closely linked genes in yeast (Jackson and Fink, 1981; Klein and Petes, 1981) and evidence for gene conversion in higher eukaryotes has been obtained by DNA sequence analysis for a number of multigene families (Slightom et al., 1980; Shen et al., 1981; Cohen et al., 1982; Hayashida and Miyata, 1983; Krawinkel et al., 1983; Powell et al., 1983; Iatrou et al., 1984).

Gene conversion can homogenize dispersed genes (as well as those arranged in tandem) and is predicted to be the most frequent form of recombination affecting members of dispersed gene families (Baltimore, 1981; Jackson and Fink 1981; Shen et al., 1981), albeit at possibly a lower level than in those gene families or subfamilies arranged in tandem.

The feather keratin genes have highly homologous protein coding and 5' non-coding regions, yet with the exception of some short segments, the intron and 3' non-coding regions are not well conserved (Figure 1.7; Gregg et al., 1983). Thus, gene correction mechanisms such as gene conversion have acted on the protein coding and 5' non-coding regions to maintain sequence homogeneity, while allowing the 3' non-coding, intron and 5' flanking regions to diverge to some extent. Perhaps the most striking example of gene correction in the feather keratin genes is the strict conservation of the 37 bp exon in the 5' non-coding region, where there is only one base change in 5 known sequences (see Section 1.4.4 and Section 5.3.5). Southern analysis using a 20 base oligonucleotide probe containing the sequence of part of this exon demonstrated that it is well conserved in all except one of the 15 complete genes in cosmids 4 and 33 (Section 4.3.4.1). The functional significance of this segment is unknown.

7.3.2 THE EVOLUTION OF THE AVIAN KERATIN AND FAST PROTEIN GENES

The proposal that feathers evolved from the scales of reptiles is almost a century old (for reviews see, Spearman, 1966; Lucas and Stettenheim, 1972). It is only recently, with the availability of cloned avian keratin sequences, that the

probable molecular mechanisms by which these changes occurred have been defined.

DNA sequence analyses of embryonic chick scale β-keratins have revealed strong regions of sequence homology to the feather keratin genes (Gregg et al., 1984) and indicate that the feather β-keratins probably arose via the deletion of the gly-gly-x region and several other smaller segments from one of the scale β-keratin genes (see Section 1.4.6, Figure 1.9b). The scale to feather transition could have occurred just before the appearance of the first known feathered creatures some 170 Myr ago (Ostrom, 1976).

Preliminary sequence data suggests that claw keratins show even stronger homology to scale keratins (see Section 4.4.2.3). One of the claw keratin genes located in cosmid 33 has been partially sequenced (L. Whitbread and K.Gregg, unpublished results). It encodes a protein of 126 amino acids which, like the scale keratins, has a gly-gly-x region but differs in that the tripeptide repeat in the claw gene is 9 amino acids in length and is repeated twice, compared to scale where it is 13 amino acids and is repeated four times. Presumably, scale and claw keratin genes had a common ancestor which diverged and gave rise to the two separate gene subfamilies. Limited protein chemical data have suggested that the avian beak keratins also have a gly-gly-x region (Gibbs, P.E.M., Walker, I.D. and Rogers, G.E., unpublished data) which, if correct, would imply that these three subfamilies (scale, beak and claw) had a common ancestry and have diverged during evolution to give rise to their respective epidermal appendages.

The feather-like genes are something of a mystery. They represent a hitherto unknown keratin gene subfamily which probably arose from the feather keratin genes by mutational divergence and were subsequently duplicated by unequal crossing-over mechanisms (see Sections 5.4.2, 7.3.1). One of the feather-like genes was sequenced (Section 5.3.1) and it does not appear to be a pseudogene. Therefore it is presumably expressed in one or more of the avian epidermal tissues.

Besides the considerable sequence homology observed between the protein coding regions of the various keratin gene subfamilies, the only other ubiquitous feature of them is the presence of an intron in the 5' non-coding region 20-23 bp prior to the initiation codon (Gregg et al., 1983; Wilton, 1983; Gregg and Rogers, 1986; this work and L. Whitbread and K. Gregg, unpublished results). Together, these two features probably reflect a common origin of the keratin gene subfamilies from a single ancestral keratin gene which had an intron in the 5' non-coding region. A similar common ancestor has been proposed for the α- and β-globin genes, where the two introns are located in the same position in both gene subfamilies (see Karlsson and Nienhuis, 1985).

The unrelated fast protein genes are co-expressed with the feather keratins in embryonic feather tissue (Powell and Rogers, 1979). At least one of the fast protein genes has an intron in an identical position to the feather keratin genes, 21 bp upstream from the initiation codon (Figure 7.2) and its 5' non-coding region is the same length and has a high degree of sequence homology with the 5' non-coding region of feather keratin genes (Morris, 1984). These similarities lead Morris

(1984) to suggest that fast proteins and feather keratins have a common ancestry, with only the similarities at the 5' end of the two gene families being preserved. If this was the case, there must have been a strong positive selective advantage which allowed divergence of the protein coding regions to generate the feather keratin and fast protein gene families, while at the same time maintaining the intron position and the 5' non-coding sequences.

7.4 EXPRESSION OF THE CHICKEN KERATIN GENES

7.4.1 EXPRESSION OF KERATIN GENES IN THE FEATHER

The aim of this work was to elucidate the chromosomal arrangement of the feather keratin genes in the chicken and no attempt was made to examine the expression of these genes. However, several lines of evidence indicate that most, and possibly all, of the 21-24 feather keratin genes are expressed in the keratinizing chick feather.

At the protein level, Walker and Rogers (1976a) suggested that there were at least 19 different feather keratin proteins present in newly hatched chicks and subsequent protein sequencing studies indicated that most of the proteins were the products of highly homologous but non-identical genes (Walker and Rogers, 1976b). This number of proteins would represent a minimum estimate of the number of feather keratin genes as some genes may be silent (e.g. pseudogenes) or may only be expressed in embryonic feather tissue (see below).

Studies using mRNA hybridization kinetics suggested that there are between 25-35 keratin genes expressed in the feather of the embryonic chicken (Kemp, 1975). This would indicate that all of the 21-24 feather keratin genes (Section 7.2.1) are

expressed in the developing feather. However, this conclusion is complicated by recent analysis of keratin clones isolated from a feather cDNA library which demonstrated that at least one scale keratin gene is expressed, albeit at a lower level, in feather tissue (Morris, 1984), confirming the earlier proposal of Dhouailly et al. (1978) (see Figure 1.3). It is not known whether scale genes are expressed in feather tissue for a functional reason, or because scale and feather keratin genes are switched on by similar, but different, signals. If the latter hypothesis is correct, the expression of scale keratins in feathers would be due to 'leaky' expression of some scale genes which recognize, albeit poorly, the similar regulatory mechanisms of the feather (Morris, 1984). These findings support the view that the avian epidermis has a 'developmental plasticity', that is, its ability to alter its path of differentiation (by activating a particular set of keratin genes) according to whether the dermis is of feather or scale origin (see Section 1.2.2).

A second, largely unsolved question is whether there are some feather keratin genes which are expressed in embryonic but not adult feather tissue, or vice versa. Early work suggested that there were both qualitative and quantitative differences in the keratin proteins produced in embryonic and adult feather tissues (Kemp and Rogers, 1972). In a more recent study, Rothnagel (1979) showed that the keratin proteins present in embryonic and adult feather tissues were similar, except that there were two additional keratin bands present in adult feathers. Recently, the complete amino acid sequence of an adult fowl feather keratin was published (Arai et al., 1983)

and, when compared to the proteins encoded by the embryonic feather keratin genes of λCFK1, the one notable difference was the absence of a single arginine residue from the adult feather protein (see Figure 1.6; Gregg and Rogers, 1986). It remains to be seen whether this 'adult' protein sequenced by Arai et al. (1983) is also present in embryonic feathers.

The pattern of expression of four of the feather keratin genes in the recombinant λCFK1 has been investigated using oligonucleotide probes which were made from the 3' non-coding regions and were specific for each gene (Gregg et al., manuscript in preparation). Each of the four genes were found to be expressed in both embryonic and adult feather tissues, under the control of the same promoters.

In summary, most of the feather keratin genes appear to be expressed in both embryonic and adult feather tissues, but there are probably a few genes which are only switched on in the adult. Whether or not these putative 'adult' specific genes are contained in the feather cluster characterized in this study, or are among the few remaining genes not mapped, is not known.

7.4.2 CO-ORDINATE REGULATION OF THE FEATHER KERATIN AND FAST PROTEIN GENES

The feather keratin and fast protein genes are co-ordinately expressed during feather development, but encode protein families that are quite distinct; no amino acid sequence homology exists between them and they have totally different amino acid compositions and polypeptide chain lengths (Walker and Rogers, 1976a; Powell and Rogers, 1979; Morris, 1984). However, these genes show strong sequence and structural

homology in their 5' non-coding regions (see Section 7.3.2. and Figure 7.2). From these observations, it was proposed that the conserved sequences are involved in the co-ordinate expression of these genes (Morris, 1984). Subsequently, an 8 bp region at about -150 relative to the cap site was found to be conserved between the fast protein and feather keratin genes, with the exception of feather keratin gene B, where there is a one base change (see Section 5.4.6). Presumably, both of these conserved segments peform some regulatory role at the transcriptional and/or translational levels to control expression of these genes. At the transcriptional level, for example, a 'factor' produced by the feather dermis may act directly or indirectly on one of these conserved elements, leading to transcriptional activation of the two gene families. However, the observation that fast proteins are not detectable in adult feathers (Walker and Rogers, 1976a; Rothnagel, 1979) whereas feather keratins are (Section 7.4.1), indicates that the developmental regulation of their gene expression is quite complex.

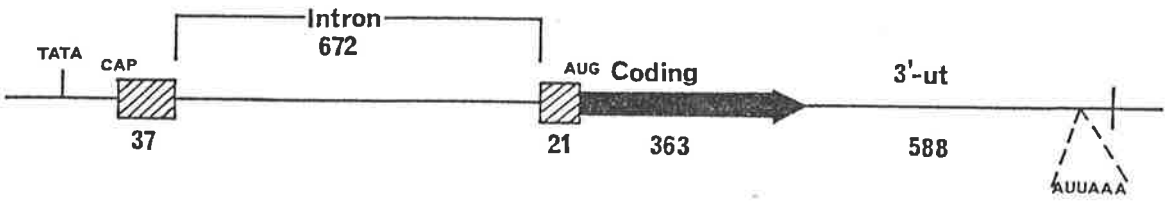
The function of the intron located in the 5' non-coding region of the fast protein gene(s), and probably all keratin gene subfamilies examined to date, is unclear. Koltunow et al. (1986) have reported that removal of the intron from a feather keratin gene significantly increased its transcription in Xenopus oocytes. The effect observed was specific for the feather keratin intron fragment and suggested that specific DNA sequences within the intron are the most likely explanation for this inhibition of transcription. Koltunow et al. (1986) proposed that these intron sequences act to keep these genes transcriptionally silent until the correct elements combine to

FIGURE 7.2 : COMPARISON OF THE STRUCTURAL FEATURES OF THE FAST PROTEIN AND FEATHER KERATIN GENE FAMILIES

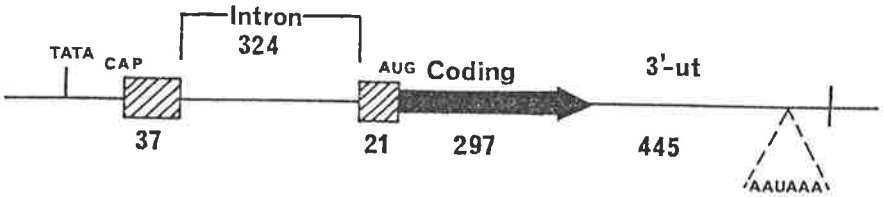
The structural features displayed by the feather keratin genes, as exemplified by gene C of the recombinant λCFK1 (Figure 1.5b; Molloy et al., 1982) are compared to the structure of the pCFK22 gene, which encodes a fast protein species (Morris, 1984).

The genes have a number of similarities at their 5' ends. They both contain an intron which splits the 5' non-coding region into two segments of 37 bp and 21 bp. The 5' non-coding region of the two genes have a high degree of sequence homology. In particular, the 37 bp 5' non-coding exon of the feather keratin and fast protein genes show 75% homology (Morris, 1984).

pCFK22 Gene



Gene C



stimulate transcription. These 'correct elements' might include protein factor(s) which bind to the '-150 box' and/or the 37 bp 5' non-coding exon sequences.

7.4.3 WHY ARE THERE SO MANY FEATHER KERATIN GENES?

The preceding discussion indicates that most, and perhaps all, of the feather keratin genes are expressed in the keratinizing embryonic chick feather. The question arises as to the biological significance of a large number of closely related but non-identical feather keratin proteins. Since no gene amplification occurs (Gibbs, 1977), the large number of genes may have arisen to permit the synthesis of large quantities of keratin proteins required during the terminal differentiation of the feather (Walker and Rogers, 1976a,b). Clearly, mutations have occurred in these multiple genes and many have been acceptable in the amino acid sequence which forms the anti-parallel β-sheet structure of the feather microfibril (Fraser et al., 1972).

However, it is not immediately clear why there should be many more genes encoding keratin proteins than other tissue-specific proteins produced in large quantities such as globin or ovalbumin. In the case of the keratins, which form the boundary between the animal and its physical surroundings, a pool of genes coding for closely related proteins may be advantageous in adjusting to a changing environment (Fraser et al., 1972). The pre-existence of genes coding for a variety of molecular forms may also confer some biological advantage and permit a better response to selection pressure. In this respect, the keratins may be compared to the chorion proteins which form the protective eggshell of silkmoth species. The

chorion proteins are encoded by well over 100 genes (for a review, see Goldsmith and Kafatos, 1984) and the complexity of this gene 'superfamily' has been ascribed to the multiple physiological functions of the eggshell, morphogenetic or structural requirements and the variation in the different microclimates to which different silkmoth species have adapted (Goldsmith and Kafatos, 1984).

In the feather, different morphological parts have different proportions of keratins (Kemp and Rogers, 1972) and it is possible that their varied expression is also evolutionarily well adapted. Additionally, in both cases, a pool of genes would ensure that if one or two became inactivated (by mutation), the others could still produce a functional protein product.

A P P E N D I X A

RESTRICTION MAPPING OF COSMID 12

This appendix describes the construction of a BamHI restriction map and the mapping of EcoRI and HindIII restriction sites in cosmid 12 which lay outside of the genomic region encompassed by cosmid 4.

A BamHI digest of cosmid 12 produced 6 fragments (Figure A.2.a, track 1) of which the 19 and 10.7 kb fragments hybridized strongly with the feather keratin probe, while the 4.2 kb fragment was detected more weakly (data not shown). This data, and the similar size of the 10.7 and 4.2 kb BamHI fragments in cosmids 4 and 12, enabled these two fragments of cosmid 12 to be mapped. The four remaining BamHI fragments of 19, 4.3, 4.2 (the 4.2 kb band of cosmid 12 is a doublet, Figure A.2.a, track 1) and 1.4 kb had no counterparts in cosmid 4. The large 19 kb BamHI fragment hybridized strongly with the feather keratin gene probe (data not shown). From the mapping of cosmid 4, there are no BamHI sites to the right of the single site located in the 1.8 kb HindIII fragment (Section 4.3.2.1, Figure 4.13). Therefore, from its size, it must span the region from the 2.15 kb HindIII fragment to the right-hand end of the clone (a total of 14 kb including the cosmid vector) and a further 5 kb at the left-hand end of the insert (Figure A.3.b).

In cosmid 4, the 3.0 kb EcoRI fragment is at the end of the insert and contains part of the vector (Figure 4.13). As discussed in Section 4.3.2.2b, this truncated 3.0 kb fragment was deduced to comprise part of the 5.0 kb EcoRI fragment of

cosmid 12. To demonstrate this directly, the 4.2 kb BamHI fragment of cosmid 4 was isolated and hybridized to EcoRI digests of cosmids 4 and 12 (Figure A.1). As predicted from the EcoRI maps of cosmids 4 and 12 (Figures 4.13, A.3.b) the probe detected the 5.0 and 3.0 kb EcoRI fragments of cosmids 12 and 4, respectively, as well as the 8.4, 0.75 and 0.6 kb fragments present in both clones. This confirmed that this 4.2 kb BamHI fragment was located at the left-hand end of the cosmid 4 insert and that the 5.0 kb EcoRI fragment of cosmid 12 was next to the two small 0.75 and 0.6 kb EcoRI fragments present in both clones. Since the 5.0 kb EcoRI fragment contains only one BamHI site (see Section 4.3.2.2d) which is 3.7 kb from the EcoRI site, either the 4.2 or 4.3 kb is next to the 4.2 kb BamHI fragment which is present in both recombinants.

To order these 6 remaining BamHI and EcoRI fragments of cosmid 12 (Figure A.3.b), the following restriction fragments from cosmid 12 were hybridized to BamHI digests of cosmid 12 (see Footnote below): the 1.0 kb EcoRI/HindIII fragment from

---

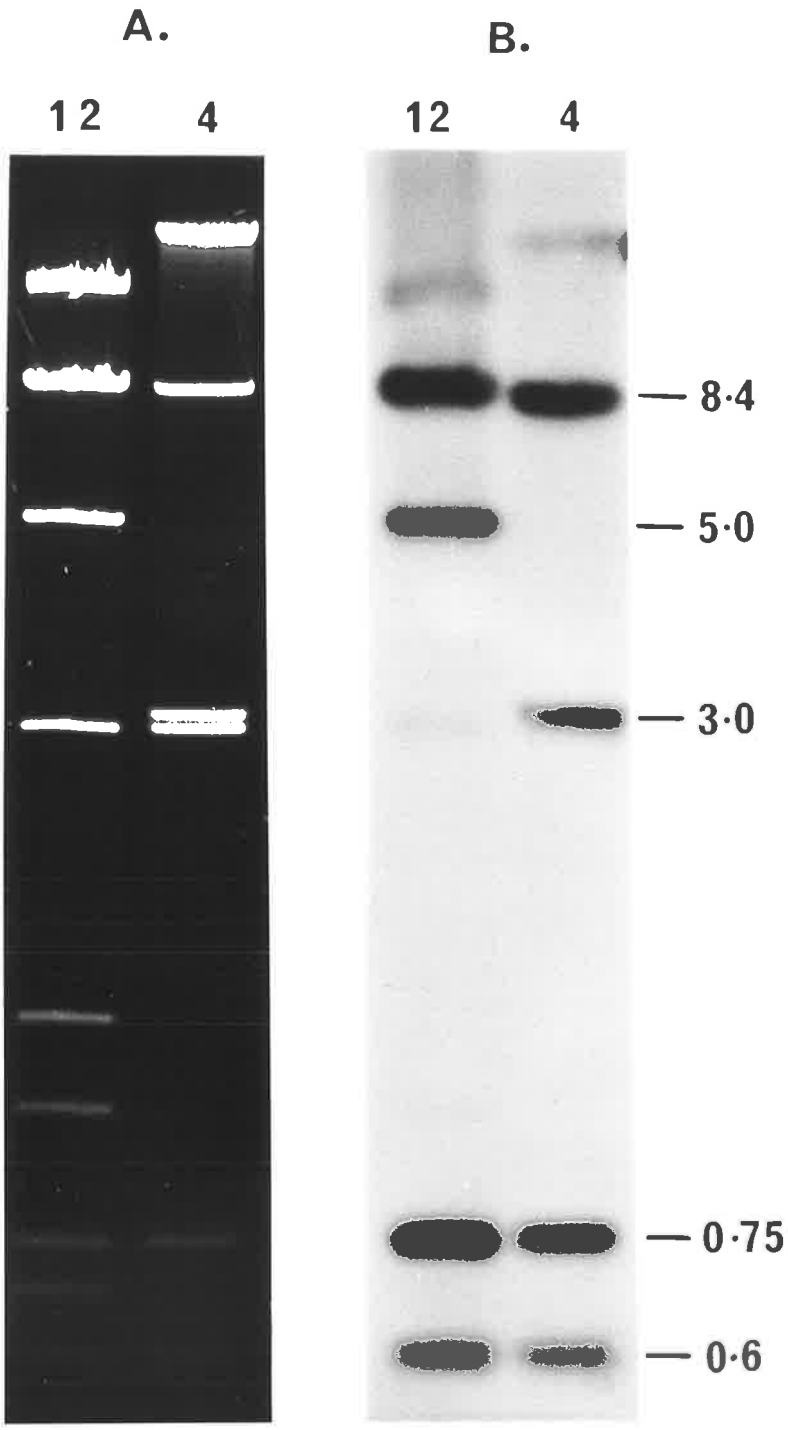
FOOTNOTE: Analysis of pBR322 subclones which contain the 1.3 and 1.1 kb EcoRI fragments of cosmid 12 (denoted as pE1.3 and pE1.1, respectively) indicated that each of these fragments contained a single BamHI site in the insert (data not shown). The fragments generated by BamHI/EcoRI digestion of these subclones were 0.7 and 0.4 kb (pE1.1) and 0.7 and 0.6 kb (pE1.3). Therefore, each of these EcoRI fragments should hybridize to two BamHI fragments in cosmid 12.

FIGURE A.1 : SOUTHERN ANALYSIS OF COSMIDS 12 AND 4 WITH THE 4.2 KB BamHI FRAGMENT

A. Cosmid 12 and 4 DNAs were digested with EcoRI and the products fractionated on a 0.8% agarose gel. The DNA fragments were stained with ethidium bromide and visualized under UV light (Section 2.2.7).

B. The gel was blotted on to nitrocellulose and the filter bound DNA probed with the 4.2 kb BamHI fragment of cosmid 4 (see Figure 4.13) which had been isolated from LGT agarose (Section 2.2.8) and labelled by nick translation (Section 2.2.11.2). The filter was washed in 0.1 x SSC, 0.1% SDS at 65°C and autoradiographed for 3 days at -80°C.

The sizes of the fragments in cosmids 4 and 12 which were detected by the probe are indicated.


A.
B.
12
4
12
4
8·4
5·0
3·0
0·75
0·6

the left-hand end of pE5.0 (see Section 4.3.2.2d); the 1.1 kb EcoRI fragment (insert of pE1.1); the 0.7 kb BamHI/EcoRI fragment from pE1.3 and the whole insert of pE1.3.

The location of the four mapping probes are shown on the deduced map of the left portion of cosmid 12 (Figure A.3.c), with the probes (numbered 2-5) referring to the blots illustrated in tracks 2-5 of Figure A.2.a. In all four mapping blots shown in Figure A.2.a except track 4, the probes bound to the 19 kb BamHI fragment which contains the right half and part of the left-hand end of cosmid 12 (Figure A.3.b). As the DNA segments contained in all of these probes is outside the portion of the clone which comprises the 19 kb BamHI fragment, it seemed likely that this hybridization signal was non-specific (e.g. due to cross-hybridization of pBR322 sequences present in the probe preparations and the 19 kb fragment) and it was subsequently disregarded when discussing and interpreting the blot data.

The 1.0 kb probe from pE5.0 hybridized to the 4.2 kb doublet (Figure A.2.a, track 2). This indicated that the 4.2 kb BamHI fragment present only in cosmid 12 was located next to the BamHI fragment present in cosmids 4 and 12 (Figure A.3.b and A.3.c). The 1.1 kb EcoRI insert hybridized very strongly to the 4.2 and 1.4 kb BamHI fragments (Figure A.2.a, track 3) indicating that, as expected, this EcoRI fragment spans a BamHI site and is partially contained in these two BamHI fragments. However, the 1.1 kb fragment also hybridizes to the 19 and 10.7 kb BamHI fragments which both contain feather keratin genes. This is thought to be due to cross-hybridization by feather-like (keratin) sequences present in the 1.1 kb EcoRI

FIGURE A.2 : SOUTHERN ANALYSIS OF COSMID 12 WITH MAPPING PROBES

A. BamHI mapping. Cosmid 12 DNA was restricted with BamHI, fractionated on a 1% agarose gel and the fragments detected in UV light after ethidium bromide staining (track 1). DNA was transferred bidirectionally on to nitrocellulose (Section 2.2.9). The filters were hybridized with the probes shown in Figure A.3.c as follows:

Track 2 - 1.0 kb EcoRI/HindIII fragment from pE5.0 (see Section 4.3.2.2d)

Track 3 - pE1.1 insert (1.1 kb EcoRI fragment)

Track 4 - 0.7 kb BamHI/EcoRI fragment from pE1.3 (1.3 kb EcoRI fragment)

Track 5 - pE1.3 insert (1.3 kb EcoRI fragment)

The origin of the probes used in these blot hybridizations are shown in Figure A.3.c. The probes (numbered 2-5) in Figure A.3.c correspond to tracks 2-5 in this figure.

B. HindIII mapping. A HindIII digest of cosmid 12 was probed with the 1.0 kb EcoRI/HindIII fragment from pE5.0 (see Section 4.3.2.2d). All four probes shown in Figure A.3.c. hybridized most strongly to the 4.5 kb HindIII fragment.

All filters were washed at 0.1 x SSC, 0.1% SDS and exposed at -80°C. Fragment sizes are shown in kilobases.

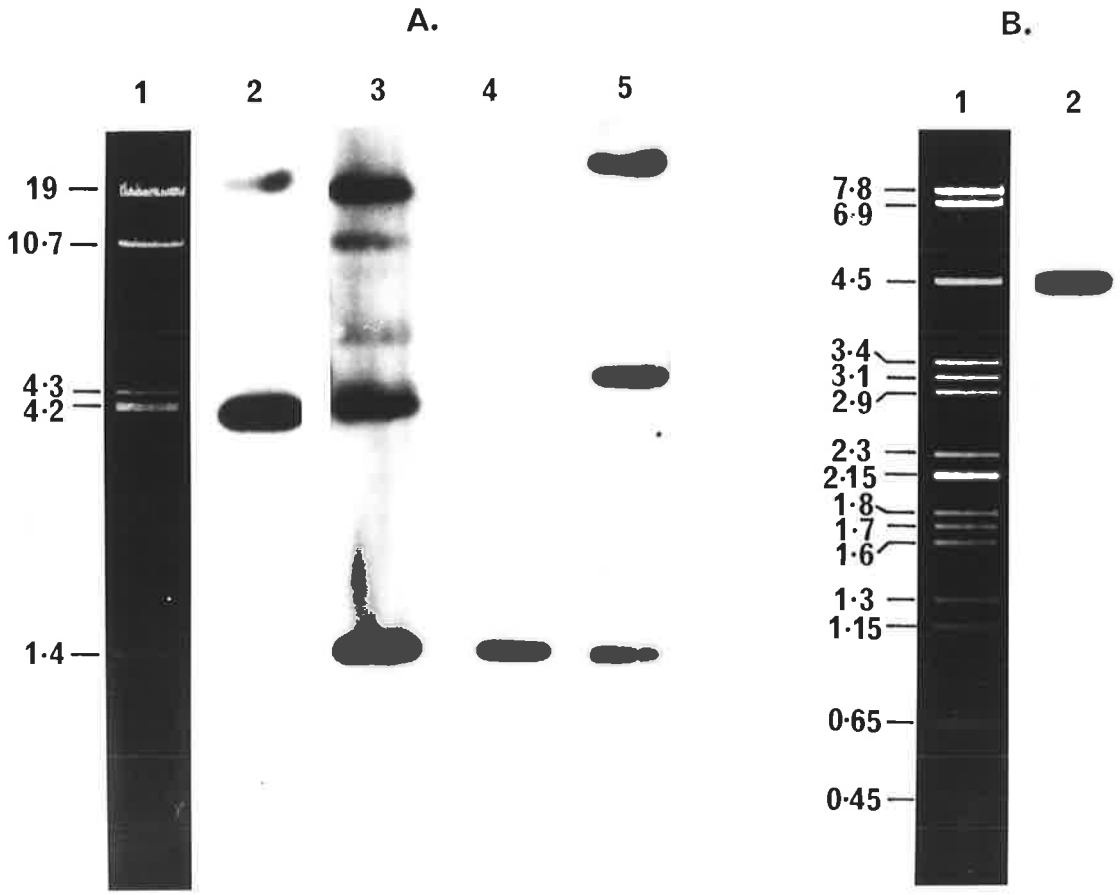
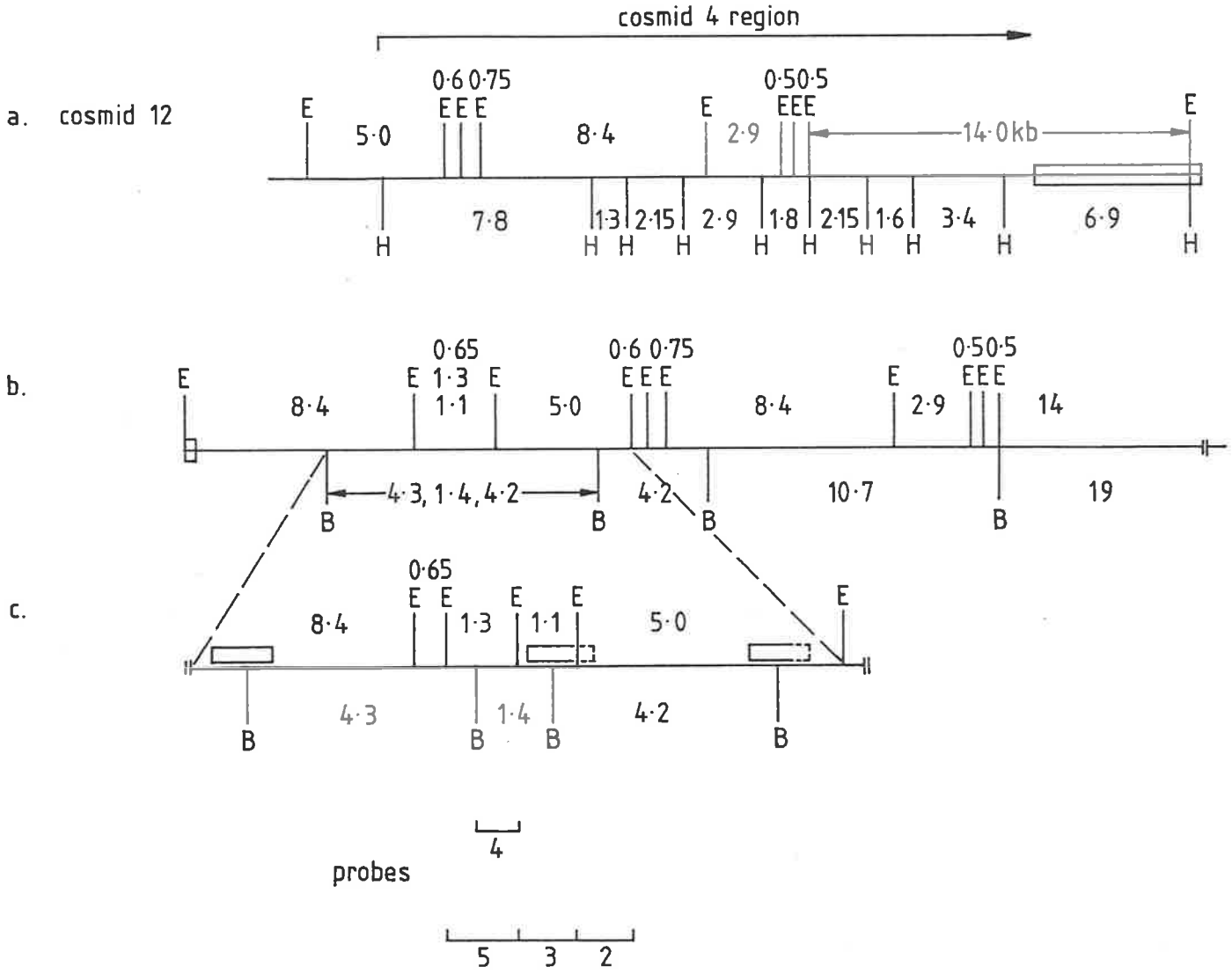

A.
1
2
3
4
5
19
10·7
4·3
4·2
1·4
B.
1
2
7·8
6·9
4·5
3·4
3·1
2·9
2·3
2·15
1·8
1·7
1·6
1·3
1·15
0·65
0·45

FIGURE A.3 : DERIVATION OF THE COSMID 12 RESTRICTION MAP

The figure depicts the derivation of the restriction map of cosmid 12, which is described in detail in Section 4.3.2.2 and Appendix A. The region of overlap with cosmid 4 and the cosmid vector (boxed region) are indicated in (a) The position of the three feather-like genes in the left-hand end segment of cosmid 12 are displayed by boxes. The gene in the 1.1 kb EcoRI fragment may span the EcoRI site (data not shown), shown as a dotted line.


cosmid 4 region
a. cosmid 12
E
0·6 0·75
E E E
0·5 0·5
E E E
E
5·0
8·4
2·9
14·0 kb
H
7·8
1·3
2·15
2·9
1·8
2·15
1·6
3·4
6·9
b.
E
0·65
E 1·3 E
1·1
0·6 0·75
E E E
E
0·5 0·5
E E E
8·4
5·0
8·4
2·9
14
B
4·3, 1·4, 4·2
4·2
10·7
19
c.
0·65
E E
E E
E
8·4
1·3
1·1
5·0
B
4·3
1·4
4·2
4
probes
5
3
2

fragment (Section 4.3.2.2 and other data not shown) with feather keratin genes in the 19 and 10.7 kb BamHI fragments. The very strong hybridization of the 1.1 kb EcoRI probe to the 1.4 and 4.2 kb BamHI fragments indicated that the 1.4 kb BamHI fragment is located next to the 4.2 kb fragment and that the 1.1 kb EcoRI fragment is located next to the 5.0 kb fragment (Figure A.3.c).

Finally, the 1.3 kb EcoRI fragment detects the 1.4 and 4.3 kb BamHI fragments (Figure A.2.a, track 5) and the 0.7 kb BamHI/EcoRI fragment from pE1.3 detects only the 1.4 kb fragment (Figure A.2.a, track 4), indicating that the 1.3 kb EcoRI fragment (which contains one BamHI site) spans the remainder of the 1.4 kb BamHI fragment and part of the 4.3 kb BamHI fragment as shown in Figure A.3.c. The remaining 0.65 kb EcoRI fragment from cosmid 12 must lie immediately to the left of the 1.3 kb EcoRI fragment.

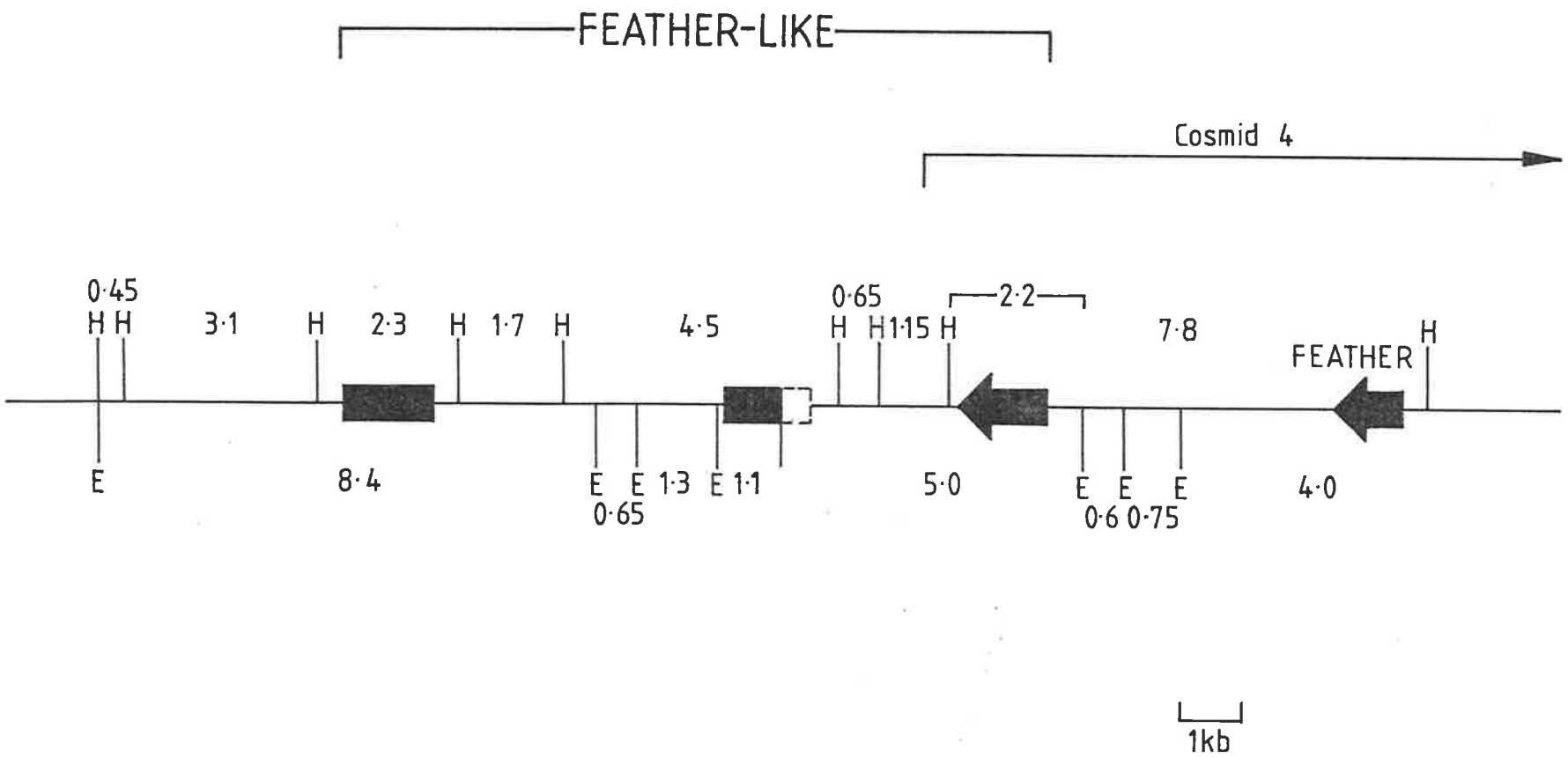
The HindIII restriction map of the segment of cosmid 12 located downstream of cosmid 4 (Figure A.4) was derived as follows. From the map of the subclone containing the 5.0 kb EcoRI fragment (Section 4.3.2.2d), the 1.15 and 0.65 kb HindIII fragments of cosmid 12 are entirely located in this fragment. The 2.2 kb EcoRI/HindIII fragment, which contains one of the feather-like genes, is at the left-hand end of the 7.8 kb HindIII fragment; this gene-containing fragment is present in cosmids 4 and 12 (Figure A.4).

The 4.5 kb HindIII fragment of cosmid 12 which hybridized with the feather keratin gene probe (arrow in Figure 4.14), is detected very strongly by all four of the probes shown in Figure A.3.c (e.g. see Figure A.2.b), indicating that it spans

this segment which includes the 1.3, 1.1 and 0.65 kb EcoRI fragments of cosmid 12. The hybridization of the 4.5 kb HindIII fragment with the feather keratin gene probe is consistent with it spanning this region, which includes the feather-like gene contained in the 1.1 kb EcoRI fragment (Figure A.4). It remained to position the 1.7, 3.1 and 2.3 kb fragments. The 2.3 kb HindIII fragment (which hybridizes with the feather keratin gene probe, arrow in Figure 4.14) is cut by BamHI, whereas the 1.7 and 3.1 kb fragments are not (data not shown), indicating that the 2.3 kb HindIII fragment spans the BamHI site near the end of the insert. The distance from the HindIII site of the 4.5 kb fragment to the BamHI site is about 3.0 kb, and as the 2.3 kb HindIII fragment is cut by BamHI, there is only space for the 1.7 kb HindIII fragment between the 4.5 and 2.3 kb fragments. The 3.1 kb HindIII fragment is therefore located at the end of the insert near the cosmid vector as shown in Figures 4.16 and A.4.

FIGURE A.4 : ARRANGEMENT OF THE FEATHER-LIKE GENES IN COSMID 12

The figures shows the restriction map of the left half of the cosmid 12 insert, indicating the position of the three feather-like genes and the last (leftmost) feather keratin gene. The genes were mapped as described in Section 4.3.2.2 and Appendix A. The orientation of the feather-like gene closest to the feather gene cluster, which is present in both cosmids 4 and 12, was determined by DNA sequence analysis (Section 5.3.1). The two remaining feather-like genes are present only in cosmid 12 and map in the 4.5 kb HindIII (1.1 kb EcoRI) and the 2.3 kb HindIII (8.4 kb EcoRI) fragments of this recombinant (Section 4.3.2.2).


FEATHER-LIKE
Cosmid 4
0·45
H H
3·1
H
2·3
H
1·7
H
4·5
0·65
H H1·15
2·2
H
7·8
FEATHER
H
E
8·4
E
0·65
E
1·3
E 1·1
5·0
E
0·6
E
0·75
E
4·0
1kb

A P P E N D I X B

RESTRICTION MAPPING OF COSMID 33

This appendix describes additional Southern blot hybridization experiments carried out during the construction of the cosmid 33 restriction map. Figure B.1 shows a number of hybridizations of HindIII restriction fragments to BamHI (Figure B.1.a) and KpnI (Figure B.1.b) digests of cosmid 33. In the case of some of the feather and claw gene containing fragments, it was necessary to subclone them into pBR322 and use fragments from the subclones as hybridization probes to obtain an unequivocal result (e.g. p2.9 and p4.65, Figure B.1). Even when this was done, there was sometimes cross-hybridization to the other restriction fragments which contain keratin sequences (e.g. p2.9, Figure B.1.a).

Another problem encountered was cross-contamination of the isolated HindIII fragments used as probes. For example, in Figure B.1.a the 7.1 kb HindIII fragment hybridizes to four BamHI fragments of the following sizes - 7.7, 6.4, 3.7 and 2.1 kb. The hybridization of the 7.1 probe to the 7.7, 3.7 and 2.1 kb fragments was due to the 7.1 kb HindIII fragment detecting homologous sequences (arrows in Figure B.1.a). The hybridization of the probe to the 6.4 kb BamHI fragment (star in Figure B.1.a) was due to the very close migration of the 7.1 and 6.7 kb HindIII fragments under the conditions used here. The 6.7 kb HindIII fragment and the 6.4 kb BamHI fragment contain most of the cosmid vector (see Figure 4.27c).

The restriction map for KpnI was determined to enable construction of a detailed map of the claw keratin gene-containing region, located at the right-hand end of cosmid 33. The following rationale led to the final restriction map of this region which is depicted in Figure B.2. (i) The 1.2 kb HindIII fragment, which contains part of the cosmid vector (Figure B.2), hybridized to the 9.0 and 1.3 kb KpnI fragments (Figure B.1.b) indicating that the 1.3 kb KpnI fragment was located at the right-hand end of cosmid 33 (Figure B.2). (ii) The 2.4 kb HindIII fragment, which contains one of the claw keratin genes (Section 4.3.3.3), hybridized to the 5.8, 1.8 and 1.3 kb KpnI fragments. Since the 1.2 kb HindIII fragment bound only the 9.0 and 1.3 kb KpnI fragments and not other KpnI fragments which contain claw keratin sequences, the 2.4 kb HindIII fragment must be located next to the 1.2 kb HindIII fragment (Figure B.2). The strong hybridization of the 2.4 kb HindIII probe to the 1.8 kb KpnI fragment and the apparent absence of any HindIII sites in the 1.8 kb KpnI fragment (data not shown) indicated that this KpnI fragment is almost entirely located in the 2.4 kb HindIII fragment. Furthermore, restriction analysis of a subclone containing the 2.7 kb HindIII fragment (which contains a claw keratin gene, Figure B.2) showed that it has a KpnI site about 0.1 kb from one of the HindIII sites (data not shown). Therefore, the 1.8 kb KpnI fragment spans this HindIII site as shown in Figure B.2.

The detection of the 5.8 kb KpnI fragment by the 2.4 kb HindIII probe is probably due mainly to the cross-hybridization of claw keratin sequences which are present in this KpnI frag-

FIGURE B.1 : EXAMPLES OF SOUTHERN BLOTS USED TO DETERMINE THE RESTRICTION MAP OF COSMID 33

A. Cosmid 33 DNA was digested with BamHI, analyzed by agarose gel electrophoresis (Section 2.2.7), blotted on to nitrocellulose (Section 2.2.9) and hybridized the following fragments used as probes:

0.8: 0.8 kb HindIII fragment of cosmid 33

0.9: 0.9 kb HindIII fragment of cosmid 33

p2.9: A PstI/HindIII subfragment from the insert of a subclone containing the keratin-positive 2.9 kb HindIII fragment of cosmid 33.

1.2: 1.2 kb HindIII fragment of cosmid 33.

7.1: 7.1 kb HindIII fragment from cosmid 33.

Arrows indicate homologous (strongly hybridizing) fragments. Star indicates hybridization of the 7.1 kb HindIII fragment to the 6.4 kb BamHI fragment. The 7.1 kb HindIII probe contains contaminating 6.7 kb fragment; the 6.7 kb HindIII fragment contains most of the cosmid vector and thus hybridized with the 6.4 kb BamHI fragment (see Figure 4.28).

B. Cosmid 33 DNA was digested with KpnI, analyzed by agarose gel electrophoresis (Section 2.2.7), blotted on to nitrocellulose (Section 2.2.9) and hybridized with the following probes:

FK: The feather keratin gene probe (Section 2.2.10)

pBR: Labelled pBR322

1.2: The 1.2 kb HindIII fragment of cosmid 33

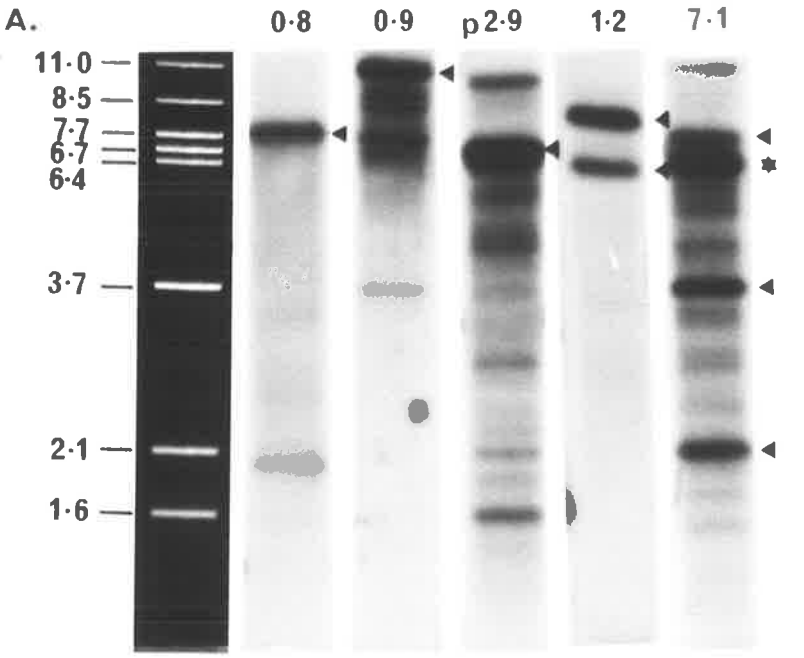
3.6: The 3.6 kb HindIII fragment of cosmid 33

2.4: The 2.4 kb HindIII fragment of cosmid 33

p4.65: The 1.6 kb BamHI fragment from the p4.65 subclone, which contains the 4.65 kb HindIII fragment of cosmid 33.

The sizes, in kilobases, of the BamHI (A) and KpnI (B) restriction fragments are shown. All fragments used as probes were isolated from LGT agarose (Section 2.2.8) and labelled by nick translation (Section 2.2.11.2).

All filters were washed in 0.1 x SSC, 0.1% SDS at 65°C and exposed to X-ray film at -80°C.


A.
0·8
0·9
p2·9
1·2
7·1
11·0
8·5
7·7
6·7
6·4
3·7
2·1
1·6

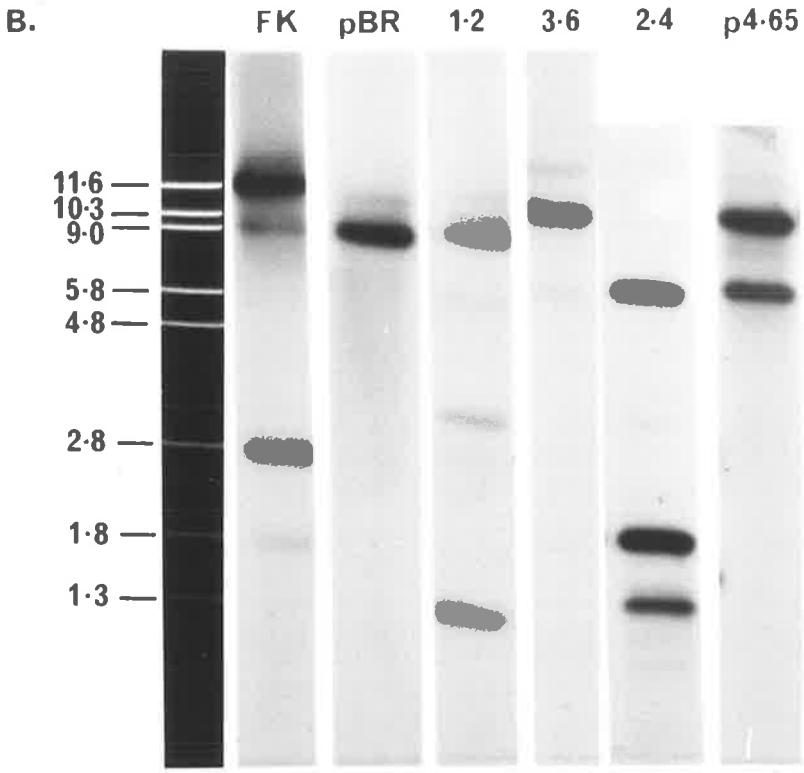

B.
FK
pBR
1·2
3·6
2·4
p4·65
11·6
10·3
9·0
5·8
4·8
2·8
1·8
1·3

TABLE B.1 : HYBRIDIZATION OF THE HINDIII FRAGMENTS OF COSMID 33 TO BamHI AND KpnI DIGESTS OF COSMID 33

The table summarizes the data obtained in Figures 4.26, 4.27, B.1 and from other data not shown. The hybridization data enabled the HindIII, BamHI and KpnI restriction maps (Figure 4.28) to be derived.

HINDIII FRAGMENT USED AS PROBE		BamHI FRAGMENT(S)	KpnI FRAGMENT(S)
7.1		7.7, 3.7, 2.1, 6.4[A]	n.d.
pBR322[B]		6.4	9.0
feather keratin gene[C]		11.0, 6.7	11.6, 9.0, 2.8
4.65[E]		8.5, 7.7, 1.6	10.3, 5.8
feather keratin positive	3.6	7.7	10.3
	2.9[E]	6.7	11.6, 2.8
2.9		11.0	n.d.
feather keratin positive	2.7[E]	6.7	9.0, 2.8
claw keratin positive	2.7[E]	8.5	n.d.
2.6		11.0, 3.7	n.d.
claw keratin positive	2.4	8.5	5.8[D], 1.8, 1.3
1.2		8.5, 6.4	9.0, 1.3
0.9		11.0	n.d.
0.8		7.7	10.3

NOTES

A: The hybridization of the 7.1 kb HindIII fragment to the 6.4 kb BamHI fragment is due to contamination of the 7.1 kb fragment with the 6.7 kb HindIII fragment, which cotains most of the cosmid vector (see Figure 4.28).

B: pBR322 was used instead of the 6.7 kb HindIII fragment, which contains most of the cosmid vector (Figure 4.27c).

C: The 6.7, 6.2, 2.9 and 2.7 kb HindIII fragments contain feather keratin sequences (Figure 4.26, 4.27b).

D: The hybridization of the 2.4 kb HindIII fragment to the 5.8 kb KpnI fragment is due to the cross-hybridization of claw keratin sequences present in this region (see text and Figure 4.29).

E: Fragments were subcloned into pBR322 and used as hybridization probes.

FIGURE B.2 : ARRANGEMENT OF THE CLAW KERATIN GENES IN COSMID 33

The figure shows the location of the claw keratin genes in the 8.5 kb BamHI fragment (Figures 4.27 and 4.28) at the right-hand end of the cosmid 33 insert. The two claw genes which form a divergently orientated gene pair have been partially characterized by DNA sequence analysis (L. Whitbread, personal communication). The other two genes have been mapped by blot hybridization (Sections 4.3.3.2, 4.3.3.3) and DNA sequence analysis (L. Whitbread and K. Gregg, unpublished results). The HindIII and BamHI restriction sites are shown above the map, while the KpnI sites are shown below the map. The sizes (kb) of the restriction fragments are indicated.

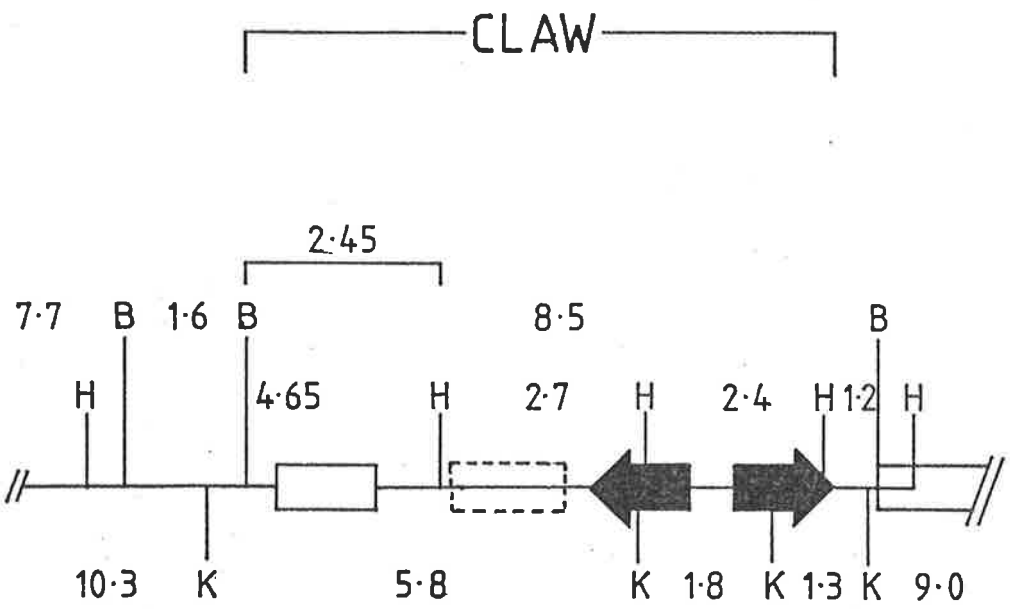
The boxed region partly shown at the righthand end of the map depicts the cosmid vector.

Restriction enzymes sites are indicated as follows:

B - BamHI

H - HindIII

K - KpnI


CLAW
2·45
7·7
B
1·6
B
8·5
B
H
4·65
H
2·7
H
2·4
H
1·2
H
10·3
K
5·8
K
1·8
K
1·3
K
9·0

ment (Figure B.2). The 4.65 kb HindIII fragment, which contains two BamHI sites (Figure B.2), lies next to the 2.7 kb HindIII fragment. The 4.65 kb HindIII fragment contains part of the large 8.5 kb BamHI fragment (which abuts the vector sequence as the 1.2 kb HindIII fragment hybridizes with this BamHI fragment, Figure B.1.a), the 1.6 kb BamHI fragment and part of the 7.7 kb BamHI fragment. The claw gene contained in the 4.65 kb HindIII fragment is located in a 2.45 kb BamHI/ HindIII fragment (Figure B.2).

The restriction map of the right-hand end of cosmid 33 is depicted in Figure B.2. The approximate location of the claw gene in the 4.65 kb HindIII fragment (open box) and the other two genes (filled boxes) are shown. The orientation and location of two of the genes have been determined by DNA sequencing from the KpnI sites (L. Whitbread, personal communication). The presence of a possible fourth gene (draw as a dotted box) has been suggested by DNA hybridization (Figure 4.29) and recent DNA sequence analysis (L. Whitbread, personal communication).

SUBCLONING OF THE 6.2 KB HINDIII FRAGMENT

In order to map the feather keratin genes in the 6.2 kb HindIII fragment of cosmid 33, it was subcloned into a HindIII-cut pBR322 vector (Section 4.2). Figure B.3 shows the restriction and Southern analysis of the pH 6.2 subclone. The 6.2 kb HindIII insert is reduced to about 5.95 kb in a BamHI/HindIII double digest (arrows in Figure B.3), indicating that the HindIII insert has a BamHI site about 0.25 kb from one end. The 0.25 kb BamHI/HindIII fragment is visible on a polyacrylamide gel (data not shown). The 3.3 and 1.8 kb EcoRI fragments

of pH 6.2 are entirely contained in the insert, since all digests including the EcoRI enzyme have these two characteristic fragments (Figure B.3.a). A 1.0 kb fragment is present in an EcoRI/HindIII digest of pH 6.2 (data not shown) but this is reduced to 0.76 kb when BamHI is included in the digest (Figure B.3.a). Therefore, the 1.0 kb EcoRI/HindIII fragment lies at the left-hand end of the clone insert as shown in Figure B.3.c, with the BamHI site 0.25 kb from the HindIII site. Both the 3.3 and 1.8 kb EcoRI fragments are detected by both the feather gene probe (Figure B.3.b) and an M13 clone which contains the 5' non-coding and flanking regions of feather keratin gene B (Section 2.2.10; data not shown). These results, and other hybridization data presented in Section 4.3.4.1, demonstrate that each of these EcoRI fragments contain a complete feather keratin gene (Figure B.3.c).

The restriction map of pH 6.2 is shown in Figure B.3.c, with the approximate location of the two feather keratin genes being depicted as black boxes. The order of the two EcoRI fragments in the clone relative to the 1.0 kb EcoRI/HindIII fragment at the left-hand end of the insert is not known. The two genes contained in the 6.2 kb HindIII fragment comprise the two 5' end genes of the feather keratin gene cluster present in these cosmids (see Figure 4.45)

FIGURE B.3

A. pH 6.2 DNA was digested with various restriction enzymes (see below), electrophoresed on 1.2% agarose, stained with ethidium bromide and visualized under UV light (Section 2.2.7). Restriction fragment sizes (kb) are shown; those underlined are derived from the pBR322 vector.

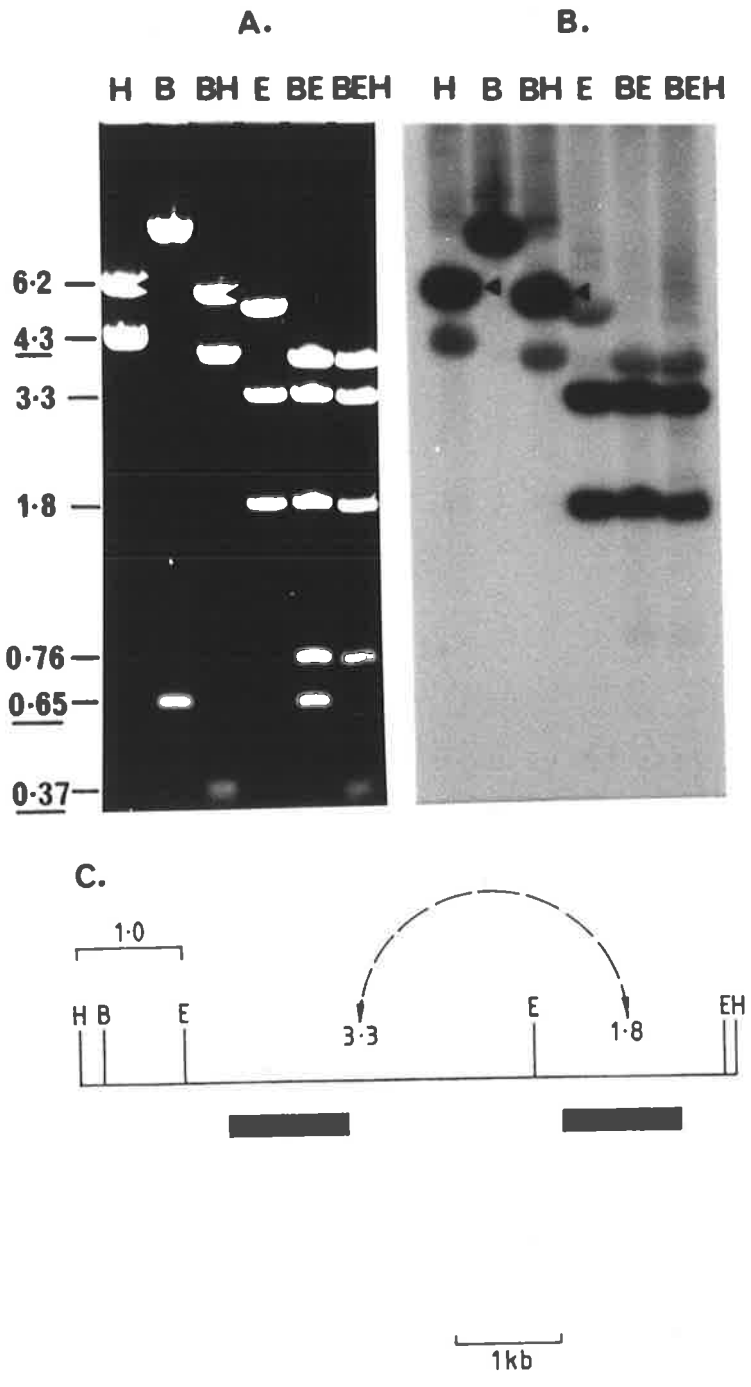
B. The gel was transferred on to a nitrocellulose filter (Section 2.2.9) and the blotted DNA fragments hybridized with the feather keratin gene probe (Section 2.2.10). The filter was washed in 0.5 x SSC, 0.1% SDS at 65°C and exposed for 8 hours at -80°C. Arrows indicate the 6.2 kb HindIII insert of pH 6.2 which is reduced to ~ 5.9 kb in a BamHI/HindIII digest. This blot is the only known example in this thesis where the feather gene probe hybridized to pBR322 (see Figure 3.6).

C. Restriction map of pH 6.2 showing the approximate location of the two feather keratin genes. The order of the two EcoRI fragments with respect to the 1.0 kb EcoRI/HindIII fragment is not known.

B - BamHI

E - EcoRI

H - HindIII


A.
B.
H B BH E BE BEH
H B BH E BE BEH
6·2
4·3
3·3
1·8
0·76
0·65
0·37
C.
1·0
H B
E
3·3
E
1·8
EH
1kb

A P P E N D I X C

RESTRICTION MAPPING OF pH 8.1

DNA from the subclone pH 8.1, a pBR322 clone which contains the 8.1 kb HindIII fragment of cosmid 4 (Section 4.3.4.2a), was restricted with BglII, HindIII and KpnI (both singly and in various combinations) and the digestion products analyzed by agarose gel electrophoresis (Figure C.1.a). The following rationale lead to the derivation of the map shown in Figure C.1.b. Three KpnI/HindIII restriction fragments were produced from the insert of the subclone (3.4, 2.85 and 1.85, Figure C.1.a). Since the 3.4 kb fragment was present in a KpnI digest as well as a KpnI/HindIII digest, the two remaining fragments must map at each end of the insert. (There are no KpnI sites in the pBR322 vector, Sutcliffe (1978)).

A BglII/HindIII digest of pH 8.1 produced three insert restriction fragments of 3.6, 3.45 and 1.05 kb, of which the 3.45 kb fragment was also present in a BglII digest (Figure C.1.a). To order the BglII and KpnI sites relative to each other, KpnI/BglII and KpnI/BglII/HindIII digests of pH 8.1 were performed (data not shown). The 1.85 kb KpnI/HindIII fragment was not cut by BglII whereas the 2.85 kb KpnI/HindIII fragment was (Figure C.1.b), indicating that the small 1.05 kb BglII/HindIII fragment was located in the 2.85 kb KpnI/HindIII fragment (Figure C.1.b). The 3.45 kb BglII fragment spans the remainder of the 2.85 kb KpnI/HindIII fragment and part of the 3.4 kb KpnI fragment. The 1.85 kb KpnI/HindIII fragment, which is cut by BglII, is entirely contained in the 3.6 kb BglII/HindIII fragment. The sizes of restriction fragments obtained in a KpnI/BglII/HindIII digest confirmed the results obtained with the double digests.

FIGURE C.1 : RESTRICTION MAPPING OF pH 8.1

A. pH 8.1 DNA was digested with a number of restriction enzymes (see below) and the digestion products analyzed by agarose gel electrophoresis (see Figure 4.36). Restriction fragment sizes are given in kilobases, including the 4.3 kb pBR322 vector fragment (underlined). The restriction enzymes are indicated as follows:

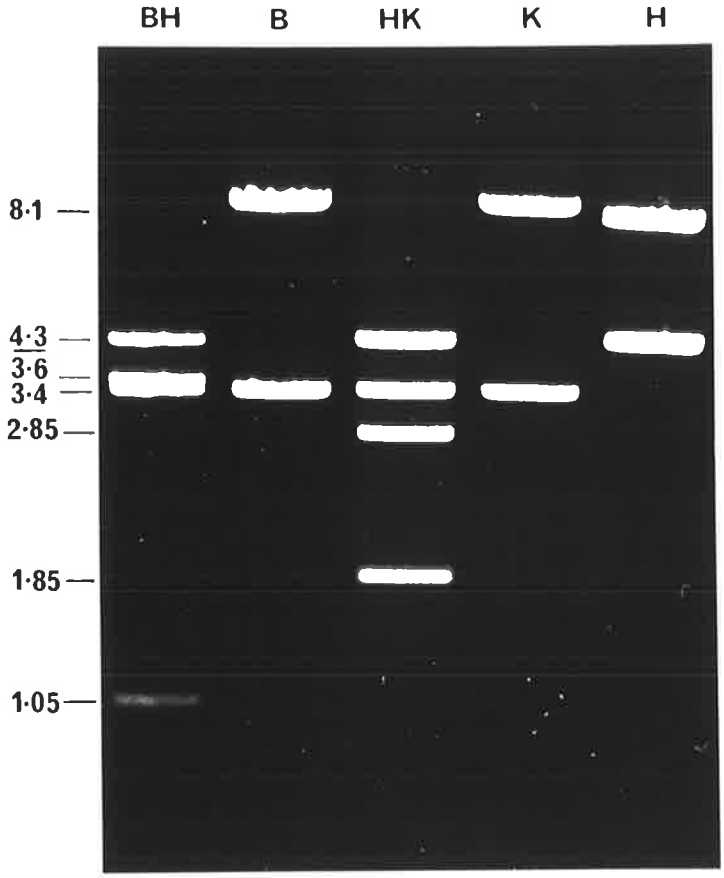
B - BglII

H - HindIII

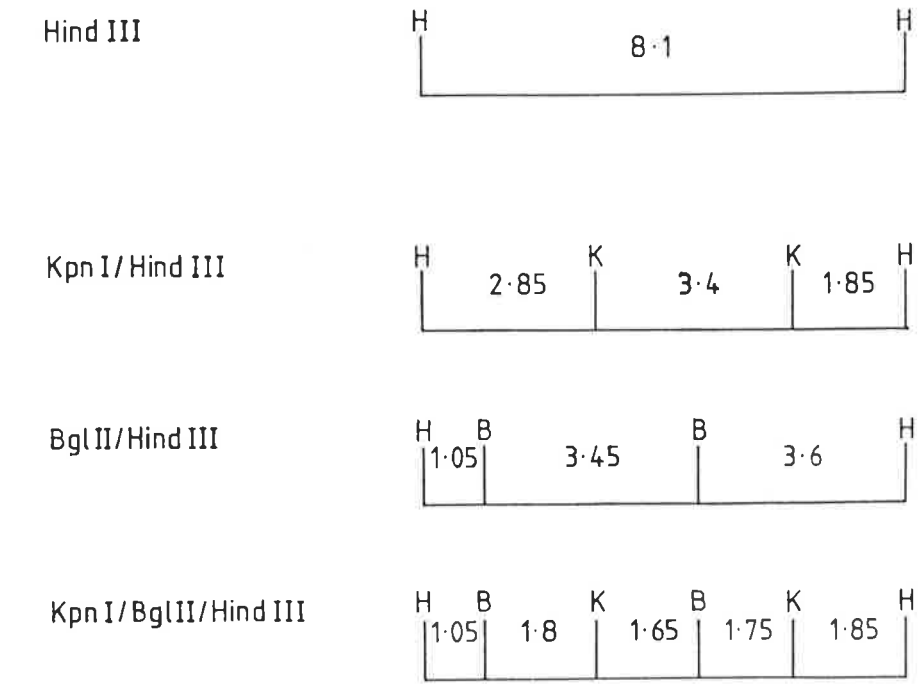
K - KpnI

B. Restriction maps for different combinations of enzymes. The maps were constructed from restriction digests shown in Figure C.1.a and from other data not shown (see text of Appendix C for a description).

A.


BH
B
HK
K
H
8·1
4·3
3·6
3·4
2·85
1·85
1·05

B.


Hind III
H
8·1
H
Kpn I / Hind III
H
2·85
K
3·4
K
1·85
H
Bgl II / Hind III
H
B
1·05
3·45
B
3·6
H
Kpn I / Bgl II / Hind III
H
B
1·05
1·8
K
1·65
B
1·75
K
1·85
H

BIBLIOGRAPHY

Adams, J.M., Webb, E., Gerondakis, S. and Cory, S. (1980). Nucl. Acids Res. 8, 6019-6032.

Arai, K.M. Takahashi, R., Yokote, Y. and Akahane, K. (1983). Eur. J. Biochem. 132, 501-507.

Astbury, W.T. and Marwick, T.C. (1932). Nature 130, 309-310.

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

Baltimore, D. (1981). Cell 24, 592-594.

Bantle, J.A. and Hahn, W.E. (1976). Cell 8, 139-150.

Bartels, E.A. (1943). Arch. F. Entwicklmech. 142, 763-816.

Bathurst, I.C., Craig, R.K. and Campbell, P.N. (1979). Biochem. J. 181, 501-504.

Bathurst, I.C., Craig, R.K., Herries, D.G. and Campbell, P.N. (1980a). Eur. J. Biochem. 109, 183-191.

Bathurst, I.C., Craig, R.K., Herries, D.G. and Campbell, P.N. (1980b). Biochem. J. 192, 489-498.

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

Beckingham-Smith, K. (1973). Dev. Biol. 30, 263-278.

Bell, E. (1964). Natl. Cancer Inst. Monogr. 13, 1-9.

Bell, E. and Merill, C. (1967). **In** The Control of Nuclear Activity (Goldstein, L., ed.), pp. 129-159, Prentice Hall, New Jersey.

Bell, E. and Thathachari, Y.T. (1963). J. Cell Biol. 16, 215-223.

Benoist, C. and Chambon, P. (1981). Nature 290, 304-310.

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

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

Berget, S.M. (1984). Nature 309, 179-182.

Bernard, O., Cory, S., Gerondakis, S., Webb, E. and Adams, J.M. (1983). EMBO J. 2, 2375-2383.

Birnboim, H.C. and Doly, J. (1979). Nucl. Acids Res. 7, 1513-1523.

Birnstiel, M., Portmann, R., Busslinger, M., Schaffner, W., Probst, E. and Kressman, A. (1979). Alfred Benson Symposium 13, 117-129.

Blake, C.C.F. (1978). Nature 273, 267.

Blake, C.C.F. (1983). Nature News and Views 306, 535-537.

Blake, C.C.F. (1985). Int. Rev. Cytol. 93, 149-185.

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.

Bolivar, F. and Backman, K. (1979). Methods Enzymol. 68, 245.

Brawerman, G. (1974). Ann. Rev. Biochem. 43, 621-642.

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

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

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

Cattaneo, R., Gorski, J. and Mach, B. (1981). Nucl. Acids Res. 9, 2777-2790.

Childs, G., Maxson, R., Cohn, R. and Kedes, L.H. (1981). Cell 23, 651-663.

Chou, P.Y. and Fasman, G.D. (1978). Adv. Enzymol. 47, 45-148.

Coen, E.S. and Dover, G.A. (1983). Cell 33, 849-855.

Cohen, J.B., Effron, K., Rechqui, G., Ben-Neriah, Y., Zakut, R. and Givol, D. (1982). Nucl. Acids Res. 10, 3353-3370.

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

Cornish-Bowden, A. (1985). Nature News and Views 313, 434-435.

Craik, C.S., Sprang, S., Fletterick, R. and Rutter, W.J. (1982). Nature 299 , 180-182.

Crewther, W.G. (1976). **In** The Fifth International Wool Textile Research Conference (Ziegler, K., ed.), Aachen, Vol. 1, pp. 1-101.

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

D'Andrea, R.J. (1985). Ph.D. Thesis, University of Adelaide.

D'Andrea, R.J., Coles, L.S., Lesnikowski, C., Tabe, L. and Wells, J.R.E. (1985). Mol. Cell. Biol. 5, 3108-3115.

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.

Dhouailly, D. (1967). J. Embryol. Exptl. Morphol. 18, 389-400.

Dhouailly, D., Hardy, M.H. and Sengel, P. (1980). J. Embryol. Exp. Morphol. 58, 63-78.

Dhouailly, D., Rogers, G.E. and Sengel, P. (1978). Dev. Biol. 65, 58-68.

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

Dodgson, J.B., Strommer, J. and Engel, J.D. (1979). Cell 17, 879-887.

Dodson, J.W. (1963). Exptl. Cell Res. 31, 233-235.

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

Eickbush, T.H. and Kafatos, F.C. (1982). Cell 29, 633-643.

Engel, J.D., Sugarman, B.J. and Dodgson, J.B. (1982). Nature 297, 434-436.

Fell, H.B. (1962). J. Embryol. Exptl. Morphol. 10, 389-409.

Fell, H.B. and Mellanby, E. (1953). J. Physiol. (Lond.) 119, 470-488.

Fietz, M.J. (1985). B.Sc. Honours Thesis, University of Adelaide.

Filshie, B.K. and Rogers, G.E. (1962). J. Cell Biol. 13, 1-12.

Fitzgerald, M. and Shenk T. (1981). Cell 24, 251-260.

Flavell, R.A. and Grosveld, F.G. (1983). In Eukaryote Genes: Their Structure, Activity and Regulation (Maclean, N., Gregory, S.P. and Flavell, R.A., eds.), pp. 299-314, Butterworths, Cambridge, U.K.

Fraser, R.D.B. and MacRae, T.P. (1980). **In** The Skin of Vertebrates (Spearman, R.I.C. and Riley, P.A., eds.) Linnean Society Symposium No. 9, pp. 67-86, Academic Press, London and New York.

Fraser, R.D.B., MacRae, T.P., Parry, D.A.D. and Suzuki, E. (1971). Polymer 12, 35-56.

Fraser, R.D.B., MacRae, T.P. and Rogers, G.E. (1972). 'Keratins, Their Composition, Structure and Biosynthesis', C.C. Thomas, Springfield, Illinois.

Frenkel, M.J. (1975). Proc. Aust. Biochem. Soc. 8, 8.

Frenkel, M.J. (1985). Ph.D. Thesis, University of Adelaide.

Frenkel, M.J. and Gillespie, J.M. (1976). Aust. J. Biol. Sci. 29, 467-479.

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

Gibbs, P.E.M. (1977). Ph.D. Thesis, University of Adelaide.

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

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

Gillespie, J.M. (1983) in Biochemistry and Physiology of the Skin (Goldsmith, L.A., ed.), Vol. 1, pp. 475-510, Oxford University Press, London.

Gitschier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.H., Vehar, G.A., Capon, D.J. and Lawn, R.M. (1984). Nature 312, 326-330.

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

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

Goldsmith, M.R. and Kafatos, F.C. (1984). Ann. Rev. Genet. 18, 443-487.

Grantham, R., Gautier, C. and Gouy, M. (1980). Nucl. Acids Res. 8, 1893-1912.

Gregg, K. and Rogers, G.E. (1986) in The Biology of the Integument (Bereiter-Hahn, J., Matoltsy, A.G. and Richards, K.S., eds.), Vol. 2, pp. 666-694, Springer-Verlag, Berlin.

Gregg, K., Wilton, S.D., Parry, D.A.D. and Rogers, G.E. (1984). EMBO J. 3, 175-178.

Gregg, K., Wilton, S.D, Rogers, G.E. and Molloy, P.L. (1983). In Manipulation and Expression of Genes in Eukaryotes (Nagley, P., Linnane, A.W., Peacock, W.J., Pateman, J.A., eds.), pp. 65-72, Academic Press, Sydney.

Grosveld, F.G., Dahl, H-H.M., de Boer, E. and Flavell, R.A. (1981). Gene 13, 227-237.

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

Guyette, W.A., Matusik, R.J. and Rosen, J.M. (1979). Cell 17, 1013-1023.

Hahn, W.E., Van Ness, J. and Maxwell, I.H. (1978). Proc. Natl. Acad. Sci. USA 75, 5544-5547.

Hanahan, D. and Meselson, M. (1980). Gene 10, 63-67.

Harrap, B.S. and Woods, E.F. (1964a). Biochem. J. 92, 8-18.

Harrap, B.S. and Woods, E.F. (1964b). Biochem. J. 92, 19-26.

Harris, S., Barrie, P.A., Weiss, M.L. and Jeffreys, A.J. (1984). J. Mol. Biol. 180, 785-801.

Hayashida, H. and Miyata, T. (1983). Proc. Natl. Acad. Sci. USA 80, 2671-2675.

Hentschel, C.C. and Birnstiel, M.L. (1981). Cell 25, 301-313.

Hohn, B. and Collins, J. (1980). Gene 11, 291-298.

Hohn, B. and Murray, K. (1977). Proc. Natl. Acad. Sci. USA 74, 3259-3263.

Holmes, D.S. and Quigley, M. (1981). Anal. Biochem. 114, 193-197.

Holt, C.A. and Childs, G. (1984). Nucl. Acids Res. 12, 6455-6471.

Hood, L., Campbell, J.H. and Elgin, S.C.R. (1975). Ann. Rev. Genet. 9, 305-353.

Hu., N-T. and Messing, J. (1982). Gene 17, 271-277.

Iatrou, K., Tsitilou, S.G. and Kafatos, F.C. (1984). Proc. Natl. Acad. Sci. USA 81, 4452-4456.

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

Jackson, J.A. and Fink, G.R. (1981). Nature 292, 306-311.

Jeffreys, A.J. and Harris, S. (1982). Nature 296, 9-10.

Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M. and Williams, J.G. (1980). Cell 21, 555-564.

Jones, C.W. and Kafatos, F.C. (1980). Cell 22, 855-867.

Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985). Cell 42, 559-572.

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

Kaplan, B.B., Bernstein, S.L. and Gioio, A.E. (1979). Biochem. J. 183, 181-184.

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

Karlsson, S. and Nienhuis, A.W. (1985). Ann. Rev. Biochem. 54, 1071-1108.

Kato, Y. (1969). J. Exptl. Zool. 170, 229-252.

Kedes, L.H. (1975). Cell 8, 321-331.

Keller, W. (1984). Cell 39, 423-425.

Keller, E.B. and Noon, W.A. (1984). Proc. Natl. Acad. Sci. USA 81, 7417-7420.

Kemp, D.J. (1975). Nature 254, 573-577.

Kemp, D.J., Dyer, P.Y. and Rogers, G.E. (1974a). J. Cell Biol. 62, 114-131.

Kemp, D.J., Partington, G.A. and Rogers, G.E. (1974b). Biochem. Biophys. Res. Commun. 60, 1006-1014.

Kemp, D.J. and Rogers, G.E. (1972). Biochemistry 11, 969-975.

Kischer, C.W. (1963). J. Ultrastr. Res. 8, 305-321.

Kischer, C.W. and Furlong, N.B. (1967). Proc. Soc. Exp. Biol. Med. 124, 1188-1190.

Kitano, Y. and Kuroda, Y. (1967). Exptl. Cell Res. 48, 350-360.

Klein, H.L. and Petes, T.D. (1981). Nature 289, 144-148.

Koltunow, A.M.G. (1986). Ph.D. Thesis, University of Adelaide.

Koltunow, A.M., Gregg, K. and Rogers, G.E. (1986). Nucl. Acids Res. 14, 6375-6393.

Kourilsky, P. (1986). TIG 2, 60-63.

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

Kozak, M. (1986). Cell 44, 283-292.

Krawinkel, U., Zoebelein, G., Brüggemann, M., Radbruch, A. and Rajewsky, K. (1983). Proc. Natl. Acad. Sci. USA 80, 4997-5001.

Kuczek, E.S. (1980). B.Sc. (Hons.) Thesis, University of Adelaide.

Kuczek, E.S. (1985). Ph.D. Thesis, University of Adelaide.

Kuczek, E. and Rogers, G.E. (1985). Eur. J. Biochem. 146, 89-93.

Kwan, S-P., Wood, T.G. and Lingrel, J.B. (1977). Proc. Natl. Acad. Sci. USA 74, 178-182.

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

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

Lauer, J., Shen, C-K.J. and Maniatis, T. (1980). Cell 20, 119-130.

Leder, A., Swan, D., Ruddle, F., d'Eustachio, P. and Leder, P. (1981). Nature 293, 196-200.

Lillie, F.R. (1965). 'Lillie's Development of the Chick' (Hamilton, H.L., ed.) pp. 546-606, Holt, Rinehart and Winston, New York.

Linsenmayer, T.F. (1972). Dev. Biol. 27, 244-271.

Lockett, T.J., Kemp, D.J. and Rogers, G.E. (1979). Biochemistry 18, 5654-5663.

Lowenhaupt, K. and Lingrel, J.B. (1978). Cell 14, 337-344.

Lucas, A.M. and Stettenheim, P.R. (1972). Avian Anatomy-Integument Part II 362, U.S. Department of Agriculture, Washington.

Lund, T., Grosveld, F.G. and Flavell, R.A. (1982). Proc. Natl. Acad. Sci. USA 79, 520-524.

Macnaughton, M., Freeman, K.B. and Bishop, J.O. (1974). Cell 1, 117-125.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning - A Laboratory Manual. Cold Spring Harbor 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.

Marshall, A.J. and Burgoyne, L.A. (1976). Nucl. Acids Res. 3, 1101-1110.

Marzluff, W.F. and Graves, R.A. (1984). In Histone Genes, Structure, Organization and Regulation (Stein, G.S., Stein, J.L. and Marzluff, W.F., eds.), pp. 281-315, Wiley and Sons, New York.

Mathis, D.J. and Chambon, P. (1981). Nature 290, 310-315.

Matulionis, D.H. (1970). Z. Anat. Entwicklungs Geschichte 132, 107-157.

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

Maxson, R., Mohun, T. and Kedes, L. (1983). In Eukaryote Genes: Their Structure, Activity and Regulation (Maclean, N., Gregory, S.P. and Flavell, R.A., eds.), pp. 277-298, Butterworths, Cambridge, U.K.

McBride, L.J. and Caruthers, M.H. (1983). Tetrahedron Letters 24, 245-248.

Meinkoth, J. and Wahl, G. (1984). Anal. Biochem. 138, 267-284.

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

Messing, J. (1983). Methods Enzymol. 101, 20-78.

Messing, J., Crea, R. and Seeburg, P.H. (1981). Nucl. Acids Res. 9, 309-321.

Messing, J. and Vieira, J. (1982). Gene 19, 269-276.

Meyerowitz, E.M., Guild, G.M., Prestidge, L.S. and Hogness, D.S. (1980). Gene 11, 271-282.

Molloy, P.L., Powell, B.C., Gregg, K., Barone, E.D. and Rogers, G.E. (1982). Nucl. Acids Res. 10, 6007-6021.

Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983). Nature 305, 600-605.

Morris, C.P. (1984). Ph.D. Thesis, University of Adelaide.

Morris, C.P. and Rogers, G.E. (1979). Mol. Biol. Reports 5, 145-149.

Mount, S.M. (1982). Nucl. Acids Res. 10, 459-472.

Müller, F. and Clarkson, S.G. (1980). Cell 19, 345-353.

Murray, N.E., Brammar, W.J. and Murray, K. (1977). Mol. Gen. Genet. 150, 53-61.

Norrander, J., Kempe, T. and Messing, J. (1983). Gene 26, 101-106.

Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohen, S.N. and Schimke, R.T. Cell 19, 355-364.

O'Donnell, I.J. (1973). Aust. J. Biol. Sci. 26, 415-437.

O'Donnell, I.J. and Inglis, A.S. (1974). Aust. J. Biol. Sci. 27, 369-382.

Ostrom, J.H. (1976). Linnaean Soc. Lond. Biol. J. 8, 91-182.

Parker, C.S. and Topol, J. (1984). Cell 37, 273-283.

Partington, G.A, Kemp, D.J. and Rogers, G.E. (1973). Nature New Biology 246, 33-36.

Petes, T.D. (1980). Cell 19, 765-774.

Powell, B.C. (1979). Ph.D. Thesis, University of Adelaide.

Powell, B.C., Cam, G.R., Fietz, M.J. and Rogers, G.E. (1986). Proc. Natl. Acad. Sci. USA 83, 5048-5052.

Powell, B.C., Kemp, D.J., Partington, G.A., Gibbs, P.E.M. and Rogers, G.E. (1976). Biochem. Biophys. Res. Commun. 68, 1263-1271.

Powell, B.C. and Rogers, G.E. (1979). Nucl. Acids Res. 7, 2165-2176.

Powell, B.C. and Rogers, G.E. (1986). **In** Biology of the Integument (Bereiter-Hahn, J., Matoltsy, A.G. and Richards, K.S., eds.), Vol. 2, pp. 695-721, Springer-Verlag, Berlin.

Powell, B.C., Sleigh, M.J., Ward, K.A. and Rogers, G.E. (1983). Nucl. Acids Res. 11, 5327-5346.

Proudfoot, N.J. and Maniatis, T. (1980). Cell 21, 537-544.

Rabbitts, T.H. (1978). Nature 275, 291-296.

Radding, C.M. (1978). Ann. Rev. Biochem. 47, 847-880.

Rawles, M.E. (1963). J. Embryol. Exptl. Morphol. 11, 765-789.

Rawles, M.E. (1965). **In** 'Biology of Skin and Hair Growth' (Lyne, A.G. and Short, B.F., eds.) pp. 105-128, Angus and Robertson, Sydney.

RayChaudhury, A., Marchuk, D., Lindhurst, M. and Fuchs, E. (1986). Mol. Cell. Biol. 6, 539-548.

Reddy, R., Henning, D. and Busch, H. (1981). J. Biol. Chem. 256, 3532-3538.

Revel, M. and Groner, Y. (1978). Ann. Rev. Biochem. 47, 1079-1126.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). J. Mol. Biol. 113, 237-251.

Roeder, G.S., Smith, M. and Lambie, E.J. (1984). Mol. Cell. Biol. 4, 703-711.

Rogers, G.E. (1959a). Ann. N.Y. Acad. Sci. 83, 378-399.

Rogers, G.E. (1959b). J. Ultrastruct. Res. 2, 309-330.

Rogers, G.E. (1984). Biochem. Soc. Symp. 49, 85-108.

Rogers, J. (1985). Nature News and Views 315, 458-459.

Romanoff, A.L. (1960). 'The Avian Embryo' MacMillan, New York.

Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M-J., O'Malley, B.W. (1978). Cell 15, 671-685.

Ross, J. and Knecht, D.A. (1978). J. Mol. Biol. 119, 1-20.

Rothnagel, J.A. (1979). B.Sc. Honours Thesis, University of Adelaide.

Rusconi, S. (1986). J. Cell Biochem. Suppl. 10D, UCLA Symposium on Transcriptional Control Mechanisms, 179.

Saint, R.B. (1979). Ph.D. Thesis, University of Adelaide.

Sambrook, J. (1977). Nature News and Views 268, 101-104.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980). J. Mol. Biol. 143, 161-178.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Sawyer, R.H. (1979). Dev. Biol. 68, 1-15.

Sawyer, R.H., Knapp, L.W. and O'Guin, W.M. (1986). **In** Biology of the Integument (Bereiter-Hahn, J., Matoltsy, A.G. and Richards, K.S., eds.), Vol. 2, pp. 194-238, Springer-Verlag, Berlin.

Schibler, U., Marcu, K.B. and Perry, R.P. (1978). Cell 15, 1495-1509.

Schwartz, D.C. and Cantor, C.R. (1984). Cell 37, 67-75.

Sengel, P. (1958). Année Biologique 34, 29-52.

Sengel, P. (1971). Advances in Morphogenesis 9, 181-230.

Sengel, P. (1986). **In** 'Biology of the Integument' (Bereiter-Hahn, J., Matoltsy, A.G. and Richards, K.S., eds.), Vol. 2, pp. 374-408, Springer-Verlag, Berlin.

Shain, W.G., Hilfer, S.R. and Fonte, V.G. (1972). Dev. Biol. 28, 202-218.

Shen, S-L., Slightom, J.L. and Smithies, O. (1981). Cell 26, 191-203.

Slightom, J.L., Blechl, A.E. and Smithies, O. (1980). Cell 21, 627-638.

Smith, G.P. (1976). Science 191, 528-535.

Smith, G.E. and Summers, M.D. (1980). Anal. Biochem. 109, 123-129.

Southern, E.M. (1975). J. Mol. Biol. 98, 503-517.

Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A. and Baltimore, D. (1986). Nature 323, 640-643.

Steiner, D.F., Chan, S.J., Welsh, J.M. and Kwok, S.C.M. (1985). Ann. Rev. Genet. 19, 463-484.

Steinert, P.M., Steven, A.C. and Roop, D.R. (1985). Cell 42, 411-419.

Spearman, R.I.C. (1966). Biol. Rev. 41, 59-96.

Stein, G.S., Sierra, F., Stein, J.L., Plumb, M., Marashi, F., Corozzi, N., Prokopp, K. and Baumbach, L. (1984). In Histone Genes, Structure, Organization and Regulation (Stein, G.S., Stein, J.L. and Marzluff, W., eds.). pp. 397-455, Wiley and Sons, New York.

Steinmetz, M., Stephan, D. and Fischer Lindahl, K. (1986). Cell 44, 895-904.

Steinmetz, M., Winoto, A., Minard, K. and Hood, L. (1982). Cell 28, 489-498.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977). Gene 1, 255-280.

Stewart, M. (1977). J. Ultrastruct. Res. 60, 27-33.

Strair, R.K., Yap, S.H., Nadal-Ginard, B. and Shafritz, D.A. (1978). J. Biol. Chem. 253, 1328-1331.

Sugimoto, M. and Endo, H. (1969). Nature 222, 1270-1272.

Sugimoto, M., Tajima, K., Kojima, A. and Endo, H. (1974). Dev. Biol. 39, 295-307.

Sutcliffe, J.G. (1978). Nucl. Acids Res. 5, 2721-2728.

Suzuki, Y., Gage, L.P. and Brown, D.D. (1972). J. Mol. Biol. 70, 637-649.

Szostak, J.W. and Wu, R. (1980). Nature 284, 426-430.

Thomas, P.S. (1980). Proc. Natl. Acad. Sci. USA 77, 5201-5205.

Thomas, P.S. (1983). Methods Enzymol. 100, 255-266.

Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Matija Peterlin, B., Sullivan, M., Maizel, J.V. and Leder, P. (1978a). Proc. Natl. Acad. Sci. USA 75, 725-729.

Tilghman, S.M., Curtis, P.J., Tiemeier, D.C., Leder, P. and Weissmann, C. (1978b). Proc. Natl. Acad. Sci. USA 75, 1309-1313.

Townes, T.M., Shapiro, S.G., Wernke, S.M. and Lingrel, J.B. (1984). J. Biol. Chem. 259, 1896-1900.

Voitkevich, A.A. (1966). 'The Feathers and Plumage of Birds', Sidgwick and Jackson, London.

Wahl, G.M., Stern, M. and Stark, G.R. (1979). Proc. Natl. Acad. Sci. USA 76, 3683-3687.

Walker, I.D. and Bridgen, J. (1976). Eur. J. Biochem. 67, 283-293.

Walker, I.D. and Rogers, G.E. (1976a). Eur. J. Biochem. 69, 329-339.

Walker, I.D. and Rogers, G.E. (1976b). Eur. J. Biochem. 69, 341-350.

Wang, S-W., Robins, A.J., D'Andrea, R.J. and Wells, J.R.E. (1985). Nucl. Acids Res. 13, 1369-1387.

Weiss, E.H., Cheah, K.S.E., Grosveld, F.G., Dahl, H-H.M., Solomon, E. and Flavell, R.A. (1982). Nucl. Acids Res. 10, 1981-1994.

Wessells, N.K. (1961a). Dev. Biol. 3, 355-389.

Wessells, N.K. (1961b). Exptl. Cell Res. 24, 131-142.

Wessells, N.K. (1962). Dev. Biol. 4, 87-107.

Wessells, N.K. (1965). Dev. Biol. 12, 131-153.

Whitbread, L.A. (1985). B.Sc. Honours Thesis, University of Adelaide.

Williams, B.G. and Blattner, F.R. (1979). J. Virol. 29, 555-575.

Wilton, S.D. (1978). B.Sc. Honours Thesis, University of Adelaide.

Wilton, S.D. (1983). Ph.D. Thesis, University of Adelaide.

Wilton, S.D., Crocker, L.A. and Rogers, G.E. (1983). Biochim. Biophys. Acta 824, 201-208.

Winter, F. and Fields, S. (1980). Nucl. Acids Res. 8, 1965-1974.

Yatvin, M.B. (1966a). Science 151, 1001-1003.

Yatvin, M.B. (1966b). Science 153, 184-185.

Young, R.A., Hagenbuchle, O. and Schibler, U. (1981). Cell 23, 451-458.