CHICK

EMBRYONIC FEATHER

GENES

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

by

CHARLES PHILLIP MORRIS B.Sc.(Hons.)

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

October, 1984.

To Debra

for caring.

TABLE OF CONTENTS

		PAGE
THESIS SUMMARY		i
STATEMENT		v
ACKNOWLEDGEMENTS		vi

CHAPTER 1

INTRODUCTION.

1.1	PRELIMINARY COMMENTS.	1
1.2	CHICK EMBRYONIC FEATHER DEVELOPMENT.	2
1.2.1	FEATHER MORPHOGENESIS.	2
1.2.2	THE ROLE OF THE DERMIS.	4
1.2.3	OTHER FACTORS AFFECTING FEATHER DIFFERENTIATION.	6
1.3	THE NATURE OF EMBRYONIC FEATHER PROTEINS.	8
1.3.1	FEATHER SPECIFIC β-KERATINS.	8
1.3.2	FAST PROTEINS.	10
1.3.3	SCALE LIKE β-KERATINS.	12
1.3.4	OTHER PROTEINS FOUND IN EMBRYONIC FEATHERS.	13
1.3.5	SUMMARY.	14
1.4	GENE EXPRESSION IN THE EMBRYONIC FEATHER.	14
1.4.1	THE TIME COURSE OF PROTEIN SYNTHESIS.	14
1.4.2	THE TIME COURSE OF mRNA SYNTHESIS.	15
1.4.3	THE NATURE OF THE mRNAs EXPRESSED IN EMBRYONIC FEATHERS.	15
1.4.4	THE NATURE OF THE GENES EXPRESSED IN EMBRYONIC FEATHERS.	17

1.4.5 SUMMARY. 20
1.5 THE MAMMALIAN EPIDERMAL TISSUES. 20
1.6 AIMS OF THE PROJECT. 23
1.6.1 GENERAL AIMS. 23
1.6.2 SPECIFIC AIMS OF THIS THESIS. 24

CHAPTER 2

MATERIALS AND GENERAL METHODS.

2.1 MATERIALS. 25
2.1.1 FINE CHEMICALS FOR SPECIFIC PROCEDURES. 25
2.1.1a ELECTROPHORESIS. 25
2.1.1b CELL-FREE TRANSLATION. 25
2.1.1c CELL-FREE TRANSCRIPTION. 25
2.1.1d COLUMN CHROMATOGRAPHY. 26
2.1.1e FLUOROGRAPHY. 26
2.1.1f SCINTILLATION COUNTING. 26
2.1.2 ENZYMES. 27
2.1.3 RADIOCHEMICALS. 27
2.1.4 BUFFERS. 28
2.1.5 BACTERIAL MEDIA. 28
2.1.5a LIQUID MEDIA. 28
2.1.5b SOLID MEDIA. 29
2.1.6 ANTIBIOTICS. 29
2.1.7 BACTERIAL STRAINS. 29
2.1.8 BACTERIOPHAGE STRAINS. 30
2.1.9 TISSUE. 30
2.1.10 MISCELLANEOUS MATERIALS. 30
2.1.11 MISCELLANEOUS FINE CHEMICALS. 31

2.2 GENERAL METHODS. 31
2.2.1 PREPARATION OF GLASSWARE AND SOLUTIONS. 31
2.2.2 PREPARATION OF FEATHER LYSATE. 32
2.2.3 PREPARATION OF FEATHER RNA. 32
2.2.3a PREPARATION OF FEATHER POLYSOMES. 32
2.2.3b PHENOL EXTRACTION OF POLYSOMES. 33
2.2.3c EDTA DISSOCIATION OF POLYSOMES. 33
2.2.3d OLIGO(dT) CELLULOSE CHROMATOGRAPHY OF RNA. 33
2.2.4 PREPARATION OF PURIFIED KERATIN mRNA. 34
2.2.5 CELL-FREE TRANSLATION OF RNA. 34
2.2.6 S-CARBOXYMETHYLATION OF TRANSLATION PRODUCTS. 35
2.2.7 GEL ELECTROPHORESIS OF PROTEINS AT pH 2.7. 36
2.2.8 SDS GEL ELECTROPHORESIS OF PROTEIN. 36
2.2.9 ETHANOL PRECIPITATION. 36
2.2.10 PREPARATION OF PLASMID DNA. 37
2.2.11 RESTRICTION OF DNA. 38
2.2.12 AGAROSE GEL ELECTROPHORESIS OF DNA. 38
2.2.13 POLYACRYLAMIDE GEL ELECTROPHORESIS OF DNA. 39
2.2.14 OLIGO(dT) PRIMED REVERSE TRANSCRIPTION. 39
2.2.15 DETECTION OF RECOMBINANT PLASMIDS. 40
2.2.16 TRANSFER OF DNA TO NITROCELLULOSE. 40
2.2.17 HYBRIDIZATION OF RADIOACTIVE PROBES TO IMMOBILIZED DNA. 40
2.2.18 ISOLATION OF RESTRICTION FRAGMENTS. 41
2.2.18a FROM POLYACRYLAMIDE GELS. 41
2.2.18b FROM LOW MELTING POINT AGAROSE GELS. 42
2.2.19 RADIOLABELLING OF DNA. 42

2.2.19a 5′-TERMINAL RADIOLABELLING. 42
2.2.19b 3′-TERMINAL RADIOLABELLING. 42
2.2.19c NICK TRANSLATION OF DOUBLE-STRANDED DNA. 43
2.2.19d EXTENSION OF UNIVERSAL PRIMER ON M13 SUB-CLONES. 43
2.2.20 SEPARATION OF RADIOLABELLED TERMINI. 44
2.2.20a STRAND SEPARATION. 44
2.2.20b SECONDARY RESTRICTION. 45
2.2.21 CHEMICAL CLEAVAGE DNA SEQUENCING REACTIONS. 45
2.2.22 THE M13 CLONING SYSTEM. 47
2.2.22a PREPARATION OF M13 REPLICATIVE FORM DNA. 48
2.2.22b PREPARATION OF M13 VECTORS. 49
2.2.22c LIGATION CONDITIONS. 49
2.2.22d TRANSFORMATION OF *E. coli*, JM101. 50
2.2.22e HARVESTING (+) STRAND OF RECOMBINANT BACTERIOPHAGE. 50
2.2.23 DIDEOXY SEQUENCING REACTIONS. 50
2.2.23a HYBRIDIZATION. 50
2.2.23b POLYMERIZATION. 50
2.2.23c SEQUENCING GEL ELECTROPHORESIS. 51
2.2.24 CONTAINMENT FACILITIES. 51
2.2.25 COMPUTER ASSISTED DNA SEQUENCE ANALYSIS. 51

CHAPTER 3

PRELIMINARY CHARACTERIZATION OF PUTATIVE mRNAs IN EMBRYONIC FEATHER.

3.1 INTRODUCTION. 53
3.2 SPECIFIC METHODS. 56

3.2.1 RANDOM PRIMED REVERSE TRANSCRIPTION. 56
3.2.2 RNA HYBRIDIZATION REACTIONS. 56
3.2.3 POLYACRYLAMIDE GEL ELECTROPHORESIS OF RNA. 57
3.2.3a 98% FORMAMIDE GELS. 57
3.2.3b 8 M UREA GELS. 58
3.2.4 DETECTION OF RNA SAMPLES ON POLYACRYLAMIDE GELS. 58
3.2.5 ^{3}H-BOROHYDRIDE LABELLING OF KERATIN mRNA. 59
3.2.6 SIZING OF POLY(A) TRACTS. 60
3.2.7 DETECTION OF POLY(A) USING POLY(U). 60
3.2.8 NUCLEOTIDE SEQUENCING OF RNA. 61
3.2.8a PLUS-MINUS SEQUENCING. 61
3.2.8b DIDEOXY SEQUENCING. 62
3.3 RESULTS. 62
3.3.1 ANALYSIS OF THE RNAs PRESENT IN 14-DAY FEATHERS. 62
3.3.1a PURIFICATION OF 9, 12 AND 14S RNA FROM RIBONUCLEOPROTEIN PARTICLES. 62
3.3.1b TRANSLATION OF RNA IN A CELL-FREE SYSTEM. 63
3.3.1c ANALYSIS OF TRANSLATION PRODUCTS BY ELECTROPHORESIS AT pH 2.7. 64
3.3.1d PURIFICATION OF RNA FROM SDS DISSOCIATED POLYSOMES. 65
3.3.1e OLIGO(dT) CELLULOSE CHROMATOGRAPHY OF RNA. 66
3.3.1f TRANSLATION OF RNA IN A CELL-FREE SYSTEM. 66
3.3.1g ANALYSIS OF TRANSLATION PRODUCTS BY ELECTROPHORESIS AT pH 2.7. 66

3.3.1h 14S RNA cDNA SYNTHESIS AND HYBRIDIZATION KINETICS. 68

3.3.2 DETAILED STUDIES OF 12S KERATIN mRNA. 69

3.3.2a 12S mRNA POLY(A) LENGTH. 69

3.3.2b ATTEMPTS TO OBTAIN STRUCTURAL INFORMATION ON KERATIN mRNA BY RESTRICTION OF KERATIN cDNA. 72

3.3.2c ATTEMPTS TO ANALYSE KERATIN mRNA 3'-SEQUENCES. 75

3.4 DISCUSSION. 77

3.4.1 THE NATURE OF 14-DAY EMBRYONIC FEATHER mRNAs. 77

3.4.2 KERATIN mRNA FINE STRUCTURE. 79

CHAPTER 4

KERATIN mRNA SEQUENCES.

4.1 INTRODUCTION. 83

4.2 SPECIFIC METHODS. 86

4.2.1 MINISCREEN EXAMINATION OF PLASMID RECOMBINANTS. 86

4.3 RESULTS. 86

4.3.1 DETERMINATION OF RECOMBINANT PLASMID INSERT SIZE. 86

4.3.2 DIGESTION OF RECOMBINANTS WITH Hae III. 87

4.3.3 SELECTION OF A GROUP I cDNA CLONE FOR FURTHER STUDY. 87

4.3.4 RESTRICTION MAPPING OF pCFK23. 88

4.3.5 RADIOLABELLING RESTRICTION FRAGMENTS FOR DNA SEQUENCING. 89

4.3.6	SEPARATION OF ^{32}P-LABELLED TERMINI.	90
4.3.7	CHEMICAL CLEAVAGE DNA SEQUENCING REACTIONS.	91
4.3.8	M13 CLONING AND DIDEOXY SEQUENCING.	92
4.3.9	THE NUCLEOTIDE SEQUENCE OF pCFK23.	93
4.3.10	SECONDARY STRUCTURE IN pCFK23.	95
4.3.11	SELECTION OF A FULL LENGTH cDNA CLONE.	96
4.3.12	THE PARTIAL NUCLEOTIDE SEQUENCE OF pCFK26.	97
4.3.13	pCFK26 - A TERATOGENY.	97
4.4	DISCUSSION.	99
4.4.1	GROUP I SEQUENCES CODE FOR EMBRYONIC FEATHER β-KERATIN.	99
4.4.2	THE NEW MODEL FOR KERATIN mRNA STRUCTURE.	99
4.4.3	THE CODING SEQUENCE OF KERATIN mRNA.	100
4.4.4	THE 3′-UNTRANSLATED SEQUENCE OF KERATIN mRNA.	101
4.4.5	THE 5′-UNTRANSLATED REGION OF KERATIN mRNA.	102

CHAPTER 5

NON-KERATIN mRNA SEQUENCES.

5.1	INTRODUCTION.	104
5.2	SPECIFIC METHODS.	105
5.2.1	ELECTROPHORESIS OF DENATURED RNA.	105
5.2.2	TRANSFER OF GLYOXYLATED RNA TO NITROCELLULOSE.	105
5.2.3	HYBRID SELECTION.	106
5.2.3a	BINDING DNA TO NITROCELLULOSE.	106
5.2.3b	HYBRIDIZATION AND ELUTION OF RNA.	106
5.3	RESULTS.	107

5.3.1 SELECTION OF A GROUP II CLONE FOR SEQUENCING. 107

5.3.2 THE NUCLEOTIDE SEQUENCE OF pCFK22. 108

5.3.3 pCFK22 SEQUENCES IN THE cDNA LIBRARY. 109

5.3.4 THE NUCLEOTIDE SEQUENCE OF pCFK10. 110

5.3.5 THE LENGTH OF THE pCFK22 mRNA. 111

5.3.6 HYBRID RELEASE TRANSLATION. 112

5.3.7 NUCLEOTIDE SEQUENCE OF pCFK3. 113

5.3.8 THE NUCLEOTIDE SEQUENCE OF pCFK15. 115

5.3.9 FREQUENCY OF SCALE SEQUENCES IN THE FEATHER cDNA LIBRARY. 116

5.3.10 THE PRESENCE OF pCFK15 SEQUENCES IN FEATHER mRNA. 117

5.4 DISCUSSION. 117

5.4.1 THE PRESENCE OF SCALE KERATIN mRNA IN EMBRYONIC FEATHERS. 117

5.4.2 GROUP II SEQUENCES CODE FOR FAST PROTEIN. 119

CHAPTER 6

IDENTIFICATION OF FAST PROTEIN GENES.

6.1 INTRODUCTION. 121

6.2 SPECIFIC METHODS. 123

6.2.1 SCREENING THE GENOMIC LIBRARY. 123

6.2.2 PREPARATION OF RECOMBINANT PHAGE DNA. 124

6.2.3 EXTENSION OF DNA PRIMERS ON FEATHER mRNA. 124

6.2.4 PRIMER EXTENDED SEQUENCING. 125

6.2.5 FPLC ANION EXCHANGE CHROMATOGRAPHY. 126

6.2.6 FPLC CATION EXCHANGE CHROMATOGRAPHY. 126

6.2.7	HPLC REVERSE PHASE C18 CHROMATOGRAPHY.	126
6.2.8	HPLC REVERSE PHASE DIPHENYL CHROMATOGRAPHY.	127
6.2.9	AUTOMATED PROTEIN SEQUENCING.	127
6.3	RESULTS.	128
6.3.1	CHARACTERIZATION OF THE pCFK22 GENE.	128
6.3.1a	SELECTION OF RECOMBINANTS FROM THE GENOMIC LIBRARY.	128
6.3.1b	LOCALIZATION OF THE pCFK22 GENE IN λCFP1.	129
6.3.1c	POSITIVE IDENTIFICATION OF THE pCFK22 GENE.	131
6.3.1d	RESTRICTION MAPPING OF λCFP1.	132
6.3.1e	THE NUCLEOTIDE SEQUENCE OF THE pCFK22 GENE.	133
6.3.1f	COMPARISON OF THE pCFK22 GENE WITH THE KERATIN GENES.	137
6.3.2	PURIFICATION AND SEQUENCING OF FAST PROTEIN.	139
6.3.2a	SEPARATION OF FAST PROTEINS BY ION EXCHANGE FPLC.	139
6.3.2b	PURIFICATION OF FAST PROTEINS BY REVERSE PHASE HPLC.	143
6.3.2c	PARTIAL AMINO ACID SEQUENCING OF FAST PROTEIN COMPONENTS.	145
6.4.	DISCUSSION.	147
6.4.1	ISOLATION OF SEQUENCES CODING FOR FAST PROTEINS. A FAMILY OF SEQUENCE RELATED GENES.	147
6.4.2	THE NATURE OF FAST PROTEINS.	149
6.4.3	THE BIOLOGICAL ROLE OF FAST PROTEINS.	150
6.4.4	CONTROL OF FAST PROTEIN GENE EXPRESSION.	151

CHAPTER 7

CONCLUDING DISCUSSION.

7.1	FEATURES OF THE EMBRYONIC FEATHER SYSTEM.	153
7.1.1	PROTEINS EXPRESSED DURING DIFFERENTIATION.	153
7.1.2	CONTROL OF GENE EXPRESSION.	156
7.1.3	EVOLUTION OF FEATHER GENES.	158
7.2	THE EMBRYONIC FEATHER SYSTEM — FUTURE PROSPECTS.	159

BIBLIOGRAPHY

THESIS SUMMARY.

The chicken embryonic feather, an epidermal tissue, undergoes a process of terminal differentiation typified by the synthesis of large amounts of β-keratins and an associated set of proteins called fast proteins. The fast proteins are characterized by a high electrophoretic mobility and a high content of histidine. The research presented in this thesis was directed toward the molecular characterisation of the major genes expressed during the differentiation of embryonic chick feather tissue. In the absence of recombinant DNA research facilities, the early work was directed toward the direct physical characterisation of the RNAs expressed in feathers by using classical RNA separation and *in vitro* translation techniques. When recombinant DNA research became possible, a more precise description of the genes expressed in feathers was achieved by the analysis of feather cDNA clones and chicken genomic DNA clones. The following results were obtained.

1. Sucrose gradient fractionation of feather RNA revealed several size classes in the range 9 to 14S, all of which were shown to contain some β-keratin translational activity. A low level of fast protein translational activity was demonstrated for both 9 and 12S size classes. A consistently observed 14S RNA peak was shown to contain little translation activity and it was identified as being largely derived from ribosomal RNA sequences by R_0t analysis.

2. Detailed studies on the keratin coding 12S RNA species were undertaken to examine its fine structure. The size of the keratin mRNA poly(A) tract was estimated to be 65 nucleotides by direct detection of stained poly(A), 3′-terminal RNA labelling and hybridization of poly(A) sequences to ^{3}H-poly(U). Restriction enzyme digestion of single-stranded keratin cDNA revealed that the keratin mRNA consisted of a complex family of RNA species. However, detection of 3′-terminally derived cDNA restriction fragments by end-labelling, prior to restriction cleavage, revealed only a few

fragments, indicating some degree of RNA sequence conservation at the 3′-end.

Attempts to directly sequence the 3′-ends of some keratin mRNAs revealed that the majority of the RNA molecules contained 30 to 40 residues at their 3′-termini, which were conserved and which contained the typical AAUAAA polyadenylation sequence, while the atypical AUUAAA sequence was shown to be present at low frequency in 12S feather mRNA.

It was concluded that classical RNA separation techniques were unable to effectively resolve the RNA species of the embryonic feather and the available techniques for the direct analysis of RNA could not successfully distinguish components of the complex mixture of molecular species present in keratin mRNA.

3. The availability of a cDNA library made from embryonic feather mRNA allowed a more detailed molecular analysis of mRNAs expressed in the feather. Two classes of sequences were present in the library (termed Group I and Group II). DNA sequence analysis of several Group I clones revealed that this sequence class coded for embryonic feather β-keratins. The sequence data extended from the poly(A) sequence at the 3′-end to the last nucleotide at the 5′-end of the RNA and allowed the first elucidation of the major proportion of a feather keratin protein amino acid sequence as well as providing a very high resolution description of the structure of keratin mRNA molecules. A comparison of the coding and 3′ and 5′-untranslated sequences between particular keratin mRNAs and between keratin gene sequences revealed a pattern of gene structure consistent among these sequences, as follows. Keratin genes were found to be a family of 8 to 40 genes, each gene containing coding regions for 97 amino acids which are strongly conserved. The long (430 to 440 bp) 3′-untranslated regions of these genes were found on average to be less conserved than the coding regions, although two

strongly conserved blocks of sequence were observed, one of which was found at the 3′-end of the mRNA. The 5′-untranslated region, which contained the single intron, revealed rigid homology between the genes only in the region 5′ to the position of the intron. It is proposed that the rigidly conserved region of the 5′-untranslated region is the site of interaction of a tissue or stage specific effector molecule involved in the initiation of transcription of stage or tissue specific sets of genes in chick epidermal tissues.

4. The nucleotide sequences of four Group II cDNA clones are presented, one of these, pCFK15, was identified as coding for the typical β-keratin protein found in embryonic scale tissue. This was confirmed by Northern transfer analysis using pCFK15 as a probe. The nucleotide sequences of two of the remaining Group II cDNA clones pCFK10 and 22 revealed that they were derived from the same gene and the sequence data indicated that they coded for a protein with a similar amino acid composition to that of fast proteins. A positive identification was not possible since these clones were not full length copies of the mRNA and no protein sequencing data was available for fast proteins, however, hybrid release data indicated that these clones and another clone, pCFK3, selected mRNAs coding for a protein which co-migrated with fast protein on SDS gels. The fourth Group II cDNA clone, pCFK3, which was presumed to be derived from a fast protein mRNA was different in sequence from pCFK10 and 22 and contained only 3′-untranslated sequence.

5. The full protein coding potential (120 amino acids) of the gene from which pCFK10 and 22 were derived was revealed by the isolation and DNA sequencing of a portion of a λ-Charon 4A chick genomic clone. The gene was shown to contain an intron in the 5′-untranslated region at the same position as the keratin genes. Strong sequence homology between the keratin genes and this gene was also shown for the 5′-untranslated region, while the

remainder of the gene showed no such homology to the keratin genes.

6. The gene described above was positively identified as a fast protein gene when the sequence of the protein for which it coded was compared to the N-terminal twenty amino acids of a fast protein component which, for the first time, was isolated in a pure state from embryonic feather protein using FPLC and HPLC techniques and then sequenced using a gas-liquid solid phase protein sequenator.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge it contains no material that has been previously published by any other person, except where due reference is made in the text.

Signed :

C. Phillip Morris

ACKNOWLEDGEMENTS

I wish to thank Professor W.H. Elliott for permission to undertake the research described in this thesis in the Department of Biochemistry, University of Adelaide.

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

I thank my colleagues in the keratin project for their stimulating discussions and for creating an enthusiastic atmosphere in which to work — Dr. Graham Cam, Dr. John Forrest, Dr. Peter Gibbs, Dr. Keith Gregg, Anna Koltonow, Rob Lock, Dr. Peter Molloy, Phil Pain, Richard Presland and Dr. Lindsay Taylor, in particular I would like to thank Morry Frenkel, Elizabeth Kuczek, Dr. Trevor Lockett, Dr. Barry Powell, Joe Rothnagel, Dr. Rob Saint and Dr. Steve Wilton who gave freely of their support, criticisms and friendship.

I would like to thank Jo Crowe for technical assistance during some of the early DNA sequencing of the cDNA clone pCFK23, presented in Chapter 4, Mark Snoswell for advice and technical assistance with the protein sequencing work, presented in Chapter 6 and Dr. D. Parry for assistance in the interpretation of fast protein primary amino acid sequence data.

I am grateful to the staff of the University of Adelaide Computing Centre where this thesis was printed using the TEX typesetting program. I am particularly indebted to Cathy Hilditch and Sandy Lomman for entering the text and to Andrew Trevorrow and Terry Fanning for their assistance and advice.

I am especially grateful to Lesley Crocker for expert technical assistance, for preparing the figures in this thesis and for her friendship and encouragment.

Finally I am deeply indebted to my wife Debra for help with proof-reading and for her continued support, understanding and tolerance.

I was supported during this work by a Commonwealth Postgraduate Research Award. The work was supported by grants from the Australian Research Grants Scheme and the Australian Wool Corporation (to G.E. Rogers).

CHAPTER 1.

INTRODUCTION.

UNIVERSITY OF ADELAIDE LIBRARY

1.1 PRELIMINARY COMMENTS.

Epidermal structures of higher organisms are characteristically durable, pliable and insoluble and serve to protect the soft body tissues from environmental pressures such as chemical attack, bacterial infection, abrasion, dehydration and heat loss. Such structures include skin, hair, horn, nail and hoof in mammals, and beak, claw, scale and feather in birds and reptiles. These structures generally consist of dead dehydrated tissues which achieve their remarkable chemical and physical properties during a terminal differentiation process which involves the synthesis of large amounts of protein, mainly keratins, a large diverse group of intracellular disulphide-crosslinked proteins which tend to form long filaments. This process continues until the cells fill with protein, dehydrate and die, resulting in the characteristic cornified structures of the epidermis.

The general aim of the work presented in this thesis is to approach an understanding of the processes involved in the terminal differentiation of avian epidermal keratinocytes, to elucidate the molecular mechanisms involved in the control of gene expression and to examine the relationship between gene product and morphological structure, in these tissues. The chicken embryonic feather has proven to be an amenable tissue for this purpose and was adopted for the studies described in this thesis.

This introduction is intended to provide some background information on chick embryonic feather development and cornification at the levels of morphological development, control of gene expression and protein interaction. These processes will be compared and contrasted with those occuring in other avian epidermal tissues and mammalian tissues. It is not intended that this should be an extensive literature survey of the current knowledge of eukaryote gene expression, although some aspects of gene expression must necessarily be considered both here and in other chapters.

1.2 CHICK EMBRYONIC FEATHER DEVELOPMENT.

1.2.1 FEATHER MORPHOGENESIS.

The development of embryonic chick feathers was first mentioned by Coiter in 1573, but it was not until the end of the 19th century that researchers approached an understanding of feather growth (Davies, 1889). Since then, the morphogenesis of embryonic chick feathers has been studied extensively using both light microscopy (for reviews see, Romanoff, 1960; Lillie, 1965; Voitkevich, 1966; Lucas and Stettenheim, 1972), and electron microscopy (Kischer, 1963; Matulionis, 1970).

Skin of embryos entering the fifth day of incubation, consists of a layer of epithelial cells overlain by peridermal cells and underlain by mesenchymal cells. The primordia of feathers appear late in the 5th day of incubation in a process which starts with the condensation of mesenchymal cells into a layer of dermis (Wessells, 1961a). This coincides with the appearance of clusters of epidermal cells, called placodes, which are more elongate than before and represent the first visible signs of feathers. During the next two days of incubation dermal cells congregate beneath these placodes resulting in rudimentary feathers consisting of a thickening of both dermal and epidermal layers. These "feather germs" not only give rise to the natal feathers, which are the subject of this study, but also juvenile and adult down feathers, contour feathers and semiplumes.

During the next few days ($7\frac{1}{2}$ days to $9\frac{1}{2}$ days) rapid cell division results in the elongation of feather germs into conical cylinders of 0.5 to 1.0 mm in length, as shown in Figure 1.1a, composed of three cell layers: an outer epidermally derived layer which is destined to become the feather sheath; an epidermally derived intermediate layer from which the final structures of the feather, such as barbs and barbules, are formed; and a dermally derived inner layer which forms the pulp of the growing feather. Espinasse (1964) described these young feathers as a finger of dermis covered by a thimble of epidermis.

FIGURE 1.1

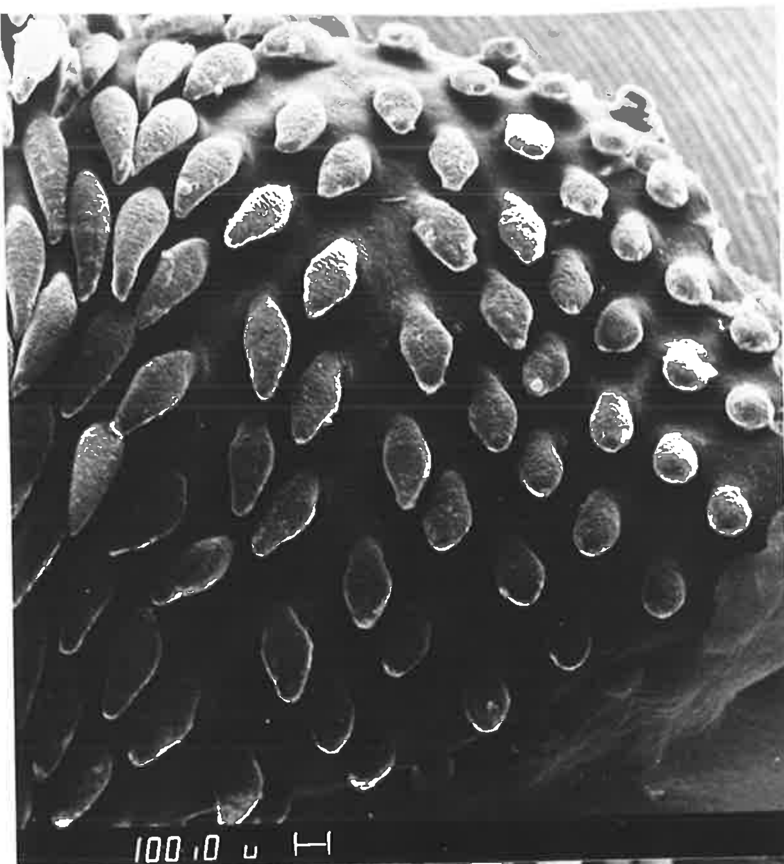
SCANNING ELECTRON MICROGRAPHS OF CHICK EMBRYONIC FEATHERS.

a. 9-day feathers.

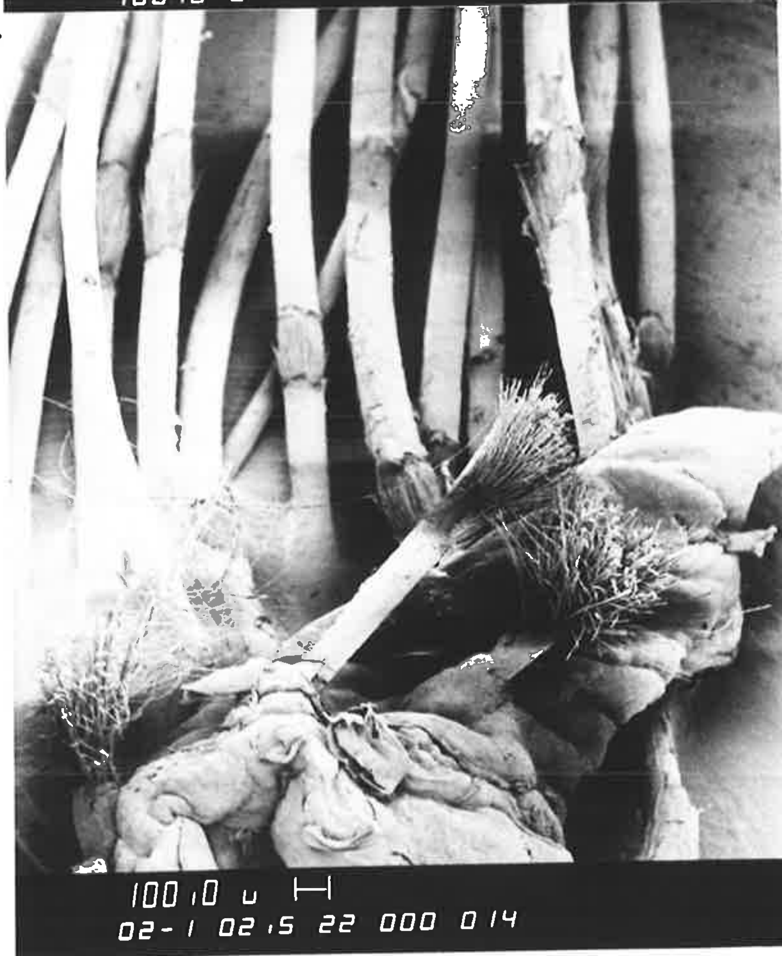
b. 19-day feathers.

(photographs kindly donated by G.E. Rogers)

a.



b.



About the 10th day of incubation, during a period of rapid growth, the cells of the epidermal intermediate layer become arranged into a pair of ridges, called barb ridges, parallel to the long axis of the feather, and on the 11th day the epidermis at the base of the feather begins to push downward into the dermis to form the feather follicle. By the 12th day cells become grouped into barb ridges on either side of the original pair until there are 9 to 11 of these around the circumference (Figure 1.2a and a′). Then all barb ridges lengthen toward the base of the feather (Hamilton, 1952) where cell division accounts for the rapid feather growth and results in an upward movement of differentiating cells and thus produces a feather which is more differentiated at its tip than its base.

Growth after 13 days is attributed to cell growth and movement, rather than further cell division so that cell division has largely ceased by 13 days when the first signs of keratinization are evident in feather tips (Bell and Thathachari, 1963; Kemp *et al.*, 1974a). Over the next few days the downy barbules form as columns of cells in the barb ridges, as shown in Figure 1.2b and b′, and schematically in Figure 1.3, and the process of keratinization accelerates until all structures of the feather are heavily keratinized. During this period the feather develops rami along the length of the barb ridges and in some feathers these can join to form a short rachis or even a typical calamus. Before this is completed the pulp, which has served to supply oxygen and nutrients to the feather, becomes entirely resorbed and a pulp cap is formed over the dermal papilla.

Keratinization and morphological differentiation of embryonic feathers is complete by day 19. All cells have filled with the fibrous keratins and are dead and nuclei and cell boundaries have disappeared in all parts of the barbs (Figure 1.2c and c′). The completed 19 day feather is also shown in Figure 1.1b where, in places, the feather sheath has broken away to reveal the bundles of downy barbules.

Shortly after a chicken hatches (21 days) the sheath of each feather dries and flakes away from the bundle of barbs. The barbs spread out, the barbules are

FIGURE 1.2

CELLULAR ORGANIZATION IN EMBRYONIC CHICK FEATHERS DURING DEVELOPMENT.

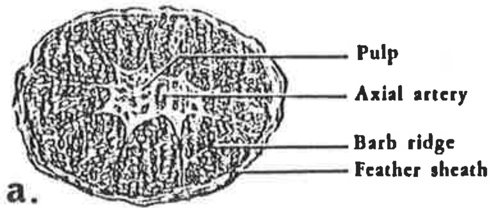
Two drawings are shown at each age, one depicts a section of an entire feather (eg., a.) and the other a section of one or two barbs at higher magnification (eg., a′).

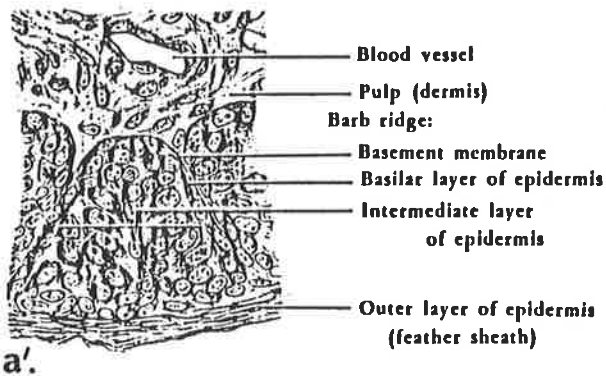
a and a′. 12-day feathers.

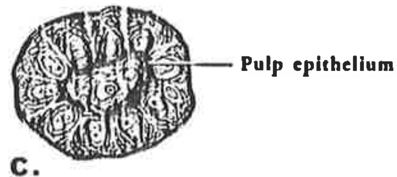
b and b′. 16-day feathers.

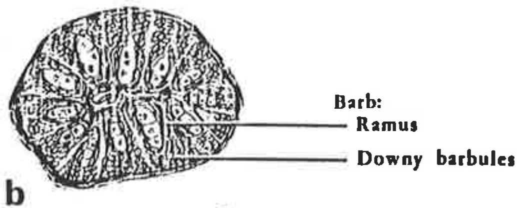
c and c′. 19-day feathers.

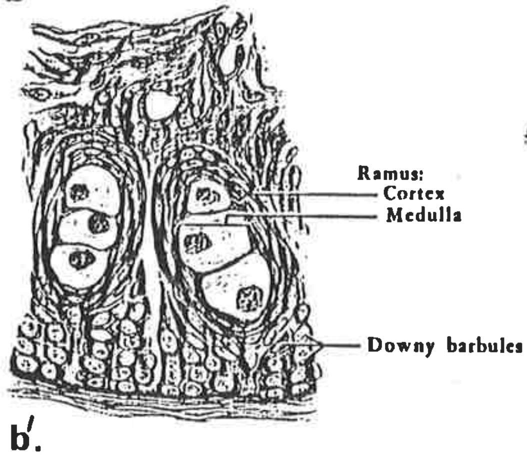
(Modified from Lucas and Stettenheim, 1972).


Pulp
Axial artery
Barb ridge
Feather sheath
a.


Blood vessel
Pulp (dermis)
Barb ridge:
Basement membrane
Basilar layer of epidermis
Intermediate layer
of epidermis
Outer layer of epidermis
(feather sheath)
a'.


Pulp epithelium
c.


Barb:
Ramus
Downy barbules
b


Ramus:
Cortex
Medulla
Downy barbules
b'.

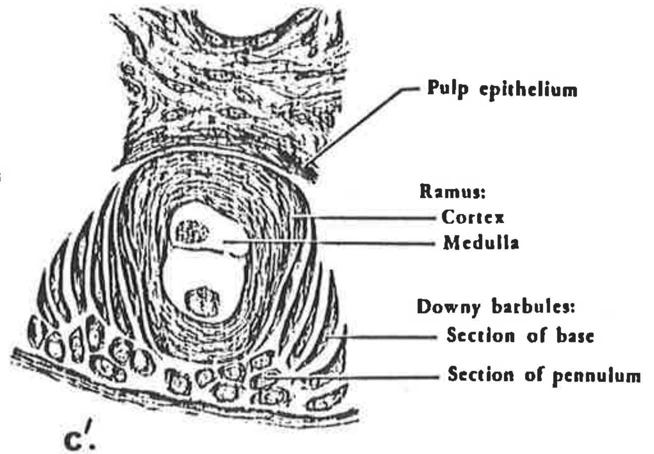

Pulp epithelium
Ramus:
Cortex
Medulla
Downy barbules:
Section of base
Section of pennulum
c'.

FIGURE 1.3

SCHEMATIC DRAWING OF

AN 11-DAY EMBRYONIC CHICK FEATHER.

ax. pl., Barb ridge axial plate.

B'bl., Barbule.

B'bl. pl., Barbule plate.

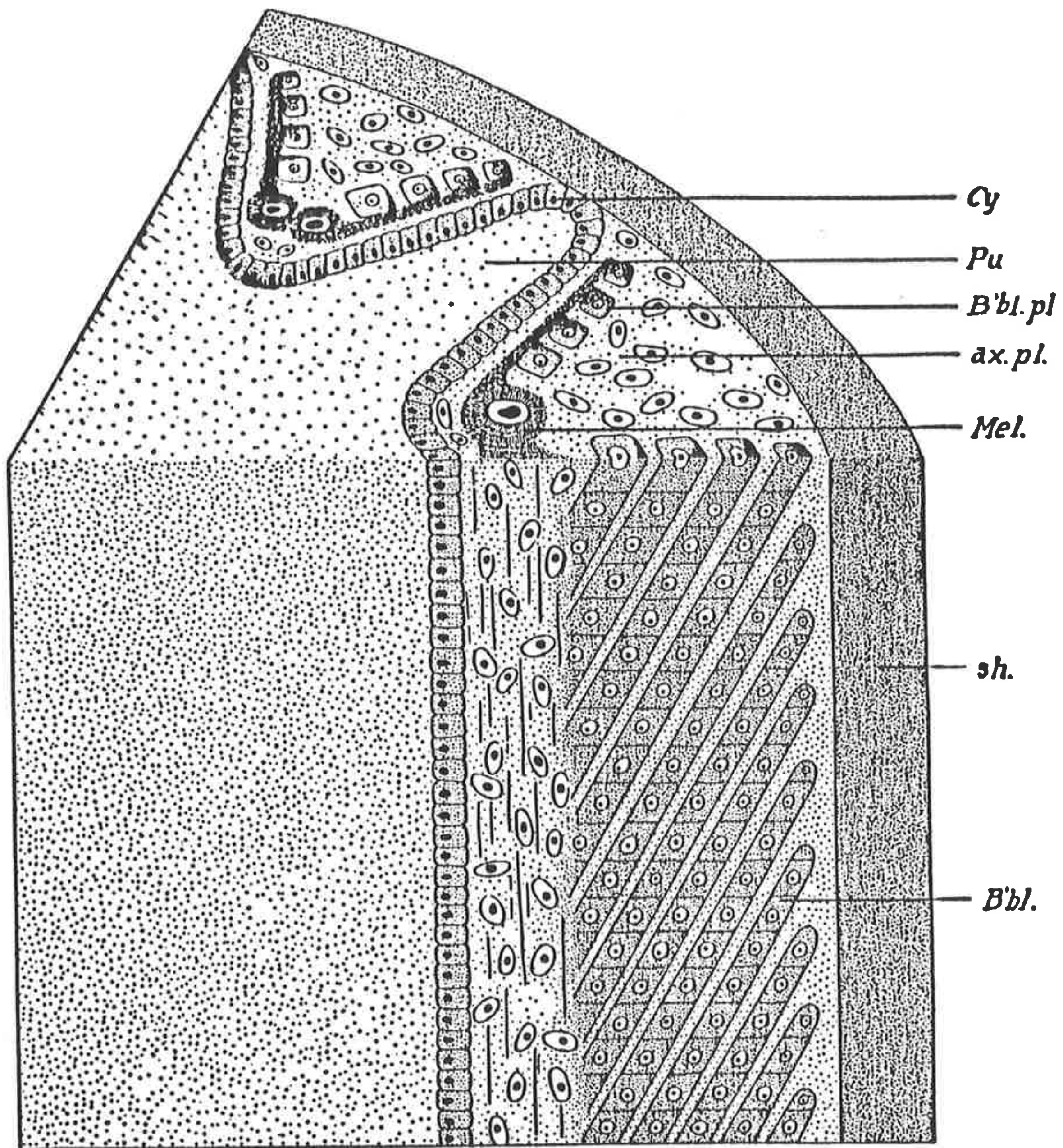
Cy., Cylinder cells.

Mel., Melanophore.

Pu., Pulp.

sh., Feather sheath.

(From Lillie, 1965).



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

unbound from the rami and the natal down feather becomes fluffy.

1.2.2 THE ROLE OF THE DERMIS.

Interactions between embryonic chick dermis and epidermis have been clearly demonstrated using embryonic skin which has been separated into its component dermis and epidermis, then cultured after recombining dermis and epidermis from different developmental stages, different species or different regions of the body (Sengel, 1957, 1958b, 1971, 1983; Rawles, 1963, 1965; Wessells, 1962, 1965). Studies by Sengel (1958a) on heterochronic recombinants (recombinants of dermis and epidermis derived from different developmental stages) revealed that development of feather primordia proceeds in a two step process of bidirectional dermal, epidermal communication. First, the dermis exerts a morphogenetic activity of short duration on the overlying epidermis and induces the formation of feather placodes. Then as the placodes grow out, they induce dermal condensations to form and grow into the elongating feather primordia, constituting their core of vascularized pulp. Although it is the dermis which "signals" the epidermis to form placodes, the secondary dermal condensations have no inductive activity.

Studies using heterotopic recombinants (formed using dermis and epidermis derived from different regions of the body) which are allowed to develop in culture (Sengel *et al.*, 1969; Kollar, 1970; Dhouailly, 1973; Kato, 1969; Sawyer, 1979; Fisher and Sawyer, 1979) revealed that the epidermis develops structures characteristic of those normally associated with the dermis used. Dhouailly *et al.* (1978) demonstrated that the keratins and other proteins expressed in these heterotopic recombinants were identical to those found in normal tissue or homotopic recombinants. Thus the recombination of presumptive scale dermis from the leg of the embryo with presumptive feather epidermis from the back of the embryo causes the epidermis to form scales and to express the set of scale proteins normally found in scale, rather than the set of feather proteins normally expressed in this epidermal tissue. The reverse recombination of presumptive feather dermis and presumptive

scale epidermis results in the formation of normal feathers expressing the normal feather proteins. These effects, exerted by the dermis, only occur at specific stages of development (Rawles, 1963, 1965) and precede the onset of keratinization in those tissues (Matulionis, 1970; Beckingham-Smith, 1973a; Kemp *et al.*, 1974a).

Wessells (1962) and Dodson (1963) observed that cultured epidermis failed to keratinize, and subsequently degenerated, when separated from the dermis, thus preventing dermal signals from stimulating the epidermis. It has been shown that the dermis is essential for the maintenance of epidermal cell proliferation, but it has been observed that non-living artificial substrates of various kinds can substitute for this function of the dermis (Wessells, 1963).

Sengel (1983) concluded that at least two requirements, normally satisfied by the dermis, must be fulfilled for the continuation of basal cell proliferation in epidermis. Firstly, an adequate physical texture of the substrate, whether it be living dermis or a non-living substitute, and an appropriate supply of nutrients. In summary, the dermis appears to control the morphology of epidermal structures by regulating the mitosis of epidermal cells, then determines, by some unknown signal, which genes are to be expressed in the differentiating epidermal tissue. It is not clear whether the dermally derived signal to initiate expression of the correct set of feather specific genes in the epidermis is mediated by a diffusable extracellular substance or by direct pseudopodal 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. However, the epidermis did not develop normally and he did not rule out the possibility of dermal pseudopodal contact with the epidermis through the filter.

Studies using heterospecific dermal-epidermal recombinants (Dhouailly, 1967) showed that duck dermis could direct chick epidermis to produce feathers which resembled those of the duck. Also, Dhouailly *et al.*, (1978) showed that mouse footpad dermis could direct the formation of a footpad-like structure in presumptive

feather-producing chick epidermis. The keratins of the induced footpad structures were of the scale type, i.e. those expected to be expressed in the skin of chicken feet. These results suggest that the dermal signal to the epidermis is recognized even when these tissues are derived from evolutionarily distantly related species since the information directing the specific cellular differentiation in the chick epidermis could only have come from the mouse dermis. It may be that the dermal signal, whether cell to cell contact or a diffusable extra cellular substance, acts as a primary event to define the morphological development of the epidermis into a feather which causes secondary, epidermally derived, signals to initiate the transcription of the correct epidermal gene set. These epidermally derived secondary signals may be a set of tissue or stage specific effector molecules which themselves have been "switched on" by a dermal signal, possibly the same one which initiates morphological development.

1.2.3 OTHER FACTORS AFFECTING FEATHER DIFFERENTIATION.

It has been demonstrated that keratinization is inhibited if immature (Fell and Mellanby, 1953) or highly differentiated (Fell and Mellanby, 1957) embryonic chick skin is cultured in the presence of high levels of Vitamin A. It was shown that keratin synthesis is completely repressed under these conditions (Beckingham-Smith, 1973b) and that for the older tissues at least, the effects are reversible on the removal of Vitamin A (Fitton-Jackson and Fell, 1963).

Voitkevich (1966) showed that several hormones of the pituitary and thyroid glands were implicated in the development of avian keratinocytes, but nothing is known of the molecular mechanisms involved. It is interesting that thyroxine has been shown to accelerate epidermal keratinization (Bartels, 1943; Wessells, 1961b; Kitano and Kuroda, 1967) since the thyroid gland attains maximum thryoxine secretion (Shain *et al.*, 1972) at about the same time as the onset of keratin synthesis in the feather.

A protein specific stimulatory activity has been demonstrated for hydrocortisone which hastens α-keratin synthesis in skin cultures (Fell, 1962; Sugimoto and Endo, 1969) and causes feather primordia, which are destined to synthesize primarily β-keratins, to abort. The presence of hydrocortisone during the culturing of chick scale epidermis results in the stimulation of the scale α-keratins while the β-keratins normally synthesized in scale are not synthesized (Sugimoto *et al.*, 1974).

Epidermal growth factor, a small protein of known sequence (Savage *et al.*, 1972) isolated from submaxillary glands of male mice (Cohen, 1962) and having a human counterpart (Cohen and Carpenter, 1975), was found to enhance the growth and maturation of newborn mouse epidermis when administered *in vivo* and in culture (Cohen, 1965; Starkey *et al.*, 1975). Epidermal growth factor stimulated an increase in RNA and protein synthesis (Hoober and Cohen, 1967) and this was found to be due to an increase in the proportion of ribosomes associated with polysomes (Cohen and Stastny, 1968). This factor has also been shown to be a powerful stimulator of cell division in cultured fibroblasts (Cohen *et al.*, 1975).

Recently, it has been possible to maintain human and mouse epidermal keratinocytes in culture and they express many of the differentiated functions found in native skin, i.e. they form stratified colonies, make cornified envelopes (Rheinwald and Green, 1975a, 1975b; Sun and Green, 1976) and actively synthesize keratins (Sun and Green, 1978; Fuchs and Green, 1978). Rheinwald and Green (1977) found that epidermal growth factor extended the life of cultured keratinocytes and suggested that it was acting to delay the ultimate fate of these cells, i.e. their terminal differentiation.

The complicated and confusing assemblage of findings presented above gives no indication of a simple mechanism controlling epidermal development in general and feather development in particular. The various factors mentioned above may be affecting mitosis in general or specific gene expression or both and may not reflect the situation *in vivo*. Clearly a molecular approach to differentiation and

gene expression in chick embryonic feathers is required for a proper understanding of the role of these various factors.

1.3 THE NATURE OF EMBRYONIC FEATHER PROTEINS.

As mentioned earlier the terminally differentiating embryonic feather goes through a stage of rapid protein synthesis which continues until the feather cells fill with protein, mostly keratins, and eventually die. The recognized variety of proteins expressed during this phase of rapid synthesis has increased with subsequent studies on the molecular events occuring in these cells. Kemp and Rogers (1972) first described the molecular nature of the proteins expressed in the developing embryonic chick feather and were only able to detect a large family of homologous, avian specific keratins. Although these avian specific keratins, termed β-keratins throughout this thesis, are not, strictly crystalographically speaking, of a β type, they contain regions of β-pleated sheet and, for this reason, they are referred to as β-keratins in this laboratory. Later studies (Walker and Rogers, 1976a; Powell, 1979; Powell and Rogers, 1979) analysed the synthesis of another group of proteins, termed "fast proteins" because of their high electrophoretic mobility on pH 2.7 gels, which appeared to comprise about 10% of total feather protein (Walker and Rogers, 1976a). Apart from the feather β-keratins and fast proteins, the results reported in this thesis (Chapter 5) indicate that scale specific β-keratins are also expressed in the developing embryonic chick feather.

1.3.1 FEATHER SPECIFIC β-KERATINS.

Keratin proteins synthesized during the differentiation of the embryonic chick feather are fibrous insoluble intracellular proteins rich in cysteine, serine, glycine and the aliphatic amino acids, with little if any lysine, histidine, tyrosine or methionine (Walker and Rogers, 1976a) and aggregate into fibrillar masses which fill the embryonic feather cells (Fraser *et al.*,1972). These proteins were reported by Walker and Rogers (1976a) to have a molecular weight of about 11,000 Mr while

the amino acid sequences of several of these proteins which were determined from the gene sequences (Molloy *et al.*, 1982; Gregg *et al.*, 1983) revealed molecular weights of about 10,000 Mr. This is close to the molecular weight estimate of 10,500 to 11,500 Mr obtained by Harrap and Woods (1964a) for the adult feather keratins. Protein sequencing data of keratins from adult feather calamus of emu (O'Donnell, 1973b) and silver gull (O'Donnell and Inglis, 1974) and from barbs of adult chick feather (Arai *et al.*, 1983) revealed protein chains of about the same length (100 residues) with homologous amino acid sequences, and molecular weights of about 10,000 Mr. These proteins, particularly the adult chick feather keratin, were remarkably similar to embryonic chick feather keratins in both protein size and sequence, yet on some SDS gel systems (S.Wilton, 1984; personal communication) adult chick feather keratins have a mobility intermediate between that of the embryonic chick feather keratins and the larger proteins of avian scale, beak and claw (14,000 to 14,500 Mr; Harrap and Woods, 1964a; Walker and Bridgen, 1976).

All of these avian β-keratin proteins (ie. scale, beak, claw, adult feather and embryonic feather keratins) share the property *in vivo* of forming oriented filaments about 4 nm in diameter which give a characteristic β-type X-ray diffraction pattern (Fraser *et al.*, 1972; Stewart, 1977). A molecular model of the structure of the feather keratin filament has been proposed (Fraser *et al.*, 1971) from a consideration of the detailed X-ray diffraction pattern data. These authors proposed that the filament consists of twisted β-sheets each of four antiparallel chains, each chain being eight residues long (Figure 1.4). Two of these twisted β-sheets constitute a unit repeating in the axial direction. It is probable that each sheet is formed from the region of hydrophobic amino acids in the middle of the feather keratin molecule, although this has not yet been demonstrated.

Due to the *in vivo* stabilization of these keratin aggregates by disulphide crosslinks (Goddard and Michaelis, 1934) their molecular study requires disruption of the disulphide bonds by reducing agents, followed by stabilization by alkylation,

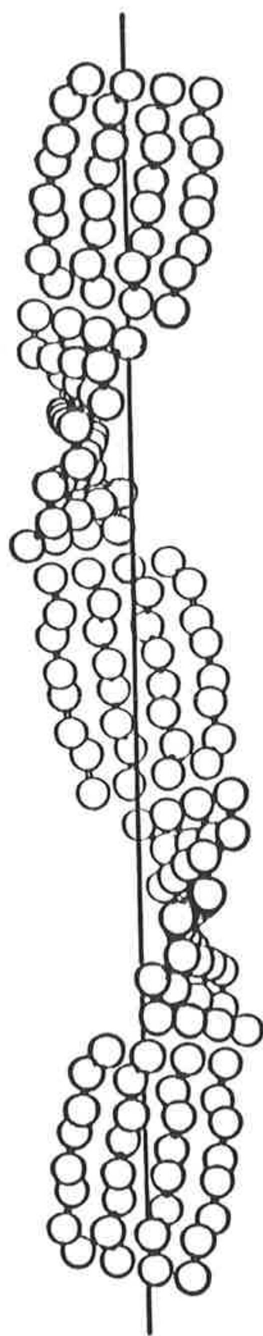
FIGURE 1.4

MOLECULAR STRUCTURE OF
THE CORE FILAMENT OF FEATHER KERATIN.

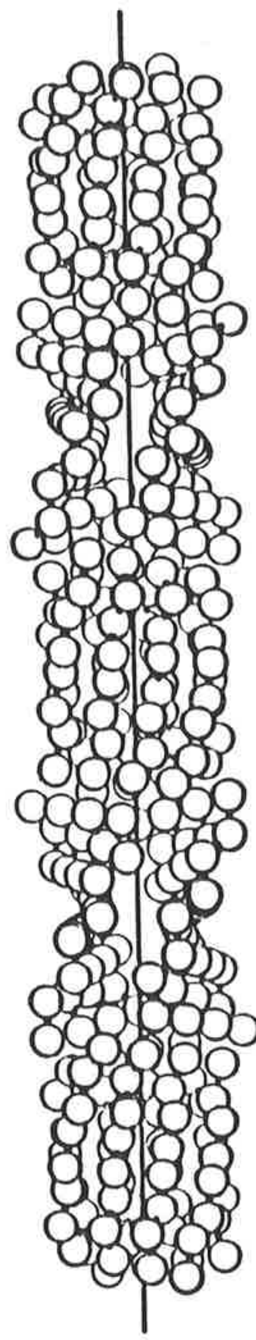
a. A right-handed strand of four twisted β-pleated sheets, each sheet being made up of units of four anti-parallel chains of eight residues.

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

(after Frazer *et al.*, 1971).



a



b

e.g. carboxymethylation (Harrop and Woods, 1964a; Kemp and Rogers, 1972). Feather keratins were originally thought to be essentially a single protein type without the complexity of the mammalian α-keratins of skin and hair (O'Donnell, 1973a; Crewther *et al.*, 1983; Dowling *et al.*, 1983; Fuchs and Hanukoglu, 1983). Although this is true in that all feather keratins appear to be conserved in polypeptide chain length and amino acid sequence it has been demonstrated that the feather keratins are a large family of homologous proteins. Walker and Rogers (1976a) were able to find an estimated twenty two protein chains by separating embryonic feather β-keratins on their charge characteristics by ion exchange and polyacrylamide gel electrophoresis. Direct, partial protein sequence analysis (Walker and Rogers, 1976b) and DNA sequencing of keratin cDNA clones (this thesis) and genes (reviewed by Rogers, 1984) indicates that each protein chain is the product of a separate gene, and that all of these proteins are closely related in primary structure.

Although embryonic feathers and specific parts of adult feather contain a series of closely related protein chains, each tissue has its own distinct population of protein species (Harrap and Woods, 1964a; Kemp and Rogers, 1972). In addition feather proteins display fundamental differences to those of other avian keratinized tissues such as scale, beak and claw (Walker and Bridgen, 1976). However, despite these differences, avian β-keratins from all of these tissues show major regions of sequence homology, indicating that they evolved from a common ancestral protein presumably by gene duplication and divergence leading to the observed group of sequence-related proteins.

1.3.2 FAST PROTEINS.

In addition to the family of embryonic feather keratins another group of proteins were found, termed "fast proteins" because of their high electrophoretic mobility on pH 2.7 gels. They were first detected by Walker and Rogers (1976a) in embryonic but not adult feathers and were shown to be co-ordinately synthesized with (Powell, 1979) but different in composition from the keratins. Until this study, little attention

has been paid to fast proteins at the molecular level, although Powell and Rogers (1979) were able to isolate an RNA fraction from denaturing gels which essentially produced just fast proteins when translated in a cell-free system.

Walker and Rogers (1976a) estimated that fast proteins comprise about 10% of the proteins in 21-day chick feathers, although it is the author's experience that they comprise up to 40% of total feather protein. Analysis of the amino acid composition of fast proteins revealed that they are strikingly different from feather keratins, being rich in histidine, glycine, arginine, tyrosine and phenylalanine but deficient in cysteine, alanine, valine and isoleucine compared to keratins (Walker and Rogers, 1976a). These authors found that fast proteins were a family of at least three related polypeptide chains with a molecular weight of 8,400 Mr. This size estimate was later confirmed by Powell and Rogers (1979) and Brush and Wyld (1982) but results presented in Chapter 5 and 6 of this thesis indicate that this is an erroneous estimate due to a gel artefact, as explained in the Discussion to Chapter 5 of this thesis.

Fast proteins are interesting in that they appear to be expressed only in embryonic tissues, being present at low levels in embryonic scale and even lower levels in embryonic beak and claw while they are undetectable in the corresponding adult tissues (Walker and Rogers, 1976a; S.Wilton, personal communication). This finding is in conflict with a study by Brush (1983) who used what was reported to be fast protein isolated from adult claw for filamentation studies, although he did not specify whether the protein was derived from chick or turkey claw. It is however, difficult to imagine that fast proteins would be expressed in adult turkey tissues when, in the chick, they are limited to embryonic tissues.

Results presented in this thesis show that at least one, and probably all fast proteins, are polypeptides of 119 amino acids with a molecular weight of 13,950 Mr and, like the keratins, they exist as a family of proteins exhibiting strong sequence homology. The role or localization of fast proteins in the feather is unknown but

it is speculated (in Chapter 6) that they serve as an interfilament "glue" much like the proteins found associated with mammalian intermediate filaments (Steinert *et al.*, 1984).

1.3.3 SCALE LIKE β-KERATINS.

Dhouailly *et al.* (1978) observed, with interest, that at least four bands of differing mobilities on SDS gels were present in embryonic chick feather β-keratins and they concluded, contrary to the accepted view (Fraser *et al.*, 1972; Walker and Rogers, 1976a), that the embryonic β-keratins were not all of identical molecular weight (11,000 Mr). These authors observed that, although the major protein component of feathers was of this molecular weight, a minor band was observed which co-migrated with chick scale β-keratins (14,500 Mr; Walker and Bridgen, 1976). This finding was subsequently verified by Powell (1979) and Brush and Wyld (1982). However, none of these authors speculated that the same scale gene(s) were being expressed in both scale and feather tissues.

The results presented in Chapter 5 of this thesis reveal that a scale like protein known to be expressed in scales (Wilton, 1983), is also expressed in embryonic feathers.

The difference in size between the scale and feather keratins can be accounted for by an insoluble tryptic peptide of about 50 amino acids containing a roughly repeating motif of the form gly.gly.X, the presence of which was demonstrated by partial amino acid sequencing of scale proteins (Walker and Bridgen, 1976) and by direct sequencing of scale cDNA clones (Wilton and Rogers, 1981; Wilton *et al.*, 1984) and genes, as reviewed by Rogers (1984). As previously stated, these scale proteins have regions of sequence homology with the feather keratins and, using the Chou and Fasman (1978) and related procedures for structural predictions, it can be demonstrated that the scale sequence adopts the same highly regular conformation of 4 to 5 β-strands of twisted β-sheet structure predicted for the feather keratins

(Fraser *et al.*, 1971). The amino and carboxyl ends probably form less structured tails and the gly.gly.X repeat region is probably excluded from the compact core of the 4 nm filament and should exist as a patch of hydrophobic peptide in a β-turn conformation on the surface of each segment of keratin filament, as described by Gregg *et al.* (1984). This is diagrammatically illustrated in Figure 1.5 and can be compared to the situation predicted for feather keratins (Figure 1.4). As already stated, these β-keratins form, with the β-keratins from feathers and other tissues, a family of sequence related proteins which presumably arose via gene duplication and divergence from a common ancestral gene (Gregg *et al.*, 1983; Rogers, 1984).

1.3.4 OTHER PROTEINS FOUND IN EMBRYONIC FEATHERS.

Analysis of the proteins expressed at very early stages of feather development by SDS electrophoresis (Powell, 1979) reveal that cysteine-containing proteins in the molecular weight range of 40,000 to 70,000 Mr are present as major species. This molecular weight range is similar to that of the mammalian intermediate filament proteins. As the feather differentiates and the lower molecular weight keratins and fast proteins are expressed these larger proteins comprise proportionately less of the proteins in feather, until at 21 days of development, when the chick hatches, they are almost undetectable. It is probable therefore, that these proteins do not form part of the set of embryonic feather proteins which undergo a massive increase in synthesis and characterize the terminal differentiation process occuring in the feather.

These proteins were seen in 19-day feathers, by Dhouailly *et al.* (1978), to roughly co-migrate with the α-keratin proteins of scale and these authors 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. Alternatively, it is possible that these proteins are not keratins and are derived from

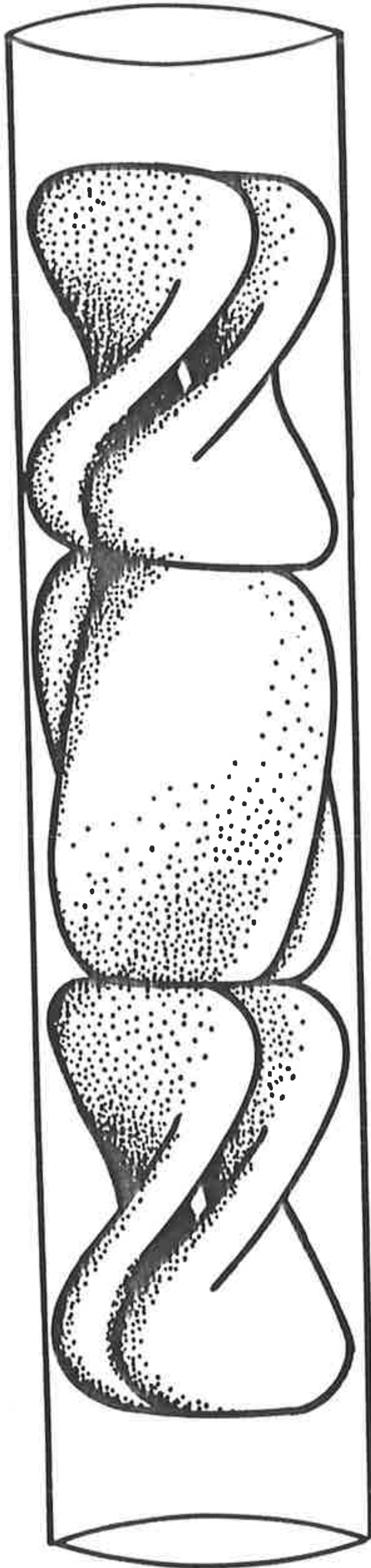
FIGURE 1.5

CARTOON REPRESENTATION OF THE TWISTED β-PLEATED SHEETS CONSTITUTING THE CORE STRUCTURE OF AVIAN KERATINS.

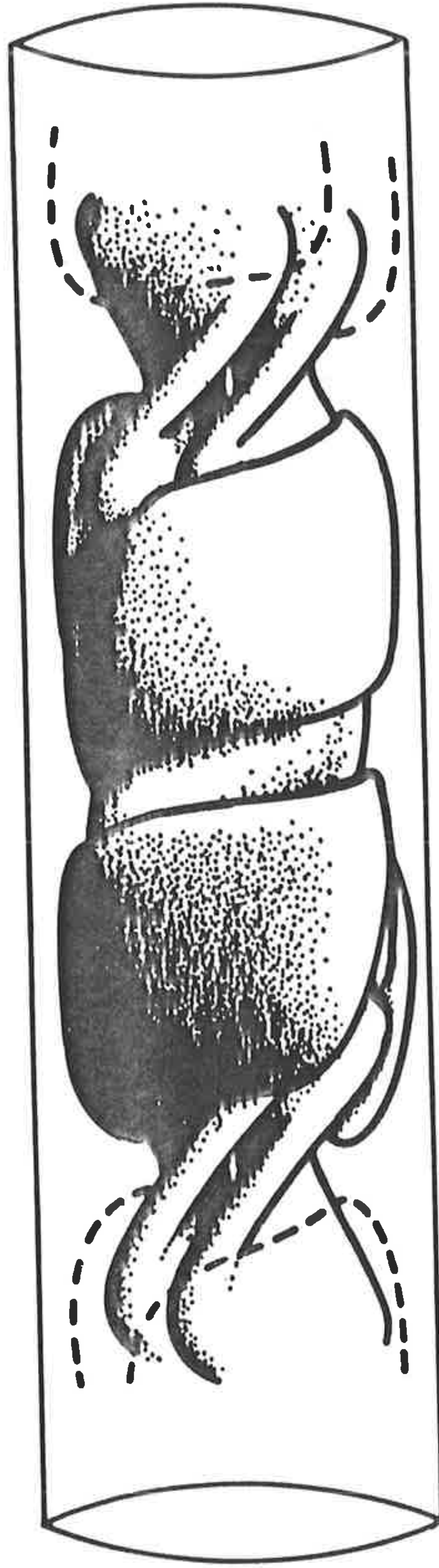
a. Feather keratin.

b. Scale keratin.

(diagram drawn by K. Gregg).



a



b

the mesenchymally-derived pulp of the embryonic feather which is resorbed as the feather matures.

1.3.5 SUMMARY.

In summary there appears to be two classes of structural proteins involved in the differentiation of the embryonic feather, i.e. the β-keratins and fast proteins. The fast proteins seem to be limited to embryonic tissues and are most abundant in embryonic feathers although also present in embryonic scale, beak and claw. The β-keratins of feather can be separated into 4 to 5 bands on SDS gels ranging in apparent molecular weight from 10,000 to 14,500 Mr, although this observation is complicated by the fact that the cysteine content of these proteins affects their mobility when they are S-carboxymethylated, as explained in the Discussion to Chapter 5 of this thesis. It is proposed, in Chapter 7, that these bands represent protein families differing either in molecular weight, cysteine content or both and that in all chick epidermal structures, i.e. embryonic and adult feather, scale, beak and claw, they are all expressed to some degree but the major protein component of these tissues is different for the various tissues.

1.4 GENE EXPRESSION IN THE EMBRYONIC FEATHER.

1.4.1 THE TIME COURSE OF PROTEIN SYNTHESIS.

The β-keratins and fast proteins are found to be co-ordinately expressed, although the control of this expression is not necessarily mediated by the same factor(s). Protein synthesis begins at about the twelfth day of embryonic development, as judged by the appearance of keratin fibrils in the cells (Matulionis, 1970) and of keratin and fast proteins, as detected by polyacrylamide gel electrophoresis (Kemp *et al.*, 1974a; Powell, 1979). Major feather keratin protein bands, which were resolved on pH 9.5 gels (Kemp *et al.*, 1974a) and the fast protein band observed on SDS gels (Powell, 1979), were both present in trace amounts in 11 and 12-day feather extracts but rapidly increased in quantity after day 12 and by 15 days they

comprised over 95% of the total proteins detected in feathers. Kemp *et al.* (1974a) used labelling studies and gel electrophoresis of proteins to show that the different keratin proteins of feathers were synthesized co-ordinately. By the eighteenth day of development all protein synthesis has ceased and the feather is completely cornified.

1.4.2 THE TIME COURSE OF mRNA SYNTHESIS.

A purified 12S fraction of mRNA coding largely for keratin proteins in a cell-free system was isolated and used as a template for the synthesis of complementary DNA (Partington *et al.*, 1973; Kemp *et al.*, 1974b). Powell *et al.* (1976) used labelled cDNA to this purified mRNA fraction to detect the presence of homologous mRNA in feather tissues of varying ages. They demonstrated that 12S mRNA species were synthesized initially from about day 11, increasing to maximal levels at 14 and 15 days and decreasing thereafter. Thus keratin mRNA synthesis occured at about the same time as the synthesis of keratin proteins, demonstrating that the massive onset of keratin synthesis after day 12 is due to an increased rate of synthesis of keratin mRNA and not to differential translation of a pre-formed mRNA pool.

No such study has been performed for the fast protein mRNAs but Powell and Rogers (1979) followed the appearance of fast protein mRNA in embryonic feathers of differing ages by assaying the capacity of mRNA isolated from such tissues to direct the synthesis of fast proteins in a cell-free translation system. They found that translatable fast protein mRNA followed the same temporal course of protein synthesis as did the keratin mRNAs.

1.4.3 THE NATURE OF THE mRNAs EXPRESSED IN EMBRYONIC FEATHERS.

The primary structure of the mRNAs expressed in feather has now been totally described by the isolation and complete DNA sequencing of 7 cDNA clones derived from 14-day embryonic feather mRNA (Chapters 4 and 5, this thesis) and the DNA

sequencing of $4\frac{1}{2}$ β-keratin feather like genes (Molloy *et al.*, 1982; Gregg *et al.*, 1983) and one fast protein gene (Chapter 6, this thesis).

Early work by Partington *et al.* (1973) and Kemp *et al.* (1974b, 1974c), who first identified and partially purified the mRNA coding for embryonic feather β-keratins, revealed that this mRNA species sedimented at 12S on sucrose gradients and had a mobility on denaturing polyacrylamide gels consistent with a length of 760 nucleotides. This mRNA was found to have a poly(A) tract at the 3′-end with an average length of 65 nucleotides (Morris and Rogers, 1979; this thesis, Chapter 3). The mRNA has a "cap" structure at the 5′-end consisting of 7-methylguanosine at the 5′-terminus coupled via a 5′–5′ triphosphate linkage to the penultimate 2′-O-methylated residue (Morris, 1976; Morris and Rogers, 1979).

Hybridization of this mRNA to its cDNA revealed that keratin mRNA is a heterogeneous mixture of species (Kemp, 1975). From this analysis it was estimated that there were 25 to 35 different keratin mRNA species, an estimate consistent with the minimum of 22 protein chains found in the feather (Walker and Rogers, 1976a). Lockett *et al.* (1979) showed that each mRNA species appeared to contain a sequence of about 150 nucleotides at the 3′-end which is unique, i.e. represented only once in the chick genome. A reiterated sequence was found in keratin mRNA attached 5′ to the unique sequence. It was proposed that the unique 3′-ends of keratin mRNA corresponded to the 3′-untranslated region of the mRNA and the repeated portions corresponded to the coding sequences.

A complete picture of the keratin mRNAs was finally obtained from the sequences of several cDNA clones and $4\frac{1}{2}$ genes (this thesis, Chapter 4; Molloy *et al.*, 1982; Gregg *et al.*, 1983). These data reveal that the keratin mRNAs exist as a family of genes, each containing coding regions for 97 amino acids which are strongly conserved. The long (430 to 440 bp) 3′-untranslated regions of these genes were found on average to be less conserved than the coding regions, although two strongly conserved blocks of sequence were observed, one of which was found at

the 3′-end of the mRNA. The mRNA is transcribed as a precursor of about 1200 nucleotides including poly(A) and is processed to the cytoplasmic form of about 860 nucleotides by the removal of the only intron (about 340 nucleotides) which is situated in the 5′-untranslated region of the mRNA. The 5′-untranslated region shows rigid homology between mRNAs but only 5′ to the position of the intron.

Very little work has been done on the fast protein mRNAs since they are difficult to separate from the keratin mRNAs, although Powell and Rogers (1979) did manage to purify a fraction on partially denaturing polyacrylamide gels which produced largely fast proteins when translated in a cell-free system. This study resulted in a broad estimate of 11 to 13S for the size of the fast protein mRNAs.

Again, a more accurate description of fast protein mRNAs had to await the use of recombinant DNA techniques. DNA sequences of three cDNA clones (this thesis, Chapter 5) revealed that the 3′-untranslated regions of these mRNAs are extremely long (590 bp) and only poor homology was observed between the two different 3′-untranslated sequences obtained. The DNA sequence of a fast protein gene (this thesis, Chapter 6) has revealed that the mRNA is transcribed as a 1,750 nucleotide precursor, assuming a poly(A) tract which is the same length as the keratin mRNAs'. This precursor is processed to the 1,080 nucleotide cytoplasmic form by the removal of the only intron of 670 nucleotides which is found in the 5′-untranslated region, in the same position as the keratin intron. Strong sequence homology exists between the conserved region of the keratin mRNA 5′-untranslated region and the corresponding region in this fast protein mRNA.

1.4.4 THE NATURE OF THE GENES EXPRESSED IN EMBRYONIC FEATHERS.

Early work on the genes coding for keratin revealed, by reassociation kinetic analysis, that between 100 and 250 genes were present in the chick genome (Kemp, 1975). Considering the degree of sequence homology between scale and feather

genes (Gregg *et al.*, 1984), which presumably extends to the genes expressed in other tissues, this estimate of the number of keratin genes probably includes the keratin genes expressed in other tissues. Lockett *et al.* (1979) used isopycnic caesium chloride gradient ultracentrifugation to demonstrate that the keratin-coding DNA is associated with long stretches of G-C rich DNA, suggesting that the keratins may be linked as a cluster of keratin genes.

Recent studies on λ-recombinants (Molloy *et al.*, 1982) has confirmed that the keratin genes are clustered and the isolation of cosmid clones containing keratin genes (R. Presland, 1984, unpublished observations) revealed that this cluster may extend beyond 15 genes. It is not yet clear whether these genes are all feather-like β-keratin genes but the $4\frac{1}{2}$ genes in the λ-clone have been sequenced and are all of the feather type (Gregg *et al.*, 1983). These sequences indicate that the genes are arranged in a regular tandem array, each gene separated from the next by about 2.3 Kb of intergenic DNA with all genes transcribed in the same direction. The arrangement of the gene cluster suggests that it evolved through a series of tandem gene duplications. Detailed analysis of the patterns of nucleotide sequences reveal regions of sequence homology crossover which may be "hot spots" for recombination. This can be compared to the models suggested for duplication of globin genes (Efstratiadis *et al.*, 1980).

As described in the previous section, each gene contains an intron in the 5′-untranslated region 37 bp from the cap site. The occurrence of the intron is therefore not related to the separation of functional protein domains brought together during evolution to form a single protein (Gilbert, 1979), as is the case for many genes (eg. the immunoglobulin genes, Sakano *et al.*, 1979a, b; and the ovomucoid gene, Stein *et al.*, 1980). Generally the introns of eukaryotic genes occur in the coding region of the gene (Berget *et al.*, 1977; Chow *et al.*, 1977; Tilghman *et al.*, 1978a, b; Roop *et al.*, 1978; Mandel *et al.*, 1978; Wahli *et al.*, 1980; Schaefer *et al.*, 1980) and not the untranslated regions, although some genes have been found to

contain introns in their 5′-untranslated regions, e.g. conalbumin (Breathnach *et al.*, 1978), insulin (Perler *et al.*, 1980) and histones (R. Sturm; 1984, personal communication). Gilbert (1978) suggested another explanation for the presence of introns in eukaryotic genes might be to provide a means for rapid gene evolution by changes in mRNA processing. This is also unlikely to be the case for the keratins since any change in the 5′-untranslated region of the gene would not affect the gene product, unless of course the 5′-untranslated region contained controlling elements, which in any case would only affect the levels of gene expression and not the gene product itself. Molloy *et al.*, (1982) suggested a likely explanation for the presence of the intron is that the same gene may be expressed in a different tissue (eg. claw or scale) or at a different stage of development (adult feather) using a different promotor and 5′-leader sequence. This is analogous to the expression of mouse α-amylase in liver and salivary gland (Young *et al.*, 1981) and numerous examples of differential splicing of 5′-ends of viral mRNAs (Reddy *et al.*, 1978a, b).

A chick genomic λ-clone containing a fast protein gene has been isolated and sequenced (this thesis, Chapter 6) and the essential structural features of the gene were described above (1.4.3). During the isolation and localization of this gene it was noticed that the 16 Kb insert contained fragments flanking the fast protein gene which weakly hybridized to the fast protein coding specific probe. 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, i.e. it is probable that homologous fast protein genes are clustered in much the same way as the keratin genes are organized.

It is also interesting that the fast protein gene contains an intron in an identical position to the intron found in the keratins, yet this is the only structural or sequence similarity which exists between these gene families.

The scale genes which are expressed in the feather have only been characterized by the DNA sequencing of a cDNA clone, derived from feather mRNA, which did not extend to the 5′-untranslated region of the mRNA (this thesis, Chapter 5). Using

scale cDNA to select clones from a genomic library, several scale genes have been isolated and partially sequenced (Wilton, 1983). One λ-clone contained four scale genes which were spaced roughly 3 Kb apart and were all transcribed in the same direction. Two of these genes which were sequenced were found to contain only one intron, located in the 5′-untranslated region of the genes in approximately the same position as the only intron of the feather keratin and fast protein genes (Wilton, 1983).

1.4.5 SUMMARY.

As a gene family the embryonic feather keratin and fast protein gene systems have a number of features of particular interest when considering an examination of the control of gene expression in a terminally differentiating tissue. They comprise a relatively large number of genes both structurally and evolutionarily related and unrelated (the fast proteins are unlike feather keratins) and are co-ordinately expressed. The conservation through evolution of non-coding sequences, which may be important in the control of gene expression or involved, for example, in protein-RNA or protein-DNA interaction, can be examined in a larger sample of genes than is possible with most gene families. Another advantage of studying the family of embryonic feather genes is that this family is one of a number of related gene families, e.g. the adult feather keratins and scale, beak and claw keratins, all of which are expressed in different tissues and at different times during development.

1.5 THE MAMMALIAN EPIDERMAL TISSUES.

A great deal of interest has been generated recently in the structure and synthesis of epidermal keratins (Roop *et al.*, 1983, 1984; Fuchs and Marchuk, 1983; Steinert and Cantieri, 1983; Steinert and Roop, 1984) and their relationship to the large family of intermediate filament proteins (Fuchs and Hanukoglu, 1983; Kim *et al.*, 1983; Steinert *et al.*, 1984).

The keratin genes of mammalian epidermis are analogous to the feather keratin genes in that they form a large multigene family existing as two related but distinct types of keratins encoded by two multigene families (Fuchs and Marchuk, 1983; Roop *et al.*, 1983). These keratins comprise a group of 10 to 20 different polypeptide species that are differentially expressed in different epithelial tissues (Moll *et al.*, 1982) and form 8nm cytoplasmic filaments in these epithelial cells (Steinert and Idler, 1975; Brysk *et al.*, 1977). These are one of the group of intermediate filaments which compose the cytoskeletal architecture of higher eukaryote cells.

Apart from the keratins in epithelial tissues, intermediate filaments are composed of a family of related proteins the members of which are differentially expressed in different tissues, although the intermediate filaments in any one tissue are usually assembled from only one or two different subunits (Steinert *et al.*, 1984). This family of proteins was originally subdivided according to their tissue origin: neurofilament protein (Liem and Hutchinson, 1982; Zackroff *et al.*, 1982); glial filament protein (Schachner *et al.*, 1977; Rueger *et al.*, 1979); vimentin or decamin (Franke *et al.*, 1978; Steinert *et al.*, 1982); desmin (Lazarides and Hubbard, 1976; Steinert *et al.*, 1981) and the keratins (Woodcock-Mitchell *et al.*, 1982; Roop *et al.*, 1983; Sun *et al.*, 1983). There exists a high degree of homology between the subunits of the intermediate filaments except for the keratins which show only poor homology with the other intermediate filament subunits (Hanukoglu and Fuchs, 1982).

It is now apparent from amino acid sequencing and biophysical studies that the α-keratins of wool and stratum corneum are a sub-set of the cytokeratin intermediate filament proteins (Crewther *et al.*, 1983; Dowling *et al.*, 1983; Fuchs and Hanukoglu, 1983) and share with the intermediate filament proteins a common molecular arrangement. Ward *et al.* (1983) have begun investigations of the genes coding for these α-keratin proteins of wool.

Much as the fast proteins are associated with β-keratins in the embryonic

feather and perhaps for the same reasons, specific proteins are found associated with intermediate filaments. Some of these proteins which have been identified as possible static or regulatory mediators of the interactions of intermediate filaments with themselves or other cytoplasmic systems, seem to have limited distribution and may be tissue specific, e.g. synemin and paranemin are found in muscle and some mesenchymal cells (Lazarides, 1982; Price and Lazarides, 1983), while some other intermediate filament associated proteins are found in several tissues (Green and Goldman, 1983; Wiche *et al.*, 1983). One of the best characterized families of intermediate filament associated proteins is the high sulphur keratins found in hair and wool (Fraser *et al.*, 1972). These proteins form disuphide cross-links with the cysteine-rich non-helical domains of the intermediate filament α-keratins, and result in a rigid matrix of insoluble protein (Crewther, 1976). Three of these high sulphur keratin genes have been isolated and completely sequenced (Powell *et al.*, 1983). Filaggrin which is an intermediate filament associated protein found in cornified epidermal cells is postulated to have a similar role in these cells as the high sulphur keratins, i.e. it serves to link intermediate filaments into a stable matrix in cornified epidermis (Steinert *et al.*, 1981; Steinert, 1984).

It should be made clear that although feather β-keratins do form filaments, they are unrelated to the mammalian α-keratins. The feather β-keratins contain mainly β-sheet structure and associate to form 4nm filaments, while mammalian α-keratins contain large sections of α-helix and associate to form the 8nm intermediate filaments. However, despite being unrelated, both proteins have a similar structural role.

1.6 AIMS OF THE PROJECT.

1.6.1 GENERAL AIMS.

The work presented in this thesis forms part of a larger project with the broad aims of elucidating the processes involved in the terminal differentiation of epidermal

tissues. The project involves work on two model systems, the first and best characterized being the developing embryonic chick feather and the second more recently initiated study being the mammalian hair follicle, in particular sheep wool and guinea pig hair follicles. The work presented in this thesis was directed solely toward a characterization of the first of these model systems.

Two processes occuring in the developing embryonic feather which are characteristic of the processes occuring during its terminal differentiation are the morphological development and differentiation of the epidermis into the feather and the phase of rapid protein synthesis which results in the maturation of the feather. This system is amenable to the study of factors affecting feather morphogenesis as described by Sengel (1983) but a molecular approach to this question awaits the development of the appropriate technology (e.g. cell-cell interaction studies). However, an approach can be made to the characterization of the molecular events involved in the co-ordinate expression of the genes involved during feather differentiation.

Before such questions can be considered, a full understanding must be reached of the genes expressed in the embryonic feather. The advent of recombinant DNA techniques has meant that these feather genes can be quickly identified, isolated and fully characterized at the level of their nucleotide sequences. Once isolated and characterized these genes can be manipulated and used in expression studies to examine controlling elements, both in the gene and in their tissue of origin. It is these approaches which are currently being employed.

1.6.2 SPECIFIC AIMS OF THIS THESIS.

The research presented in this thesis was aimed at a full characterization of the major genes expressed during the terminal differentiation of the embryonic chick feather tissue. In the absence of recombinant DNA facilities during the early phase

of the work, initial approaches to this question involved the use of classical RNA separation and *in vitro* translation techniques. The specific aims were:

1. To characterize the nature of the mRNA species present in 14-day embryonic feathers, in particular the 12S mRNA shown to code for the abundant β-keratins.

2. Having obtained feather cDNA clones, to determine the nucleotide sequences of several clones in order to describe the primary structure of the mRNAs expressed in feathers. Co-incidentally this would provide information on the primary structure of the proteins expressed in this tissue.

3. To relate the mRNA sequences obtained from the cDNA clones to the sequences of the genes from which they were derived.

4. To examine these mRNA and gene sequences in order that common features might indicate their evolutionary relationships and any potential regions involved in the modulation of gene expression during the process of differentiation.

CHAPTER 2

MATERIALS AND GENERAL METHODS.

2.1 MATERIALS.

2.1.1 FINE CHEMICALS FOR SPECIFIC PROCEDURES.

2.1.1a ELECTROPHORESIS.

Acrylamide : *E. Merck, Darmstadt, twice recrystallized from chloroform.*

Agarose, type I : *Sigma Chemical Co., St. Louis, Missouri, U.S.A.*

Agarose, low gelling temperature : *B.D.H. Chemicals Ltd., Poole, Dorset, England.*

Ammonium persulphate : *Sigma.*

Bicine, N,N-bis(2-hydroxyethyl)glycine : *Sigma.*

Bromophenol blue : *B.D.H.*

Coomassie brilliant blue R : *I.C.I. Ltd., England.*

Ethidium bromide : *Sigma.*

Formamide : *B.D.H., deionized as described by Pinder et al., (1974).*

N',N'-methylenebisacrylamide : *B.D.H.*

N,N,N',N'-tetramethylethenediamine (TEMED) : *Eastman Organic Chemicals, Rochester, New York, U.S.A.*

Toluidine blue : *George T. Gurr, London, England.*

Xylene Cyanol FF : *Tokyo Kasei, Tokyo, Japan.*

2.1.1b CELL-FREE TRANSLATION.

Amino acids : *Mann Research Laboratories, N.Y., U.S.A.*

ATP, equine muscle, sodium salt : *Sigma.*

GTP, equine muscle, sodium salt : *Sigma.*

HEPES : *Sigma.*

Phosphocreatine, disodium salt, high grade : *Sigma.*

2.1.1c CELL-FREE TRANSCRIPTION.

Deoxyribonucleoside triphosphates : *Sigma.*

Dideoxyribonucleoside triphosphates : *Sigma.*

Dithiothreitol : *Sigma.*

Oligodeoxythymidylic acid, free acid : *P.L. Biochemicals Inc., Milwaukee, Wisconsin, U.S.A.*

2.1.1d COLUMN CHROMATOGRAPHY.

Acetonitrile, HPLC grade : *Waters Associates, Millford, Massachusetts, U.S.A.*

C18 reverse phase HPLC column (SP-C18, 3 μ) : *Varian, Walnut Creek, California, U.S.A.*

Diphenyl reverse phase HPLC column (Protesil-300, 10 μ) : *Whatman Ltd., England.*

Mono Q strong anion exchange FPLC column : *Pharmacia Fine Chemicals Ltd., Uppsala, Sweden.*

Mono S strong cation exchange FPLC column : *Pharmacia.*

Oligo(dT) cellulose : *P.L. Biochemicals.*

Oligo(dA) cellulose : *P.L. Biochemicals.*

Propanol, HPLC grade : *Waters.*

Sephadex, grades G-50 and G-100 fine : *Pharmacia.*

Trifluoroacetic acid, sequencing grade : *Pearce Chem. Co., Rockford, Illinois, U.S.A.*

2.1.1e FLUOROGRAPHY.

DMSO (dimethylsulphoxide) : *Crown Zellerbach Corp., Camas, Washington, U.S.A.*

PPO (2,5-diphenyloxazole) : *Sigma, initially, and later reclaimed from DMSO by the procedure of Laskey and Mills (1975).*

X-ray film, Fuji Rx : *Fuji Photo Film Co. Ltd., Tokyo, Japan.*

2.1.1f SCINTILLATION COUNTING.

NCS tissue solubilizer : *Amersham/Searle Corp., Arlington Heights, Illinois, U.S.A.*

PCS solubilizer and phase-combining system : *Amersham/Searle.*

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

PPO (2,5-diphenyloxazole) : *Sigma.*

2.1.2 ENZYMES.

The enzymes used in this study were obtained from the sources listed below.

Calf intestinal phosphatase : *Sigma.*

Calf-thymus terminal deoxynucleotidyl transferase : *Boehringer, Mannheim.*

Creatine phosphokinase : *Sigma.*

E. coli Deoxyribonuclease I : *Sigma.*

E. coli DNA-polymerase I (Klenow fragment) : *Boehringer, Mannheim.*

Lysozyme, chicken : *Sigma.*

Polynucleotide kinase : *Boehringer, Mannheim.*

Proteinase K : *Sigma.*

Ribonuclease A : *Sigma.*

Ribonuclease T_1 : *Sigma.*

Ribonuclease T_2 : *Sigma.*

RNA-dependent DNA polymerase (reverse transcriptase) : *a gift from Dr. J.R.E. Wells, prepared from avian myeloblastosis virus, donated by Dr. J.W. Beard and the N.I.H. Cancer Programme.*

S_1 nuclease : *Boehringer, Mannheim.*

T4 DNA-Ligase : *Boehringer, Mannheim.*

All restriction enzymes used during the course of this study were obtained from either New England Biolabs, Beverly, M.A. U.S.A. or Boehringer, Mannheim.

2.1.3 RADIOCHEMICALS.

L-4,5-^{3}H-Leucine, 105 Ci/mmole : *Radiochemical Centre, Amersham.*

L-3-^{3}H-Serine, 28 Ci/mmole : *Radiochemical Centre, Amersham.*

^{3}H-Polyuridylate, 22 mCi/mmole : *Schwarz Mann, Orangeburg, New York, U.S.A.*

^{3}H-Potassium borohydride, 12.4 Ci/mmole : *Radiochemical Centre, Amersham.*

^{3}H-dCTP, 26.2 Ci/mmole : *Radiochemical Centre, Amersham.*

^{3}H-dGTP, 25 Ci/mmole : *Radiochemical Centre, Amersham.*

α-^{32}P-dCTP, 450 Ci/mmole : *Gift of Dr. R.H. Symons.*

α-^{32}P-dGTP, 450 Ci/mmole : *Gift of Dr. R.H. Symons.*

α-^{32}P-dATP, 2400 Ci/mmole : *BRESA, Adelaide, Australia.*

α-^{32}P-dCTP, 2400 Ci/mmole : *BRESA.*

γ-^{32}P-ATP, 1000 Ci/mmole : *Radiochemical Centre, Amersham.*

γ-^{32}P-ATP, 2400 Ci/mmole : *BRESA.*

2.1.4 BUFFERS.

Buffers commonly used in this study are listed below.

MSB : *200 mM KCl, 5.3 mM $MgCl_2$, 10 mM Tirs-HCl pH 7.4.*

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

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

SET : *30 mM Tris-HCl pH 7.8, 5 mM EDTA, 0.5% SDS.*

SSCE : *150 mM NaCl, 15 mM sodium citrate, 1 mM EDTA.*

TAE : *40 mM Tris-Acetate pH 8.2, 20 mM sodium 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 : *10 mM Tris-HCl pH 7.5, 0.1 mM EDTA.*

2.1.5 BACTERIAL MEDIA.

2.1.5a LIQUID MEDIA.

All liquid media were prepared in glass-distilled water and were sterilized by

autoclaving for 25 minutes at a temperature of 120°C and a pressure of 15 lb/inch^2.

The compositions of the various media were as follows:

L-Broth : *1% Bacto-Tryptone, 0.5% yeast extract, 1% NaCl.*

2X YT : *1.6% Amine A, 1% yeast extract, 0.5% NaCl.*

Minimal medium : *1.05% K_2HPO_4, 0.45% KH_2PO_4, 0.1% $(NH_4)_2SO_4$, 0.05% sodium citrate.*

NZCYM : *1% Amine A, 0.5% yeast extract, 0.5% NaCl, 0.1% Casaminoacids, 0.246% $MgSO_4$, 0.2% maltose.*

2.1.5b SOLID MEDIA.

L-soft agar : *L-broth containing 0.7% Bacto-agar.*

All plates were prepared from 30 ml of the relevant mixture, dried (with the lids on) overnight at 37°C and stored at 4°C until required.

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

L plus Tet. : *L-Broth, 1% Bacto-agar, tetracycline (20 μg/ml).*

L plus Amp. : *L-Broth, 1% Bacto-agar, ampicillin (50 μg/ml).*

Minimal medium plus glucose : *Minimal medium, 1% Bacto-agar, 0.02% $MgSO_4$, 0.2% glucose, 0.0005% Thiamine-HCl.*

NZCYM : *1% Amine A, 0.5% yeast extract, 0.5% NaCl, 0.1% Casaminoacids, 0.246% $MgSO_4$, 1% Bacto-agar, 0.2% maltose.*

2.1.6 ANTIBIOTICS.

Ampicillin : *Sigma.*

Chloramphenicol : *Sigma.*

Tetracycline : *Sigma.*

2.1.7 BACTERIAL STRAINS.

The *E.coli* K12 strains used in this study are described below.

ED8654 : r_{κ}^{-}, m_{κ}^{+}, *trp BE∇, met, trp R, rec* A_{56}, *sup E. (Murray et al., 1977).*

JM101 : *lac, pro. sup E, thi. F′ tra D36, pro AB, lac* I^{q} *Z ΔM15. (Messing, 1979).*

LE392 : F^{-}, *hsd R514* ($r_{\kappa}4^{-}$, m_{κ}^{+}), *sup E44, sup F58, lac* Y_1 *or Δ(lac IZY)*$_6$, *gal K2, gal T22, met Bl, trp R55,* λ^{-}. *(Murray et al., 1977).*

Mc1o61 : *ara D 139, Δ(ara, leu)7697, ΔlacX74, gal* U^{-}, *gal* K^{-}, *hsr*$^{-}$, *hsm*$^{+}$, *str A. (Casadaban and Cohen, 1980).*

2.1.8 BACTERIOPHAGE STRAINS.

The bacteriophage strains used in this study are described below

λ-Charon 4A : *Aam 32, Bam 1, lac 5, bio 256, ∇KH54, ∇NIN5, ϕ80 QSR. (Williams and Blattner, 1979).*

M13mp8 : *Messing and Vieira (1982).*

M13mp9 : *Messing and Vieira (1982).*

2.1.9 TISSUE.

Fertilized eggs of White Leghorn fowls (Gallus domesticus), strain Para 3 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). Feathers designated, for example, as "14-day feathers" were from embryos incubated for a total of 14 days.

Embryos were removed from the eggs and washed with Hanks balanced salt solution. They were then immersed in the same solution, and the body feathers removed by plucking with jewellers' forceps into ice-cold MSB (200 mM KCl, 5.3 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.4). The feathers were then washed in ice-cold MSB by repeated low speed centrifugation performed at 4°C.

2.1.10 MISCELLANEOUS MATERIALS.

Cellophane dialysis tubing : *B.D.H. The tubing was boiled in 1% w/v $NaHCO_3$ before use (Thompson and O'Donnell, 1965).*

Cellulose nitrate centrifuge tubes : *Beckman, Palo Alto, California, U.S.A.*

GF/A glass filter discs : *Whatman Ltd., England.*

Nitrocellulose filter paper : *Sartorius.*

X-ray film : *Fuji Rx X-ray film. Fuji Photo Film Co. Ltd.*

2.1.11 MISCELLANEOUS FINE CHEMICALS.

Caesium chloride, optical grade : *Harshaw Chemical Co., Cleveland, Ohio, U.S.A.*

Diethylpyrocarbonate : *Sigma.*

Phenol : *B.D.H., redistilled under nitrogen and reduced pressure, stored at −20° C under nitrogen prior to use.*

Sarkosyl (sodium dodecyl sarcosinate) : *Ciba-Geigy, Basle, Switzerland.*

SDS (sodium dodecyl sulphate) : *B.D.H., 99% pure.*

Sodium Azide : *B.D.H. Ltd.*

Sucrose, Ultrapure, RNase free : *Schwarz-Mann, Orangeburg, N.Y., U.S.A.*

All other chemicals used were of Analytical Reagent grade, or of the highest available purity.

2.2 GENERAL METHODS.

2.2.1 PREPARATION OF GLASSWARE AND SOLUTIONS.

All solutions were prepared using glass-distilled or double-deionized water and rendered nuclease-free and microbiologically sterile either by treatment with diethylpyrocarbonate or autoclaving at 20 psi for 25 min at 140°C. Glassware was rendered nuclease-free by either autoclaving, incubation at 110°C overnight, washing with 1M KOH and rinsing with sterile glass-distilled water, or by a combination of these procedures. Spatulas, etc., were washed with alkali and rinsed in sterile

water, as described. Pipettes and micropipettes were washed in glass-distilled water containing diethylpyrocarbonate and dried for 16 hours at 110°C, or autoclaved after treatment with diethylpyrocarbonate and then dried. Where possible centrifuge tubes were autoclaved, otherwise they were alkali-washed, as described.

2.2.2 PREPARATION OF FEATHER LYSATE.

Feathers were plucked and washed as described in 2.1.9 and, after the final wash, were pelleted by centrifugation at 3,000 r.p.m. in a bench MSE centrifuge at 4°C. All subsequent procedures were carried out at 4°C. The pelleted feathers were resuspended in two volumes of MSB containing 10 mM dithiothreitol and allowed to stand on ice for 10 min, prior to homogenization by about ten strokes of a 10 ml teflon-glass homogenizer. Cell debris and nuclei were pelleted by centrifugation at 12,000 r.p.m. for 10 min at 4°C using a Sorvall HB-4 swingout rotor. The supernatant, termed the lysate, was carefully withdrawn by pasteur pipette ensuring that the pellet remained undisturbed.

2.2.3 PREPARATION OF FEATHER RNA.

In this laboratory feather RNA was routinely prepared from feather lysate by several alternative procedures, as outlined below:

2.2.3a PREPARATION OF FEATHER POLYSOMES.

Feather polysomes were prepared essentially as described by Partington *et al.* (1973).

The polysomes were pelleted from the feather lysate (see 2.2.2 above) by centrifugation in a Beckman Ti 50 rotor at 45,000 r.p.m. for 90 min at 4°C. Each tube contained 1 ml of 40% w/v sucrose in TK buffer (15 mM KCl, 10 mM Tris-HCl, pH 7.4), layered under the lysate, and during centrifugation the sucrose formed a pad which minimized protein contamination of the polysome pellet. The supernatant was decanted and the pellet rinsed with TK buffer.

2.2.3b PHENOL EXTRACTION OF POLYSOMES.

Polysome pellets, prepared as described above (see 2.2.3a), were resuspended in 2.7 ml of 0.1 M Tris-HCl, pH 9, after which 0.3 ml of 1 M Tris-HCl, pH 9, 1% SDS was added. Then an equal volume of water-saturated phenol was added, mixed thoroughly and allowed to stand for 5 min. The mixture was centrifuged at 3,000 r.p.m. for 10 min in a bench MSE centrifuge, the aqueous phase was carefully removed and re-extracted with an equal volume of water-saturated phenol and the phenol phase was re-extracted with 0.5 volumes of buffer containing 0.1 M Tris-HCl, pH 9, 1% SDS (Brawerman, 1976). The aqueous phases were carefully removed, pooled, and extracted with water-saturated phenol until the interface remained clear. Usually only one more extraction was required.

2.2.3c EDTA DISSOCIATION OF POLYSOMES.

Polysomes, prepared as described above (see 2.2.3a), were dissociated as described by Kemp *et al.* (1974b) and the EDTA-treated polysomes were centrifuged on 10 to 40% w/v linear sucrose gradients, prepared in 0.015 M KCl, 0.01 M Tris-HCl, pH 7.4, in a Beckman SW41 rotor for 16 hours at 4°C. The gradients were fractionated using an ISCO model 640 gradient fractionator and the mRNP particles were collected, ethanol precipitated, and the RNA which was liberated from the protein moiety by resuspension in an SDS-containing buffer (Kemp *et al.*, 1974b), was centrifuged on 10 to 40% w/v sucrose gradients.

2.2.3d OLIGO(dT) CELLULOSE CHROMATOGRAPHY OF RNA.

Poly(A) containing RNA was isolated in enriched form by selecting the fraction of SDS dissociated polysomes retained on an oligo(dT) cellulose chromatography column at high salt (500 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5% SDS), as described by Krystosek *et al.*, (1975), and eluting at low salt (10 mM Tris-HCl pH 7.5, 1mM EDTA). Alternatively polysome pellets were phenol-chloroform extracted prior to chromatography. For phenol extraction polysomes were dissolved

in 100 mM NaCl, 100 mM Tris-HCl pH 9.0, 1% sarkosyl, and extracted three times with phenol-chloroform (1:1), under conditions which favour retention of poly(A) containing RNA in the aqueous phase (Georgiev and Martieva, 1962; Brawerman *et al.*, 1972; Perry *et al.*, 1972; Brawerman, 1973, 1974, 1976). The aqueous phase was precipitated with ethanol and dissolved in 500 mM NaCl, 0.5% SDS, 10 mM Tris-HCl pH 7.5, 1 mM EDTA for subjection to oligo(dT) cellulose chromatography, performed as described above.

2.2.4 PREPARATION OF PURIFIED KERATIN mRNA.

RNA was prepared from the RNP particles which were produced by EDTA dissociation of polysomes, as described by Kemp *et al.* (1974b), and as outlined in 2.2.3c.

The RNP fraction, collected off sucrose gradients (e.g., see Figure 3.2), was ethanol precipitated, dried and resuspended in 0.1 M Tris-HCl, pH 9, 0.1% SDS. The sample was then heated at 65°C for 10 min to reduce RNA aggregation, quick-chilled on ice and centrifuged on 11.2 ml 10 to 40% w/v sucrose gradients, (prepared in 0.1 M NaCl, 0.01 M Tris-HCl, pH 7.4, 0.001 M EDTA), at 38,000 r.p.m. in a Beckman SW41 rotor for 16 hours at 4°C. The gradients were fractionated using an ISCO model 640 gradient fractionator and the 12S RNA peak, containing keratin mRNA (Kemp *et al.*, 1974c), was purified by another cycle of centrifugation.

2.2.5 CELL-FREE TRANSLATION OF RNA.

The cell-free wheat germ system used during these experiments was prepared from commercial wheat germ (Adelaide Milling Co., Port Adelaide, South Australia) using the method of Marcu and Dudock (1974). All the S-30 wheat germ extracts were dispensed into sterile glass vials, frozen in liquid nitrogen and stored at −80°C.

Cell-free translations were initially carried out using the conditions described by Roberts and Paterson (1973). The 50 μl reaction mix contained 20 mM HEPES-KOH pH 7.5, 70 mM KCl, 3 mM magnesium acetate, 2 mM DTT, 1 mM ATP, 20 μM

GTP, 8 mM creatine phosphate, 4 μg/ml creatine phosphokinase, 5 μCi ^{3}H-Leucine (S.A. 105 Ci/mmole) and/or ^{3}H-Serine (S.A. 28 Ci/mmole), 25 μM of the other unlabelled amino acids and 15 μl of the S-30 wheat germ extract. Generally 1.0 to 2.0 μg of RNA was added to each 50 μl reaction mix and the incubations were carried out at 28°C for 90 minutes. The incorporation of ^{3}H-Leucine into protein during cell-free translation was determined by assaying duplicate 2 μl aliquots of each sample on Whatman GF/A filters using the procedure of Bollum (1968). After drying, toluene scintillant was added to the filters and the radioactivity was determined in a Packard liquid scintillation spectrometer.

2.2.6 S-CARBOXYMETHYLATION OF TRANSLATION PRODUCTS.

To permit a comparison of the newly synthesized translation products with SCM-feather β-keratins on polyacrylamide gel systems, it was necessary to reduce and alkylate the translation products. In order to minimize any losses of the translation products during this procedure, the following method was employed. 16 μl of 2 mM EDTA, 10% w/v Ficoll, 2% SDS, 0.01 M Tris-Bicine was added to the 46 μl translation reaction and mixed thoroughly. The reduction process began with the addition of 5 μl of 1 M DTT, incubated for 15 minutes at 65°C and then allowed to cool to room temperature. 6 μl of 30% iodoacetic acid (pH 8.0) was added and the alkylation allowed to proceed for 30 minutes at 37°C. Bromophenol blue was then added as a tracker dye and this mixture could either be loaded directly onto a polyacrylamide gel or stored frozen until required. The radiolabelled protein bands were detected in the gel using the fluorographic procedure of Bonner and Laskey (1974), as modified by Laskey and Mills (1975), or when the products were separated on tube gels, radioactivity in gel slices (1 mm) was determined after solubilizing each slice with 0.2 ml NCS tissue solubilizer plus 0.025 ml 8 M NH_4OH. The slices were either left over-night at room temperature or heated at 55°C for 2 hours. Then 2 mls of 15% (v/v) PCS toluene scintillation fluid was added to

each slice, mixed thoroughly, and the samples were counted in a Packard liquid scintillation spectrometer.

2.2.7 GEL ELECTROPHORESIS OF PROTEINS AT pH 2.7.

15% polyacrylamide tube gels (0.5 cm internal diameter x 10 cm long) with a running pH of 2.7 were prepared in 2.5 M urea as described by Panyim and Chalkley (1969), except that the cross-linking agent was ethylene diacrylate (Cain and Pitney, 1968) instead of bisacrylamide. Protein bands were stained with Coomassie Brilliant Blue R in 10% TCA (Chrambach *et al.*, 1967), and either scanned with a Gilford Gel Scanner coupled to a W & W Recorder or photographed.

2.2.8 SDS GEL ELECTROPHORESIS OF PROTEIN.

This was performed in 'Tris-bicine-SDS' slab gels using two minor modifications of the procedure of Weeds (1976) in that 0.1% SDS was included in the reservoir buffer and the samples were heated at 60°C for 10 min prior to loading.

The samples were loaded into the wells of 12.5% acrylamide, 0.15% bisacrylamide SDS-slab gels and electrophoresed for about 6 hours at 35 mA. Gels were either stained and destained by the method of Swank and Munkres (1971), or fixed in 10% acetic acid for 60 min.

2.2.9 ETHANOL PRECIPITATION.

In all cases samples were made 300 mM with respect to sodium acetate using a 3M 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 10 minutes in an ethanol-dry ice bath or left for at least 2 hours at –20°C. Precipitates were collected by centrifugation at 12,000 rpm for 5 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 ethanol, dried briefly under vacuum and resuspended in the appropriate solution.

2.2.10 PREPARATION OF PLASMID DNA.

Recombinant plasmid DNA was prepared by a modified procedure of Birnboim and Doly (1979). Recombinant clones were grown overnight in L-broth supplemented with tetracycline (20 μg/ml), diluted 1:100 into 500 ml of fresh L-broth and grown to stationary phase. The cells were pelleted by centrifugation and resuspended in 4 ml of ice-cold 25 mM Tris-HCl pH 8.0, 10 mM EDTA and 15% sucrose. After lysozyme was added to a final concentration of 3 mg/ml, the solution was kept on ice for 40 minutes, 8 ml of 200 mM NaOH, 1.0% SDS was added and the mixture was left on ice for another 10 minutes. 5 ml of 3.0 M Na-acetate pH 6.0 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 rpm for 45 minutes at 4°C. The supernatant was decanted with care taken to avoid disturbing the relatively soft pellet and treated with 50 μg of RN'ase A (heat denatured at 80°C for 20 minutes to inactive any contaminating DN'ases) for 2 hours at 45°C to degrade the contaminating RNA. The solution was treated with 5 mg of Proteinase K for 2 hours at 45°C, extracted once with an equal volume of SET-buffered phenol and chloroform (1:1) and the nucleic acids were precipitated by the addition of 2 volumes of ethanol, chilling at –20°C for 2 hours and centrifugation at 12,000 rpm for 20 minutes at 4°C. The pellet was resuspended in 1.6 ml of H_2O, to which was added 400 μl 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 at 10,000 rpm for 10 minutes at 4°C, washed with ice-cold 70% ethanol, briefly dried *in vacuo* and resuspended in 500 μl of H_2O.

Alternatively a gentle lysis caesium chloride gradient procedure was employed as follows. Recombinant plasmid clones were inoculated into media as described above and grown to an A_{600} of 0.7 to 0.8. Chloramphenicol was added to a final concentration of 50 μg/ml and the culture was left incubating with aeration at 37°C overnight. The cells were chloroform treated to kill the bacteria, centrifuged at

10,000 r.p.m. for 10 min, and washed in 40 ml of TE. The cells were centrifuged and resuspended in 12 mls of 25% sucrose in 0.05 M Tris-HCl pH 8.0 at 4°C. 4 ml of fresh 10 mg/ml lysozyme was added and the mixture swirled gently for 5 min. 4 ml of 0.5 M EDTA pH 9.0 was then added and the mixture swirled gently for a further 5 min. 20 ml of a solution of 0.1% Triton X-100, 0.0625 M EDTA and 0.05 M Tris-HCl pH 8.0 was then added and the mixture swirled gently for 10 min before being centrifuged at 40,000 r.p.m. for 30 min in a Beckman Ti50 rotor and L5-50 centrifuge. DNA in the supernatant was precipitated at 4°C overnight by the addition of solid NaCl to 0.5 M and solid polyethyleneglycol to 10% (w/v). The DNA was pelleted by centrifugation at 12,000 g for 10 min and the pellet resuspended in TE. One gm of solid caesium chloride and 0.1 ml of a 10 mg/ml solution of ethidium bromide were added per ml of TE and the solution centrifuged at 40,000 r.p.m. for 60 hours. The lower (plasmid) band was removed by side puncture with a syringe, extracted with iso-amyl alcohol until the aqueous phase was free of ethidium bromide and finally dialyzed extensively against TE. All DNA solutions were stored in TE at 4°C.

2.2.11 RESTRICTION OF DNA.

Restriction endonuclease digestion of DNA was carried out using the conditions appropriate for each enzyme detailed in the New England Biolabs enzyme catalogue. Generally, one unit of enzyme was added for each microgram of DNA to be digested and the reaction mix was incubated for at least a two fold excess of time to ensure complete digestion.

2.2.12 AGAROSE GEL ELECTROPHORESIS OF DNA.

Electrophoresis of DNA for analytical purposes was usually carried out on small submarine gels. The electrophoretic fractionation of DNA for transfer to nitrocellulose was carried out on 14 x 14 x 0.3 cm vertical slab gels of 1.0 to 1.5% agarose in TAE buffer. Electrophoresis was carried out until the bromophenol blue

tracker dye had migrated an appropriate distance to ensure that adequate separation of the DNA fragments had taken place. DNA was visualized under U.V. light after the gel had been briefly stained in 0.05% ethidium bromide.

2.2.13 POLYACRYLAMIDE GEL ELECTROPHORESIS OF DNA.

All polyacrylamide solutions (acrylamide to bisacrylamide ratio of 20:1) were deionized using Amberlite MB-1 Mixed Bed Resin before use, as this was found to remove anions and cations which seriously inhibited many enzymes and adversely affected the chemical cleavage sequencing reactions. The TBE pH 8.3 buffer (Peacock and Dingman, 1968) was used routinely in analytical and preparative polyacrylamide gels as well as the sequencing gels on which the chemical cleavage sequencing reactions were fractionated. The TBE pH 8.8 buffer was used for the fractionation of the products of the dideoxy sequencing reactions.

2.2.14 OLIGO(dT) PRIMED REVERSE TRANSCRIPTION.

Oligo(dT) primed reverse transcription was carried out in a 25 μl reaction mix containing 1 to 2 μg RNA, 660 μM of dATP and dTTP, 6 μM α-^{32}P-dGTP and α-^{32}P-dCTP (S.A. 450 Ci/mmole) or 6 μM α-^{32}P-dATP and α-^{32}P-dCTP (S.A. 2400 Ci/mmole), 60 μM of the dNTPs not included as labelled nucleotides, and 8 mM DTT, 10 mM $MgCl_2$, 50 mM Tris-HCl pH 8.3, oligo(dT) (final concentration 10 μg/ml) and Actinomycin D (100 μg/ml final concentration). 16 units of AMV reverse transcriptase (0.5 μl) was added and the reaction mix was incubated at 42°C for 20 minutes. The reaction was terminated by the addition of an equal volume of 1.0% SDS. Following a 2 minute incubation at 37°C to disrupt the protein-nucleic acid complexes, the RNA template was removed by alkaline hydrolysis with 300 mM NaOH for at least 35 minutes at 37°C. The solution was neutralized by the addition of HCl to 300 mM and Tris-HCl pH 7.5 to 100 mM. The mixture was loaded onto a 0.6 cm x 20 cm Sephadex G.50 (Fine) column and eluted with water to remove unincorporated nucleotides.

2.2.15 DETECTION OF RECOMBINANT PLASMIDS.

Detection of pBR322 plasmids carrying DNA complementary to various DNA probes was carried out using a modification of the procedure of Grunstein and Hogness (1975). Colonies previously found to be Tet^{res} and Amp^{sens} were transferred by toothpick to a masterplate and to a thrice-boiled sterile sheet of nitrocellulose overlaid on a L plus tet agar plate. These plates were incubated overnight. The masterplates were stored at 4°C for up to 6 weeks before it was necessary to prepare a fresh masterplate. DNA from the colonies grown on the nitrocellulose filters was immobilized onto this support by the sequential transfer of the nitrocellulose onto 3 MM paper saturated with 0.5 M NaOH for 7 minutes, 1 M Tris-HCl pH 7.5 for 2 minutes, 1 M Tris-HCl pH 7.5 for 2 minutes, 1.5 M NaCl, 0.5 M Tris-HCl pH 7.5 for 4 minutes. The nitrocellulose filter was then washed with ethanol, air dried and baked at 80°C *in vacuo* for 2 hours.

2.2.16 TRANSFER OF DNA TO NITROCELLULOSE.

Restricted DNA, fractionated on either polyacrylamide or 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 DNA with 0.25 N HCl. In Chapter 6, a bi-directional transfer procedure (Smith and Summers, 1980) using 1 M ammonium acetate 0.02 M NaOH in the final washing step instead of 0.5 M Tris-HCl pH 7.5, 3 M NaCl was employed in place of the unidirectional Southern transfer.

2.2.17 HYBRIDIZATION OF RADIOACTIVE PROBES TO IMMOBILIZED DNA.

The prehybridization and hybridization of radioactive probes to DNA immobilized on nitrocellulose was carried out exactly as described by Wahl *et al.*, (1979). After at least a 16 hour hybridization step, the filters were washed at high, medium or low stringencies (0.1 x SSCE 0.1% SDS, 0.5 x SSCE 0.1% SDS or 2 x SSCE

0.1% SDS respectively) for at least 60 minutes at 65°C, blotted dry, covered with plastic wrap, placed in contact with X-ray film and exposed at −80°C in the presence of a tungsten intensifying screen.

2.2.18 ISOLATION OF RESTRICTION FRAGMENTS.

2.2.18a FROM POLYACRYLAMIDE GELS.

Restriction fragments were detected under UV light after the polyacrylamide gel was briefly stained with ethidium bromide. The desired bands were excised from the gel and the DNA was eluted essentially as described by Maxam and Gilbert (1980). The gel slice was placed in a 1 ml disposable syringe and pushed through sterile gauze into a silanized P1000 pipette tip that had been heat sealed at one end and plugged with a small wad of silanized glass wool. 400 μl of elution buffer (0.5 M ammonium acetate, 10 mM $MgCl_2$, 1 mM EDTA, 0.1% SDS) was added, the P1000 tip covered with Parafilm and incubated overnight at 37°C with shaking. The bottom of the plugged tip was cut off and the buffer with the eluted DNA was centrifuged through the glass wool plug into another sealed P1000 tip. The eluate was then transferred to an Eppendorf tube and precipitated by the addition of 2 volumes of ethanol, chilling at −80°C for 10 minutes and centrifugation at 12,000 rpm for 10 minutes. The pellet was resuspended in 200 μl of 300 mM Na-acetate pH 6.0 and ethanol precipitated again, as described above. Generally, the pellet was then washed once with ice-cold ethanol. However, when the precipitated DNA was ready for direct DNA sequencing, it was considered imperative that this sample was salt-free and the final ethanol wash was discarded in favour of resuspending the pellet in 25 μl of H_2O, adding 1 ml of ethanol, chilling and centrifugation as described above.

2.2.18b FROM LOW MELTING POINT AGAROSE GELS.

Restriction fragments were detected under UV light after the agarose gel was briefly stained with ethidium bromide. The desired bands were excised from the gel

and the DNA extracted as described in the BRL catalogue. The gel slice was melted at 68°C and diluted by the addition of several volumes of pre-warmed buffer (50 mM Tris-HCl pH 8.5, 1 mM EDTA). The melted gel slice was then transferred to 37°C and allowed to equilibrate before the addition of an equal volume of buffer saturated phenol. After vortexing for 30 sec the phases were separated by centrifugation at 12,000 rpm for 2 min in an eppendorf centrifuge. The aqueous phase was removed, taking care to avoid the agarose which banded as a thick layer at the buffer phenol interface. The DNA was phenol extracted at least once more, or until there was no visible interface. After extraction with two volumes of ether to remove phenol, the DNA was precipitated by the addition of 0.1 volumes of 3M sodium acetate pH 5.5 and 2.5 volumes of ethanol.

2.2.19 RADIOLABELLING OF DNA.

2.2.19a 5′-TERMINAL RADIOLABELLING.

Restriction fragments which had been purified as described above were labelled at the 5′-terminus with γ-^{32}P-ATP (S.A. 1000 or 2400 Ci/mmole) using polynucleotide kinase exactly as described by Maxam and Gilbert (1980).

2.2.19b 3′-TERMINAL RADIOLABELLING.

It was not necessary to isolate single fragments for the 3′ radiolabelling procedure. Two μg of recombinant DNA was digested with the appropriate restriction enzymes in a 200 μl reaction mix. The NaCl concentration was adjusted to 50 mM where necessary and the mix was transferred to a tube in which 50 pmoles of α-^{32}P-dATP and α-^{32}P-dCTP (S.A. 2400 Ci/mmole) had been dried *in vacuo*. dGTP and dTTP were added to a final concentration of 5 μM and the reaction mix was incubated with 0.5 units of Klenow (large fragment of *E. coli* DNA Pol I) at 37°C for 20 minutes. dATP and dCTP were added to a final concentration of 10 μM and the reaction was chased for 10 minutes to ensure that all termini had been completely filled. The radiolabelled DNA fragments were ethanol precipitated

and fractionated on a TBE pH 8.3 polyacrylamide gel which was then covered with Plasticwrap and exposed to X-ray film. The exact position of the radiolabelled fragments was determined by placing radioactive ink markers adjacent to the gel to orientate the autoradiograph. The radiolabelled fragments were excised from the gel and eluted as described in 2.2.18.

2.2.19c NICK TRANSLATION OF DOUBLE-STRANDED DNA.

Labelling double-stranded DNA using *E. coli* DNA polymerase I (Rigby *et al.*, 1977) was carried out essentially as described by Maniatis *et al.*, (1975). The 50 μl incubation mix contained 50 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 50 μg/ml of bovine serum albumin, 5 μM of each of α-^{32}P-dATP, α-^{32}P-dCTP, α-^{32}P-dTTP (1000 to 2000 Ci/mmole each) and unlabelled dGTP. The DNA was nicked by the addition of 20 pg of *E. coli* DNase I and incubation at room temperature for 1 min before the reaction was started by the addition of 2 units of *E. coli* polymerase I. The solution was incubated at 15°C for 90 min. Plasmid pBR322 DNA labelled in this way was full length on a 1% agarose gel (data not shown) and labelled to a specific activity of 1 to 5 x 10^8 c.p.m. per μg.

2.2.19d EXTENSION OF UNIVERSAL PRIMER ON M13 SUB-CLONES.

DNA fragments subcloned into M13 were labelled by the hybridization of 10 μl of prepared M13 subclone single-stranded DNA (2.2.22e) to 4 units (one unit is sufficient for one set of 4 dideoxy sequencing reactions) of universal primer DNA as described in 2.2.23a. This subclone/primer hybrid was then added to a 50 μl incubation mix which contained final concentrations of 50 mM NaCl, 10 mM Tris-HCl pH 8.5, 10 mM $MgCl_2$, 2 mM DTT, 3 μM of α-^{32}P-dCTP and α-^{32}P-dATP (S.A. 2400 Ci/mmole) and 20 μM of unlabelled dGTP and dTTP. The primer/subclone template was elongated through the inserted DNA segment by incubating this mix at 37°C for 15 min in the presence of 2 units of Klenow (large fragment of *E. coli*

DNA pol I). The labelled insert was then released from the M13 vector DNA by appropriately adjusting the salt concentration of the incubation mix and digesting with several units of the required restriction enzyme(s) for 30 min, e.g. if a Bam HI fragment was cloned into the Bam HI site of M13, thus regenerating the terminal Bam HI sites, the subclone was digested with Bam HI, however if the cloning sites were not regenerated, e.g. blunt-ended ligation of a Hgi AI fragment into the Sma I site, the subclone was digested with a pair of enzymes which would cleave very close to, but outside, the ends of the inserted DNA, in this example Eco RI and Sal I were used. The labelled insert DNA was then purified away from contaminating vector DNA by quick electrophoresis on a low melting point submarine agarose gel and extracted from the gel as described in 2.2.18b after detection of the probe DNA by autoradiography for 30 sec. Generally subclones containing fragments, ranging in size from 100 to 450 bp, incorporated 2.0 to 4.0 x 10^7 Cherenkov c.p.m. into resected insert DNA.

2.2.20 SEPARATION OF RADIOLABELLED TERMINI.

2.2.20a STRAND SEPARATION.

The separation of radiolabelled complementary strands was carried out essentially as described by Maxam and Gilbert (1980). The radiolabelled fragment was resuspended in 40 μl of strand separation buffer (30% DMSO, 1 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol), heated to 90°C for 2 minutes, quick chilled in an ethanol/dry ice bath for 5 seconds and left on ice. The sample was loaded onto a pre-electrophoresed TBE pH 8.3 polyacrylamide gel (6% acrylamide : 0.1% bis-acrylamide) and electrophoresed until adequate separation between the single strands had occurred (determined empirically). The radiolabelled bands were detected by autoradiography and purified as described in 2.2.18a.

2.2.20b SECONDARY RESTRICTION.

The terminally labelled fragment was resuspended in 100 μl of the appropriate

buffer, digested with an appropriate endonuclease, ethanol precipitated and fractionated on a TBE pH 8.3 polyacrylamide gel. The radiolabelled bands were prepared as described in 2.2.18a.

2.2.21 CHEMICAL CLEAVAGE DNA SEQUENCING REACTIONS.

The chemical cleavage sequencing reactions were carried out essentially as described by Maxam and Gilbert (1980) except for the purine specific reaction where piperidine formate pH 2.0 was replaced by 100% formic acid. Five reactions specific for purines, (P), guanine, (G), strong adenine weak cytosine, (A>C), pyrimidines, (Y), and cytosine, (C), bases were used in the analysis of 5′-end labelled DNA. Adenine and thymine were identified by their presence in the purine and pyrimidine reactions respectively, and absence in the guanine and cytosine reactions.

End-labelled DNA was dissolved in 35 μl of water and divided into five aliquots, G (5 μl), P (10 μl), A>C (5 μl), Y (10 μl) and C (5 μl). 200 μl of cacodylate buffer and 1 μl of dimethyl sulphate were added to G and the reaction mixture incubated at 21°C for 30 seconds to 1 minute. The reaction was stopped by the addition of 50 μl of G stop mix, (3 M sodium-acetate pH 6.0, 2.5 M 2-mercaptoethanol, 1 mM EDTA, 0.1 mg/ml *E. coli* tRNA), 1 ml of ethanol and the DNA precipitated at –70°C for 1 hour.

25 μl of formic acid was added to the P sample and incubated at 21°C for 2 minutes followed by the addition of P stop mix, (0.3 M sodium-acetate pH 6.0, 0.1 mM EDTA, 25 μg/ml *E. coli* tRNA), 1 ml of ethanol and the DNA precipitated at –70°C for 1 hour. 15 μl of water and 20 μl of 5 M NaCl were added to Y and C respectively. 30 μl of hydrazine was added to both with incubation at 21°C for 2 minutes (Y) or 1 minute (C), followed by the addition of 300 μl of either Y stop mix, (0.3 M NaCl, 0.1 mM EDTA pH 8.0, 25 μg/ml *E. coli* tRNA), or C stop mix (0.1 mM ETDA pH 8.0, 25 μg/ml *E. coli* tRNA), 1 ml of ethanol and the DNA

precipitated at –70°C for 1 hour. 100 μl of 1.2 N NaOH, 1 mM EDTA was added to the A>C reaction and the mixture incubated at 90°C for 5 to 15 minutes. The reaction was stopped by the addition of 150 μl of 1 M acetic acid, 5 μg of *E. coli* tRNA, 1 ml of ethanol and the DNA precipitated at –70°C for 1 hour.

After centrifugation, all samples were reprecipitated by the addition of 300 μl of 0.3 M sodium-acetate pH 6.0 and 1 ml of ethanol at –70°C for 1 hour, then centrifuged, washed with 1 ml of ethanol and dried under vacuum.

All samples were redissolved in 100 μl of freshly prepared 1 M piperidine, heated at 90°C for 15 minutes, (in Eppendorf microfuge tubes), then dried under vacuum. After the addition of 25 μl of water, samples were again dried and then dissolved in 4 μl or multiples thereof of formamide loading buffer, (90% (v/v) deionized formamide, 0.01% (w/v) bromophenol blue, 0.01% (w/v) xylene cyanol FF, 0.1 mM EDTA pH 8.0).

Products of the chemical degradation sequencing reactions were separated by electrophoresis on polyacrylamide gels which included 8.3 M urea as a denaturant.

For 20% gels, a 100 ml mixture containing 20 ml of 10 x TBE, (0.5 M Tris-borate pH 8.3, 10 mM EDTA), 19 g of acrylamide, 1 g of N,N'-methylenebisacrylamide, 42 g of area, 1 ml of 10% (w/v) ammonium persulphate was filtered and de-gassed, 20 μl of TEMED added and poured into a 30 x 40 x 0.5 cm gel mould, and allowed to polymerize. Reservoir tanks contained 2 x TBE and the gels were pre-electrophoresed for 1 to 4 hours.

Samples were heated to 90°C for 2 minutes then chilled on ice before loading. Debris and urea were removed from sample wells prior to loading by flushing with electrophoresis buffer from a syringe. 10% and 8% gels were also used in sequencing reactions. All gels were run at 1,000 to 1,200 volts. Gels were autoradiographed in the presence of an intensifying screen at –80°C for varying times.

2.2.22 THE M13 CLONING SYSTEM.

M13 is an *E. coli* F^+ specific single-stranded DNA bacteriophage which exists in a double-stranded intracellular replicative form (RF). The RF resembles a plasmid vehicle in several aspects e.g., size, copy number and autonomous replication, and can be used as a vector exactly as certain bacterial plasmids. The DNA to be cloned, for example for eventual sequencing, is digested with restriction endonucleases and the fragments of DNA may be ligated into a unique restriction site in the RF of M13 or one of the genetically engineered derivatives of bacteriophage M13. (Messing *et al.*, 1977, Messing and Vieira, 1981). The recombinant RF, i.e. with inserted foreign DNA, is then used to transform competent cells which are plated out for single plaques indicating foci for phage infection in a feeder lawn of bacterial cells. The (+) strand is packaged into the protein coat and extruded from the cell without cell lysis and this particle becomes the extracellular and infective form of M13. Thus it is possible to separate the complementary strands of the foreign DNA inserted into the RF so that the orientation of the cloned DNA determines which strand will be incorporated in the (+) strand of M13.

The recombinant M13 phage are detected on the basis of a complementation assay. The vector molecule contains a portion of the β-galactosidase gene from the lac operon of *E. coli*, in particular the region encoding the amino-terminal 145 amino acids, the α-peptide which is capable of associating with and complementing another type of defective β-galactosidase, synthesized in the host cell, which results in a functional enzyme. The presence or absence of an active β-galactosidase is determined using the lactose analogue 5-bromo-4-chloro-3-indoyl-β-galactoside, which upon hydrolysis yields the blue dye, bromo-chlorindole. Hence the M13 plaques are blue and the recombinant phage, in which the insertion of DNA into coding region of the α-peptide has destroyed any possible complementation, produces white plaques. The recombinant phage are amplified in small liquid cultures and the DNA from the extruded phage can be harvested and used as a template in

the ensuing DNA sequencing reactions.

A short, commercially available primer, annealed to a specific region 3′ to the inserted DNA, can be extended in the 5′ to 3′ direction using the large subunit of DNA Pol I (Klenow fragment) and deoxyribonucleotides (dNTPs). The incorporation of a dideoxynucleotide (ddNTP) into a newly synthesized DNA strand will result in the termination of that strand at that nucleotide. The lack of a 3′ hydroxyl group on the ribose moiety of the ddNTP prevents further polymerization. Thus the ddNTPs act as specific chain terminators and can be used to determine the nucleotide sequence of the inserted DNA 3′ to the annealed primer.

The M13 "Shotgun strategy" (Messing *et al.*, 1981) was not used in this study as generally, DNA sequence was only required from short DNA fragments (of the order of 1 Kb) which were not amenable to the "Shotgun" approach.

2.2.22a PREPARATION OF M13 REPLICATIVE FORM DNA.

M13 replicative form DNA was prepared essentially as described by Winter and Fields (1980).

To 3 mls of 0.7% agar at 45°C was added 30 μl of BCIG, (5-bromo-4-chloro- 3-indoyl-β-galactoside), (20 mg/ml in dimethylformamide), 20 μl of IPTG, (isopropyl-β-D-thio-galactopyranoside), (20 mg/ml in water), 0.2 ml of exponential, (A_{600} of 0.6), JM101, and 0.1 ml of diluted M13 phage (enough to give about 200 pfu). This mixture was poured onto a minimal (+ glucose) plate and incubated at 37°C for 9 hours.

A blue plaque was selected, toothpicked into 1 ml of 2 x YT broth and grown with shaking for 6 hours. Meanwhile a 10 ml culture of JM101 from a single colony on a minimal glucose plate was grown to an A_{600} of 0.5, and added to 1 litre of 2 x YT. When the A_{600} of this culture reached 0.5, the 1 ml of phage solution

was added and grown for 4 hours. Replicative form M13 DNA was prepared from pelleted cells by the alkali lysis method previously described (2.2.10).

2.2.22b PREPARATION OF M13 VECTORS.

The replicative form of M13 mp8 and 9 was prepared as described above. The RF was digested with the appropriate enzyme(s) to generate the desired termini. The linearized RF was fractionated on a 1% agarose gel (low gelling temperature) to remove any intact molecules, extracted as described in 2.2.18b and resuspended in H_2O to a final concentration of 20 ng/μl. A loss of about 25% of the linearized RF was anticipated during these extraction procedures. The enzyme Sma I recognizes the sequence 5′-CCCGGG-3′ and produces blunt ended termini. This enzyme was found to contain a low level of exonuclease activity and it was necessary to remove the 5′ terminal phosphate groups from the Sma I linearized vector with phosphatase as described by Maxam and Gilbert (1980). The failure to phosphatase this vector led to the production of "false-whites", that is M13 plaques which were "white" not from the insertion of foreign DNA into the α-peptide coding region, but the removal of one or two bases from either terminus had altered the reading frame and thus prevented any possible complementation between the α-peptide and the other defective β-galactosidase gene synthesized by the host.

2.2.22c LIGATION CONDITIONS.

The DNA fragment and appropriate M13 vector were combined in a molar ratio of approximately 5:1 in a 20 μl reaction mix containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 1 mM DTT. All ligations, regardless of the termini, were carried out using 0.5 units of T4 DNA ligase for at least 3 hours at 15°C.

2.2.22d TRANSFORMATION OF *E. coli*, JM101.

E. coli, strain JM101, was made competent essentially using the method of Messing *et al*., (1981) except that the cells were harvested after reaching an A_{600}

of 0.75 and finally resuspended in one twentieth of the growth volume of ice-cold 100 mM $CaCl_2$. The cells were left on ice for at least one hour before use and the density of these cells was such that no additional bacteria were needed to act as a feeder lawn. The transformation was carried out as described by Messing *et al.*, (1981) except that no feeder lawn was added to the top agarose and only 20 μl of 5-bromo-4-chloro-3-indoyl-β-galactoside (20 mg/ml in dimethylformamide) was used in each plate.

2.2.22e HARVESTING (+) STRAND OF RECOMBINANT BACTERIOPHAGE.

The white plaques were harvested from 1 ml cultures essentially as described by Messing *et al.*, (1981) except that the PEG pellet was resuspended in 100 μl of SET buffer and extracted once with an equal volume of SET-buffered phenol and extracted twice with five volumes of ether. The single-stranded recombinant DNA was resuspended in 30 μl of H_2O and stored at −20°C.

2.2.23 DIDEOXY SEQUENCING REACTIONS.

2.2.23a HYBRIDIZATION.

4 μl of the single-stranded M13 template, 1 μl of primer (1 unit/μl of Biolabs 15-mer), 1 μl of 10 x Hin buffer (60 mM Tris-HCl pH 7.5, 60 mM $MgCl_2$, 500 mM NaCl) and 4 μl of H_2O were combined in an Eppendorf tube which was placed in a small boiling water-bath and left to cool to room temperature (about 45 minutes). The tubes were centrifuged for 5 seconds to collect the condensate and left at 4°C until required.

2.2.23b POLYMERIZATION.

The chain termination reaction (Sanger *et al.*, 1977) was carried out as described by Messing *et al.*, (1981) except that α-^{32}P-dGTP, α-^{32}P-dCTP or α-^{32}P-dATP was the radiolabelled nucleotide and the incubations were carried out at

37°C. The final dideoxynucleotide concentrations used in this study were 0.08 mM ddGTP, 0.6 mM ddATP, 1.3 mM ddTTP and 0.4 mM ddCTP. It was convenient to combine the G° and ddGTP solutions, etc. of Messing *et al.*, (1981) and to store these mastermixes at –20°C until required.

2.2.23c SEQUENCING GEL ELECTROPHORESIS.

The samples were electrophoresed through a 6% polyacrylamide gel containing 50% urea. TBE pH 8.8 was used in these gels as the buffer provided greater resolution for long electrophoresis runs than the TBE pH 8.3 buffer used elsewhere in this study. In cases of prolonged electrophoresis of the samples through the sequencing gel, the electrophoresis buffer was changed every time the BPB tracker dye migrated 40 cm. After electrophoresis, the gel was fixed for at least 10 minutes with 12% acetic acid, dried in an oven at 110°C for at least 30 minutes, covered with plastic wrap and exposed to X-ray film. Usually an overnight exposure at room temperature was sufficient to detect all the bands.

2.2.24 CONTAINMENT FACILITIES.

All work involving recombinant DNA was initially carried out under C3/EKI containment conditions for work involving viable organisms amd C0 containment conditions for work not involving viable organisms, as defined and approved by the Australian Academy of Science Committee on Recombinant DNA and by the University Council of the University of Adelaide.

2.2.25 COMPUTER ASSISTED DNA SEQUENCE ANALYSIS.

Analysis of DNA sequence data was facilitated by the application of several DNA sequence handling programs (Staden, 1977, 1978, 1979, 1980a, 1980b), listed below, on a Disc Computer Systems DS-23 microcomputer fitted with a 40 megabyte hard disc drive. The programs were generously donated by R. Staden and slightly modified before use on the DS-23 microcomputer. Two programs, TESTCD and

COMSTR were written and generously donated by I. Dodd and A.V. Sivaprasad respectively.

BASSUM : *Counts base totals for regions of a DNA sequence.*

BATIN : *Used to enter DNA sequence data.*

BPFIT : *Looks for regions in two sequences which could base pair.*

CODSUM : *Counts codon totals for regions of a DNA sequence.*

COMSTR : *Searches for and displays possible secondary structures present in DNA sequence files.*

CUTSIT : *Searches a DNA sequence for a list of restriction sites.*

DBCOMP : *Searches for overlaps between DNA sequences.*

FILINS : *Creates a file containing sections of sequence from other DNA sequence files.*

HAIRPN, HAIRGU : *Search for palindromes, inverted repeats and potential hairpin loops in nucleic acid sequences.*

MWCALC : *Calculates the molecular weight, polarity index and amino acid composition of protein sequences.*

SEARCH, SRCHMU : *Search for nucleotide strings in a DNA sequence.*

SEQEDT : *Stores and edits sequence data.*

SEQLST : *Produces formatted copies of DNA sequence files.*

SQRVCM : *Reverses and complements a DNA sequence file.*

TESTCD : *Searches a DNA sequence for possible coding regions.*

TRANMT, TRNTRP, TRANDM : *Translate a DNA sequence into a protein sequence.*

CHAPTER 3.

PRELIMINARY CHARACTERIZATION OF PUTATIVE mRNAs IN EMBRYONIC FEATHER.

3.1 INTRODUCTION.

The embryonic chick feather is a terminally differentiating tissue and as such offers an opportunity to study some of the processes associated with differentiation. During development the embryonic feather grows and senesces, eventually filling with protein, mostly keratin, shortly before hatching. Analysis of the proteins from newly-hatched chick feathers on SDS polyacrylamide gels (Figure 3.1) reveals the presence of four groups of proteins. The major group of proteins are the β-keratins, a family of 19 to 25 homologous polypeptide chains, each with a molecular weight of 10,000 to 11,000 Mr, and reported to represent about 90% of the total protein in embryonic feathers (Kemp and Rogers, 1970, 1972; Kemp, 1972; Walker, 1974; Walker and Rogers, 1976a, 1976b). At the time the work reported in this chapter was initiated very little work on the other protein species of feather had been conducted. A subsequent study (Powell, 1979; Powell and Rogers, 1979) examined a second class of homologous proteins of 8,000 to 8,500 Mr, termed "fast proteins" because of their high mobility on pH 2.7 polyacrylamide gels and reported to comprise the remaining 10% of feather protein (Walker and Rogers, 1976a; Powell and Rogers, 1979). This would appear, from Figure 3.1 to be an underestimate of their abundance.

Another group of proteins of about 50,000 Mr and larger also appear to be present in embryonic feathers, but at a very low level, which may be why they were not considered in previous studies. These larger proteins could be similar to the group of α-keratins found in chick scales (Wilton, 1983). The fourth group of proteins, which are present at significant levels, co-migrate with the β-keratins of scale and have a molecular weight of about 14,500 Mr. Powell (1979, personal communication) believes these proteins to be β-keratins which migrate at an artefactually slow rate as a consequence of S-carboxymethylation.

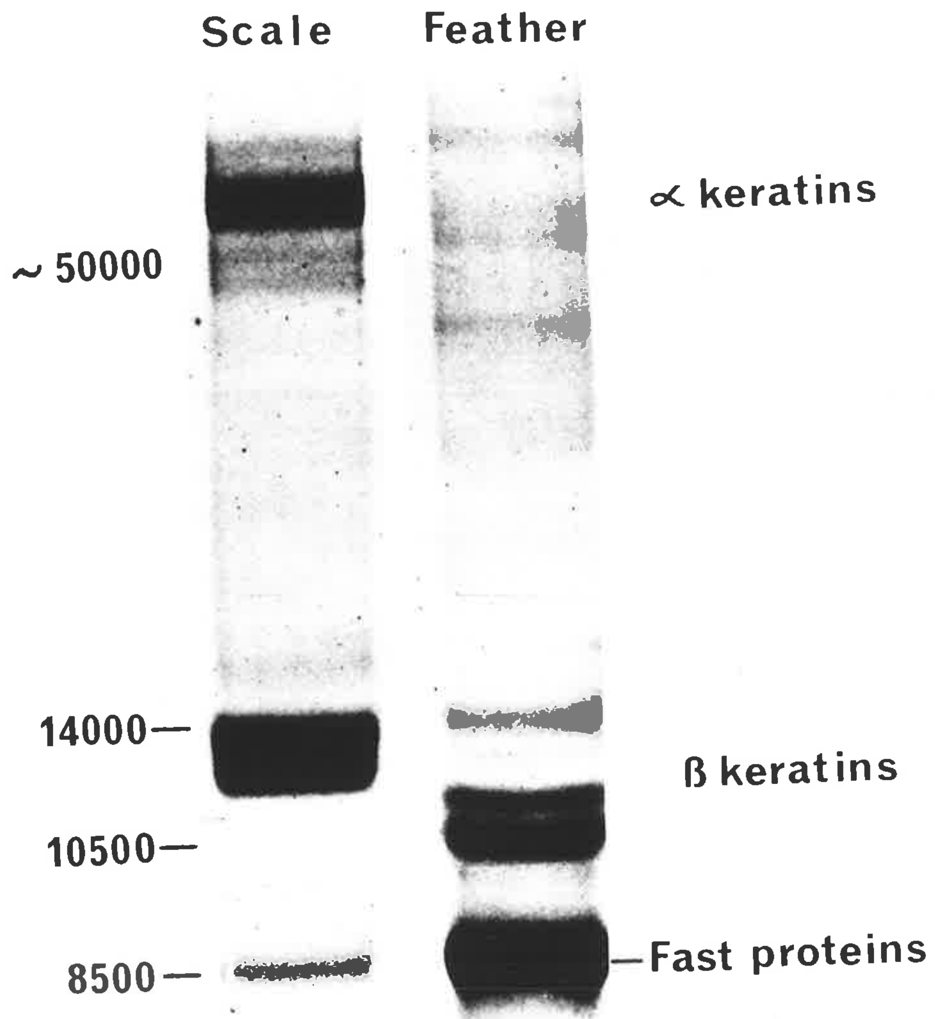
The 12S mRNA coding for keratin was isolated (Partington *et al.*, 1973; Kemp *et al.*, 1974a) and appeared to produce only β-keratins upon *in vitro* translation

FIGURE 3.1

ELECTROPHORETIC ANALYSIS OF SCALE AND FEATHER PROTEINS ON AN SDS POLYACRYLAMIDE GEL.

S-carboxymethylated proteins from 17-day feathers and 18-day anterior shank scales fractionated by electrophoresis on a 10% polyacrylamide SDS 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
∝ keratins
~ 50000
14000
10500
8500
ß keratins
Fast proteins

(Partington *et al.*, 1973; Kemp *et al.*, 1974c). This mRNA was found, in the presence of an oligo(dT) primer, to act as an efficient template for the synthesis of complementary DNA by avian myelobastosis virus (AMV) reverse transcriptase (Kemp, 1975; Kemp *et al.*, 1975). A comparison of the rate of reannealing of this cDNA to the mRNA from which it was transcribed, with that of globin mRNA to its cDNA indicated that there are 25 to 35 different mRNAs coding for keratin in the embryonic feather. These keratin mRNA to cDNA hybridizations revealed rapid transitions when assayed on hydroxyapatite, compared to assay by S_1 nuclease, indicating the presence of both unique and repetitive sequences in the keratin mRNA. This result led Kemp (1975) to propose that these repetitive sequences are the keratin coding sequences and the unique sequences correspond to untranslated regions. This hypothesis was supported by mRNA to cDNA hybridization kinetic studies where hybridizations using short oligo(dT) primed cDNA from the 3′-end of the mRNA were compared to hybridizations using full length oligo(dT) primed cDNA, which indicated that the 3′-untranslated regions of keratin mRNA are unique (Lockett and Rogers, 1979). Presumably the large number of mRNA species, indicated by the above hybridization characteristics, code for the multiple polypeptide chains described by Walker and Rogers (1976a), some of which have been partially sequenced (Walker and Rogers, 1976b).

Like the majority of eukaryotic mRNAs, embryonic feather keratin mRNA was found to have 7-methylguanosine triphosphate at its 5′-terminus (Morris, 1976; Morris and Rogers, 1979), i.e. it was "capped" (Rottman *et al.*, 1974). The cap structure apparently plays a role in the correct formation of stable initiation complexes between eukaryotic mRNA and ribosomes (Both *et al.*, 1975a, 1975b; Rose, 1975; Kemper, 1976; Roman *et al.*, 1976). At the time of this study the capping of eukaryotic mRNAs was popularly being discussed as having a role in the regulation of gene expression at the level of mRNA synthesis, translation and degradation (for a review see Shatkin, 1976). Another structural feature involved in the regulation of eukaryotic mRNA, popular at that time, was the size distribution

of polyadenylate sequences at the 3′-terminus. Although, like the cap structure, some eukaryotic mRNAs have been shown to be devoid of 3′-poly(A) sequences, for example, histones (Adesnik and Darnell, 1972; Greenberg and Perry, 1972), it appears that it does play a role in potentiating initiation of translation (Doel and Carey, 1976). Poly(A) has also been ascribed a role in mRNA stability (Marbaix *et al.*, 1975) and the transport of mRNA into the cytoplasm (Mendecki *et al.*, 1972), although this role has not been clearly demonstrated (for a review see Brawerman, 1976). Although keratin mRNA is polyadenylated no serious attempt to determine its size distribution prior to this study has been made.

This study attempts to approach a fuller understanding of the control of gene expression in the embryonic chick feather at 14 days of embryonic development, the time of maximal protein and mRNA synthesis (Kemp *et al.*, 1974b). In accord with the trend of the literature at the time, and the earlier work carried out on the characterization of the proteins and mRNAs of the embryonic feather, this question has been approached by an examination of the mRNAs expressed at that time and, in particular, the 12S mRNA coding for the β-keratins, the fine structure of which was only poorly understood.

The cloning and nucleotide sequencing of double-stranded cDNA copies of mRNA was an obvious approach to the purification and characterization of a mixture of homologous mRNAs. Unfortunately at the time of this study the appropriate containment facilities were not available. When such facilities became available, this approach was used and the results of those studies are presented in later chapters.

This chapter describes studies on several RNA species present in 14-day embryonic chick feathers and distinguishable by sedimentation through sucrose gradients, i.e. 9S, 12S and 14S, of these only the 12S, which had previously been shown to code for β-keratins (Kemp *et al.*, 1974a), could unequivocally be described as a mRNA. The 9S RNA directed the synthesis of more fast protein in a wheat germ cell-free system than did 12S but it could not be described as a fast protein

messenger RNA. The 14S RNA was shown to be largely composed of fragments of degraded ribosomal RNA.

More detailed study of 12S mRNA indicated the presence of polyadenylate tracts ranging in size from 45 to 165 nucleotides with a number-average length of 65 nucleotides and a weight-average length of 85 nucleotides. The ability of the restriction endonuclease Hae III to cleave single-stranded cDNA was used to demonstrate that the 3′-ends of 12S mRNAs show some degree of homology and this was confirmed by initial attempts at sequencing the 3′-ends of 12S mRNAs by elongating specific dT_8 oligomers in the presence of specific chain terminating nucleotides (Sanger *et al.*, 1977; Hamlyn *et al.*, 1978). It should be pointed out that the RNA sequencing work was performed prior to the publication of the technique and at a very early stage of its development.

3.2 SPECIFIC METHODS.

3.2.1 RANDOM PRIMED REVERSE TRANSCRIPTION.

Priming of 18S and 28S ribosomal RNA lacking a 3′-terminal poly(A) tract was achieved by the random hybridization of oligonucleotides of salmon sperm DNA, prepared as described by Taylor *et al.*, (1976).

Conditions for the synthesis of this cDNA were as described for the oligo(dT) primed reaction (2.2.14), except that oligo(dT) was replaced by a final concentration of 2 mg/ml of oligonucleotide and the synthesis proceeded for 60 min at 37°C. The cDNA synthesized was isolated as described for oligo(dT) primed synthesis.

3.2.2 RNA HYBRIDIZATION REACTIONS.

Hybridization reactions were performed either in small glass tubes or in sealed capillary tubes. All glass-ware was siliconized to minimize the adhesion of nucleic acids to it, and when hybridizations were performed in glass tubes the reaction mixture was overlaid with sterile paraffin oil in order to reduce evaporation.

Hybridizations were carried out in 0.18 M NaCl, 0.01 M Tris-HCl, pH 7.4, 0.001 M EDTA, 0.5% SDS. The reaction mixtures were first denatured at 100°C for 5 minutes then incubated at 60°C to the desired R_0t value (the product of ribonucleotide concentration (mol/litre) and time (sec); Bishop *et al.*, 1974). The hybridization reactions were carried out under conditions where the unlabelled species (RNA) was in excess over the cDNA probe. In experiments where the kinetics of hybridization were investigated, different R_0t values were achieved by using a series of dilutions of the nucleic acid solution and incubating all reaction tubes for the same length of time.

After incubation, the extent of hybridization was assayed using the single-strand specific nuclease of *Aspergillus oryzae* (Ando, 1966). The hybridization mixtures were diluted in low salt S_1 assay buffer (0.05 M NaCl, 0.03 M sodium acetate, pH 4.6, 0.001 M $ZnSO_4$, 5% glycerol) containing 12 μg of calf thymus DNA included as carrier, divided into equal parts, and 4 units of partially purified S_1 nuclease (Vogt, 1973) was added to one half. (The enzyme had been purified to the end of step 4 as described by Vogt, 1973). Digestion was carried out at 45°C for 30 minutes and the samples were then TCA precipitated, with the addition of 25 μg of BSA as carrier. After 60 minutes on ice the acid-insoluble material was collected on Whatman GF/A disks, washed with cold 5% TCA followed by ether. The samples were then dried at 110°C and counted in toluene-based scintillation fluid. The extent of hybridization was calculated from the acid-insoluble radioactivity in the nuclease S_1 treated and untreated samples.

3.2.3 POLYACRYLAMIDE GEL ELECTROPHORESIS OF RNA.

3.2.3a 98% FORMAMIDE GELS.

Electrophoresis of RNA in 98% formamide was based on the system initially devised by Staynov *et al.*, (1972) with the modifications suggested by Gould and Hamlyn (1973) and Pinder *et al.*, (1974), except that Zeokarb 225 was used

instead of Amberlite to deionize the formamide. Gels used had a total acrylamide concentration of either 4% or 10%. The 4% gels were made 3.4% w/v acrylamide and 0.6% w/v bis-acrylamide, while 10% gels were 9% w/v acrylamide and 1% w/v bis-acrylamide. Cylindrical gels (8 cm x 6 mm diam.) were prepared in Perspex tubes and overlaid with 70% formamide to give level gel surfaces during polymerization. The gels were allowed to polymerize for 30 min then stored at 4°C until required. The 70% formamide was then replaced by 1 cm of buffered formamide (0.02 M diethylbarbituric acid, NaOH pH 9.0), which was carefully overlaid with electrode buffer (16 mM NaCl 8 mM Tris-HCl pH 8.7). The gels were pre-electrophoresed for 15 minutes at 5 mA/gel before use. The RNA samples were dissolved in 25 μl of buffered formamide plus 5% sucrose and a trace of bromophenol blue, and heated to 95°C for 1 minute before loading. Electrophoresis was carried out at 5 mA/gel until the bromophenol blue band was about 1 cm from the bottom of the gel. The gels were then placed in sterile water ready for staining, scanning or slicing.

3.2.3b 8 M UREA GELS.

Electrophoresis of RNA in 8 M urea gels was based on the system of Sanger and Coulson (1975).

The tube gels contained either 4% or 6% acrylamide and 8 M urea in 25 mM tris-glycine pH 8.3. Electrophoresis was carried out at room temperature for 4 hours at 2 mA per gel using 25 mM Tris-glycine pH 8.3 electrophoresis buffer. The gels were stained with toluidene-blue and destained in distilled water, in the dark at 4°C, to prevent degradation of the RNA, (Popa and Bosch, 1969).

3.2.4 DETECTION OF RNA SAMPLES ON POLYACRYLAMIDE GELS.

The position of RNA bands on unstained gels was determined by scanning at 280 nm in a Gilford linear transport attachment (Gilford Instrument Laboratories, Oberlin, Ohio, U.S.A.) coupled to a W & W recorder (W & W Electronic Inc., Basle,

Switzerland). Alternatively gels were stained in 0.05% toluidine blue (dissolved in 55 mM sodium acetate, 0.1 mM EDTA, pH 5.5) for 30 minutes and destained overnight against two changes of distilled water. The mobility of RNA species was either determined directly from these gels or they were scanned at 600 nm as described earlier.

Radioactive RNA samples were detected by soaking the gel in water to remove formamide, freezing in dry ice, and slicing into 1 mm sections using a single blade type gel slicer (Mickle Engineering Co., Gomshall, Surrey, England). Each section was placed in a scintillation vial and the RNA extracted by treatment with 0.2 ml NCS tissue solubilizer and 0.025 ml concentrated ammonium hydroxide solution overnight at room temperature. The samples were then prepared for radioactivity determination by the addition of 2 mls of toluene based scintillation fluid containing 15% v/v PCS scintillation fluid.

3.2.5 ^{3}H-BOROHYDRIDE LABELLING OF KERATIN mRNA.

The labelling of RNA by periodate oxidation followed by reduction with labelled borohydride was based on a system initially devised by Khorana (RajBhandary *et al.*, 1966, 1968, RajBhandary, 1968) and later reported by Randerath and Randerath (1973). Use of this method to detect modified nucleosides at the 5′-terminus with free 3′, 2′ hydroxyls has been reported by Symons (1975) and Yang *et al.* (1976). ^{3}H-KBH_4 (12.4 Ci/mmole) in vials containing 10 mCi, prepared by Dr. R.H. Symons were used as a source of radioactive borohydride. To 45 μg of keratin mRNA in 30 μl 1 mM EDTA, pH 5.5 was added 23 nmole $NaIO_4$ in 2 μl of water and the reaction mixture incubated at room temperature for 1 hr. The contents of one vial of ^{3}H-KBH_4 (10 mCi, 0.8 μmole) were dissolved in 10 μl of 10 mM KOH and 30 μl 0.1 M sodium borate-HCl pH 8.0, and 15 μl was added to the oxidised RNA. After 100 min at room temperature in the dark, 8 μl of acetic acid (glacial, refluxed with acetic anhydride then re-distilled) was added to decompose unreacted KBH_4. After a further 30 min the RNA was precipitated three times with two volumes of

ethanol from a solution of 0.2 M NaCl, 50 mM sodium acetate. 1 mM EDTA, pH 5.0, and the RNA obtained was dissolved in 250 μl of 1 mM EDTA, pH 8.0. At this stage the oxidized-reduced RNA contained approximately 2.4 x 10^6 cpm with an estimated contamination of 0.3% by non-ethanol precipitable material.

3.2.6 SIZING OF POLY(A) TRACTS.

Keratin mRNA was dissolved in 1 ml of digestion buffer (0.1 M NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 7.4) containing 2 μg/ml of RNAase A and 5 units/ml of RNAase T_1. This was then incubated at 37°C for 30 min during which time all RNA except the polyadenylate tracts should be degraded (Adesnik and Darnell, 1972; Darnell *et al.*, 1971). The remaining RNA was precipitated with 50 μg of *E. coli* tRNA by the addition of sodium acetate (pH 5.5) to a final concentration of 0.1 M and 2.5 volumes of ethanol. After 16 hours at -20°C the precipitate was collected by centrifugation at 18,000 rpm for 30 minutes in a Beckman JA-20 rotor. The pellet was resuspended in 25 μl of loading buffer and electrophoresis carried out in 10% polyacrylamide gels containing 98% formamide. The gel was run in parallel with two identical gels containing 5.8S, 4S and 5S RNA as molecular weight markers. The bromophenol blue dye usually ran in an identical position on all gels, which allowed the gel containing poly(A) to be exactly aligned before slicing.

3.2.7 DETECTION OF POLY(A) USING POLY(U).

The use of ^{3}H-poly(U) to detect poly(A) on gels has been described by Bishop *et al.*, (1974) and Rosbash and Ford, (1974). This method has been employed in an adapted form for use with formamide gels.

Gels containing poly(A) were frozen and sliced into 1 mm sections. Fractions consisting of two adjacent 1 mm slices were placed in sterile siliconized glass scintillation vials and the RNA eluted at various temperatures for 48 hours in various volumes of 2 x SSC (from 1 ml to 0.25 ml). A 2.5 to 5 fold excess of ^{3}H-poly(U) was then added to the eluted RNA and the solution incubated at 45°C for varying times.

After incubation, samples were placed at 0°C and 2 ml of ice cold ribonuclease A (25 μg/ml in 2 x SSC) were added to digest the single stranded poly(U), since under these conditions poly(U)-poly(A) hybrids are resistant to RNAase A (Gillespie *et al.*, 1972 and Slater *et al.*, 1972, 1973). After 20 minutes at 0°C, 100 μg of BSA and 2 ml of 20% TCA were added, the samples filtered through 2.5 cm glass filters (Whatman GF/A) and counted in toluene based scintillation fluid.

3.2.8 NUCLEOTIDE SEQUENCING OF RNA.

3.2.8a PLUS-MINUS SEQUENCING.

The plus and minus method for sequencing DNA (Sanger and Coulson, 1975) was adapted with slight modifications for use in the direct sequencing of RNA using primers of the form $(dT)_8dN$.

Between 0.5 and 1.0 μg of purified mRNA was combined in a 40 μl incubation mix containing 6.6 mM Tris-HCl pH 7.4, 6.6 mM $MgCl_2$, 1 mM DTT, 50 mM NaCl, 10 μCi α-^{32}P-dATP, 3 μM dGTP, dCTP and dTTP, 0.5 μg of the appropriate $(dT)_8dN$ primer and several units of AMV reverse transcriptase and incubated at 37°C. One third of the reaction mix was removed and added to 5 μl 0.2 M EDTA after 1, 2 and 5 min and these were pooled and phenol extracted prior to separation of unincorporated nucleotides by Sephadex G-50 chromatography. The cDNA peak was collected dried under vacuum and dissolved in 10 μl of 1 x H buffer containing a two fold excess of mRNA. This solution was heated at 100°C for 2 minutes and allowed to hybridize at 65°C for 1 hour, after which 1 μl aliquots were used in the plus and minus reactions.

Plus reactions were carried out in 10 μl reaction mixes containing 10 μM of the appropriate single dNTP, 1.5 x H buffer, 1 μl cDNA/mRNA hybrid and 1 μl of Klenow (large fragment of DNA pol I) and incubated at 37°C for 10 minutes followed by freezing and drying under vacuum. Minus reactions were carried out in 10 μl reaction mixes containing 10 μM of the appropriate three dNTPs, 1.5 x H

buffer, 1 μl cDNA/mRNA hybrid and several units of AMV reverse transcriptase and were incubated and stopped as for the plus reactions.

All reactions were then resuspended in 20 μl of formamide loading buffer, heated at 100°C for 2 minutes and 5 μl applied to a 0.5 mm thick sequencing gel prepared and run as described in 2.2.23c.

3.2.8b DIDEOXY SEQUENCING.

Dideoxy sequencing of RNA was performed using the dideoxy reactions of Sanger *et al.*, (1977) as described by Hamlyn *et al.*, (1978).

Using 0.5 μg of purified mRNA as template and 0.5 μg of the oligonucleotide primers in turn, cDNA was transcribed using reverse transcriptase in four reactions each containing a different dideoxy nucleotide triphosphate, e.g. a 10 μl "A" reaction contained 0.125 μg of mRNA, 0.125 μg of primer, 0.5 μl of AMV reverse transcriptase, 50 μM dNTPs except the ^{32}P-labelled nucleotide triphosphate which was present at 2 to 3 μM, dideoxyadenosine triphosphate at 10 μM and 10 mM Tris-HCl pH 8.3, 50 mM NaCl, 10 mM DTT and 6 mM $MgCl_2$. The reactions were incubated at 37°C for 15 minutes and added to an equal volume of formamide loading buffer and 5 μl samples treated and electrophoresed as for the plus and minus reactions.

3.3 RESULTS.

3.3.1 ANALYSIS OF THE RNAs PRESENT IN 14-DAY FEATHERS.

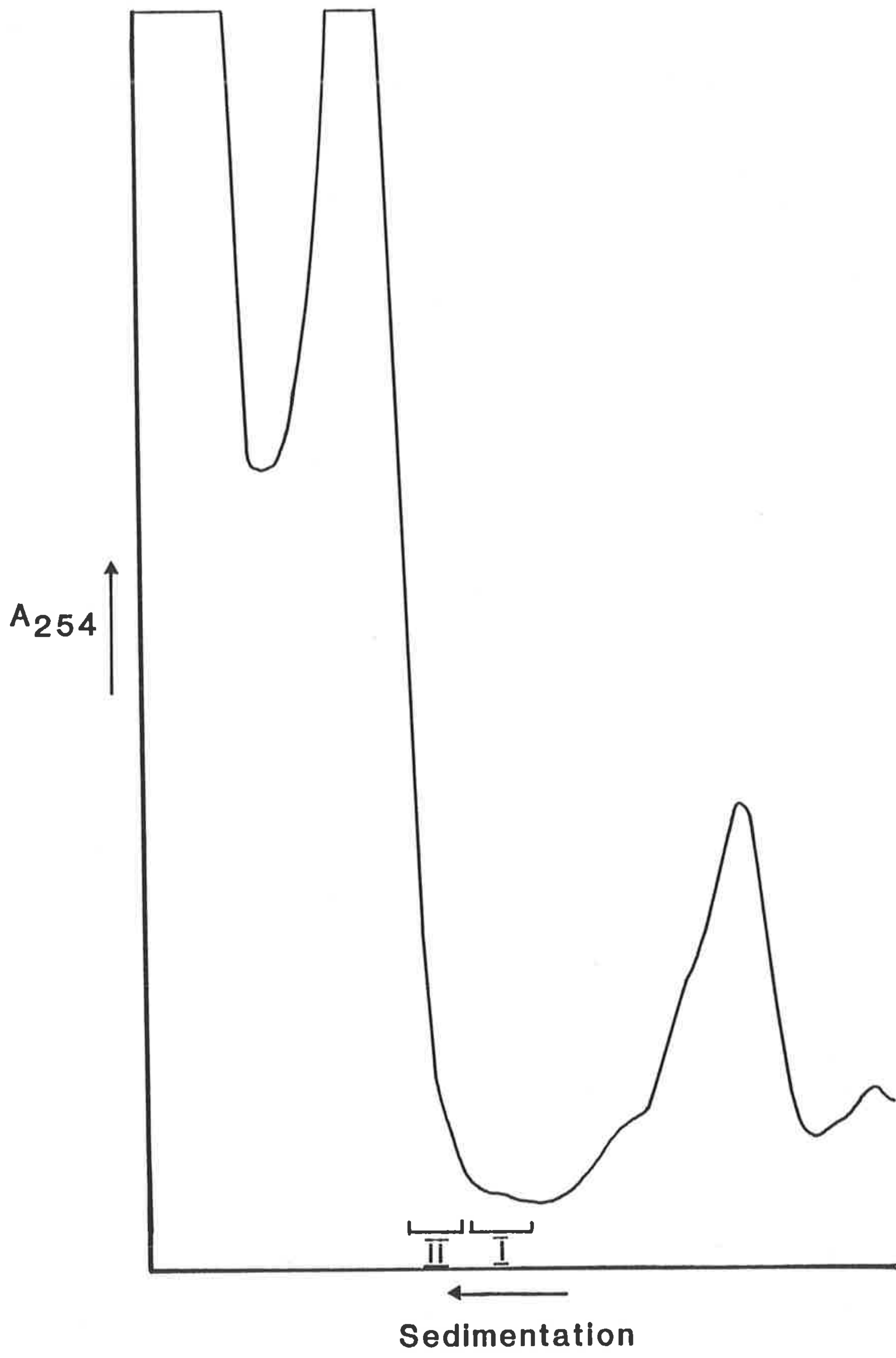
3.3.1a PURIFICATION OF 9, 12 AND 14S RNA FROM RIBONUCLEOPROTEIN PARTICLES.

Embryonic feather ribonucleoprotein (RNP) complexes were prepared from 14-day embryonic feathers as described in 2.2.1c. Ribonucleoprotein fractions I and II, as indicated in Figure 3.2 were collected, ethanol precipitated to concentrate and

FIGURE 3.2

SUCROSE GRADIENT SEDIMENTATION OF FEATHER RNP PARTICLES.

Polysomes were prepared from 14-day embryonic feathers as described in 2.2.3a and dissociated with EDTA (2.2.3c) to release ribonucleoprotein particles. The EDTA-treated polysomes were applied to 11.6 ml linear 10 to 40% w/v sucrose gradients prepared in 0.15 M KCl, 0.01 M Tris-HCl pH 7.4. The gradients were centrifuged for 16 hours at 38,000 rpm at 4°C in a Beckman SW41 rotor, then fractionated using an ISCO model 640 gradient fractionator coupled to a W & W pen recorder. The indicated fractions were collected for further analysis.



A254
II
I
Sedimentation

remove sucrose, and dissociated with SDS prior to sedimentation through linear 10 to 40% (w/v) sucrose gradients (Figure 3.3). RNP fraction I contained 5S ribosomal RNA, tRNAs and major peaks of 9S and 12S RNA. Although the 9S and 12S RNAs were present, they were not well resolved from one another (Figure 3.3a). RNP fraction II contained 5S and 18S ribosomal RNAs as well as peaks of 9S, 12S and 14S RNA, the 12S being the most abundant species (Figure 3.3b). The 14S RNA peak appeared broad and poorly resolved from the 18S peak. These results were in agreement with those obtained by Partington *et al.* (1973), Kemp *et al.* (1974a) and Powell (1979).

Fractions indicated by the bars in Figure 3.3a and b were collected and the two 9S and 12S peaks were separately pooled. The 14S peak and the separately pooled 9 and 12S peaks were ethanol precipitated and subjected to two further cycles of sedimentation through sucrose (Figure 3.4). The 9S, 12S and 14S peaks all appeared sharp and symmetrical, the 12S peak being the broadest of the three. The sucrose gradient sedimentation profile of the purified 12 and 14S RNA superimposed upon that of the 9S RNA (Figure 3.4a) showed that all three were well resolved compared to Figure 3.3b, this was presumably due to stringent size selection during the two further cycles of sedimentation.

3.3.1b TRANSLATION OF RNA IN A CELL-FREE SYSTEM.

The 9, 12 and 14S RNA peaks (Figure 3.4) were collected, ethanol precipitated and translated using a wheat embryo cell-free system, as described in 2.2.5 (Table 3.1). The 12S RNA which was shown by Kemp *et al.* (1974b) to be the mRNA for the β-keratins of embryonic feathers, was found to be the most active RNA fraction in directing the incorporation of ^{3}H-leucine into acid-precipitable protein. 9S RNA was also active, but only at about half the level of 12S RNA. 14S RNA was poorly active, if at all, only stimulating the wheat embryo system two fold when approximately 3 μg of RNA were added and, as expected, the 18S RNA was not

FIGURE 3.3

SUCROSE GRADIENT SEDIMENTATION OF SDS-DISSOCIATED RNP PARTICLES.

The two feather RNP fractions prepared and collected as described in the legend to Figure 3.2 were dissolved in 0.1 M Tris-HCl pH 9.0, 0.1% SDS and heated at 65°C for 10 min then loaded onto 11.6 ml linear 10 to 40% sucrose gradients prepared in NET. The gradients were centrifuged and fractionated as described in the legend to Figure 3.2. The fractions indicated by the bars were collected and ethanol-precipitated.

a. RNP fraction I.

b. RNP fraction II.

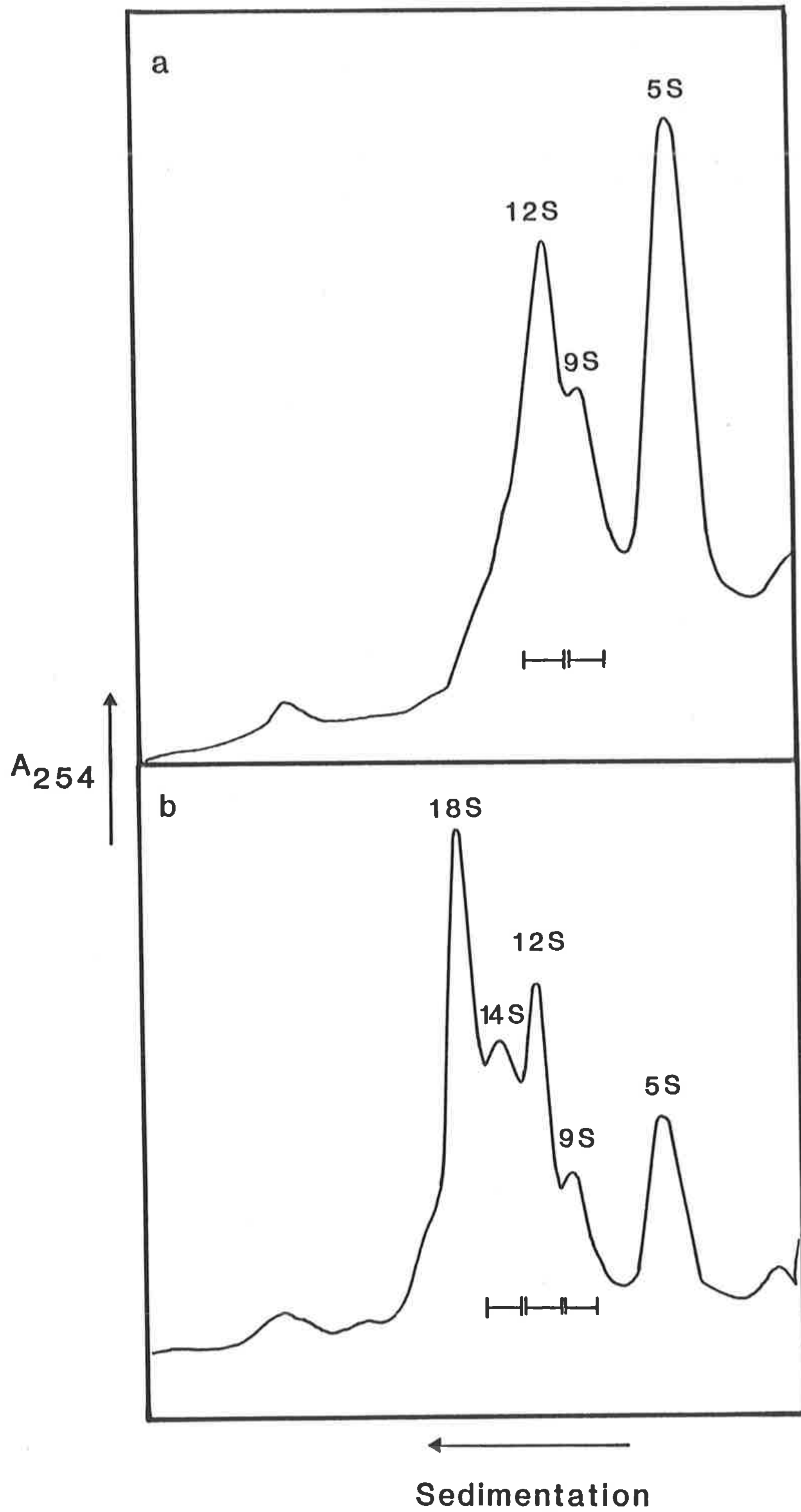

a
5S
12S
9S
A254
b
18S
12S
14S
5S
9S
Sedimentation

FIGURE 3.4

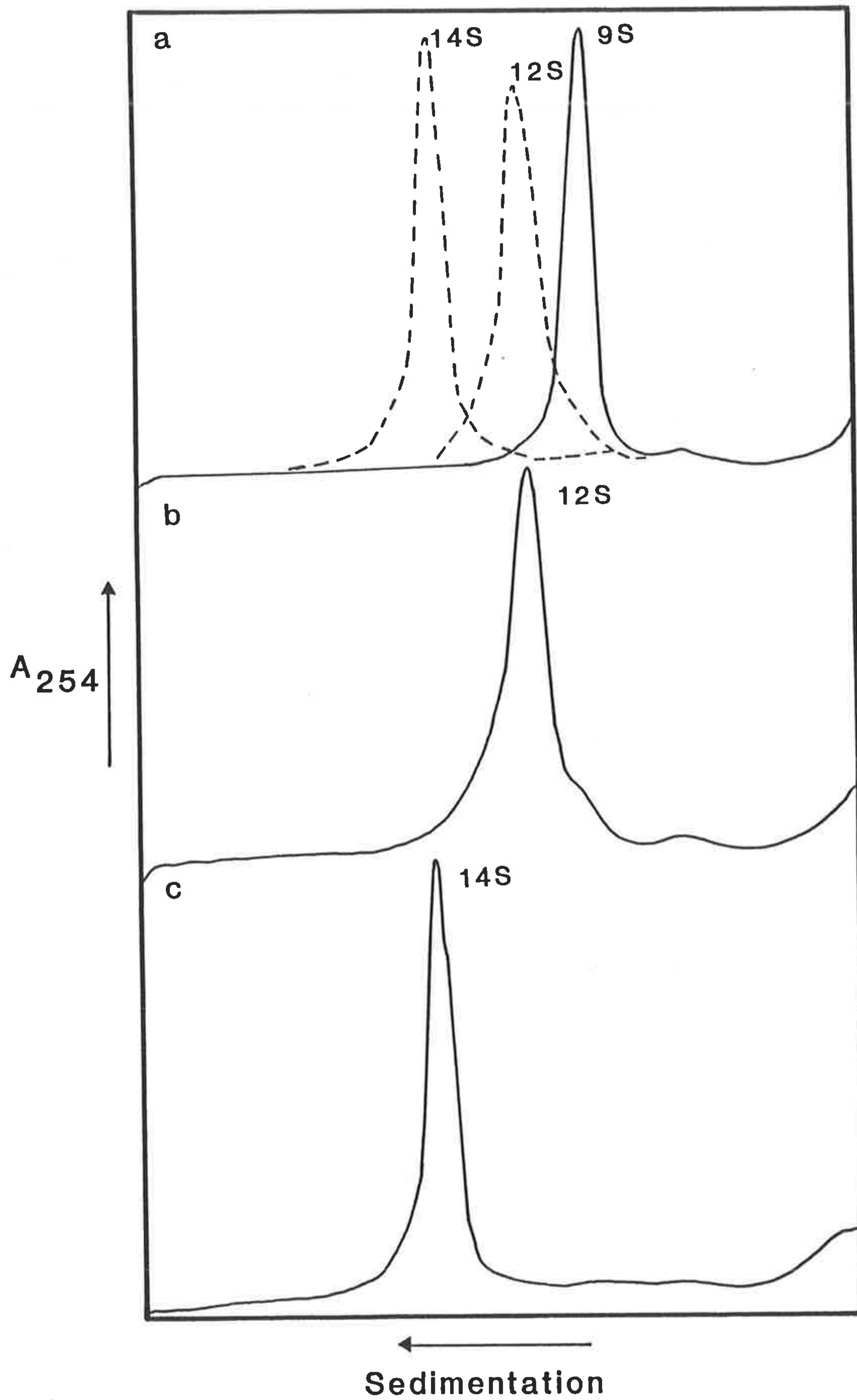
PURIFICATION OF RNA FRACTIONS BY SUCROSE GRADIENT SEDIMENTATION.

RNA fractions, collected as indicated by the bars in Figure 3.3, were purified by two further cycles of sedimentation through sucrose. Separately pooled RNA fractions were dissolved in 0.3 ml of NET and sedimented through sucrose gradients and fractionated as described in the legend to Figure 3.3.

a. 9S RNA fraction with 12S and 14S RNA profiles (- - -) superimposed.

b. 12S RNA fraction.

c. 14S RNA fraction.



a
14S
12S
9S
b
12S
c
14S
A254
Sedimentation

TABLE 3.1

INCORPORATION OF ^{3}H-LEUCINE INTO PROTEIN SYNTHESIZED BY 9, 12, 14, and 18S RNA PURIFIED FROM RNP PARTICLES.

Purified 9, 12, 14, and 18S RNA fractions prepared as described in the legend to Figure 3.4 were translated in a wheat embryo cell-free system as described in 2.2.5 and the amount of radioactivity incorporated into protein was determined by TCA precipitation of duplicate aliquots of the samples on Whatman GF/A discs by the procedure of Bollum (1968).

* Samples used for further analysis on pH 2.7 gels (Figures 3.5 and 3.6).

TABLE 3.1

INCORPORATION OF ^{3}H-LEUCINE INTO PROTEIN SYNTHESIZED BY 9, 12 AND 14S RNA PURIFIED FROM RNP PARTICLES.

RNA	µg RNA per translation	Total c.p.m.	c.p.m. minus background	c.p.m. per µg RNA
No RNA*	—	57,000	—	—
9S	0.56	102,500	45,000	80,500
9S*	2.8	263,000	205,500	73,500
12S	0.16	87,000	29,500	184,500
12S	0.5	120,000	62,500	125,000
12S*	1.0	223,000	165,500	165,500
14S	0.56	67,500	10,000	18,000
14S*	2.8	100,000	42,500	15,000
18S*	3.1	75,000	18,000	6,000

* Samples used for further analysis on pH 2.7 gels.

at all active in stimulating the wheat embryo system. In all cases the amount of radioactivity incorporated increased as the amount of added RNA increased.

3.3.1c ANALYSIS OF TRANSLATION PRODUCTS BY ELECTROPHORESIS AT pH 2.7.

The S-carboxymethylated proteins of embryonic feather can be efficiently resolved by electrophoresis on polyacrylamide gels at pH 2.7 (Walker and Rogers, 1976a), the β-keratins being resolved into about 7 bands which are particularly well separated from fast protein. S-carboxymethylated translation products of embryonic feather RNA, using wheat embryo and rabbit reticulocyte cell-free systems, have previously been analysed by electrophoresis on polyacrylamide gels at pH 2.7 (Partington *et al.*, 1973; Powell, 1979). The translation products used in these studies were S-carboxymethylated since this was found necessary to improve the solubility of the keratin proteins which were used as standards (Harrap and Woods, 1964a; Kemp and Rogers, 1972). However, Powell (1979) reported loss of incorporated radioactivity during the manipulations required for S-carboxymethylation. Analysis of products was therefore carried out directly by adding an equal volume of two-times loading solution to an aliquot of the translation mix and applying this directly to the gel prior to electrophoresis.

Figure 3.5 shows the fractionation of about 150 μg of reduced, but not S-carboxymethylated 21-day embryonic feather protein by polyacrylamide gel electrophoresis at pH 2.7, as described in 2.2.7. The positions of the seven major β-keratin bands and fast protein are marked by the horizontal bars. The profile of radioactivity incorporated into protein by the wheat embryo system, with no RNA added, is superimposed upon this profile. No obvious bands were present in the endogenous wheat embryo translation products. Figure 3.6 shows the profile of radioactivity obtained when the 9, 12, 14 and 18S translation products were fractionated on pH 2.7 gels. As expected, the 12S RNA translated to give several bands

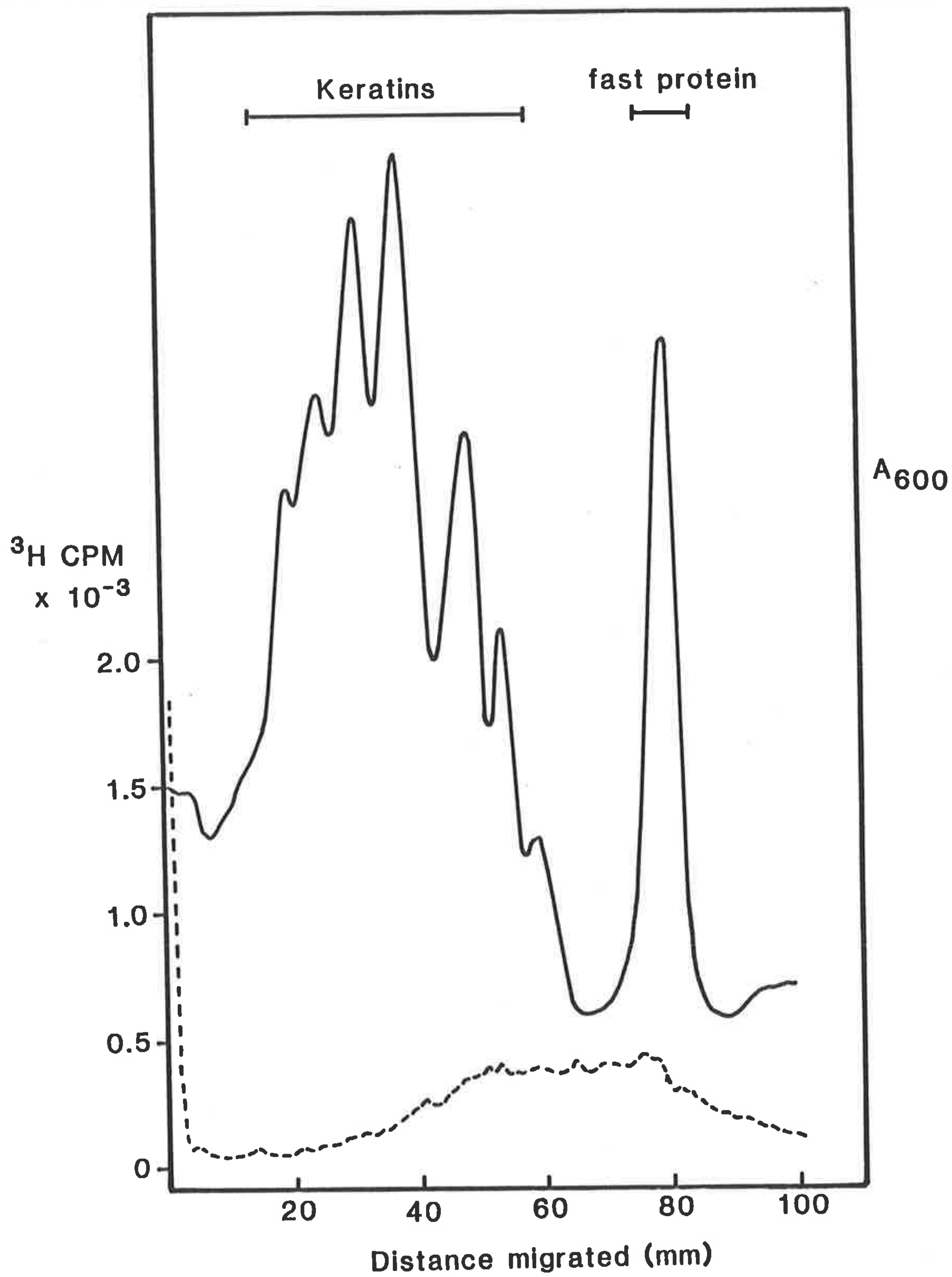
FIGURE 3.5

ELECTROPHORETIC ANALYSIS AT pH 2.7 OF REDUCED FEATHER PROTEIN AND ENDOGENOUS WHEAT EMBRYO TRANSLATION PRODUCTS.

Reduced but not S-carboxymethylated 21-day embryonic feather protein (about 150 μg) was electrophoresed on a pH 2.7 polyacrylamide gel as described in 2.2.7. The position of the major β-keratin bands and fast protein are marked by the horizontal bars. The profile of radioactivity incorporated into protein by the wheat embryo system with no added RNA is superimposed upon this profile. Directly after the translation assay an aliquot of the "no RNA" reaction mix, from Table 3.1, was loaded directly onto the gel without prior carboxymethylation. After electrophoresis radioactivity in the gel was determined as described in 2.2.6.

Reduced feather protein; ________________

Endogenous wheat embryo products; – – – – – – – – – –



Keratins
fast protein
A600
3H CPM
x 10-3
2.0
1.5
1.0
0.5
0
20
40
60
80
100
Distance migrated (mm)

FIGURE 3.6

ELECTROPHORETIC ANALYSIS AT pH 2.7 OF 9, 12, 14, AND 18S RNA TRANSLATION PRODUCTS.

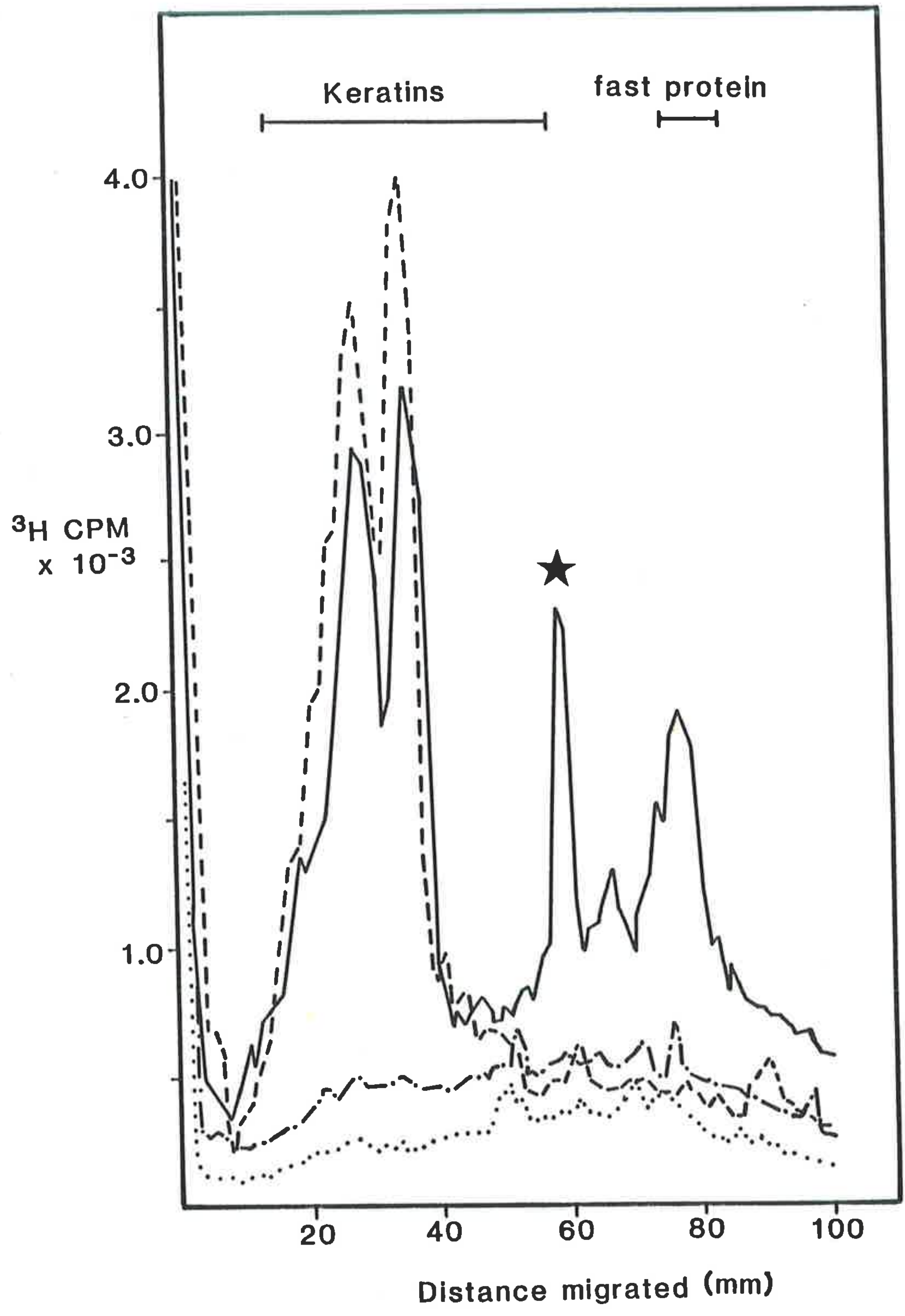
Gel electrophoresis at pH 2.7 of translation products of the RNA fractions prepared and collected as shown in Figures 3.3 and 3.4 and translated as described in Table 3.1. Translation products were analysed on pH 2.7 gels as described in Figure 3.5. The position of the fast protein and keratin peaks are indicated by the bars, as determined from a parallel gel containing reduced feather protein. A peak has been marked with a star to distinguish it from the other peaks.

9S RNA translation products; ————————

12S RNA translation products; — — — — — — — — — —

14S RNA translation products; —·—·—·—·—·—·—·

18S RNA translation products; ....



Keratins
fast protein
4.0
3.0
2.0
1.0
^{3}H CPM x 10^{-3}
20
40
60
80
100
Distance migrated (mm)

migrating with keratin and very little if any fast protein. The 9S RNA also gave several keratin bands and several bands which migrated with a higher mobility than the keratins. There was some radioactivity migrating near fast protein, but no distinct single peak was observed. The peak marked with a star (*) appeared to migrate with the seventh keratin band. The 14S and 18S RNA translation products appeared indistinguishable from the wheat embryo endogenous products, except perhaps for a small amount of keratin in the 14S RNA products.

It can be concluded from *in vitro* translation of the RNAs prepared from dissociated ribonucleoprotein particles, that the 12S RNA is the mRNA for β-keratins, the 9S RNA is probably partially degraded β-keratin mRNA, possibly containing some contaminating fast protein mRNA and the 14S RNA contains no significant mRNA activity. This may have been due to the nature of the RNA preparation procedure. Selection of ribonucleoprotein particles may have excluded some mRNA activity perhaps because some ribonucleoprotein particles are unstable, or the ribonucleoprotein complex may have sedimented in an area of the gradient not collected (Figure 3.2). For this reason it was decided to prepare RNA by a method not requiring ribonucleoprotein particles as intermediates and which minimized the chance of ribonuclease degradation of RNA.

3.3.1d PURIFICATION OF RNA FROM SDS DISSOCIATED POLYSOMES.

Polysomal RNA was prepared from 14-day embryonic feather polysomes, prepared as described in 2.2.3a except that 0.1% diethylpyrocarbonate was included in the homogenization buffer, by dissociation in a buffer containing 1.0% SDS. After sedimentation through linear sucrose gradients containing 0.2% SDS, fractions A, B and C, as indicated by the bars in Figure 3.7, were collected, ethanol precipitated and resedimented on sucrose gradients as shown in Figure 3.8. Figure 3.8a shows the presence of one symmetrical peak sedimenting at about 9S, while Figure 3.8b shows a broader peak of 12S contaminated by a small amount of 9S RNA. Figure

FIGURE 3.7

SUCROSE GRADIENT SEDIMENTATION OF SDS DISSOCIATED FEATHER POLYSOMES.

Polysomes were prepared from 14-day embryonic feathers as described in 2.2.3a, except that 0.1% diethylpyrocarbonate was included in the homogenization buffer, and dissociated by resuspension in NET containing 1.0% SDS to release RNA. The SDS dissociated polysomes were sedimented through linear 5 to 20% sucrose gradients prepared in NET, 0.2% SDS. Gradients were centrifuged and fractionated as described in Figure 3.2. Fractions A, B and C were collected, as indicated by the bars.

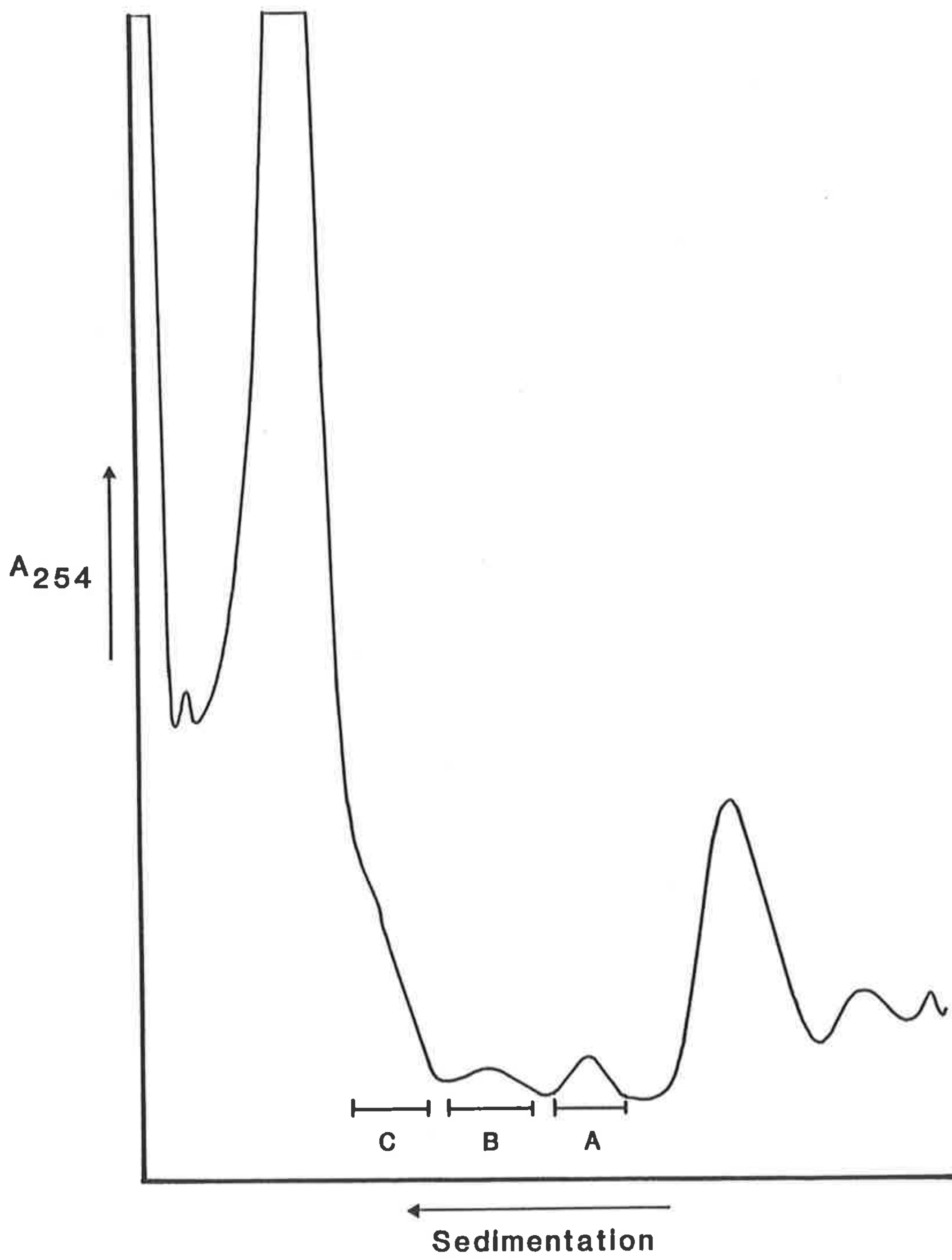

A254
C
B
A
Sedimentation

FIGURE 3.8

SUCROSE GRADIENT SEDIMENTATION OF SDS DISSOCIATED FEATHER POLYSOMAL RNA FRACTIONS.

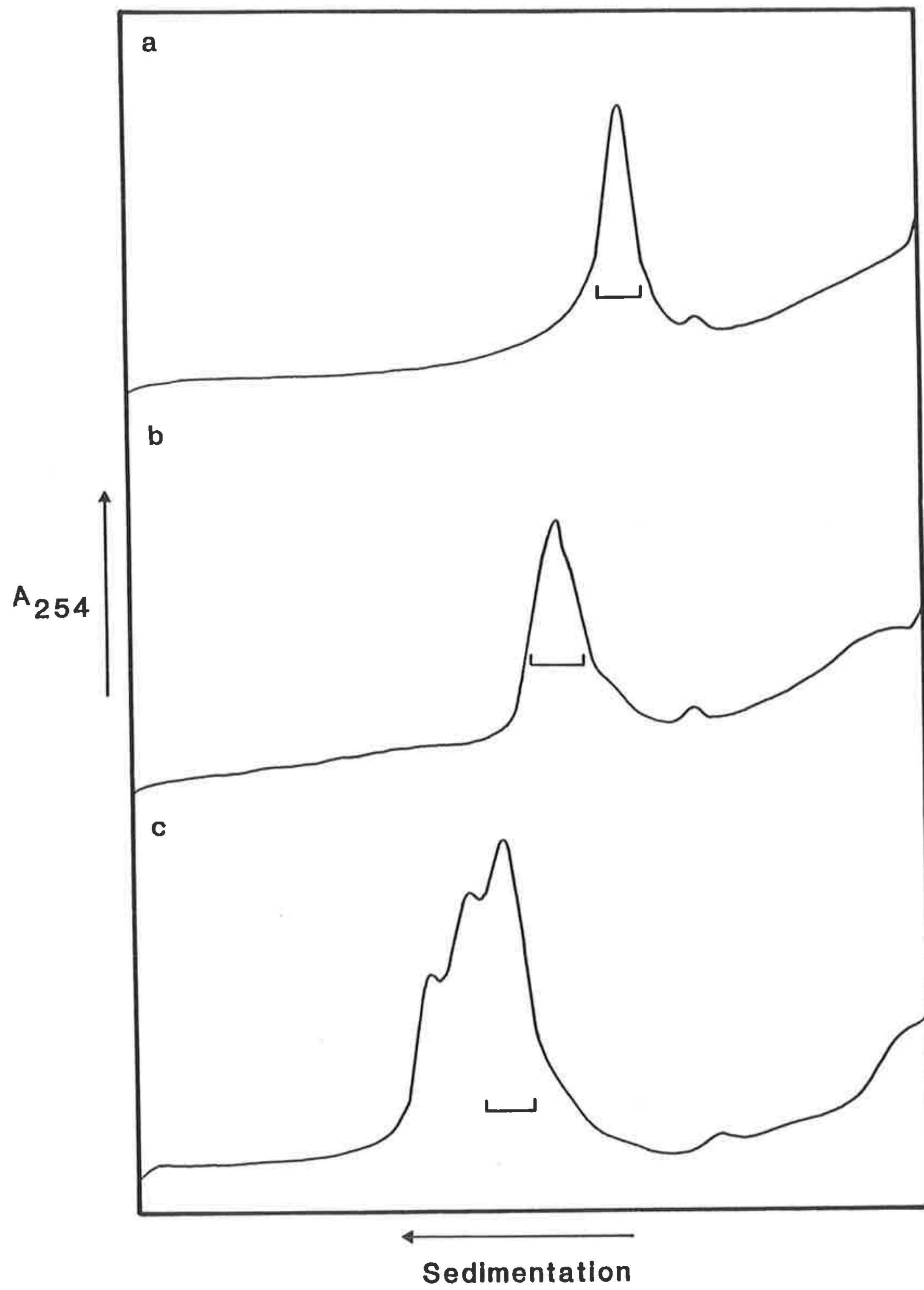
RNA fractions from SDS dissociated polysomes were collected as indicated by the bars in Figure 3.7 and resedimented through linear 10 to 40% sucrose gradients prepared in NET. Gradients were centrifuged and fractionated as described in Figure 3.2. The fractions indicated by the bars were collected and further purified by resedimentation through sucrose gradients.

RNA fractions collected in Figure 3.7

a. Fraction A.

b. Fraction B.

c. Fraction C.



a
b
c
A254
Sedimentation

3.8c shows a major peak of 14S RNA heavily contaminated by RNA peaks of 16S and 18S, and a small amount of 12S RNA. Fractions, as indicated by the bars in Figure 3.8, were collected and resedimented on sucrose gradients (14S RNA was subjected to two further cycles of sedimentation, 9 and 12S RNA required only one further cycle) until a single symmetrical peak was obtained (data not shown).

3.3.1e OLIGO(dT) CELLULOSE CHROMATOGRAPHY OF RNA.

Aliquots of 9, 12 and 14S RNA, prepared from SDS dissociated polysomes as described above, were separately applied to an oligo(dT) cellulose column in high salt, washed, and the bound RNA eluted in low salt buffer, as described in 2.2.3d (data not shown). The amount of RNA in the eluate was followed by passage of the eluate through a UV monitor. It appeared that only 10 to 20% of the 12S RNA bound to the oligo(dT) cellulose column, and little if any 9 or 14S RNA bound to the column. In the further analysis of these RNA fractions, oligo(dT) cellulose bound and unbound 12S RNA was used, but 9S and 14S RNAs were used without prior passage through oligo(dT) cellulose. Where 12S unbound RNA is indicated, this is RNA which did not bind after two passages through oligo(dT) cellulose.

3.3.1f TRANSLATION OF RNA IN A CELL-FREE SYSTEM.

The 9S, 14S, 12S bound and 12S unbound RNAs described in 3.3.1e were ethanol precipitated and translated in a wheat embryo cell-free system (Table 3.2). With the exception of 14S RNA all of the RNAs translated were active in directing the incorporation of ^{3}H-leucine into acid-precipitable protein. The 12S RNA which bound to oligo(dT) cellulose was the most active RNA, being three times as active as 9S RNA and five times as active as unbound 12S RNA. 14S RNA was only poorly active in stimulating the wheat embryo cell-free system.

3.3.1g ANALYSIS OF TRANSLATION PRODUCTS BY ELECTROPHORESIS AT pH 2.7.

The translation products of 9S, 14S, 12S bound and 12S unbound RNAs,

TABLE 3.2

INCORPORATION OF ^{3}H-LEUCINE INTO PROTEIN SYNTHESIZED BY 9, 12 AND 14S RNA PURIFIED FROM SDS DISSOCIATED POLYSOMES.

Purified 9, 12 and 14S RNA fractions prepared as described in the legend to Figure 3.8 were translated in a wheat embryo cell-free system as described in 2.2.5 and the amount of radioactivity incorporated into protein was determined as described in the legend to Table 3.1.

TABLE 3.2

INCORPORATION OF ^{3}H-LEUCINE INTO PROTEIN SYNTHESIZED BY 9, 12 AND 14S RNA PURIFIED FROM SDS DISSOCIATED POLYSOMES.

RNA	μg RNA per translation	Total c.p.m.	c.p.m. minus background	c.p.m. per μg RNA
No RNA	—	55,500	—	—
9S	1.5	192,500	137,000	91,000
12S*	0.4	243,500	187,500	469,000
12S§	1.4	257,000	201,500	144,000
14S	1.5	85,500	30,000	20,000

* 12S RNA which bound to oligo(dT) cellulose.

§ 12S RNA which did not bind to oligo(dT) cellulose.

described in 3.3.1e, were fractionated on polyacrylamide gels at pH 2.7, essentially as described in 3.3.1c except that about 100 μg of reduced but not S-carboxymethylated unfractionated 21-day embryonic feather protein was used to determine the mobility of the various protein components. Figure 3.9 shows the pH 2.7 gel profile of radioactivity incorporated into protein in the wheat embryo cell-free system by 9S and 14S RNAs and the endogenous wheat embryo RNA. Superimposed on these profiles is a densitometer scan of a parallel gel containing total 21-day feather protein.

The 14S RNA translation products were almost indistinguishable from the endogenous wheat embryo RNA products, except for a small amount of radioactivity which migrated with the keratin proteins. The 9S RNA directed the synthesis of very little full length feather keratin but labelled products were present which co-migrated with fast protein. Several other slower migrating proteins were present in the 9S products, including a peak which appeared in the translation products of 9S RNA made from EDTA dissociated polysomes (Figure 3.6). This peak has been marked by a star ($\star$), as in Figure 3.6, to distinguish it from other peaks.

Fractionation of the translation products from oligo(dT) cellulose bound 12S RNA on pH 2.7 polyacrylamide gels (Figure 3.10) revealed that this RNA, which should be enriched for poly(A) containing mRNA, was very active in directing the synthesis of β-keratin proteins in the wheat embryo cell-free system. The oligo(dT) cellulose unbound 12S RNA translation products also contained β-keratin proteins (Figure 3.10), indicating that the binding of polyadenylated keratin mRNA to oligo(dT) cellulose was not sufficiently efficient to remove all mRNA activity from the RNA preparation. Both oligo(dT) cellulose bound and unbound 12S RNAs directed the synthesis of some fast protein with the unbound 12S RNA retaining a higher proportion of fast protein mRNA activity than the bound 12S RNA.

FIGURE 3.9

ELECTROPHORETIC ANALYSIS AT pH 2.7 OF REDUCED FEATHER PROTEIN AND 9S, 14S AND ENDOGENOUS WHEAT EMBRYO TRANSLATION PRODUCTS.

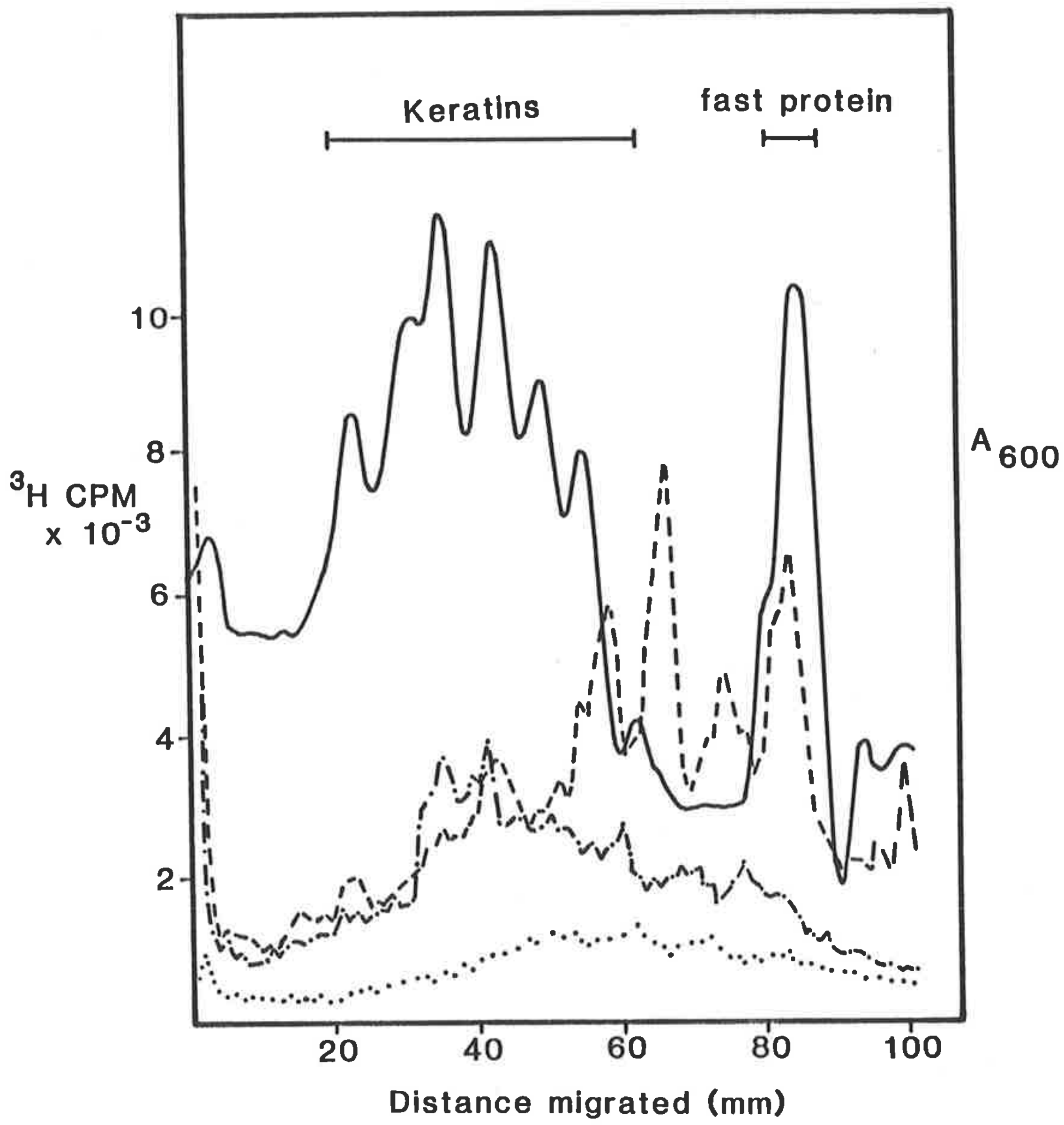
Reduced but not S-carboxymethylated feather protein, used as a mobility standard, was fractionated by electrophoresis at pH 2.7 as described in the legend of Figure 3.5. The position of the major β-keratin bands and fast protein are marked by the horizontal bars. The profiles of radioactivity incorporated into protein by the RNA fractions collected as shown in Figure 3.8 and translated as described in the legend to Table 3.2 are superimposed upon this profile. The translation products were electrophoresed and radioactivity in the gel was detected as described in the legend to Figure 3.5.

Reduced feather protein; ————————————

Endogenous wheat embryo products; ....

9S RNA products; — — — — — — — — — —

14S RNA products; —.—.—.—.—.—.—.



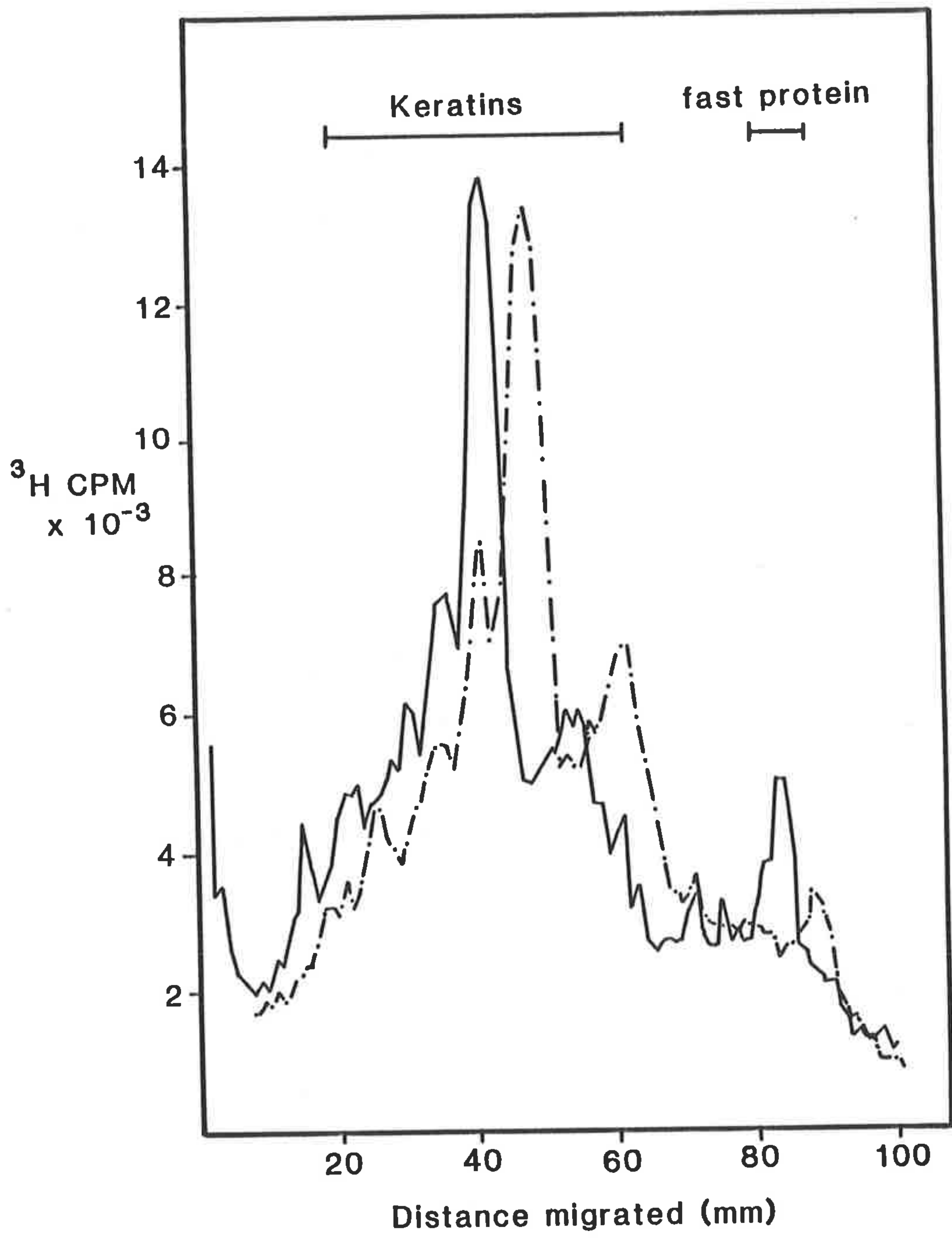
Keratins
fast protein
10
8
6
4
2
³H CPM
x 10⁻³
A 600
20
40
60
80
100
Distance migrated (mm)

FIGURE 3.10

ELECTROPHORETIC ANALYSIS AT pH 2.7 OF OLIGO(dT) CELLULOSE BOUND AND UNBOUND 12S RNA TRANSLATION PRODUCTS.

Gel electrophoresis at pH 2.7 of translation products of oligo(dT) cellulose bound and unbound 12S RNA, prepared as described in 3.3.1e and translated as described in Table 3.2. Translation products were electrophoresed on pH 2.7 gels as described in the legend to Figure 3.5. The position of fast protein and keratin bands are indicated by the bars, as determined from a parallel gel containing reduced feather protein.

Oligo(dT) cellulose RNA fractions:



Keratins
fast protein
14
12
10
8
6
4
2
^{3}H CPM
x 10^{-3}
20
40
60
80
100
Distance migrated (mm)

3.3.1h 14S RNA cDNA SYNTHESIS AND HYBRIDIZATION KINETICS.

Synthesis of DNA complementary to 14S RNA by oligo(dT) priming in the presence of ^{3}H-dCTP, as described in 2.2.14, did not result in the incorporation of radioactivity into a discernible cDNA peak after chromatography of the transcription mix on a Sephadex G-50 column (data not shown). Complementary DNA to 14S RNA was therefore made by use of random priming with salmon sperm oligo-nucleotides by the method of Taylor *et al.* (1976) and Taylor and Tse (1976), as described in 3.2.1. 14S RNA acted as an efficient template using this method, resulting in the incorporation of 2 to 5 x 10^{6} cpm/μg of RNA using ^{3}H-dCTP with a specific activity of 25 Ci/mmole. This rate of incorporation proved similar to that of 18S ribosomal RNA.

It was clear that 14S RNA prepared by the two methods described, see 3.3.1a and 3.3.1d, contained little if any mRNA activity. It seemed likely that at least some of the 14S RNA could be degraded 18 or 28S ribosomal RNA. For this reason, the kinetics of hybridization of 14S cDNA to 14S RNA and 18 and 28S ribosomal RNAs were examined.

The kinetics of hybridization of excess unlabelled RNA to its labelled complementary DNA are dependant on the sequence complexity of the RNA. Comparison of these hybridization kinetics with those of a kinetic standard of known complexity, for example 18S ribosomal RNA allows an estimate of the complexity of the RNA.

Figure 3.11 shows the kinetics of hybridization of 14S cDNA to 14S RNA as determined by the resistance of the hybrids to the single-strand specific nuclease S_1 (3.2.2). The hybridization gave a sharp transition with a mid point ($R_0t\ \frac{1}{2}$) of 5 x 10^{-3} mol.sec.l^{-1}. The kinetics of hybridization of 18S cDNA to 18S RNA (Figure 3.11) shows a sharp transition with a $R_0t\ \frac{1}{2}$ of 3.5 x 10^{-3}. Comparing the rates of hybridization of 14S cDNA *vs.* 14S RNA with 18S cDNA *vs.* 18S RNA, it

FIGURE 3.11

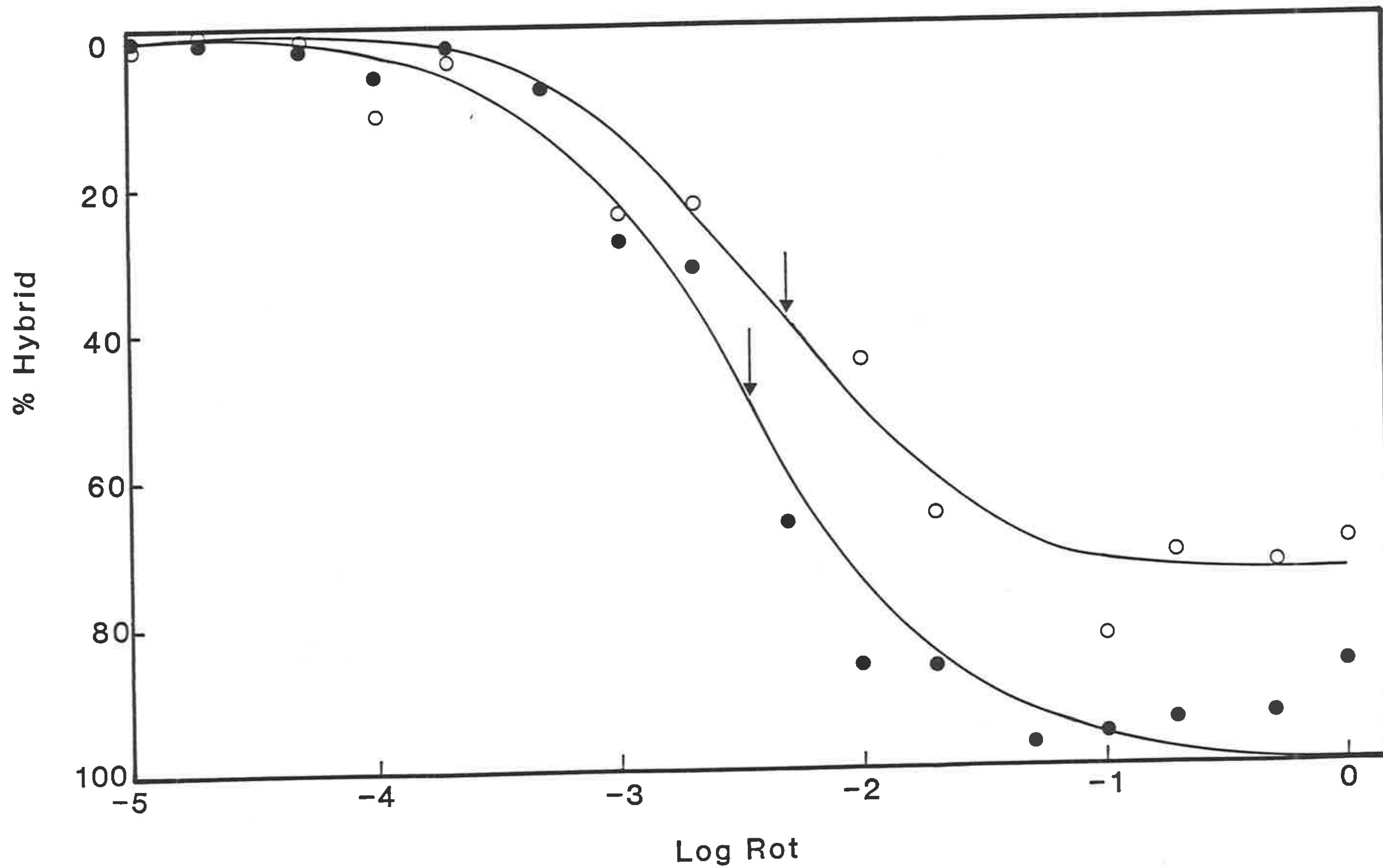
KINETICS OF HYBRIDIZATION OF 14S RNA TO 14S cDNA.

14S RNA, prepared as described in 3.3.1, was hybridized to 14S cDNA prepared by the method of Taylor *et al.* (1976) (3.2.1) and assayed using the single-strand specific nuclease S_1, as described in 3.3.2.

The hybridizations were carried out in RNA excess and the results are expressed as the percentage of input cDNA radioactivity in nuclease S_1 resistant form after hybridization. The hybridization kinetics of 18S ribosomal RNA to 18S cDNA are included as a molecular complexity standard. Arrows indicate the midpoints of the reactions. Curves are the average of two experiments.

14S RNA *vs.* 14S cDNA; ○——○

18S RNA *vs.* 18S cDNA; ●——●


% Hybrid
0
20
40
60
80
100
−5
−4
−3
−2
−1
0
Log Rot

can be calculated that 14S RNA has a complexity of greater than 2,000 nucleotides, which is the approximate molecular size of 18S RNA.

Figure 3.12 shows the kinetics of hybridization of 14S cDNA *vs.* 18S ribosomal RNA, 14S cDNA *vs.* 28S ribosomal RNA and 14S cDNA *vs.* 18S and 28S ribosomal RNAs combined in an equimolar mixture, i.e. mixed in a molar ratio of 2:5. The extent to which the hybridization has gone, i.e. the plateau level of hybridization, should reflect the proportion of 14S RNA which is ribosomal RNA. The 14S cDNA *vs.* 18S ribosomal RNA hybridization plateaued at about 75%, with a $R_0t\ \frac{1}{2}$ of 4×10^{-3}, the 14S cDNA *vs.* 28S ribosomal RNA hybridization plateaued at about 45%, with a $R_0t\ \frac{1}{2}$ of 10^{-2} and the 14S cDNA *vs.* 18S and 28S ribosomal RNAs went almost to completion, with a $R_0t\ \frac{1}{2}$ of 6×10^{-3}. The extent to which these hybridizations have gone to completion indicate that 14S RNA is composed almost entirely of a mixture of 18S and 28S ribosomal RNAs, with the mixture containing somewhat more 18S than 28S RNA. The $R_0t\ \frac{1}{2}$ values support this conclusion since the 14S cDNA *vs.* 18S RNA hybridization proceeded about 2.5 times faster than the 14S cDNA *vs.* 28S RNA hybridization, reflecting the relative sizes of 18S and 28S RNAs, and the 14S cDNA *vs.* 18S plus 28S RNA hybridization gave a $R_0t\ \frac{1}{2}$ between the two but somewhat closer to the 14S cDNA *vs.* 18S RNA hybridization than the 14S cDNA *vs.* 28S RNA hybridization.

3.3.2 DETAILED STUDIES OF 12S KERATIN mRNA.

3.3.2a 12S mRNA POLY(A) LENGTH.

The ^{32}P-labelling of embryonic chick feather RNA *in vivo* or in cultured feathers produces RNA of extremely low specific activity (G. A. Partington, unpublished observations) which does not allow analysis of the 3′-terminal polyadenylate tract size distribution. This is contrary to what has been reported for numerous other eukaryotic RNA molecules (Sanger *et al.*, 1965; Barrell, 1971; Shatkin, 1976). For this reason the more indirect method of hybridization to ^{3}H-poly(U) was used to

FIGURE 3.12

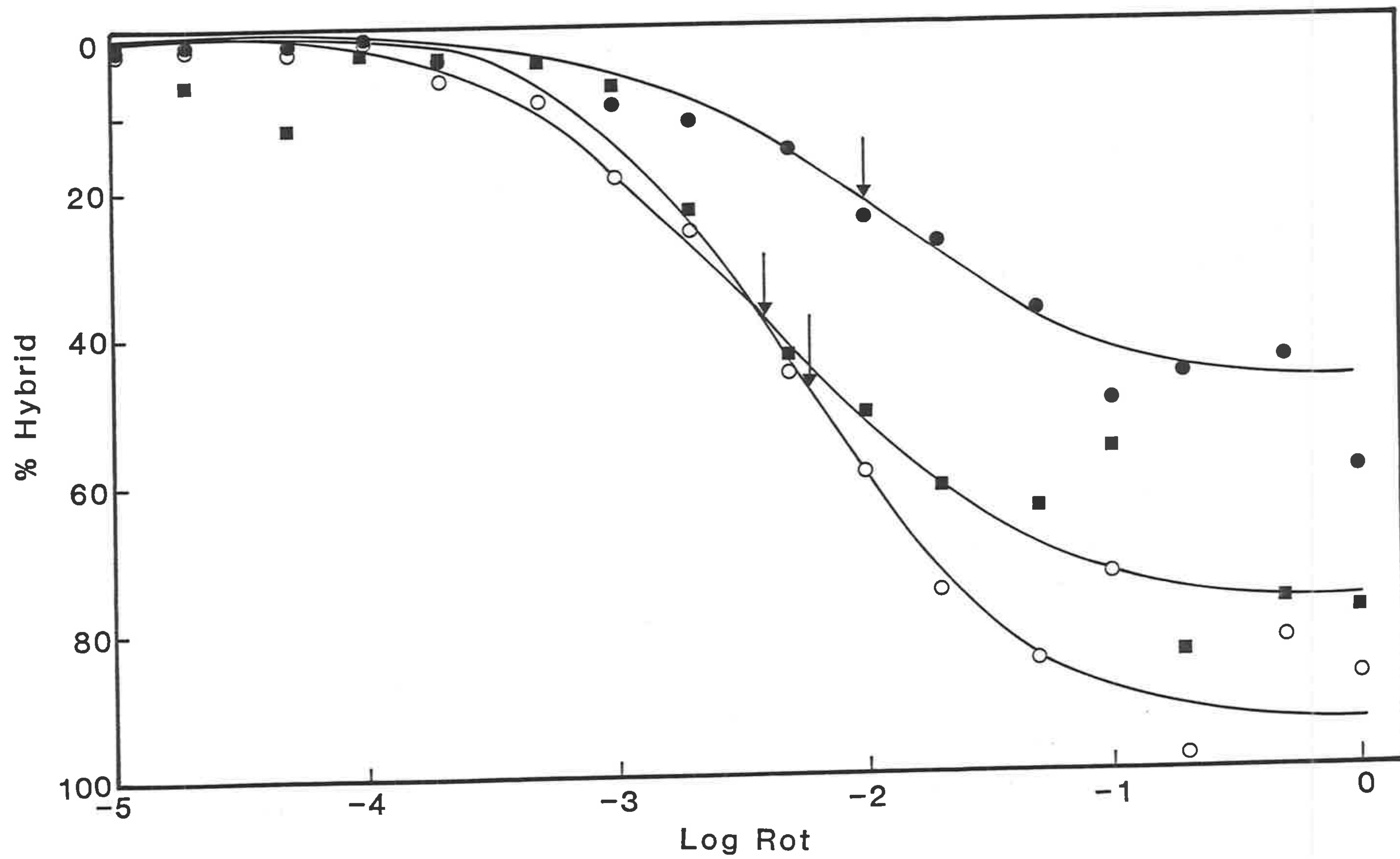
HYBRIDIZATION KINETICS OF 18S, 28S AND AN EQUIMOLAR MIXTURE OF 18 AND 28S RIBOSOMAL RNA TO 14S cDNA.

Using RNA in excess, 14S cDNA, prepared as described in the legend to Figure 3.11, was hybridized to mixtures of ribosomal RNAs. The kinetics of hybridization were determined by using nuclease S_1 to assay the amount of added radioactivity in cDNA found in hybrid form. All other conditions are as described in the legend to Figure 3.11.

18S RNA *vs*. 14S cDNA; ■——■

28S RNA *vs*. 14S cDNA; ●——●

18S and 28S RNA (equimolar ratio) *vs*. 14S cDNA; ○——○



% Hybrid
0
20
40
60
80
100
-5
-4
-3
-2
-1
0
Log Rot

detect poly(A) tracts in mRNA. Ribonucleases A and T_1 were used to release poly(A) from mRNA by reducing the remainder of the mRNA to nucleotides and small oligonucleotides (Adesnik and Darnell, 1972). This poly(A) was then sized by electrophoresis on denaturing polyacrylamide gels. Fractionated poly(A) was then detected by examining the profile of acid-precipitable radioactivity after ribonuclease A digestion of hybrids formed between ^{3}H-poly(U) and poly(A) eluted from gel slices, as described by Bishop *et al.* (1972) and in 3.2.7. This method proved unreliable for the detection of poly(A) released from keratin mRNA (data not shown). A simpler method was therefore adopted which indirectly sized the poly(A) tracts released from keratin mRNA.

Keratin mRNA was hybridized to ^{3}H-poly(U) using a ten-fold excess of poly(U) over the likely poly(A) content (approximately 100 nucleotides out of a total of 800 nucleotides), then digested with ribonucleases A and T_1 as described in 3.2.6 to release ^{3}H-poly(U)/poly(A) hybrids and digest any poly(U) not in hybrid form. These hybrids were then ethanol precipitated, taken up and denatured in a formamide loading buffer prior to fractionation on a 98% formamide polyacrylamide gel. Figure 3.13a shows the profile of radioactivity eluted from successive pairs of 1 mm slices across such a gel. The two sharp peaks migrating near the tracker dye were also present in a parallel gel on which an identical sample was run except that no mRNA was included in the initial hybridization step (Figure 3.13b). This indicates that these peaks were due to nucleotides and small oligonucleotides which were produced during the ribonuclease digestion of excess ^{3}H-poly(U). The ^{3}H-poly(U) fragments protected from digestion by the ribonucleases, due to the formation of ribonuclease resistant hybrids with poly(A) tracts in keratin mRNA, ranged in length from 40 to 180 nucleotides with a mean length of 86 nucleotides (Figure 3.13a). Since the ^{3}H-poly(U) was uniformly labelled, longer poly(A) fragments contained proportionately more radioactivity and therefore contributed more to the calculated mean length than did smaller poly(A) fragments. The calculated mean length of 86 nucleotides is therefore a weight-average length and not a number-average length.

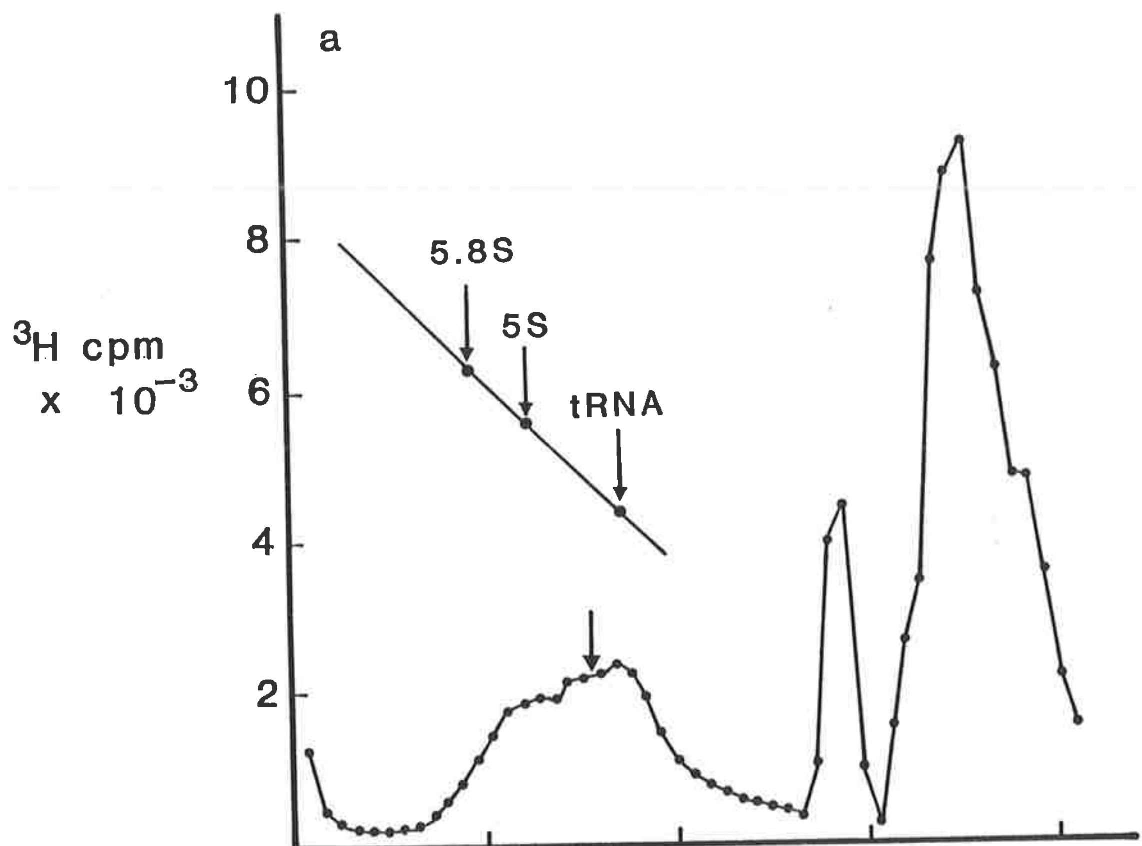
FIGURE 3.13

KERATIN mRNA poly(A) LENGTH AS DETERMINED BY HYBRIDIZATION TO ^{3}H-poly(U).

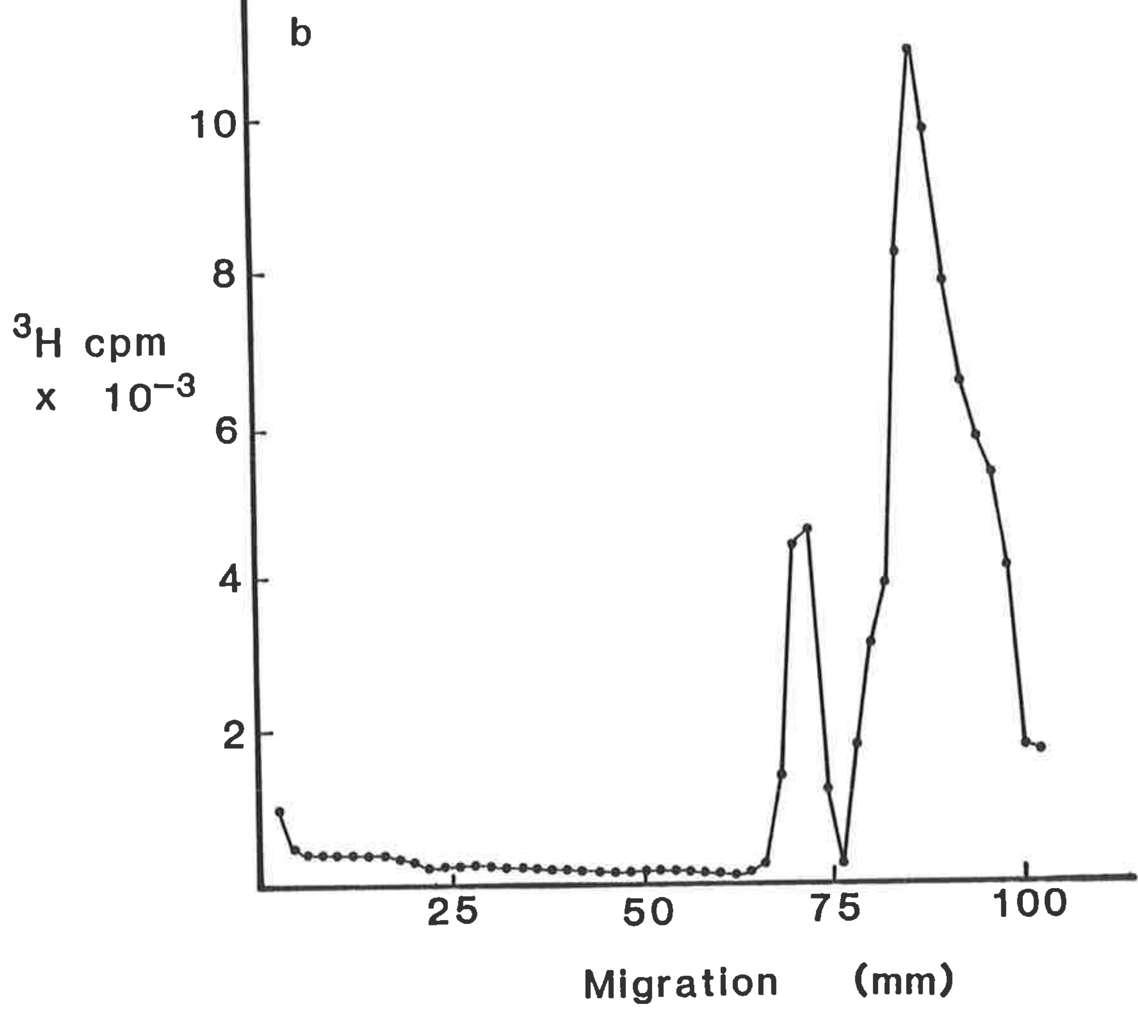
Keratin mRNA was hybridized to ^{3}H-poly(U) using a 10 fold molar excess of poly(U) over the likely poly(A) content (approximately 100 residues out of 800), then digested with ribonucleases A and T_1 as described in 3.2.6 to release ^{3}H-poly(U)/poly(A) hybrids and digest any ^{3}H-poly(U) not in hybrid form. These hybrids were ethanol precipitated, then dissolved and denatured in formamide loading buffer prior to fractionation on a 98% formamide 10% polyacrylamide gel as described in 3.2.3a. The profile of radioactivity was determined by eluting counts from successive pairs of 1 mm slices across the gel. The molecular weight of poly(A) tracts was determined by comparison with a parallel gel containing 5.8S, 5S and tRNA as molecular weight markers. The arrow indicates the weight-average poly(A) length.

a. Keratin mRNA hybridized to ^{3}H-poly(U).

b. As in a. but with no mRNA added.



a
10
8
6
4
2
3H cpm
x 10-3
5.8S
5S
tRNA



b
10
8
6
4
2
3H cpm
x 10-3
25
50
75
100
Migration (mm)

As mentioned earlier, this method of poly(A) tract length determination is indirect since it is the length of poly(U) protected by hybridization to poly(A) which is measured, but it requires much less mRNA than the direct detection of poly(A) by gel staining. The direct method requires large amounts of mRNA but it was considered essential that the previous result be confirmed.

The length of poly(A) produced by ribonuclease A and T_1 digestion of sufficient mRNA to be detected on a gel by staining was determined by electrophoresis on a polyacrylamide gel in 8 M urea as described by Sanger and Coulson (1975) and in 3.2.3b. Figure 3.14 shows the distribution of lengths so obtained. This method revealed a range of poly(A) lengths from 45 to 165 nucleotides with a mean length of 85 nucleotides. Again, this measurement of mean poly(A) length is a weight-average length since longer molecules of poly(A) bind proportionately more stain than do shorter poly(A) molecules. These results are in close agreement with the previous, less direct, method of poly(A) length determination.

There is probably greater biological significance in describing a poly(A) length distribution in terms of number-average rather than a weight-average length, i.e. the average length of poly(A) tract on each molecule of mRNA without allowing any greater weighting to mRNA molecules which have long poly(A) tracts. A value for the number-average length has been calculated from the profile of poly(A) lengths shown in Figure 3.15. This value was achieved by reaction of mRNA with ^{3}H-KBH_4, as described in 3.2.5, which only labels the 3′-terminal adenosine of the poly(A) tracts. Labelled mRNA was subsequently digested with ribonucleases A and T_1 to release ^{3}H-poly(A) labelled at its 3′-terminus which was fractionated on a denaturing polyacrylamide gel, as previously described. Since each poly(A) molecule is given equal weighting, i.e. the number of radioactive counts in poly(A) is the same, regardless of its length, the profile obtained lends itself to the calculation of a number-average length for the poly(A) sequences. As expected the number-average length of 65 nucleotides is smaller than the weight-average length of 85 nucleotides.

FIGURE 3.14

DIRECT DETERMINATION OF KERATIN mRNA POLY(A) LENGTH BY STAINING.

Keratin mRNA was digested with ribonucleases A and T_1, as described in 3.2.6., to release poly(A) tracts which were ethanol precipitated and dissolved and denatured in a formamide loading buffer prior to electrophoresis on a 10% polyacrylamide gel in 8 M urea as described in 3.2.3b. After electrophoresis the gel was stained in 0.05% toluidine blue for 30 mins, destained and scanned at 600 nm in a Gilford linear transport attachment coupled to a W & W chart recorder. The mobilities of 4S tRNA, 5S and 5.8S rRNAs were determined on a parallel gel.

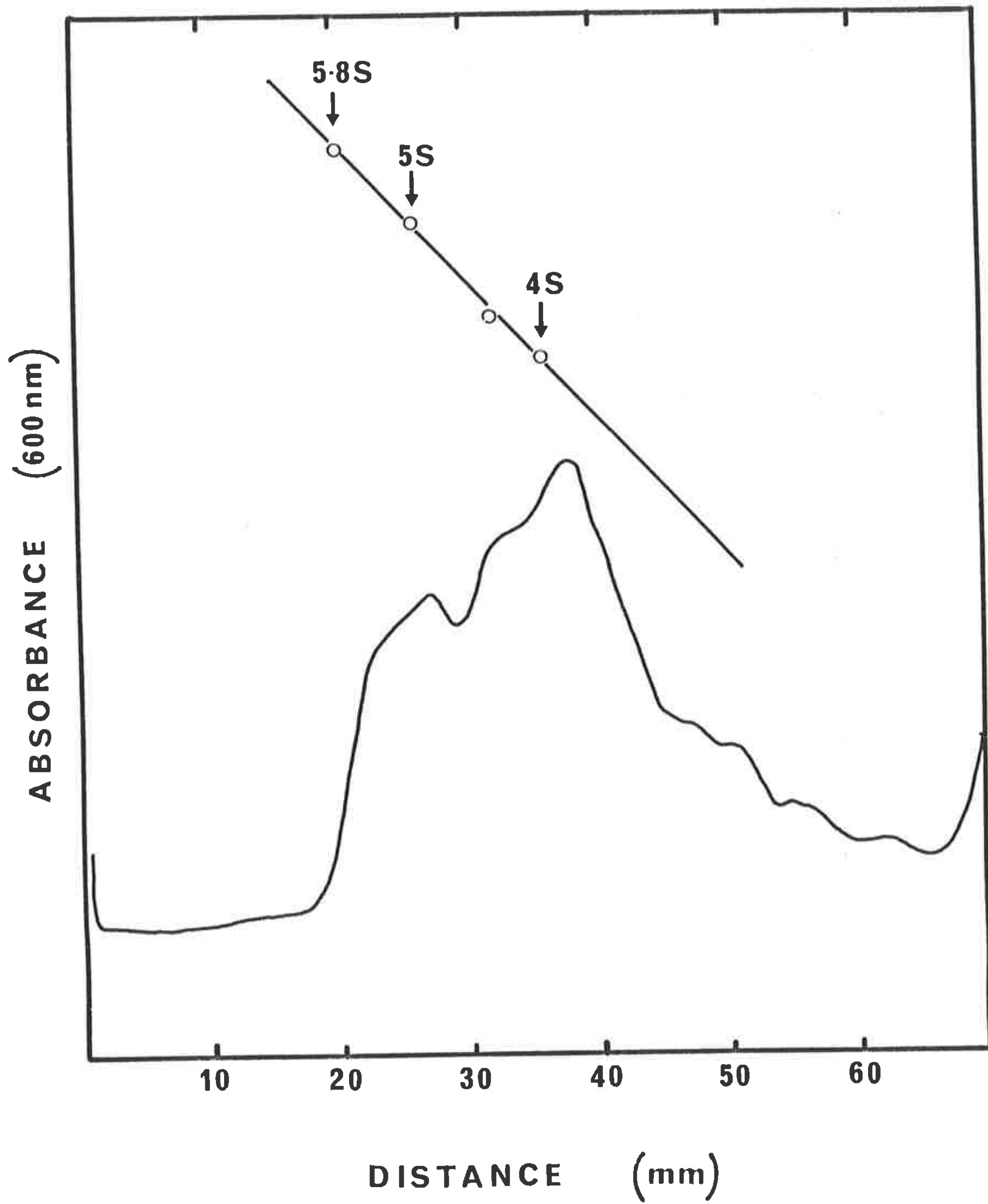
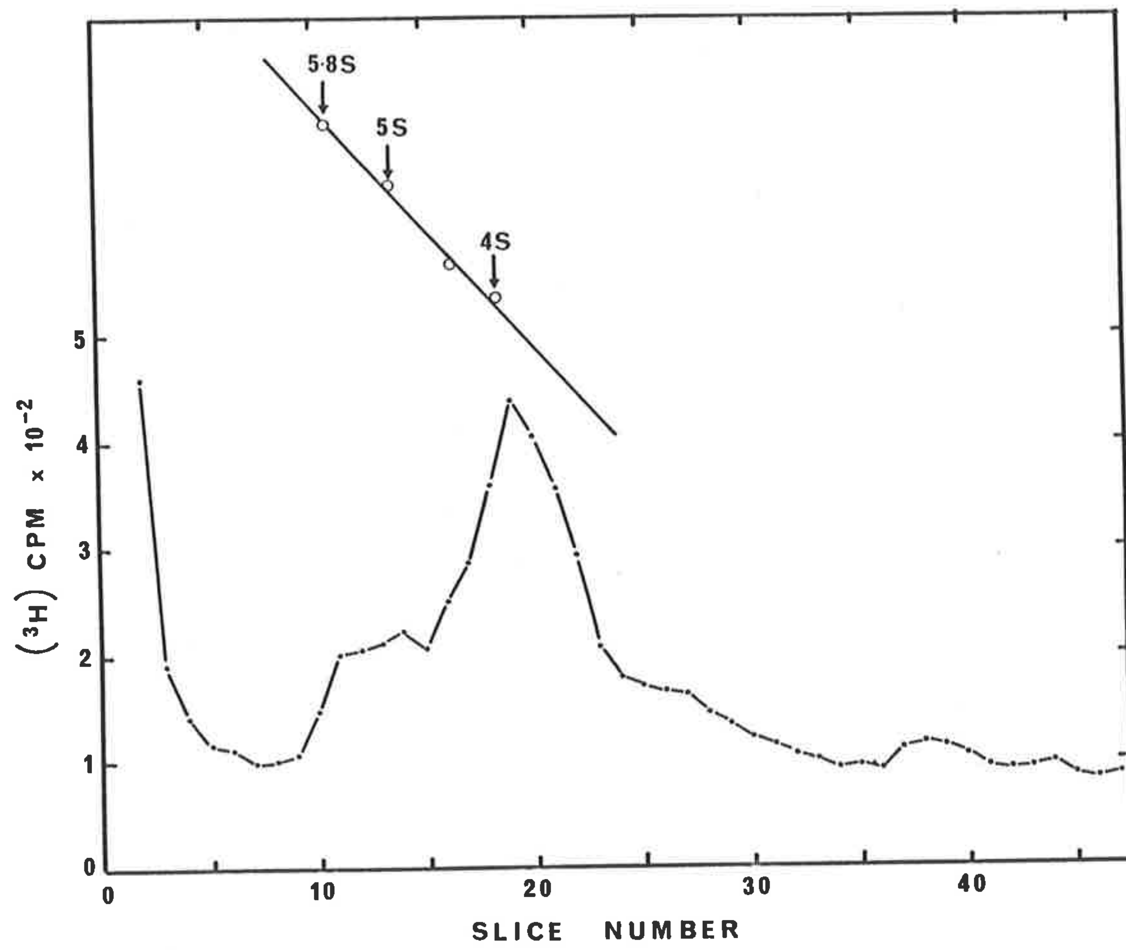

5·8S
5S
4S
ABSORBANCE (600 nm)
10
20
30
40
50
60
DISTANCE (mm)

FIGURE 3.15

DETERMINATION OF poly(A) LENGTH USING KERATIN mRNA LABELLED WITH ^{3}H-KBH$_4$.

Keratin mRNA was labelled at its termini by $NaIO_4$ oxidation of the 2′,3′-hydroxyls and their reduction with ^{3}H-KHB$_4$ as described in 3.2.5. Labelled mRNA was then digested with ribonucleases A and T_1, as described in 3.2.6, to release labelled poly(A) sequences. The poly(A) sequences were then electrophoresed on an 8 M polyacrylamide gel, as described in the legend to Figure 3.14 and the position of radioactively labelled poly(A) sequences was determined by counting samples eluted from successive pairs of 1 mm slices as described in the legend to Figure 3.13. The mobilities of 4S tRNA, 5S and 5.8S ribosomal RNA markers was determined on a parallel gel.


5·8S
5S
4S
(³H) CPM × 10⁻²
0
1
2
3
4
5
0
10
20
30
40
SLICE NUMBER

but the profile of poly(A) lengths shown in Figure 3.15 is almost identical to that shown in Figures 3.13a and 3.14.

3.3.2b ATTEMPTS TO OBTAIN STRUCTURAL INFORMATION ON KERATIN mRNA BY RESTRICTION OF KERATIN cDNA.

It is recognized that restriction endonucleases will cut single-stranded DNA when used in vast excess (Horiuchi and Zinder, 1975), in particular Hae III, Hha I and Hpa II. Since these enzymes are sequence specific and should be capable of cleaving single-stranded cDNA, it was thought likely that their use in the characterization of keratin cDNA would reveal aspects of sequence complexity and possibly sequence organization in keratin mRNA. It was thought unlikely that the direct isolation of cDNA fragments for sequencing by the chemical cleavage method (Maxam and Gilbert, 1977) would prove successful, since the mRNA is composed of many species making it difficult to obtain any one species in pure form or in sufficient quantity.

Hae III has a four base recognition sequence for double-stranded DNA and as such should on average cleave once every 250 base pairs. If this holds true for single-stranded keratin cDNA, one would expect to generate three to four fragments from the 800 nucleotide long full length keratin cDNA. If, as estimated by Kemp (1975), there are 25 to 35 different mRNA molecules, Hae III digestion of keratin cDNA should generate 75 to 140 fragments, provided that all of the mRNA species are unique.

Figure 3.16 shows an autoradiograph of a 7.5% polyacrylamide gel on which ^{32}P-labelled, oligo(dT) primed, keratin cDNA (synthesized as described in 2.2.14) digested with various amounts of Hae III, has been fractionated. It is clear that between 500 and 1,000 fold excess of restriction enzyme was required to completely digest keratin cDNA. There were obvious bands clearly resolved from the smear

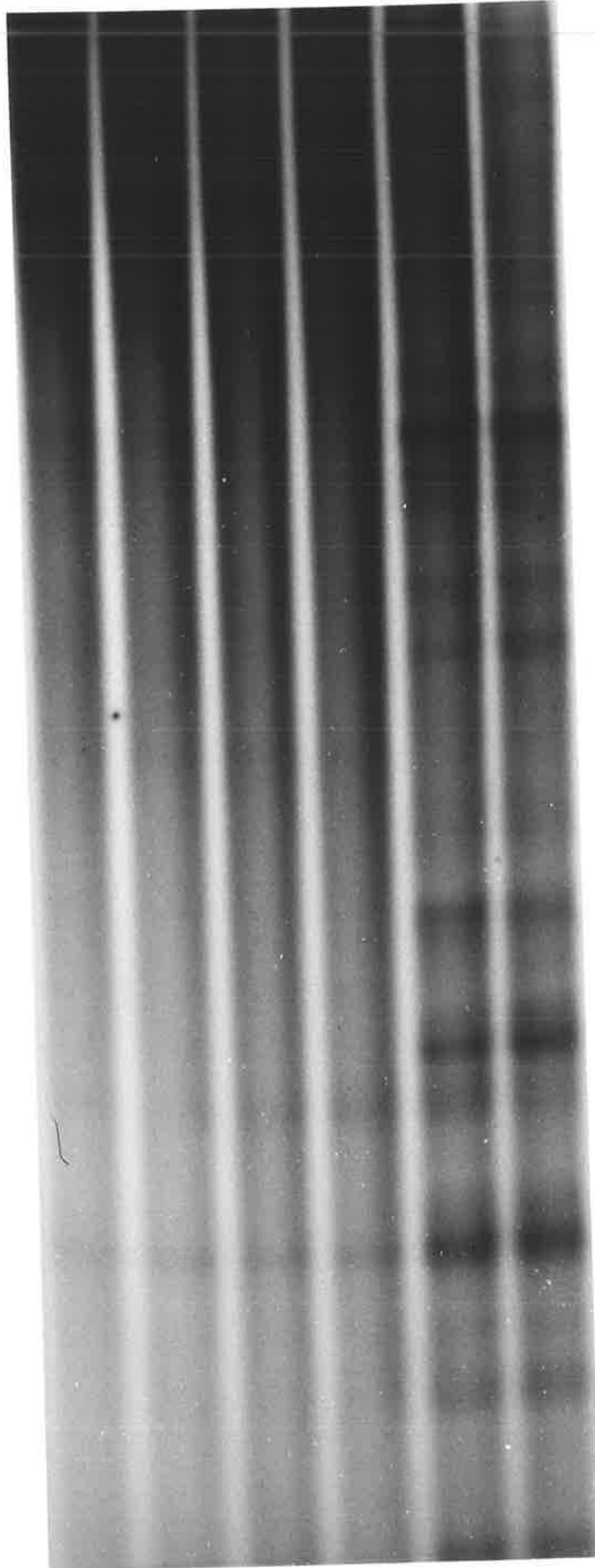
FIGURE 3.16

ELECTROPHORESIS OF Hae III CLEAVED UNIFORMLY LABELLED SINGLE-STRANDED KERATIN cDNA.

Uniformly labelled keratin cDNA was synthesized by oligo(dT) priming on keratin mRNA in the presence of ^{32}P-dGTP as described in 2.2.14. Labelled DNA was digested with from 5 fold to 1,000 fold excess of Hae III restriction enzyme (one fold excess is digestion of 1 μg of DNA with 1 unit of enzyme for 1 hr), as described in 2.2.11. Digestion products were then fractionated by electrophoresis on a 7.5% polyacrylamide slab gel in TBE, as described in 2.2.13, and the labelled DNA detected by autoradiography.

Hae III

5X 10X 50X 100X 500X 1000X



of radioactivity, which was probably due to fragments of irregular size generated by premature termination of the cDNA transcripts. The 1,000 fold excess track in Figure 3.16 shows the presence of about 15 major and many minor bands. This result is consistent with there being a large number of keratin mRNA molecules, the sequences of which differ, at least in part, from one another.

The capacity of Hha I to cut keratin cDNA was also tested, but it was found (Figure 3.17) not to produce discernible fragments whereas a parallel experiment using Hae III to digest keratin cDNA did produce fragments resolved from the smear of radioactivity (Figure 3.17). Some of the larger fragments seen in Figure 3.16 could not be seen in this experiment. Although incubation of keratin cDNA with Hha I did not produce defined fragments, it did appear that the cDNA was being shortened, perhaps by exposure to nucleases contaminating the Hha I enzyme preparation. It is possible that the variation in the banding pattern produced by the previous two Hae III digests of keratin cDNA (Figures 3.16 and 3.17), was due to different transcription efficiencies of the two mRNA preparations used in these experiments.

Several different mRNA preparations were transcribed and tested to determine the pattern of fragments produced after Hae III digestion (Figure 3.18). The experiments were performed under identical conditions to those already described, except that a fresh batch of restriction enzyme was used and the fractionation was performed at a higher voltage, resulting in a reduced electrophoresis time. Figure 3.18 shows that although the mRNAs tested incorporated different amounts of radioactivity into cDNA and produced some minor variations in the relative intensities of some bands, the pattern of bands produced was identical, irrespective of the mRNA source. The method changes mentioned above appeared to improve the resolution of the bands and increased the proportion of larger fragments present in the digests.

Further analysis of the pattern of fragments produced by Hae III cleavage of

FIGURE 3.17

ELECTROPHORESIS OF Hae III AND Hha I CLEAVED SINGLE-STRANDED KERATIN cDNA.

^{32}P-labelled, oligo(dT) primed keratin cDNA was synthesized and cleaved using Hae III and Hha I restriction endonucleases, as described in the legend to Figure 3.16. Keratin cDNA was digested using 10 fold and 1,000 fold excess of each enzyme, as indicated above each track and the digestion products were fractionated by polyacrylamide gel electrophoresis and detected by autoradiography as described in the legend to Figure 3.16.

Hae III Hha I

10X 1000X 10X 1000X

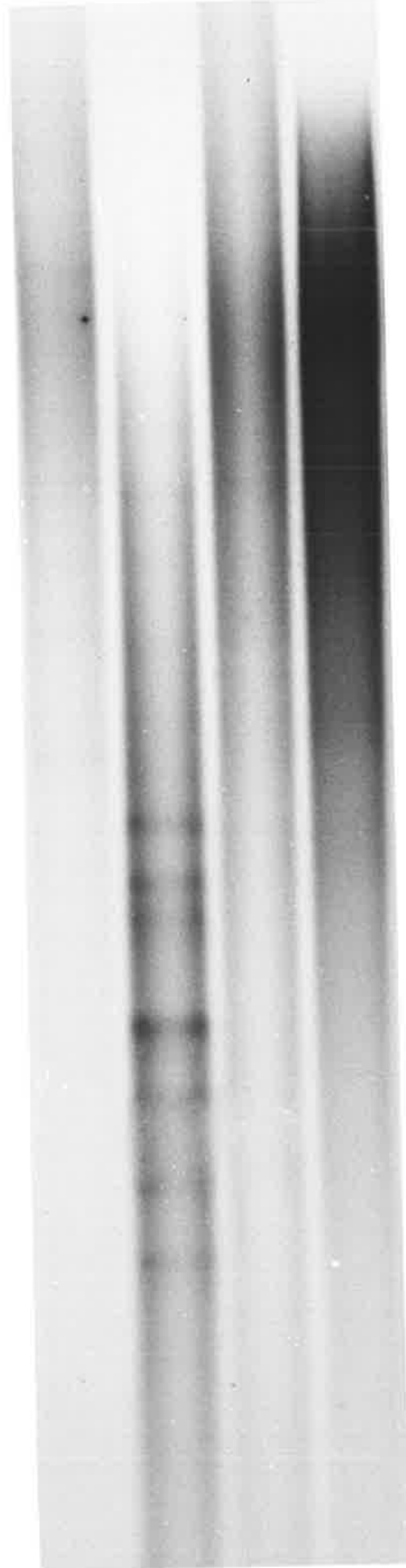


FIGURE 3.18

ELECTROPHORESIS OF Hae III CLEAVED SINGLE-STRANDED KERATIN cDNA DERIVED FROM SEVERAL mRNA PREPARATIONS.

Uniformly ^{32}P-labelled, oligo(dT) primed keratin cDNA was synthesized, cleaved with Hae III (using a 1,000 fold enzyme excess), and the products detected by autoradiography after polyacrylamide gel electrophoresis, as described in the legend to Figure 3.16, except that a fresh batch of restriction enzyme was used and the electrophoretic fractionation of digestion products was performed at a higher voltage. The four gel tracks shown, A, B, C and D, represent cleaved cDNA derived from four different mRNA preparations.

A
B
C
D

keratin cDNA appeared futile since the pattern appeared too complicated to draw further meaningful conclusions, however analysis of cDNA fragments derived from the 5′-ends of the cDNA, and therefore the 3′-end of the mRNA, should give an indication of the number of mRNA species present since the 3′-terminal sequences of the mRNAs were believed to be unique (Kemp, 1975; Lockett *et al.*, 1979).

Uniformly labelled keratin cDNA was digested with Hae III and the mixture was bound to an oligo(dA) cellulose column under conditions used for oligo(dT) cellulose chromatography of poly(A) containing RNA (2.2.1d). The eluted cDNA fragments (which should all be derived from the 3′-end of the mRNA, due to the use of oligo(dT) as a primer during cDNA synthesis) were then fractionated on a gel in parallel to cleaved cDNA which had not been bound to oligo(dA) cellulose. The oligo(dA) cellulose column appeared to uniformly bind a small amount of all of the fragments present in Hae III cleaved cDNA, and did not preferentially select oligo(dT) containing cDNA fragments (data not shown). Oligo$(dT)_{10}$ was used as a primer during cDNA synthesis, and this may not have been long enough to bind to the oligo(dA) cellulose column.

An alternative method was used to detect the cDNA fragments derived from the 3′-end of keratin mRNA. Keratin cDNA was synthesized using ^{3}H-labelled nucleotides, then incubated in the presence of γ-^{32}P-ATP and polynucleotide kinase, to label the 5′-end of the cDNA, which is derived from the 3′-end of the mRNA. This cDNA was then digested with Hae III and the products fractionated on a polyacrylamide gel in parallel with tracks containing uniformly ^{32}P-labelled, Hae III cleaved, keratin cDNA, as shown in Figure 3.19. One major and 3 to 5 minor bands were present in the end-labelled cDNA. The major band was present in the uniformly labelled cDNA track as a major band but the 3 to 5 minor bands were not detectable, although they could have been represented by minor bands present in the uniformly labelled cDNA. This result suggests that either there are not 25 to 35 keratin mRNA species, but one major and several minor ones, or the 3′-ends of

FIGURE 3.19

ELECTROPHORESIS OF 3′-END FRAGMENTS OF Hae III CLEAVED SINGLE STRANDED KERATIN cDNA.

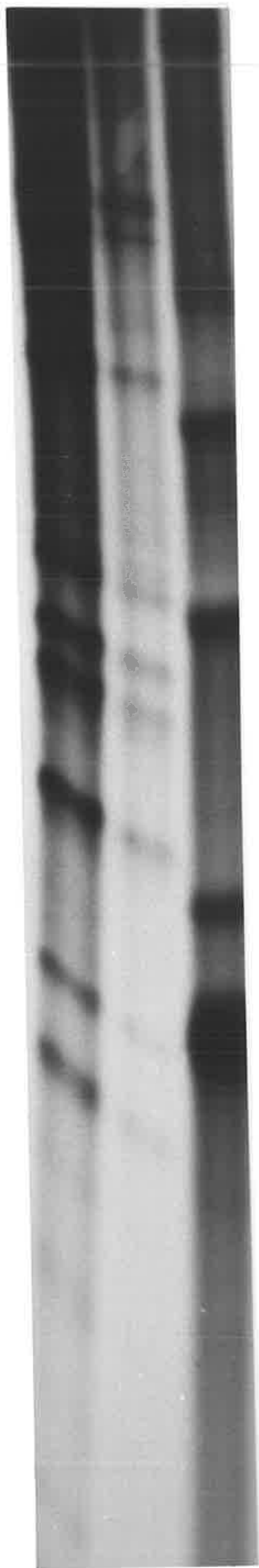
cDNA fragments derived from the 3′-end of keratin mRNA were detected by end-labelling of keratin cDNA followed by digestion with Hae III. Keratin cDNA was synthesized as described in the legend to Figure 3.16, except that ^{3}H-labelled nucleotides were used to follow the reaction. The 5′-ends of the keratin cDNA, which were derived from the 3′-ends of the mRNA, were labelled using γ-^{32}P-ATP and polynucleotide kinase, as described in 2.2.19a. The cDNA was then digested with Hae III and the products fractionated on a polyacrylamide gel in parallel with tracks containing uniformly ^{32}P-labelled, Hae III cleaved keratin cDNA, as described in the legend to Figure 3.18.

The tracks contained :

A. Hae III cleaved uniformly ^{32}P-labelled keratin cDNA.

B. As in A. but using one eighth the amount of cDNA.

C. Hae III cleaved, end-labelled keratin cDNA.


A
B
C

keratin mRNAs may have a substantial degree of homology, such that only a few fragments are produced from the many mRNA species.

3.3.2c ATTEMPTS TO ANALYSE KERATIN mRNA 3'-SEQUENCES.

At the time of this study methods were being developed for the rapid determination of DNA nucleotide sequences (Sanger and Coulson, 1975; Sanger *et al.*, 1977) by extension of primers on DNA in the presence of DNA polymerase I and labelled nucleotides. The first of these methods, termed "plus minus", produces nucleotide specific (e.g. adenosine (A) specific) stoppages in DNA transcripts of varying length, by enzymatically reducing the size of these transcripts in the presence of, for example dATP to the last A in the transcript (plus reaction), and enzymatically increasing the size of transcripts in the absence of, for example dATP but with the other three nucleotide triphosphates present, to the next A in the template DNA (minus reaction). Hence the plus reaction produces a transcript ending in A while the minus reaction ends at the residue preceding the A (similar reactions for G, C and T are also performed).

The second of these methods is simpler and requires only four reactions, not eight. The method uses A, G, C and T dideoxynucleotide triphosphates which are specific chain terminating inhibitors of transcription, to produce transcripts ending with specific nucleotides. The transcripts produced by these two methods can then be separated according to their length on denaturing polyacrylamide gels and the sequence of nucleotides in the DNA can be interpreted from the position of bands of increasing length.

Both of these methods could be applied to the direct sequencing of mRNAs provided that a specific primer was available and the mRNA contained only one molecular species or, if several species existed, the primer primed in a position where these species possessed strong sequence homology. The results presented in

3.3.2b indicated that keratin mRNAs might have a high degree of homology at their 3′-ends and might therefore be suitable for this form of sequence analysis, provided that a specific primer was available. Oligo(dT)$_8$dA, oligo(dT)$_8$dC and oligo(dT)$_8$dG were commercially available and should prime specifically at the junction of keratin mRNA and its poly(A) tract.

Figure 3.20 shows the fractionation on a denaturing polyacrylamide gel of ^{32}P-labelled cDNA fragments produced when keratin mRNA primed with either oligo(dT)$_8$dA, oligo(dT)$_8$dC or oligo(dT)$_8$dG was incubated separately in the presence of the four dideoxynucleotide triphosphates as described in 3.2.8b. Oligo(dT)$_8$dA and oligo(dT)$_8$dG acted as efficient primers of transcription on keratin mRNA as judged by the amount of radioactivity incorporated into cDNA, but oligo(dT)$_8$dA appeared to prime on a set of mRNA molecules which differed sufficiently in sequence at their 3′-ends, to produce a pattern of bands from which it was not possible to "read" a sequence. Oligo(dT)$_8$dG however, primed on a set of mRNA molecules with sufficient homology at their 3′-ends to provide some discernible sequence data (shown with each base aligned with its corresponding band in Figure 3.20).

Oligo(dT)$_8$dC acted poorly as a primer and therefore was probably priming on a subset of mRNA molecules which was only a small proportion of the total keratin mRNA population. It was not possible to read any sequence from the oligo(dT)$_8$dC primed transcription in Figure 3.20, but when higher concentrations of the dideoxynucleotide triphosphates were used, some sequence information was obtained (Figure 3.21). These data are discussed in light of subsequently derived sequence results (Chapter 4) in 3.4.

Attempts to sequence keratin mRNA using the plus minus method, as described in 3.2.8a, proved unsuccessful (Figure 3.22), probably because it was not possible to generate short cDNA fragments in the region of the 3′-end of the mRNAs where the mRNAs appear to possess strong sequence homology. Similar problems were

FIGURE 3.20

NUCLEOTIDE SEQUENCING OF KERATIN mRNA 3′-ENDS USING SPECIFIC PRIMERS.

The 3′-ends of keratin mRNA were sequenced by extension of primers (oligo(dT)$_8$dA, oligo(dT)$_8$dC and oligo(dT)$_8$dG, which should prime specifically at the junction of keratin mRNA and its poly(A) tail), which were separately incubated in the presence of the four dideoxynucleotide triphosphates followed by fractionation of the products on a denaturing polyacrylamide gel and detection by autoradiography, as described in 3.2.8b. The nucleotide sequence determined for the oligo(dT)$_8$dG primed reactions is displayed next to the gel, aligned with the appropriate bands. The three primers used are indicated above each set of four tracks, eg. the oligo(dT)$_8$dA primed reactions are indicated by "T$_8$A" above those tracks. The four specific dideoxynucleotide sequencing reactions are indicated above each track:

A. dideoxyadenosine reaction (25 μM ddATP).

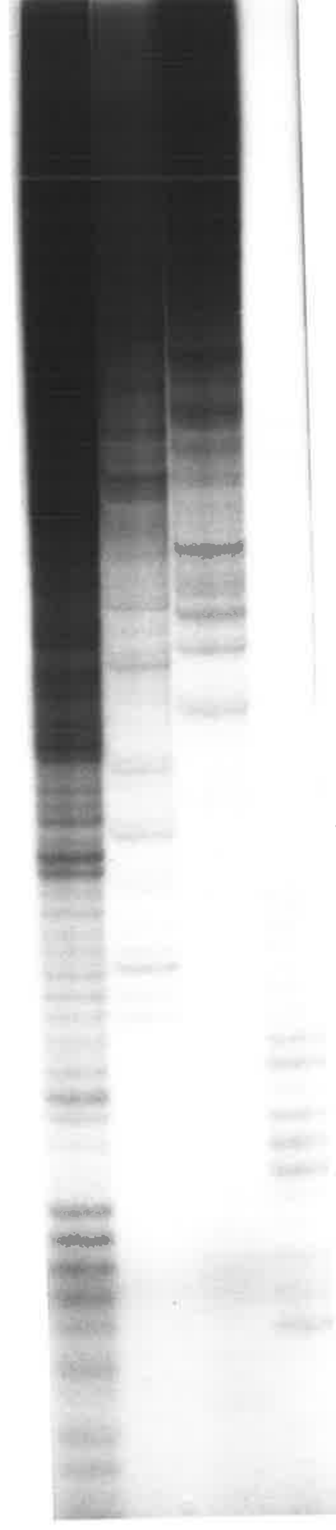
G. dideoxyguanosine reaction (60 μM ddGTP).

C. dideoxycytosine reaction (25 μM ddCTP).

T. dideoxythymidine reaction (20 μM ddTTP).

T_8C
A G C T

T_8A
A G C T



T_8G
A G C T
3′
GGAA
AATA
AAAG
GTTA
A
A
G
T
T
A
T
T
T
T
A
A
A
T
A
C
A
A
5′

FIGURE 3.21

OLIGO(dT)$_8$dC PRIMED NUCLEOTIDE SEQUENCING OF KERATIN mRNA.

The sequence shown was determined by oligo(dT)$_8$dC priming on keratin mRNA in the presence of dideoxynucleotides, exactly as described in the legend to Figure 3.20, except that the dideoxynucleotide triphosphate concentrations were doubled. All diagrammatic details are as described in the legend to Figure 3.20.

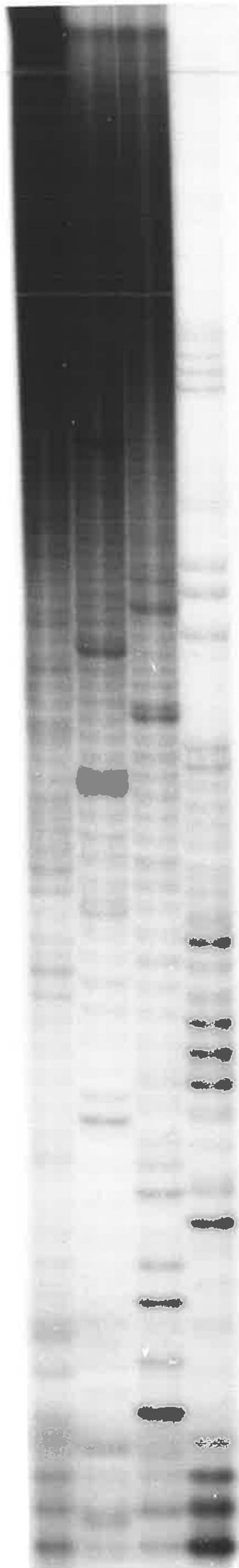
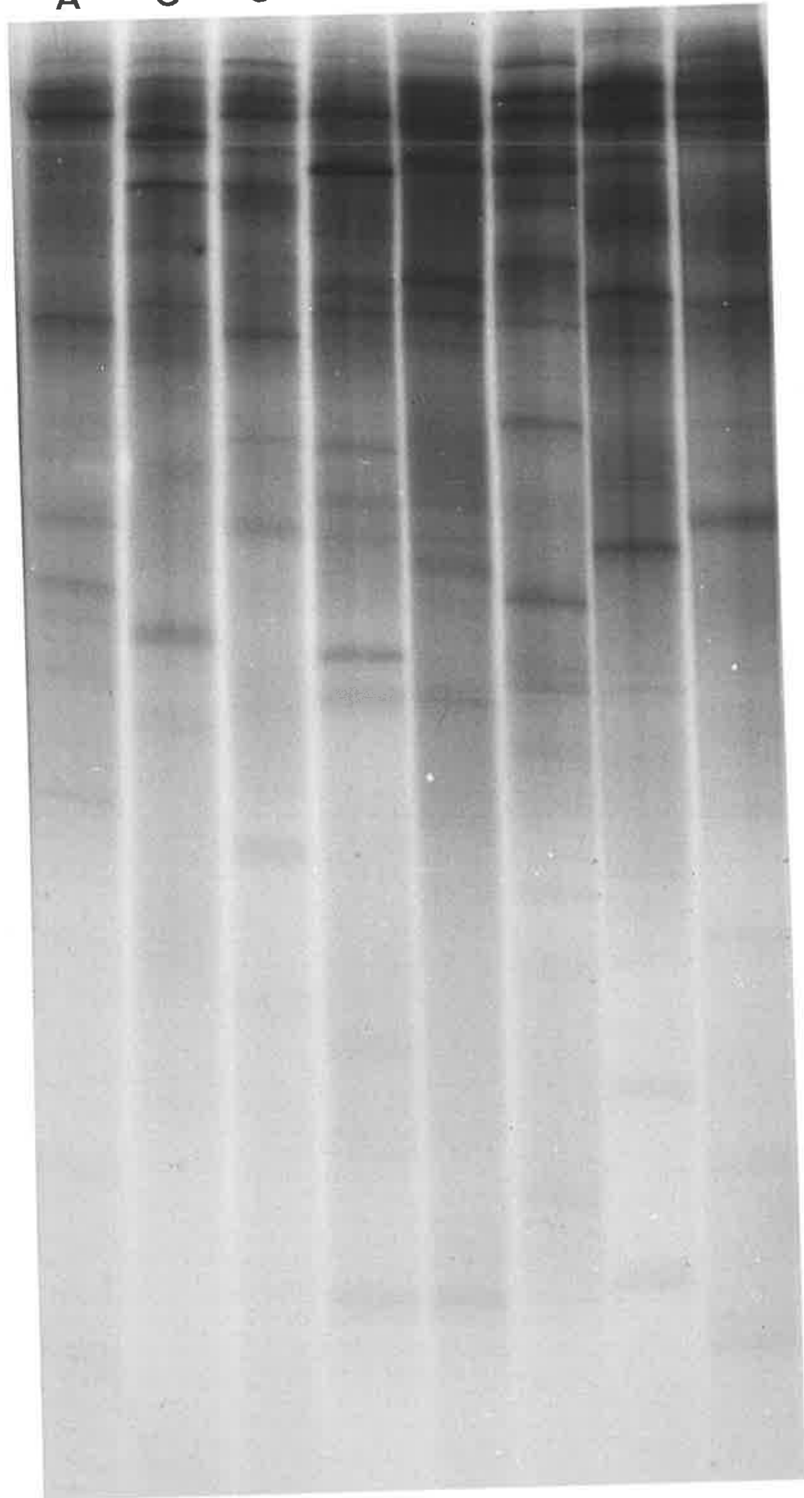

T8C
A G C T
AT
CTACA
GTACT
ACCAA
AGGTT
A
A
A
A
G
T
A
A
T
T
T
G
C
A
C
T
C
C
A
C
A
C
G

FIGURE 3.22

"PLUS MINUS" SEQUENCING OF KERATIN mRNA PRIMED WITH OLIGO$(dT)_8$dC.

The plus minus sequencing reactions of Sanger and Coulson (1975) were performed on oligo$(dT)_8$dC primed keratin mRNA, as described in 3.2.8a. The four base specific plus and minus reactions are indicated above each track.


Plus
Minus
A G C T A G C T

experienced when a method described by Brownlee and Cartwright (1977), for the sequencing of RNA by transcription from specific primers, was attempted for keratin mRNA (data not shown). This method relied on four reactions in which low concentrations of each of the four nucleotide triphosphates limited the length of transcripts and produced the required specific termination of those transcripts.

3.4 DISCUSSION.

3.4.1 THE NATURE OF 14-DAY EMBRYONIC FEATHER mRNAs.

This chapter describes two methods for the preparation and purification of the RNA species present in 14-day embryonic feathers (3.3.1a and 3.3.1d). Indeed, during the course of this study many procedures for the preparation of mRNA have been used, most of which are described in 2.2.3. The majority of these methods use 14-day polysomes as the starting point, although one method, not described in Chapter 2, used guanadinium hydrochloride to extract RNA directly from feathers (Brooker *et al.*, 1980), while another (2.2.2) extracted RNA directly from cleared embryonic feather lysate. RNA so produced was invariably further fractionated either on poly(A) content by oligo(dT) cellulose chromatography (2.2.3d), or on size by sedimentation through sucrose gradients.

Both RNA methods reported achieved some resolution of RNA species present in prepared RNA and these species could apparently be purified to the stage where they sedimented as single symmetrical peaks on sucrose gradients. However, it was clear that these peaks contained RNA which originally sedimented at higher Svedberg (S) values. The fractionation of translation products of 9 and 12S RNAs prepared from RNPs on pH 2.7 gels (Figure 3.6) indicated that although 9S RNA contained proportionately more fast protein mRNA activity than did 12S RNA, the major mRNA activity in both RNAs coded for keratins. This result was obtained despite the complete resolution of these RNA peaks on sucrose gradients (Figure

3.4), indicating that either keratin mRNA varies greatly in length, or the 9S peak is heavily contaminated with degraded 12S keratin mRNA which has retained some translational activity. The latter of these hypotheses is supported by the observation that several protein peaks in the 9S RNA translation products migrated in a position consistent with them being short translation products produced from degraded keratin mRNA (Figure 3.6 and 3.9). These proteins were more obvious in the 9S translation products of RNA made from SDS dissociated polysomes, indeed very little keratin was detected in the translation products of this RNA preparation.

The observation (Figure 3.6) that 9S RNA contained more fast protein mRNA activity than did 12S RNA, would indicate that fast protein mRNA must be more susceptible to ribonuclease degradation than keratin mRNA, since results presented in Chapter 6 demonstrate that fast protein mRNA would normally sediment at about 13.5S. Figure 3.10 confirms that fast protein mRNA is partially degraded, but still active, since more fast protein mRNA activity was contained in 12S RNA which did not bind to oligo(dT) cellulose than 12S RNA which did bind. That is to say, the RNA contained in the 12S RNA fraction which had lost the 3′-poly(A) tract and would consequently not bind to oligo(dT) cellulose (ie. RNA which had been degraded), was more likely to code for fast protein than keratin.

No other mRNA activity besides keratin and fast protein was detected in translation products of 9, 12 and 14S RNA on pH 2.7 gels. Although pH 2.7 gels resolve β-keratins from fast proteins the presence of any 14,500 Mr scale β-keratin or 50,000 Mr α-keratin found in scale, might not be seen since no study of their migration on pH 2.7 gels has been conducted. Fractionation of translation products of total poly(A)-plus mRNA from 14-day embryonic feathers on SDS gels (Figure 5.7) shows that no α-keratins were detectable but 14,500 Mr products were present.

Hybridization of 14S cDNA to ribosomal RNAs indicated that this RNA was composed almost exclusively of degraded ribosomal RNA, which explained why the RNA acted so poorly as a template for the incorporation of radioactivity into protein

in a wheat embryo cell-free system. Although this RNA should have contained fast protein mRNA activity (because fast protein mRNA should sediment at about 13.5S, as described above) it is likely that any mRNA activity was masked by the presence of vast amounts of ribosomal RNA.

It is obvious from these results that the methods employed to purify and separate mRNA species were not sufficiently selective to achieve significant resolution of mRNAs present in the embryonic chick feather. Although some resolution of RNA species might be achieved by complicated systems of electrophoresis, as described for histone mRNAs (Gross *et al.*, 1976) or by immunoprecipitation of polysomes (Scott and Wells, 1976, 1978), such systems are notoriously intractable and unreliable. These techniques would almost certainly not resolve RNA components within a family, such as the 25 to 35 sequence related keratin mRNA species.

3.4.2 KERATIN mRNA FINE STRUCTURE.

Most eukaryotic mRNAs contain a 3′-terminal polyadenylate tract but its biological function is unclear. It is not required for mRNA function *in vivo* or in cell-free systems since most histone mRNAs (Jacobs-Lorena *et al.*, 1972; Ruderman and Pardue, 1978), deadenylated protamine mRNA (Iatrou and Dixon, 1977; Gedamu *et al.*, 1977) and deadenylated globin mRNA (Breindl and Gallwitz, 1973) lack poly(A) tracks but are translated *in vitro*. However, deadenylated ovalbumin mRNA was shown to be less efficiently reinitiated than the native mRNA (Doel and Carey, 1976). It is possible that poly(A) sequences confer functional stability to mRNAs (Marbaix *et al.*, 1975; Nudel *et al.*, 1976).

The results presented in this chapter confirm the conclusion (Partington *et al.*, 1973) that keratin mRNA is polyadenylated and that the poly(A) tracts have a size distribution consistent with that of other polyadenylated eukaryotic mRNAs (Brawerman, 1974). It was also apparent that the poly(A) length distribution was the same from one mRNA preparation to another. These results did not explain

why a large proportion of keratin mRNA does not bind to oligo(dT) cellulose under standard conditions since almost no poly(A) tracts were found to be less than about 45 nucleotides, which should be long enough to efficiently bind to oligo(dT) cellulose.

Kemp (1975) described keratin mRNA as a complex set of 25 to 35 sequence related molecules and he proposed that they contained coding regions which were strongly conserved, behaving as reiterated sequences when their hybridization kinetics were examined, and untranslated regions which were unique. The restriction analysis of keratin cDNA described in this chapter was designed to investigate this hypothesis and to obtain a more accurate estimate of keratin mRNA complexity.

Restriction of uniformly labelled cDNA revealed that the mRNA was highly complex, although the number of fragments generated was less than that expected for 25 to 35 totally unique mRNA species, presumably because of the conserved sequences present in the mRNA coding regions. Analysis of 3′-terminally derived cDNA restriction fragments (Figure 3.19) revealed a much simpler pattern than that predicted by the above hypothesis and from hybridization studies of short 3′-terminally derived cDNA transcripts which indicated that the 3′-ends of keratin mRNAs were unique (Lockett *et al.*, 1979). This result appears to be in conflict with the results of Lockett *et al.* (1979) but subsequent RNA (this chapter, 3.3.2c), double-stranded cDNA (Chapter 4) and genomic clone sequencing data (Molloy *et al.*, 1982) revealed that both results are compatible with the known structure of keratin mRNA. This is further discussed later in this chapter.

Attempts to sequence the 3′-ends of keratin mRNA were only partially successful in that the sequence data which was obtained could only be determined for the 3′-terminal 30 to 40 nucleotides and even these could not be read with great confidence. The fact that any sequence could be read at all indicated that the sequence of keratin mRNA was highly conserved for those 3′-terminal 30 to 40 nucleotides.

Of the three primers used, oligo$(dT)_8$dG proved to be the most efficient since it incorporated the largest amount of radioactivity into cDNA. The sequence determined using this primer is shown in Figure 3.23a compared with the DNA sequences of a cDNA clone (pCFK23; Chapter 4) and a genomic gene (gene C; Molloy *et al.*, 1982) both of which are known to code for chick embryonic feather keratins. It is obvious from this comparison that the oligo$(dT)_8$dG primed sequence is very similar to the two keratin sequences, including the presence of the AAUAAA polyadenylation signal (Proudfoot and Brownlee, 1976) approximately the correct distance from the poly(A) tract. The oligo$(dT)_8$dG primed sequence cannot be read past about 35 nucleotides (about 45 nucleotides from the poly(A)) which indicates that this primer is priming on a set of mRNA molecules with strong sequence homology at their 3′-ends which breaks down after about 40 nucleotides. This pattern of sequence homology was confirmed by the sequence analysis of keratin cDNA and genomic clones (see Figure 4.13). It is also interesting that all cloned keratin DNA sequences examined so far have a 3′-terminal cytosine, confirming that oligo$(dT)_8$dG would prime on these mRNAs.

If all keratin mRNAs end with cytosine, the mRNAs which primed with oligo$(dT)_8$dA and oligo$(dT)_8$dC must code for proteins other than keratins. This suggests that the 12S keratin mRNA preparation was not as pure as previously thought (95% keratin coding; Kemp *et al.*, 1974a) since oligo$(dT)_8$dA acted efficiently as a primer, possibly accounting for 30 to 40% of the total template capacity of 12S mRNA (Figure 3.20). It was not possible to recognize any sequence in the oligo$(dT)_8$dA primed sequencing reactions, meaning that the mRNAs on which it was priming must contain quite diverse 3′-terminal sequences. It is interesting that the 3′-terminal residue for one confirmed fast protein mRNA (pCFK22; Chapters 5 and 6) and one suspected fast protein mRNA (pCFK3; Chapter 5) is uridine and the 3′-ends of these two mRNAs show very poor sequence homology.

Oligo$(dT)_8$dC didn't prime well on 12S mRNA indicating that the mRNA(s) on

FIGURE 3.23

COMPARISON OF KERATIN mRNA 3′-END NUCLEOTIDE SEQUENCES WITH SEQUENCES OF SUBSEQUENTLY DETERMINED KERATIN cDNA AND GENOMIC CLONES.

a. The sequence determined by oligo$(dT)_8$dG priming on keratin mRNA (Figure 3.20) is shown compared to sequences of a cDNA clone (pCFK23) and a genomally derived gene (gene C).

b. The oligo$(dT)_8$dC primed sequence (Figure 3.21) is shown compared to the sequence of a cDNA clone, pCFK3.

In both a. and b. the cDNA sequence, determined by priming on mRNA, has been converted to the complementary mRNA sequence (shown, 5′ to 3′) to facilitate comparison with the keratin DNA sequences. Matches between the cDNA sequence and the cloned sequences are indicated by an asterisk. Hyphens indicate gaps in the sequences which were introduced to assist in their alignment.

a.

Oligo(dT)$_8$dG Primed Sequence

```
        3'                                         5'
cDNA         AAGGATAAGAAAATTGAAGTTATTTTAAATACAA
        5'                                         3'
mRNA         TTCCTATTCTTTTAACTTCAATAAAATTTATGTT
             ***   ****** *  **************** *
pCFK23     GATTC---TTCTTT-AGATTCAATAAAATTTATGCTGCATT
             ****  ****** *   *************** *
Gene C     GATTCC--TTCTTTGAGC-TCAATAAAATTTATGCTGCATT
```

b.

Oligo(dT)$_8$dC Primed Sequence

```
        3'                                                5'
cDNA         TAACATCTCATGAACCATT-GGAAAAAGTAATTTGCAC-TCCACACG
        5'                                                3'
mRNA         ATTGTAGAGTACTTGGTAA-CCTTTTTCATTAAACGTG-AGGTGTGC
               ***   * * *   *   *****   ******  *  * *  * *
pCFK3      TTTCTGTTCTGGAATATATCTGCCTTTCATATTAAAA-TCTACGC-TTCATC
```

which it primed represented a small proportion of the total RNA. The sequence obtained using this primer was quite readable for about 30 nucleotides and is shown compared with sequences from two fast protein cDNA clones (pCFK3 and pCFK10; Chapter 5) in Figure 3.23b. The oligo(dT)$_8$dC primed sequence exhibits only poor sequence homology with either cDNA clone but all three contain the unusual polyadenylation signal AUUAAA (Jung *et al.*, 1976; Hagenbuchle *et al.*, 1980; Goeddel *et al.*, 1981). The origin of this mRNA sequence is unclear since all fast protein and keratin mRNAs examined so far do not have guanosine as their 3′-terminal nucleoside nor does a scale-like keratin cDNA clone found to be expressed in feathers (pCFK15, Chapter 5). The only mRNA sequence so far examined, which is found in the epidermal appendages of the chick and ends in a guanosine, is a scale cDNA clone (pCSK9; Wilton, 1983) which contains the AUUAAA polyadenylation sequence found in the oligo(dT)$_8$dC primed mRNA sequence but shows no other homology with this sequence.

CHAPTER 4.

KERATIN mRNA SEQUENCES.

4.1 INTRODUCTION.

Chapter 3 described attempts to purify and characterize mRNA species expressed in the embryonic feather, by classical physico-chemical techniques, i.e. RNA size fractionation and analysis of translational activity. 12S mRNA, which predominantly codes for keratin, was further characterized by physical means and attempts at direct sequencing of 12S mRNA yielded limited data which was equivocal because of the heterogeneous nature of the messenger RNA molecules in 12S RNA. It became clear that individual mRNA species from the mixed population of mRNA molecules would be required if any study were to be conducted on the nature and fine structure of mRNAs expressed in 14-day feathers.

During the course of the work presented in Chapter 3 the moratorium on recombinant DNA experiments in Australia was lifted under the surveillance of a R-DNA committee (ASCORD), and the newly developed techniques for molecular cloning of double-stranded cDNA copies of mRNA (Efstratiadis *et al.*, 1975; Rougeon and Mach, 1976; Maniatis *et al.*, 1978) enabled the purification of individual species of 14-day embryonic feather mRNA by molecular cloning of double-stranded cDNA into *E. coli* (Saint, 1979). Recombinant DNA molecules were constructed by annealing poly(dC)-tailed double-stranded cDNA to the well characterized (Sutcliffe, 1978) bacterial plasmid pBR322 (Bolivar *et al.*, 1977), which had been cut with Pst I and poly(dG)-tailed. This procedure regenerates the Pst I cloning site after introduction of the chimeric molecules into *E. coli*, thus allowing convenient excision of the inserted DNA.

Preliminary analysis of the cloned cDNAs revealed the existence of two distinct groups of species represented in the RNA from which the clones were made. The two classes of sequences were termed Group I (the more abundant class) and Group II (the less abundant) by Saint (1979).

It was found that filter bound DNA of individual Group I species bound more cDNA made from partially purified keratin mRNA than equal amounts of Group II species DNA, indicating the presence of individual Group II species as contaminants of partially purified keratin mRNA. It was found that individual Group I species and keratin cDNA annealed to exactly the same complex set of Hind III digested chick genomic DNA fragments (Figure 4.1). Group II species annealed to restricted chick genomic DNA fragments which were totally different to those annealing to either Group I species or keratin cDNA. The restricted genomic DNA band patterns produced by Group II species appeared to have several minor bands in common although most Group IIs had a different pattern of major bands. This study lead to the conclusion that Group I species represented keratin coding sequences while the less abundant Group II species were postulated to comprise a family of sequence-related molecules which were not simply mRNAs coding for a background of "housekeeping" functions, although the exact nature of this group was uncertain.

Non-identical Group I species were shown to anneal to just a few, different restricted chick DNA fragments under stringent (low salt) hybridization and washing conditions while under low stringency (high salt) conditions they annealed to the same complex set of 20-30 DNA fragments, indicating that differences must exist in the sequences of adjacent non-coding and/or intervening sequences for these species. These findings further indicated that keratin-coding (Group I) species are members of a family of homologous, but different genes present in the chick genome. The finding of different Group I species which hybridized to the same genomic DNA fragment suggested that this family of keratin genes is linked in the chick genome, although the extent of gene clustering was not determined.

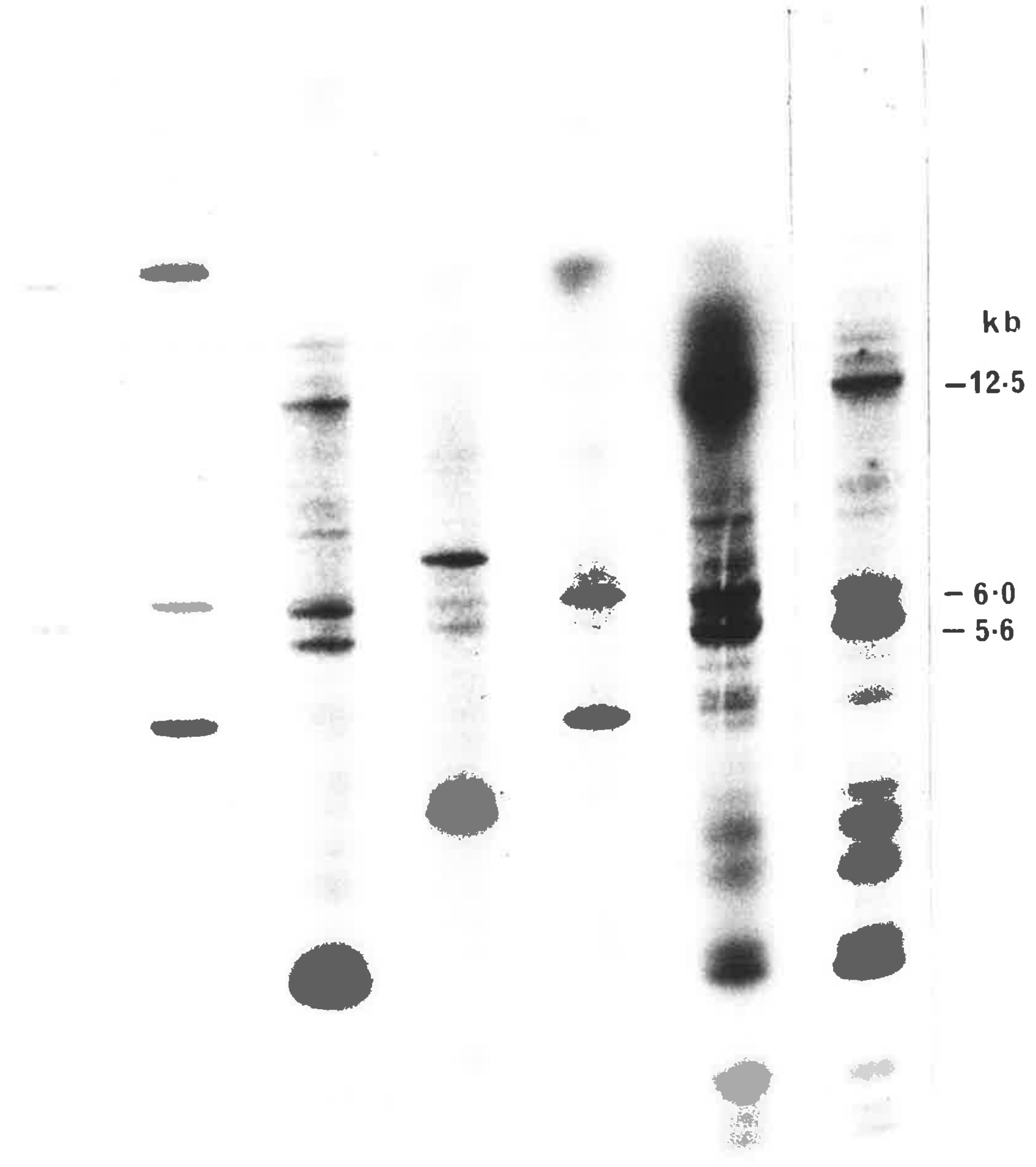
The "library" of cloned embryonic feather double-stranded cDNAs provided an excellent opportunity for the study of the nature and fine structure of mRNAs expressed in the embryonic feather. Such a study was made possible by the generous donation (by R.B. Saint) of this double-stranded cDNA library after completion of

FIGURE 4.1

SOUTHERN PATTERNS OF Hind III CLEAVED CHICK GENOMIC DNA PROBED WITH GROUP I AND GROUP II cDNA CLONE SEQUENCES.

The Southern patterns shown resulted from agarose gel fractionation of Hind III digested chick DNA and hybridization with nick-translated keratin cDNA (kcDNA), Group I cDNA clone sequences (pCFK12 and 23) and Group II cDNA clone sequences (pCFK3, 10, 15 and 22).

(From Saint, 1979).


pk 3
10
12
15
22
23
k cDNA
kb
—12·5
— 6·0
— 5·6

the preliminary investigation described above.

After the work presented in this chapter was initiated, studies on the genes for embryonic keratins were begun (Molloy *et al.*, 1982; Gregg *et al.*, 1983). These studies concentrated on a λ-Charon 4A clone containing 15 kilobases of chick DNA and shown to contain four unique complete keratin genes and the 3′-end of a fifth keratin gene. The genes are all transcribed in the same direction and are regularly spaced 3.3 Kb apart, although the intergenic sequences differ from one another. The genes all have the same structure, containing a protein coding region of 297 bp (coding for a protein of 97 amino acids, plus initiation and termination codons), a 3′-untranslated region of about 435 bp, and a 5′-untranslated region of 58 or 59 bp which is interrupted by an intron of about 325 bp, 37 bp from the putative "cap" site.

DNA sequencing and the consequent mapping and endonuclease restriction work described in this and the following two chapters, necessarily generated data too copious to be included in a thesis. For this reason the process of restriction mapping and DNA sequencing is described and illustrated by examples in this chapter, but the full data have not been included for the sake of brevity. Once these processes have been illustrated, subsequent DNA sequencing data will be presented without further description, unless a particular point is to be made.

This chapter describes the DNA sequencing of two Group I clones derived from embryonic feather mRNA. These sequences are compared with the genomic DNA sequences which have become available since the majority of the work presented in this chapter was completed (Gregg *et al.*, 1984). The DNA sequence presented in this chapter was obtained initially by use of the chemical cleavage method of DNA sequencing, described by Maxam and Gilbert (1977, 1980). Later sequences were generated by using the combined techniques of cloning into the single-stranded DNA bacteriophage M13 (Messing *et al.*, 1977; Messing and Vieira, 1982) to produce

single stranded templates suitable for the dideoxy DNA sequencing method of Sanger *et al.* (1977), briefly described in 3.1, 2.2.22 and 2.2.23.

4.2 SPECIFIC METHODS.

4.2.1 MINISCREEN EXAMINATION OF PLASMID RECOMBINANTS.

The isolation of plasmid DNA from small cultures for size estimation of the inserted sequence and for preliminary restriction analysis was carried out as follows. 1.5 ml cultures of the recombinants were grown overnight at 37°C in L-broth. The cells were pelleted by centrifugation for 2 minutes in an Eppendorf centrifuge, washed once with 10 mM Tris pH 9.0, 1 mM EDTA and resuspended in 150 μl of 15% sucrose, 50 mM Tris-HCl pH 9.0 and 50 mM EDTA. 50 μl of a freshly made solution of lysozyme (4 mg/ml) was added and the mixture incubated at room temperature for 15 minutes and then on ice for 30 minutes. 150 μl of ice-cold water was added and the mixture was left on ice for a further 10 minutes followed by heating at 70°C for 15 minutes. Cellular debris was removed by centrifugation at 20,000 rpm for 90 minutes at 4°C. The supernatant was carefully decanted, extracted once with an equal volume of phenol : chloroform (1:1), ethanol precipitated and resuspended in 100 μl of H_2O.

4.3 RESULTS.

4.3.1 DETERMINATION OF RECOMBINANT PLASMID INSERT SIZE.

DNA prepared (as described in 2.2.10) from several randomly selected cDNA clones was digested with Pst I to excise the inserted DNA. These digests were fractionated on a 6% polyacrylamide gel, as shown in Figure 4.2a and b. Plasmid clones pCFK (plasmid chicken feather keratin) 1, 26 and 28 did not produce an excised insert fragment when digested with Pst I, indicating that one or both of

FIGURE 4.2

ELECTROPHORESIS OF Pst I AND Hae III CLEAVED RECOMBINANT cDNA PLASMID DNAs.

Recombinant cDNA plasmid DNA was prepared from several cDNA clones, using the method described in 2.2.10, and digested with either Pst I or Hae III prior to fractionation on a poly acrylamide gel as described in 2.2.13. DNA was detected by staining with ethidium bromide and visualization under UV. Molecular weights were calculated from marker tracks (indicated by M above the track) containing Pst I digested pBR322 and Hae III digested pBR322 (a. and b.) or Hae III digested pBR322 alone (c. and d.).

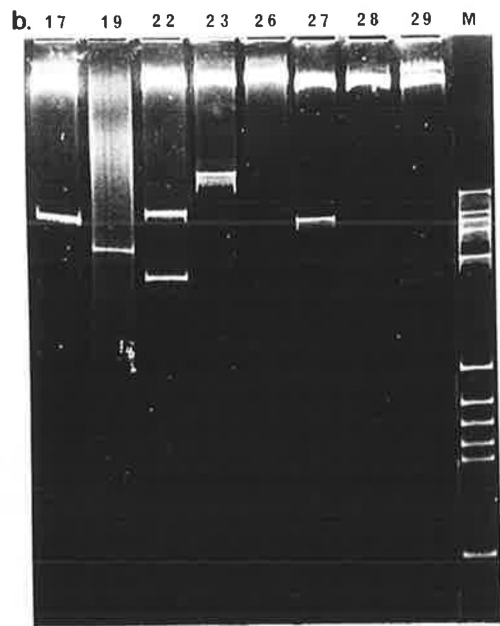
a. Pst digested pCFK1, 3, 4, 7, 9, 10, 12, 15.

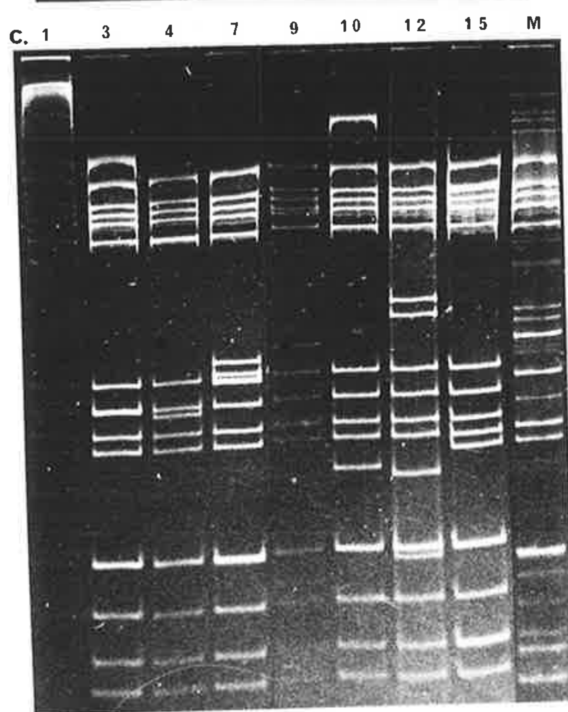
b. Pst digested pCFK17, 19, 22, 23, 26, 27, 28, 29.

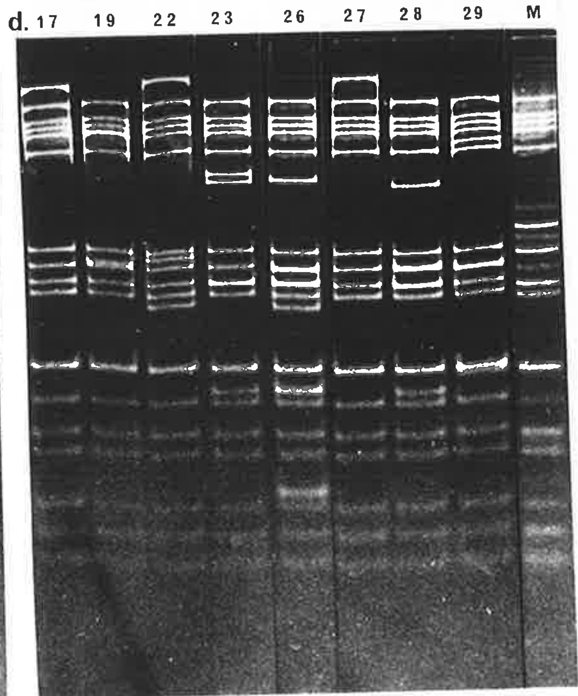
c. Hae III digested pCFK1, 3, 4, 7, 9, 10, 12, 15.

d. Hae III digested pCFK17, 19, 22, 23, 26, 27, 28, 29.

a.
1 3 4 7 9 10 12 15 M


b.
17 19 22 23 26 27 28 29 M


c.
1 3 4 7 9 10 12 15 M


d.
17 19 22 23 26 27 28 29 M

the terminal Pst I cloning sites were not regenerated during the cloning process. All other clones tested had intact Pst I sites and therefore produced resected insert fragments. Two clones pCFK15 and 22 had internal Pst I sites resulting in more than one insert fragment. pCFK15 had two internal Pst I sites resulting in fragments of 265, 190 and 150 bp (a total of 605 bp), the latter of these not being visible on this gel, but its presence was indicated by a partial digestion product of 415 bp which has subsequently been found to be due to incomplete cleavage between the 265 and 150 bp fragments. pCFK22 had one internal Pst I site and therefore produced two fragments (370 and 515 bp) with a total length of 885 bp. The inserted DNA from these clones varied in size from 885 bp (pCFK22) to 370 bp (pCFK19) with the majority of inserts being between 500 to 700 bp in length.

4.3.2 DIGESTION OF RECOMBINANTS WITH Hae III

In order to further characterize the selected recombinants, plasmid DNA from these clones was digested with Hae III. The pattern of bands produced when the products were fractionated on polyacrylamide gels (Figure 4.2 c and d) can be used as a "fingerprint" to facilitate recognition of recombinants derived from the same mRNA species. The banding patterns obtained indicated that most of the recombinants appeared to be derived from fundamentally different mRNA species, however pCFK23 and 28 appeared to have several bands in common, possibly because they represented independant isolates of the same mRNA species, or perhaps the mRNAs from which they were derived had closely related sequences.

4.3.3 SELECTION OF A GROUP I cDNA CLONE FOR FURTHER STUDY.

Of the cDNA clones shown in Figure 4.2, pCFK 9, 12, 17, 23, 28 and 29 were demonstrated by Saint (1979) to be Group I sequences, and were therefore most likely to code for keratins, as proposed by Saint (1979). Of these clones only pCFK23 (770 bp) contained an excisable insert approaching the size of the

full length copy of a keratin mRNA. The mRNA for embryonic feather keratin has been estimated to be 760 nucleotides (Kemp *et al.*, 1974b) containing an average poly(A) length of 65 nucleotides (Morris and Rogers, 1979). Assuming an average of 20 for the length of each string of G + C base pairs joining the insert to the vector, resected inserts of full length copies of keratin mRNA should be of the order of 740-800 bp in length. Consequently pCFK23 was the only cDNA clone examined which had the potential of containing a full or nearly full length copy of keratin mRNA, it was therefore chosen for further examination, including DNA sequencing.

4.3.4 RESTRICTION MAPPING OF pCFK23.

The chemical cleavage method of DNA sequencing is facilitated by a good understanding of the number and position of restriction endonuclease recognition sequences, i.e. a restriction map is useful, particularly since secondary restriction of terminally labelled fragments is required to separate labelled termini. An indication of the length and number of restriction fragments contained within the insert DNA can be obtained by digestion of plasmid DNA followed by fractionation on polyacrylamide gels. Figure 4.3 shows the fractionation of pCFK23 DNA digested with Alu I, Hae III, Hha I and Msp I, pBR322 was also digested and the products fractionated in parallel tracks to indicate the digestion pattern resulting from the vector molecule. Any extra bands over and above those produced from pBR322 appearing in the pCFK23 tracks were attributed to foreign DNA inserted into pBR322. As expected the pBR322 fragment which contains the Pst I site disappeared to produce sections of DNA, each of which is attached to the two terminal insert fragments. Unfortunately, the data shown in Figure 4.3 gave no indication which of the extra bands were derived from the insert termini, and some fragments were obscured because they migrated with vector fragments. Another problem with ultraviolet detection of DNA fragments stained with ethidium bromide, is that very short fragments cannot be seen easily.

All of these problems were overcome by preparing excised insert DNA by cutting

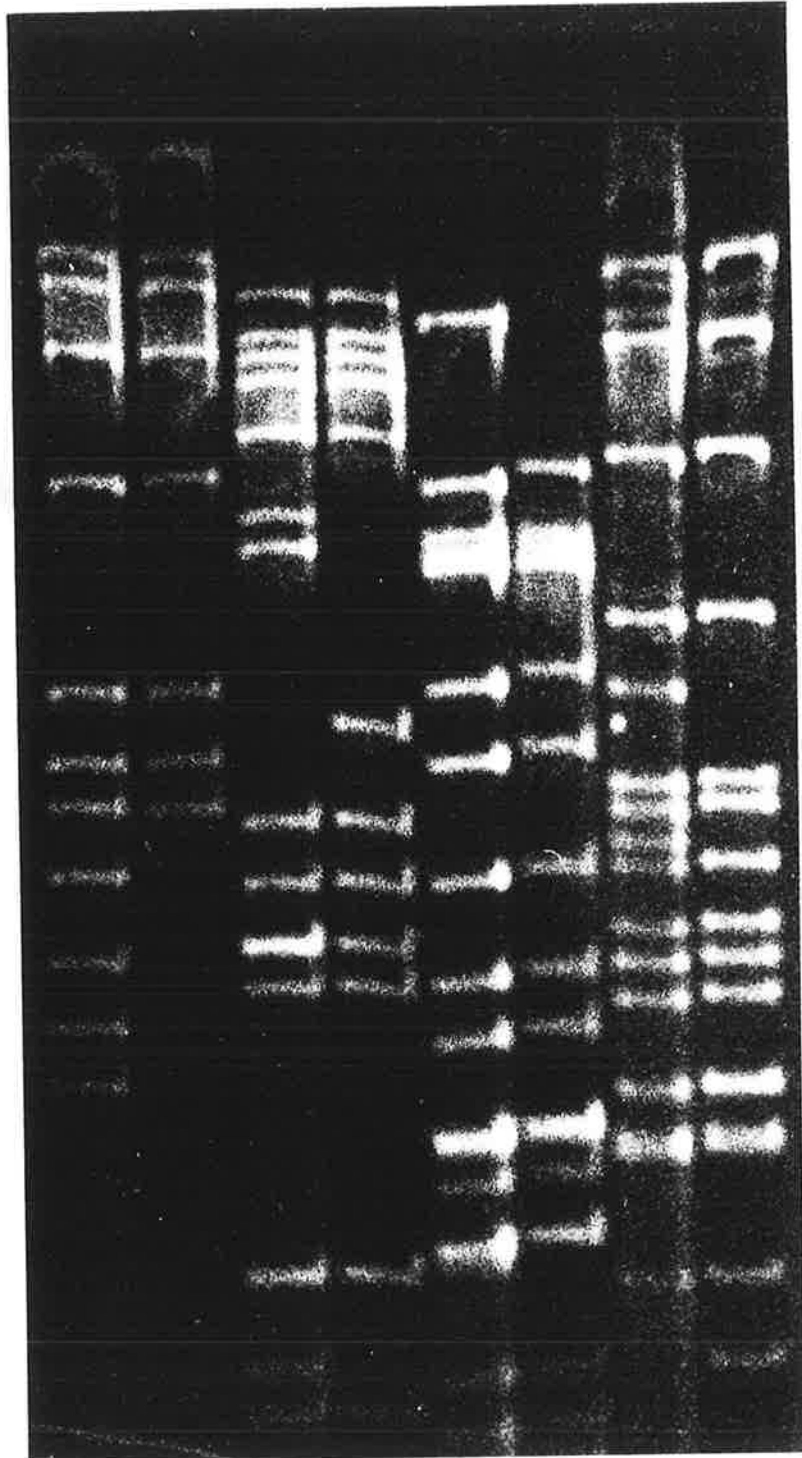
FIGURE 4.3

IDENTIFICATION OF RECOMBINANT INSERT FRAGMENTS IN pCFK23 AFTER DIGESTION WITH Alu I, Hae III, Hha I AND Msp I.

The DNA fragments produced by digestion of DNA from the cDNA clone pCFK23 with Alu I, Hae III, Hha I and Msp I are shown fractionated on a 6% TBE polyacrylamide gel in parallel with tracks containing pBR322 DNA digested with the same enzymes. The conditions for electrophoresis and DNA detection were as described in the legend to Figure 4.2.

Alu I Hae III Hha I Msp I

23 pBR 23 pBR 23 pBR 23 pBR



pCFK23 with Pst I and eluting the resected insert DNA after fractionation on an acrylamide gel (data not shown). The prepared insert DNA was end-labelled (2.2.19b) by incubation with the Klenow fragment of DNA polymerase I and ^{32}P-labelled dGTP under conditions which allowed the exchange of the 3′-terminal nucleotide for a radioactive one (Englund, 1972). End-labelled pCFK23 insert DNA was then digested with Alu I, Hae III, Hha I and Msp I and the digestion products fractionated on an acrylamide gel (Figure 4.4), which allowed the detection of the terminal insert fragments. The remaining insert fragments were detected by end-labelling, by terminal exchange, or end-filling (2.2.19) after digestion with the appropriate restriction enzyme and fractionation of the products on an acrylamide gel (Figure 4.4). The identity of terminal fragments produced from the digested plasmid could then be inferred. Comparison of the length of terminal fragments from insert DNA with the length of terminal fragments from whole recombinant plasmid DNA allowed the orientation of the insert DNA in pBR322, since the terminal fragments from the insert increased in size by an amount equivalent to the length of the section of pBR322 from the flanking restriction sites to the Pst I site. Provided that the flanking sections of pBR322 DNA were of unequal length, the insert DNA could be orientated with respect to the vector molecule.

The ordering of restriction fragments contained within the pCFK23 insert was achieved by fractionation of terminally labelled fragments produced after digestion of the insert with various combinations of several restriction enzymes prior to labelling. This also allowed the positioning of restriction sites for one enzyme within a fragment produced by a different enzyme. This approach resulted in the production of a restriction map of pCFK23 for the enzymes : Bam HI; Alu I; Hae III; Hha I; Mbo I; Msp I; Taq I and Sau96 I (Figure 4.5).

4.3.5 RADIOLABELLING RESTRICTION FRAGMENTS FOR DNA SEQUENCING.

The large number of restriction sites within the pCFK23 insert allowed

FIGURE 4.4

ELECTROPHORESIS OF END-LABELLED pCFK23 INSERT DNA DIGESTED WITH Alu I, Hae III, Hha I and Msp I.

pCFK23 insert DNA was prepared by digestion of pCFK23 plasmid DNA with Pst I and elution of the insert fragment after polyacrylamide gel electrophoresis (data not shown). Pst I ends were labelled by terminal exchange as described in 2.2.19b followed by digestion with the appropriate restriction enzyme and polyacrylamide gel electrophoresis (tracks marked Pst I) to detect the terminally derived restriction fragments. The internal restriction fragments were detected by end-labelling after digestion with the restriction enzyme (tracks marked "All"). End-labelled Hae III cleaved pBR322 DNA was included as a molecular weight marker (tracks marked "M"). All other conditions were as described in the legend to Figure 4.2 except that DNA was detected by direct autoradiography.

Alu I M Hha I Hae III M Msp I

Pst I All Pst I All Pst I All Pst I All Labelled ends

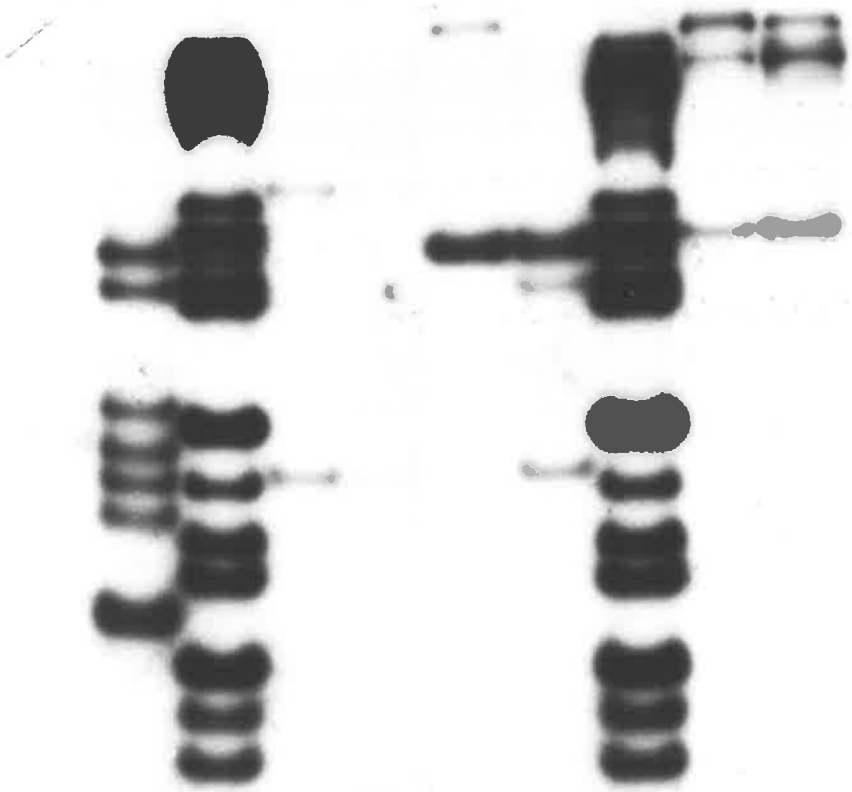


FIGURE 4.5

MAP OF THE RESTRICTION ENDONUCLEASE SITES IN THE INSERT OF THE cDNA CLONE pCFK23.

Mapping of the restriction endonuclease sites in the pCFK23 insert and orientation of the insert in the pBR322 vector DNA was achieved using the strategies described in the text (4.3.4).

The orientation of the insert is indicated by the distance of each end of the insert from the unique pBR322 Eco RI site. The boundaries of the insert are delineated by the downward arrows which show the position of the regenerated terminal Pst I sites.

The restriction endonuclease sites in pCFK23 are:

Alu I ..... A

Bam HI ..... B

Hinf I ..... f

Hae III ..... H

Hha I ..... h

Mbo I ..... M

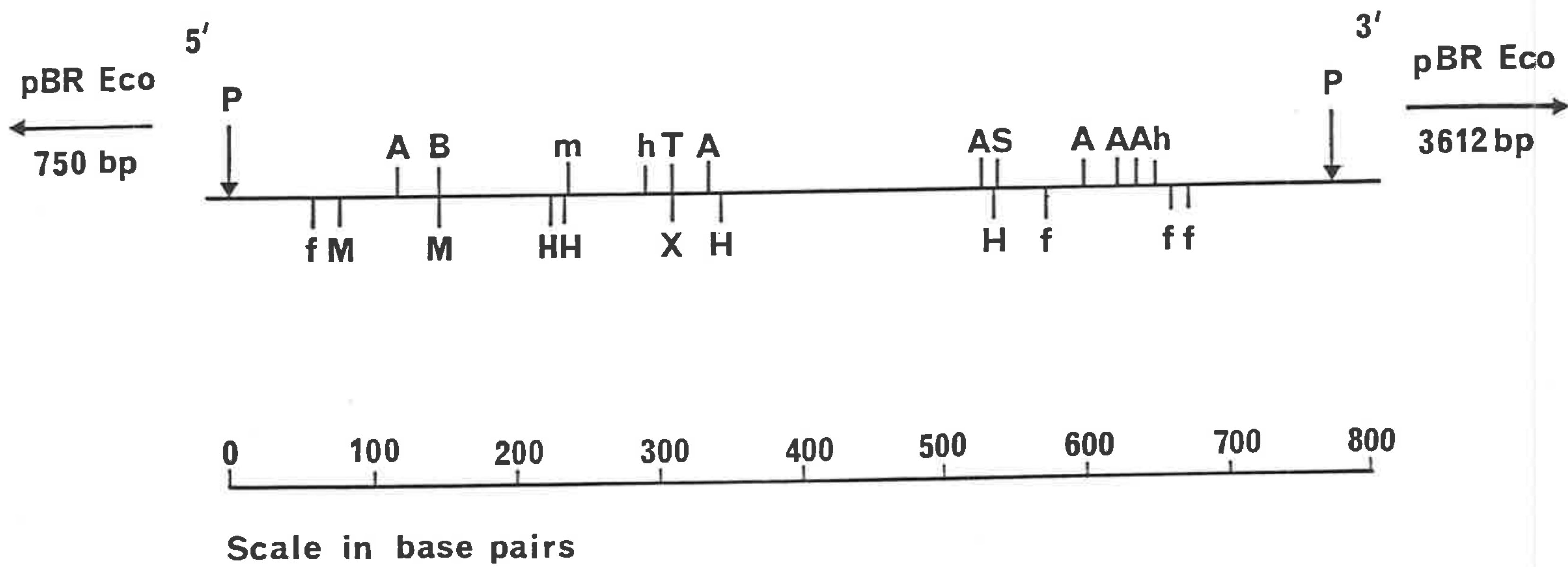
Msp I ..... m

Pst I ..... P

Sau96 I ..... S

Taq I ..... T

Xho I X



pBR Eco
750 bp
5′
P
A B
m
h T A
AS
A AAh
3′
P
pBR Eco
3612 bp
f M
M
HH
X H
H
f
f f
0
100
200
300
400
500
600
700
800
Scale in base pairs

sufficient radiolabelling of fragments, using either polynucleotide kinase for 5′-terminal labelling (2.2.19a) or 3′-terminal labelling with the Klenow fragment of DNA polymerase I, to enable overlapping DNA sequences to be determined throughout the length of the pCFK23 insert by the chemical cleavage method. Generally insert DNA was cut with an appropriate restriction enzyme and the resulting fragments terminally labelled using one of the two methods mentioned above. Labelled fragments were then fractionated on a polyacrylamide gel, detected by autoradiography and eluted as described in 2.2.18a. Occasionally the whole plasmid DNA was digested and labelled, when it was known that the resulting insert fragments could be resolved from labelled vector fragments.

4.3.6 SEPARATION OF ^{32}P-LABELLED TERMINI.

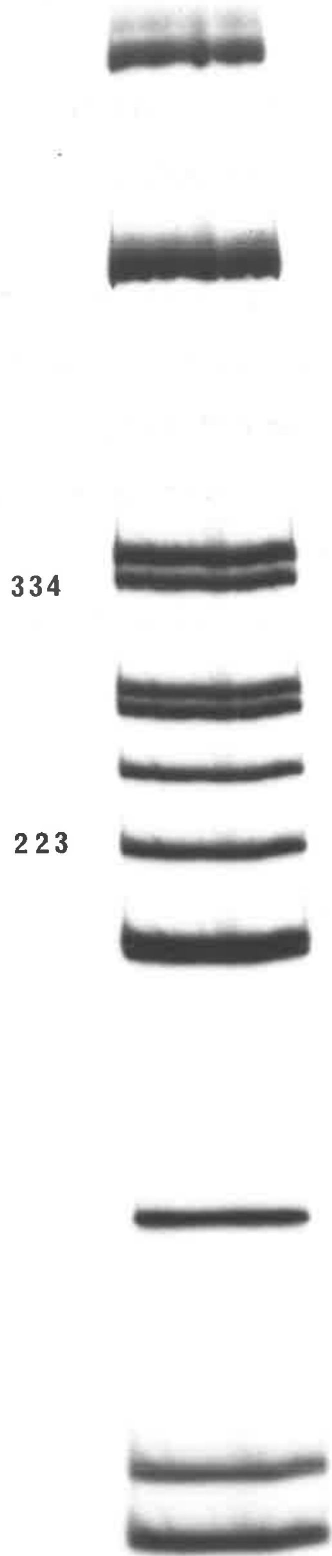
The chemical cleavage method of DNA sequencing requires that only one end of the DNA fragment to be cleaved should be labelled, while both the 3′ and 5′-terminal labelling procedures result in both termini being labelled. The two methods used to achieve a separation of labelled termini were either secondary restriction of labelled DNA or dissociation and separation of DNA strands.

The availability of an accurate restriction map greatly facilitated the use of the secondary restriction approach to separating labelled fragment ends. Ideally a second restriction enzyme would be chosen such that the resulting fragments could easily be resolved from each other. This approach is illustrated in Figure 4.6. Figure 4.6 shows the polyacrylamide gel fractionation of pCFK23 plasmid DNA after digestion with Sau96 I and end-labelling of the resulting fragments so that two insert fragments are resolved from the vector fragments, one of 223 bp and the other of 334 bp. The 223 bp fragment was excised, eluted from the gel and redigested with Hinf I resulting in the formation of 48 bp and 175 bp fragments, suitable for sequencing since they could be resolved from each other on another gel. This method could also be used for DNA fragments which initially migrated with vector fragments provided that, either the vector DNA was not cut by the second

FIGURE 4.6

AN EXAMPLE OF SECONDARY RESTRICTION OF A DNA FRAGMENT IN ORDER TO SEPARATE LABELLED TERMINI.

Total pCFK23 plasmid DNA was digested with Sau96 I and the fragments were labelled by end filling (2.2.19b) prior to electrophoresis on a 6% polyacrylamide gel. Two insert derived bands are shown separated from the pBR322 restriction fragments and are marked at the side of the gel by their length in base pairs (223 and 334). The 223 bp insert fragment was excised from the gel, the DNA was eluted (2.2.18a), and cut with a second restriction enzyme (Hinf I) to release two fragments which were separated by electrophoresis on a polyacrylamide gel. The autoradiography of this gel shows a small amount of the uncut 223 bp insert fragment and the 175 and 48 bp fragments resulting from Hinf I cleavage.


23
Sau 96 I
334
223
Hinf I
223
175
48

restriction enzyme, or the resulting vector fragments could be resolved from the insert fragments after secondary cutting.

An alternative method for the separation of labelled termini was to prepare dissociated DNA strands (Haward, 1972; Maxam and Gilbert, 1977, 1980). The procedure involves denaturation of the DNA fragments in DMSO followed by electrophoresis on a polyacrylamide gel, under conditions which do not allow reannealing of the complementary strands. As an example Figure 4.7 shows the separation of the component strands of the 222 bp and 334 bp Sau96 I fragments of pCFK23 shown in Figure 4.6a but prepared in a separate experiment.

4.3.7 CHEMICAL CLEAVAGE DNA SEQUENCING REACTIONS.

Base specific modification and cleavage were carried out as described in 2.2.21. After strand scission, samples were denatured in formamide loading buffer and applied to a denaturing urea polyacrylamide gel. Generally the base specific modification reactions were reliable and only guanine, purine, pyrimidine and cytosine specific base modification reactions were required to elucidate the sequence of most DNA fragments. Occasionally however, problems were encountered with the cytosine reaction and it became necessary to employ a strong adenine, weak cytosine cleavage reaction which easily overcame this problem. The separation of base specific cleavage products on a denaturing urea, polyacrylamide gel is exemplified by Figure 4.8 which shows the five reaction sequencing ladder generated from a 342 bp Taq I/Hha I insert fragment prepared by digestion of an end-labelled Taq I fragment with Hha I. From the long (80 cm) 6% gel more than 170 residues were read (not all of which are shown adjacent to the sequencing ladder), the last of these being more than 230 bases away from the labelled Taq I terminus. This gel shows the use of the strong adenine, weak cytosine reaction to overcome the problem with the cytosine track which is almost indistinguishable from the pyrimidine track. The extra reaction also aids in the discrimination of adenosine from guanosine residues.

FIGURE 4.7

AN EXAMPLE OF STRAND SEPARATION OF END-LABELLED DNA FRAGMENTS.

The 223 and 334 bp Sau96 I fragments from the pCFK23 insert were prepared and end-labelled exactly as described in the legend to Figure 4.6 but they were prepared in a separate experiment. The fragments were then denatured and their component strands separated by electrophoresis under conditions which do not allow reannealing of complementary strands (2.2.20a). The arrows at the side of the gel indicate the position of the component strands separated from double-stranded material which migrates faster under these conditions.

Sau 96I
334

Sau96I
223

FIGURE 4.8

AN EXAMPLE OF CHEMICAL CLEAVAGE DNA SEQUENCING REACTIONS.

A 342 bp Taq I/Hha I (see map, Figure 4.5) fragment labelled at the Taq I end was prepared from pCFK23 by secondary restriction, as described in the legend to Figure 4.6, and subjected to the five base specific chemical cleavage sequencing reactions described in 2.2.21. These products were then separated on an 80 cm 6% polyacrylamide 8 M urea gel and detected by autoradiography. For ease of handling the 80 cm gel was cut in half prior to autoradiography, thus resulting in a discontinuous sequencing ladder. The determined sequence is shown aligned with the appropriate band.

The base specific reactions were:

G. guanine

A. purine

A>C. strong adenine weak cytosine

T. pyrimidine

C. cytosine

G A A>C T C

GTAGT
GAGTA
AAGAA
GACGA
AAGAG
AGAAG
TTTGT
GGTCG
GTGTC
T
T
T
C
C
A
G
A
C
A
G
C
A
C
A
G
G
C
T
G
G
C
A
C
C
C
T
G
G
C
T
T
C
C
C
T
C
T
C
T
G
C
T
T
T
G
C
A
G
G
A
G
G

4.3.8 M13 CLONING AND DIDEOXY SEQUENCING.

Although the majority of the pCFK23 DNA sequence was obtained by use of the chemical cleavage method, some sequence was obtained, and much of it confirmed by use of M13 cloning and dideoxy sequencing. The dideoxy sequencing strategy employed for sequencing pCFK23 and all other clones presented in this thesis, is described below (see also 2.2.22 and 23).

Generally the plasmid DNA to be sequenced was digested with a range of restriction enzymes to determine which enzymes produced fragments of a suitable size for subcloning into M13 (100 to 500 bp), without the need for prior mapping of restriction sites. Suitable fragments were eluted from the gels and ligated into the appropriate M13 vector, e.g. a Bam HI cut vector for a Sau 3A fragment, or a Sma I cut vector for a blunt ended fragment, prior to transformation of *E. coli*. DNA from clones so obtained, were then sequenced by the dideoxy sequencing method as described in 2.2.23, and the complete recombinant plasmid sequence deduced by comparing the sequence of overlapping fragments.

Figure 4.9 illustrates a typical example of the fractionation of DNA from several cDNA clones (including pCFK23), after digestion with either Sau 3A or Hae III. The insert fragments, were excised, eluted from the gel and cloned into the appropriate vector, either cleaved with Bam HI or Sma I. Figure 4.10 shows a sequencing gel on which the products of sequencing reactions for seven of the M13 clones obtained from five of the insert fragments shown in Figure 4.9, have been fractionated. Although this is a short gel (40 cm) it was possible to read over 200 bases where the sequence was not interrupted by homopolymeric guanosine and to a lesser extent cytosine, which often tend to cause the sequence beyond the G-tail to be unreadable. Sequence ladder A, derived from the pCFK3 Sau 3A fragment of 660 bp (Figure 4.9) extends from the disrupted Bam HI site of M13, through 279 bases of pBR322 to the Pst I site, followed by a string of Gs, poly(T) and the insert sequence. Sequence ladder B shows the sequence of a 170 bp Sau 3A fragment from within the insert

FIGURE 4.9

AN EXAMPLE OF PREPARATION OF FRAGMENTS FROM cDNA CLONES FOR SUBCLONING INTO M13.

The cDNA clones pCFK15, 10 and 3 were digested with Sau3A and pCFK22 and 23 were digested with Hae III and all were end-labelled as described in 2.2.19b prior to electrophoresis and autoradiography. Only one fifth of each Sau3A digest was end-filled then it was combined with the remainder of the unlabelled digest, thus ensuring that the Sau3A sticky end was available for ligation into the Bam HI site of M13. The spots indicate insert fragments which were excised and eluted prior to subcloning into M13.

Sau3AI HaeIII
15 10 3 22 23

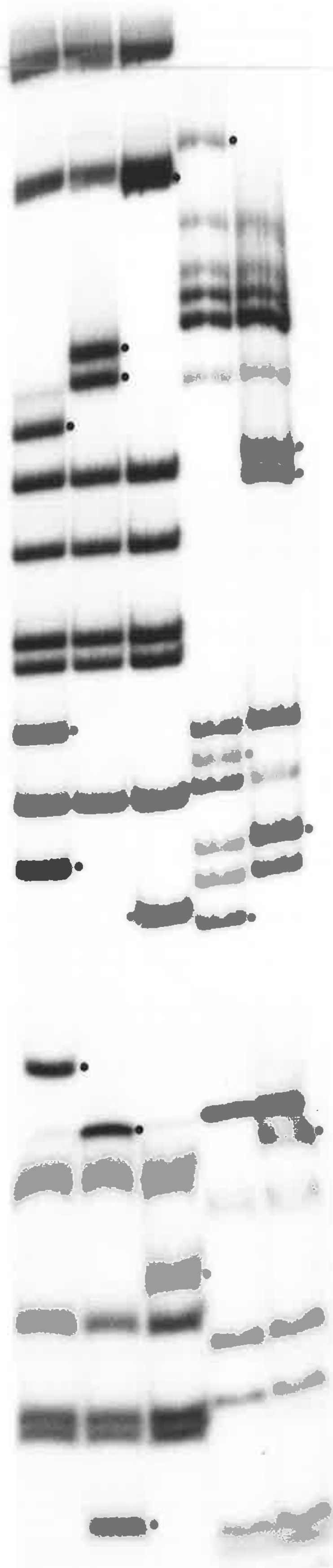


FIGURE 4.10

AN EXAMPLE OF A DIDEOXY SEQUENCING GEL.

Fragments eluted from the gel shown in Figure 2.9 were subcloned into M13 and subjected to the dideoxy sequencing reactions as described in 2.2.22 and 23.

The sequencing ladders are of M13 subclones of cDNA fragments as follows:

A. 660 bp Sau3A fragment of pCFK3.

B. 170 bp Sau3A fragment of pCFK3.

C, D and E. 96 bp Sau3A fragment of pCFK3.

F. 220 bp Hae III fragment of pCFK22.

G. 365 bp Hae III fragment of pCFK23.

(also see text 4.3.8).

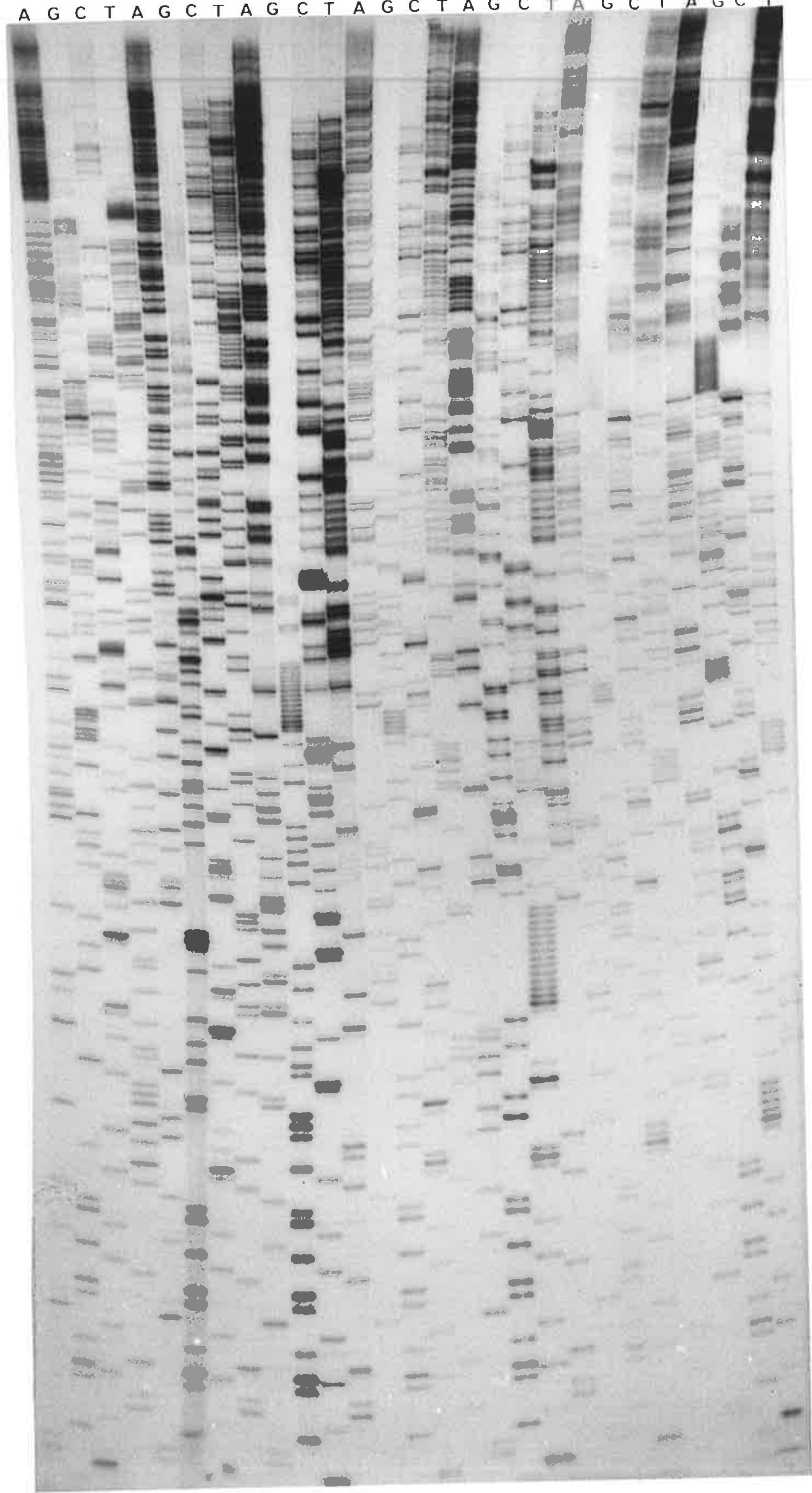
The four base specific chain termination reactions were:

A. dideoxyadenosine.

G. dideoxyguanosine.

C. dideoxycytosine.

T. dideoxythymidine.


A B C D E F G
A G C T A G C T A G C T A G C T A G C T A G C T A G C T

of pCFK3. Sequencing ladders C, D and E show the sequence of clones which all resulted from the cloning of the 96 bp Sau 3A fragment (Figure 4.9) of pCFK3. Ladders C and E result from opposite orientations of this fragment, C reads from outside the insert, through 62 bases of pBR322, to a G-tail with the insert beyond, while E reads through 20 bases of insert, to a 15 base C-tail followed by pBR322. Ladder D shows the sequence of an unknown fragment 1 to 2 bases shorter than the 96 bp pCFK3 fragment. The fragment was not derived from either the original vector pBR322 or the M13 vector (as determined by computer-assisted sequence comparisons), but probably came from *E. coli* chromosomal DNA contaminating the plasmid preparation of pCFK3. Sequencing ladders F and G were derived from a 220 bp pCFK22 Hae III fragment and a 365 bp pCFK23 Hae III fragment, respectively and show the same sequence from the Sma I cloning site, through 147 bases of pBR322 to G-tails, which lead into the respective insert sequences. Of the seven sequences displayed on the gel shown in Figure 4.10 only two, B and E, provided useful sequence data within the insert of a cDNA clone. This is often the case with initial approaches to the sequencing of several clones, but useful sequences invariably came quickly when M13 subclones of opposite orientation to those already sequenced were obtained using the C-test described by Messing and Vieira (1982), e.g. opposite orientations of the clones sequenced in ladders A, B, C, F and G would yield useful sequences within the insert.

4.3.9 THE NUCLEOTIDE SEQUENCE OF pCFK23.

The complete nucleotide sequence of pCFK23 is shown in Figure 4.11, including G and C-tails, poly(A) tract and intact terminal Pst I cloning sites. It was concluded that the mRNA from which pCFK23 was derived coded for embryonic feather β-keratin by comparing conceptual translations of the pCFK23 nucleotide sequence with short embryonic feather keratin peptide sequences (Walker and Rogers, 1976b) and the protein sequences of emu (O'Donnell, 1973) and silver gull (O'Donnell and Inglis, 1974) feather keratins. This comparison revealed that, including the AUU

FIGURE 4.11

THE COMPLETE NUCLEOTIDE SEQUENCE OF THE cDNA CLONE pCFK23.

The complete nucleotide sequence of the pCFK23 insert was determined using the techniques described in the text (4.3.5 to 8). The sequence of the antisense (mRNA) strand is presented showing the intact Pst I cloning sites, the G and C tails, polyadenylation sequence (AATAAA), poly(A) tract and the protein translation termination codon (TAA). The partial restriction map, shown below the nucleotide sequence, shows the restriction sites used in determining the sequence. The arrows indicate the direction and length of sequence obtained from a particular restriction site.

Restriction sites are:

H. Hae III

h. Hha I

Mbo. Mbo I

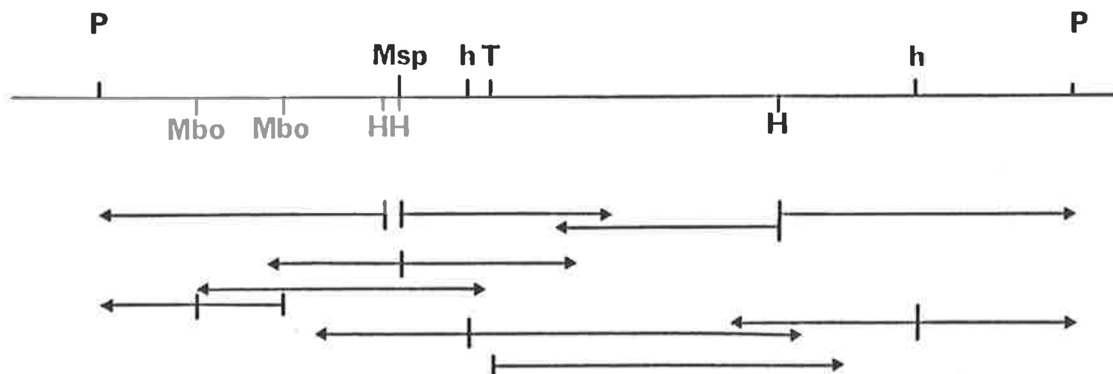
Msp. Msp I

P. Pst I

T. Taq I

pCFK23

```
         10         20         30         40         50         60
CTGCAGGGGG GGGGGGGGGG GGGGGGGGGG GGGGAGCCCT GTGTGAGGCA GTGCCAGGAC

         70         80         90        100        110        120
TCCCGTGTGG TGATCCAGCC CTCTCCCGTG GTGGTGACCC TGCCTGGTCC CATCCTCAGC

        130        140        150        160        170        180
TCCTTCCCCC AGAACACTGC TGTGGGATCC AGCACCTCTG CTGCTGTTGG CAGCATCCTG

        190        200        210        220        230        240
AGCCAGGAGG GAGTGCCCAT CTCTTCTGGA GGCTTTGGCT TCTCTGGCCT AGGTGGCCGG

        250        260        270        280        290        300
TTCTCTGGCA GGAGGTGCCT GCCATGCTAA AGCCAAGGTG AACGTCCTCT GAGCGCATCC
                       Termination Codon

        310        320        330        340        350        360
CAGTGATGCT CGAGCCAGCA CCAGACTGAG GATGTAGCTG CTGGCCGGGG TTTCCGGATG

        370        380        390        400        410        420
GGCTGACCAC CCTCCTGCCC TCCTGCAAAG CAGAGAGGGA AGCCAGGGTG CCAGCCTGTG

        430        440        450        460        470        480
CTGTCTGGAA ACACAGCCAG CAAACATCTT CTTCTCCTGC TTTCTTCTCA TCATCACAAG

        490        500        510        520        530        540
TCTTCGTTGT CTCGTTGCTG TCCTGTGCCA TGGGTTCATC CTGAAGCAAG CTGAGAGGGC

        550        560        570        580        590        600
CCTACTTCTT CCTCTCGCCA CATGAGGGAG GAAGACTCGC ACATCCTATT ATGCAGTTGC

        610        620        630        640        650        660
AGCTGATGCC AATCTGTTTA ACAGCTGCCT TGTAGCAGCT TTAAACTATG CGCTGCTTTG

        670        680        690        700        710        720
ATTCTTCTTT AGATTCAATA AAATTTATGC TGCATTGTAA TCTCAAAAAA AAAAAAAAAA
        Polyadenylation Signal

        730        740        750        760        770
AAAAACCCGG CTTCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCTGC AG
```



termination codon, pCFK23 contained 237 bases of coding region, coding for 78 amino acids from the carboxyl-terminus of an embryonic feather keratin. It can be noted that, although small regions of embryonic feather keratin proteins had been sequenced (Walker and Rogers 1976b), this was the first demonstration of the major proportion (78 out of 97 amino acids) of an embryonic feather keratin. A comparison of the coding potential of pCFK23 and a subsequently sequenced (Molloy *et al.*, 1982) chromosomal keratin gene (gene C), with embryonic feather keratin peptides and the protein sequences of emu and silver gull feather keratins (Figure 4.12), revealed a remarkable degree of sequence conservation at both the nucleotide and protein sequence level. It is probable that this degree of conservation extends throughout the embryonic feather keratin family, since other cDNA clones (presented later in this chapter) and genomic sequences continue this trend.

The codon usage in the coding region of pCFK23 shows a strong bias toward the use of C (34 times) over G and U (22 and 17) and particularly A (5) as the third base of a codon. This bias is even more evident when the genomic sequence, gene C, is examined (C, 50; G, 21; U, 25 and A, 4; Molloy *et al.*, 1982).

pCFK23 contains a long 3′-untranslated region of 434 bases. The entire 3′-untranslated region of the mRNA was represented in pCFK23, unlike pCFK17, a cDNA clone sequenced by E. Kuczek (B.Sc. Honours Thesis, unpublished data), which was 10 bases short of the polyadenylation site and, incidentally, only contained sufficient coding region for 11 amino acids. It is obvious from Figure 4.13 which shows the 3′-untranslated regions of pCFK23, gene C and pCFK17 aligned to maximize homology, that a remarkable degree of homology does exist between sections in the 3′-untranslated regions of the embryonic feather keratin genes. In some parts of the 3′-untranslated region the homology breaks down to between 40 and 60%, but there are two large blocks near the middle of the sequences and at the 3′-ends of about 70 bases and 60 bases, respectively, where the homology exceeds 90%. These blocks are more evident in Figure 4.14, where the degree

FIGURE 4.12

COMPARISON OF THE NUCLEOTIDE AND PROTEIN SEQUENCES OF pCFK23 WITH OTHER KERATIN SEQUENCES.

The nucleotide sequence of, and the protein encoded by the pCFK23 coding region is compared with the corresponding sequences from a genomic clone (gene C; Molloy *et al.*, 1982), emu (O'Donnell, 1973) and silver gull (O'Donnell and Inglis, 1974) feather keratins, an amino terminal (T1 b) and two carboxyl terminal (T1 c and T1 d) chick embryonic feather keratin peptides (Walker and Rogers, 1976b). The protein sequences are presented in single letter amino acid code and matches between the two nucleotide sequences are indicated by an asterisk.

```
        E  P  C  V  R  Q  C  Q  D  S  R  V  V  I  Q  P  S  P  V  V
pCFK23  GAGCCCTGTGTGAGGCAGTGCCAGGACTCCCGTGTGGTGATCCAGCCCTCTCCCGTGGTG
        ************ * ***************** ***************************
GeneC   GAGCCCTGTGTGCGCCAGTGCCAGGACTCCCGGGTGGTGATCCAGCCCTCTCCCGTGGTG
        E  P  C  V  R  Q  C  Q  D  S  R  V  V  I  Q  P  S  P  V  V
Emu     E  P  CL F  R  Q  C  Q  D  S  T  V  V  I  E  P  S  P  V  V
Gull    E  P  C  V  R  Q  C  E  A  S  R  V  V  I  Q  P  S  T  V  V
T1 b    Z  P  C  V  R  Q  C  Q  N  S  R  V  V  I  Z  P  S


        V  T  L  P  G  P  I  L  S  S  F  P  Q  N  T  A  V  G  S  S
pCFK23  GTGACCCTGCCTGGTCCCATCCTCAGCTCCTTCCCCCAGAACACTGCTGTGGGATCCAGC
        ************** ***************************** **** *** ******
GeneC   GTGACCCTGCCTGGACCCATCCTCAGCTCCTTCCCCCAGAACACCGCTGCGGGCTCCAGC
        V  T  L  P  G  P  I  L  S  S  F  P  Q  N  T  A  A  G  S  S
Emu     V  T  L  P  G  P  I  L  S  S  F  P  Q  N  T  V  V  GG S  S
Gull    V  T  L  P  G  P  I  L  S  S  F  P  Q  S  T  A  V  G  G  S


        T  S  A  A  V  G  S  I  L  S  Q  E  G  V  P  I  S  S  G  G
pCFK23  ACCTCTGCTGCTGTTGGCAGCATCCTGAGCCAGGAGGGAGTGCCCATCTCTTCTGGAGGC
        ******************** ********  ******************* ***** ***
GeneC   ACCTCTGCTGCTGTTGGCAGTATCCTGAGTGAGGAGGGAGTGCCCATCTCCTCTGGTGGC
        T  S  A  A  V  G  S  I  L  S  E  E  G  V  P  I  S  S  G  G
Emu     T  S  A  A  V  G  S  I  L  S  S  Q  G  V  P  I  S  S  G  G
Gull    A  S  A  A  V  G  N  E  LL A  S  Q  G  V  P  I  F  S  G  G


        F  G  F  S  G  L  G  G  R  F  S  G  R  R  C  L  P  C  *
pCFK23  TTTGGCTTCTCTGGCCTAGGTGGCCGGTTCTCTGGCAGGAGGTGCCTGCCATGCTAA
        ****** ********** *** **** ***************** ***** ******
GeneC   TTTGGCATCTCTGGCCTGGGTAGCCGCTTCTCTGGCAGGAGGTGTCTGCCCTGCTAA
        F  G  I  S  G  L  G  S  R  F  S  G  R  R  C  L  P  C  *
Emu     F  N  L  S  G  L  S  G  R  Y  S  G  A  R  C  L  P  C
Gull    F  G  L  G  G  L  G  C  Y     S  G  R  RG C  Y  P  C  P
T1 c                   L  S  G  R  F  C  G  R  R  C  L  P  C
T1 d                   L  G  G  R  Y  S  G  R  R  C  L  P  C
```

FIGURE 4.13

COMPARISON OF 3′-UNTRANSLATED SEQUENCES OF pCFK23 WITH GENE C AND pCFK17.

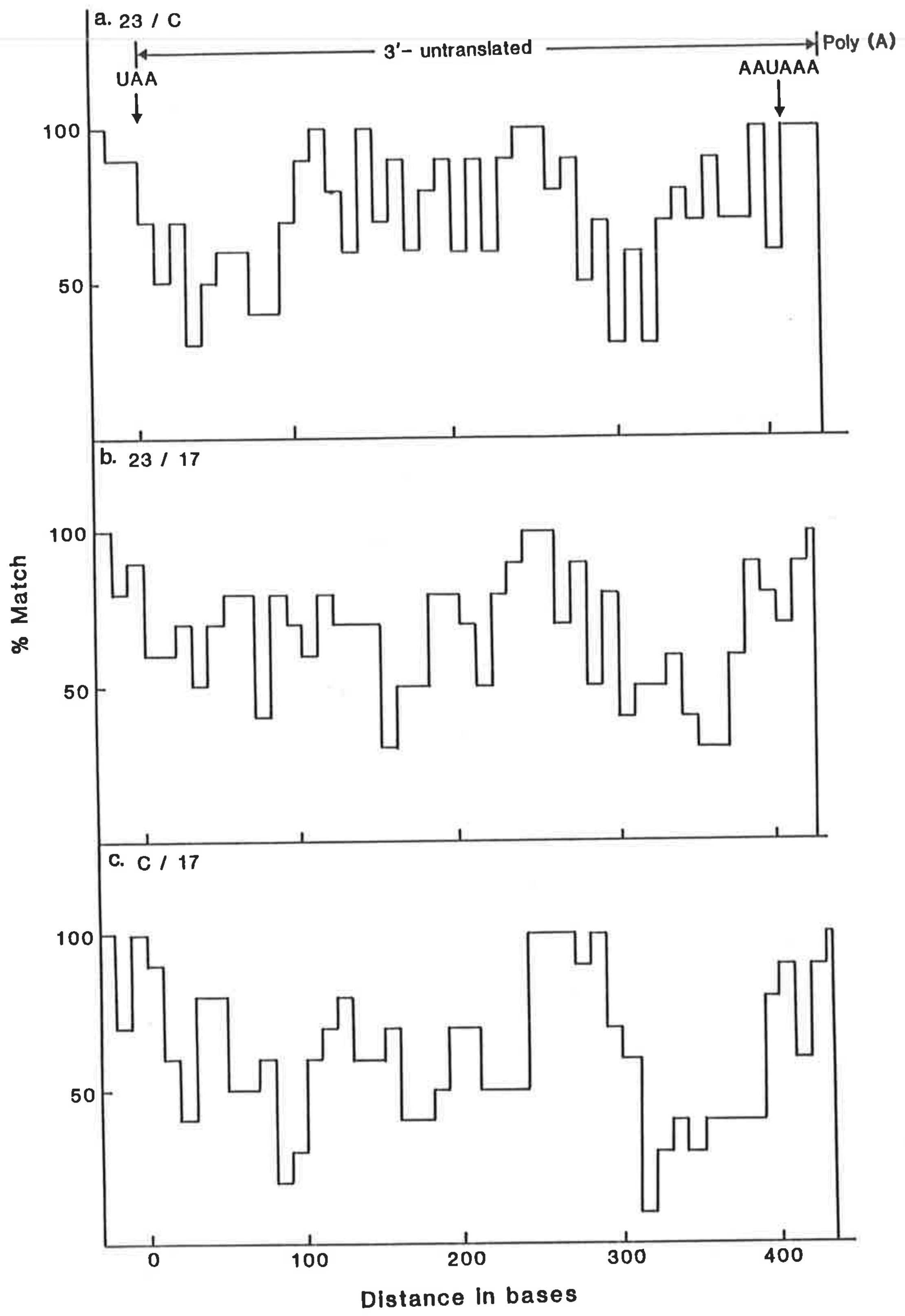
The nucleotide sequences from the 3′-untranslated regions of pCFK23, gene C (see Figure 4.12) and pCFK17 (E. Kuczek, B.Sc. Honours Thesis, unpublished data) are shown aligned to maximize homology. Matches between the pCFK23 sequence and the other two sequences are indicated by an asterisk. Hyphens indicate gaps in the sequences which were introduced to assist in their alignment. The positions of the protein translation termination codon and the polyadenylation sequence are highlighted by boxes.

```
          Coding Region    Termination Codon    3'-untranslated Region
Gene C    TTCTCTGGCAGGAGGTGTCTGCCCTGCTAAGGACGAGGTGTTCATCCCATGGATGCATCC
          **************** ***** ****** * * *****  * ***  **   ******
pCFK23    TTCTCTGGCAGGAGGTGCCTGCCATGCTAAAGCCAAGGTGAACGTCCTCTGAGCGCATCC
          *********** **** ****** ****** *    ***** * ****  ***** ***
pCFK17    TTCTCTGGCAGTAGGTACCTGCCCTGCTAAGGATGAGGTGGATGTCCCATGAGCCCATTG


Gene C    TCAGGAAACCCAAAGCTTGGTGCTGGACTGCTGACTGAGCTTCTGAGCA---------GGA
              **  * * * **  *   *  *****  **   **** *** *          ***
pCFK23    CAGTGATGCTCGA-GCCAGCA-CCAGACTGAGGATGTAGCTGCTGGCCGGGGTTTCCGGA
          *   ** ** * * *****   *** * ********  ** *** * * *    **** **
pCFK17    CCAGGAAGCCCAAAGCCAGATGCCATATTGAGGATGGAG-TGCAGCCAGCA--TTCCAGA


Gene C    TCCACTGAGCACCCTCCTGCTCTCCTGCAAAGCAA-AGAGGGAATTCAAG----TTGCC-
          *   **** *********** ************* ********  ** *        ****
pCFK23    TGGGCTGACCACCCTCCTGCCCTCCTGCAAAGCAG-AGAGGGAAGCCAGG----GTGCC-
          **** *** ** *** ** *  ** ***** **** **   *** ** **      ****
pCFK17    TGGGATGAGCAACCTGCT-CTGTC-TGCAA-GCAGGAG--TGAATCCTGGATGCCTGCCT


Gene C    AGCCTGTGCTGCCTGTAGACACAGACAGCAGCTGTCTTCTTCTTTTCCTTCTTTCTT-TC
          *********** *** * ****** *****          ******* **** ******* **
pCFK23    AGCCTGTGCTGTCTGGAAACACAGCCAGCAAACA---TCTTCTTCTCCTGCTTTCTTCTC
           * ****  ** *        ** ** *    *      ***********  *****  ***
pCFK17    TGTCTGTC-TGGCAATGCTGCCA-CCTGATGTCT---TCTTCTTCTCCCACTTTCCACTC


Gene C    ATCATTAGGGG--TTCTT-GTTGTGTCCTT--TGTCCTGTGCCTTGGGTTCATCCTGAAG
          ***** *   *     **** ***** ** **   ************************
pCFK23    ATCATCACAAG---TCTTCGTTGTCTCGTTGCTGTCCTGTGCCATGGGTTCATCCTGAAG
          *** ***   **     **       * *****  ************** ****************
pCFK17    ATCGTCATTAGGAGTCCATCTGGTCTC-CTGCTGTCCTGTGCCTTGGGTTCATCCTGAAG


Gene C    CAAGTTGAGATGGCCCTGCTTCTTCTTCCACTTGTCT---TGTGATG-GGGAAGACATGC
          **** ***** ****** ******* **       **        *** **** **
pCFK23    CAAGCTGAGAGGGCCCTACTTCTTCCTC------TCG---CCACATGAGGGAGGAAGACT
          **** ***** ** *** ********       *         *     ******** **  *
pCFK17    CAAGTTGAGATGGTCCTGCTTCTTCTTC------TTCTTGTCTTCTGAGGGAGAAA--CA


Gene C    ATCCCATCTTCCTGTAGTTTCCCTCCTTATG---------GCCAATATGTTTGCCAGCTGT
            * **** *   *  *        *  *  ** **        ****** *****  ******
pCFK23    CGCACATCCTATTATGCAGTTGCAGCTGAT----------GCCAATCTGTTTAACAGCTGC
          *     ****  *  **  *  ** ** **  **          **     * ***     *  **
pCFK17    CCTGAATCCCGTGGTGGTGGTGGAGTGGAGTGTCCGAGTGCTGGCTTTTTTCGTCGAAGC

                                                Polyadenylation sequence
Gene C    ATTTG-TAGCAGCCTTTAACT-TTCACTTTTTTGATTCCTTCTTTGAGCTCAATAAAATT
            *** ******* ** ****  * *  * **   ******* *******  ***********
pCFK23    C-TTG-TAGCAGCTTTAAACTATGCGCTGCTTTGATT-CTTCTTTAGATTCAATAAAATT
            *   **     **   *   **** **********  ***** ***   *  ****** **
pCFK17    ATGGGCAAGATCACTTCTATCATGCACTGCTTTGATTTCTTCTGTAGTCTGAATAAAGTT


Gene C    TATGCTGCATTGTAATCTC
          *******************
pCFK23    TATGCTGCATTGTAATCTC
          *********
pCFK17    TATGCTGCA
```

FIGURE 4.14

GRAPHICAL REPRESENTATION OF HOMOLOGY IN THE 3′-UNTRANSLATED REGIONS OF pCFK23 AND OTHER KERATIN SEQUENCES.

The nucleotide sequences from the 3′-untranslated regions of pCFK23, pCFK17 and gene C were aligned to maximize homology (see Figure 4.13), the sequences were taken in pairs, ie. pCFK23/gene C, pCFK23/pCFK17 and pCFK17/gene C and the percent homology was scored for 10 base blocks of sequence. These scores are represented by the histograms shown.


a. 23 / C
3'- untranslated
Poly (A)
UAA
AAUAAA
100
50
b. 23 / 17
100
50
% Match
c. C / 17
100
50
0
100
200
300
400
Distance in bases

of homology between the 3′-untranslated region of pCFK23 and the two other sequences is represented graphically.

4.3.10 SECONDARY STRUCTURE IN pCFK23.

When the nucleotide sequence of pCFK23 was examined for the presence of possible secondary structures, many were found in both the 3′-untranslated regions and to a lesser extent in the coding region. Understandably any conceptually determined secondary structures found in the coding region of pCFK23 appeared to be conserved when gene C was examined for the same structures but it is probably the protein coding function of the mRNA, rather than any constraint placed on the sequence by the need to produce a secondary structure, which was responsible for this observation. Generally, following the same logic as that presented in the previous sentence, secondary structures found in the 3′-untranslated region of pCFK23 were not conserved when the appropriate regions from gene C and pCFK17 were examined.

Two interesting structures in the 3′-untranslated region of pCFK23 were discovered however, one was interesting because of its magnitude and the other because of the degree to which the sequence was conserved in the appropriate region of gene C and pCFK17. The first structure, Structure A, shown in Figure 4.15, is close to the termination codon. It involves almost 175 bases of sequence, 122 of which are base paired, in one structure which appears to be quite stable (Δ G = –277 kilojoules). However, only portions of this structure can be found in gene C and pCFK17, so the overall structure is not conserved amongst the keratin genes. The other structure, Structure B found about 150 bases from the 3′-poly(A) tract, is much smaller and less stable than Structure A (Δ G = –29 kilojoules). The structure is of interest however, because it can be found in the equivalent position in gene C (Figure 4.15, Structure C), and slightly altered in pCFK17 (data not shown). If it exists in the cellular mRNA, it is possible that the structure itself, and not the nucleotide sequence, is being selected for, since the nucleotide sequences of gene

FIGURE 4.15

POSSIBLE SECONDARY STRUCTURES FOUND IN pCFK23.

The structures shown were conceptually derived from the nucleotide sequence of pCFK23.

Structure A : ($\Delta G = -277$ kilojoules) found close to the termination codon.

Structure B : ($\Delta G = -29$ kilojoules) found 150 bp from the 3′-poly(A) tract.

Structure C : A gene C structure found in the equivalent position to Structure B.

Structure A

AAA
CG
GC
TA
CG A
CG
TA
CG
CG
G CG A
TA
CG C
C C
TA
CG
CG
CG
AT
C CG C
AGC
TA
CG
GC
GC
GT
TG
A TGCT
G
GT CTG
CG
CG
TA
TA
TA
GGCA
G C
GCCGA
GC
GC
TA
CG
GC
GCTAA
ATGCA
TA
AT
GC
GT
AT
CGC T
TAT
GC
CATC
C C
AT
CG
GC
AT T
C T
CGC
AT
TCGTC
C T
GC
TA
AT
GC
TA
GCGCATCCCAGTCACAAGTCTT

Structure B

C A C A
C T
G C G
T A-G
C G-G
T A
C G
C G
T A
T A
C G
T A
CCCTACT CTC

Structure C

G T C
T T
T T
C G
A T
C G-A
C G-T
T A
T G
C G
T A
T A
C G
T A
CCCTGCT CAT

C and pCFK17 show only about 60% homology to the pCFK23 sequence in this region.

It should be emphasized that the secondary structures described here were all conceptually derived, there being no direct evidence for their presence *in vivo* or any indication of possible biological function or significance.

4.3.11 SELECTION OF A FULL LENGTH cDNA CLONE.

It was obvious since pCFK17 contained almost no coding region and pCFK23 was well short of being full length, yet contained the longest excisable insert of the cDNA clones tested, that the majority of cDNA clones would not be of full length. A full length cDNA clone, or one which extended to within a few bases of the 5′-end of the mRNA, would be extremely useful, since it would provide a direct means of confirming the 5′-untranslated region sequence and the position of the 324 base intron, which had since been indicated by the sequencing of several genomic genes (Molloy *et al.*, 1982; Gregg *et al.*, 1983). In order to select a cDNA clone which extended to the 5′-end of the mRNA, about 200 cDNA clones from the cDNA library were "screened" with a probe made by labelling an M13 subclone (as described in 2.2.19d) of a 326 bp fragment from the 5′-end of gene C, which contained 5′-flanking and intron sequences and 37 bases from the 5′-end of the mRNA. Figure 4.16 shows about 200 cDNA clones screened with this probe, at high and low stringency. The low stringency wash detected three clones, pCFK26, 32 and 72, while only pCFK26 was detected at high stringency, which indicated that pCFK32 and 72 contained sequences which were probably less related to the 5′-untranslated region of gene C than sequences in pCFK26. pCFK26 was therefore chosen for sequencing since it was most likely to extend to near the 5′-end of a keratin mRNA.

4.3.12 THE PARTIAL NUCLEOTIDE SEQUENCE OF pCFK26.

The partial nucleotide sequence of pCFK26 is shown in Figure 4.17. It was not considered essential to completely sequence pCFK26 since the sequence obtained

FIGURE 4.16

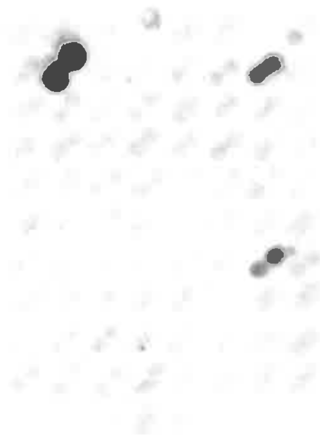
SELECTION OF A FULL LENGTH cDNA CLONE FROM THE cDNA LIBRARY.

Approximately 200 cDNA clones were grown on two nitrocellulose filters as diagonal duplicates and the plasmid DNA immobilized for use in the colony hybridization procedure (2.2.15). The filters were screened with an M13 subclone, ^{32}P-labelled as described in 2.2.19d, of a fragment containing the extreme 5′-end of a keratin gene (gene C). Two pBR322 containing colonies were grown as controls on each filter (on the extreme right of the left filter and at the bottom of the right filter). The clones were arranged in numerical order from left to right and top to bottom on the filter, though some early numbered clones were not present.

a. Filters washed at low stringency (2.0 x SSC, 65°C).

b. Filters washed at high stringency (0.2 x SSC, 65°C).

a



b

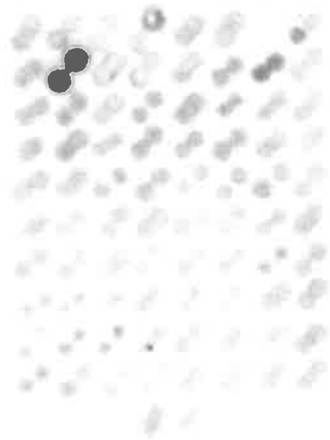


FIGURE 4.17

THE PARTIAL NUCLEOTIDE SEQUENCE OF THE cDNA CLONE pCFK26.

The partial nucleotide sequence of the pCFK26 insert was determined using the techniques described in the text (4.3.5 to 8). Base numbers 1 to 397 inclusive represent the sense strand and base numbers 398 to 1204 represent the antisense (mRNA) strand. The 5′-end of the mRNA, ie. cap site, is therefore at base position 398. The cap site, initiation codon and polyadenylation signal are boxed. The dashes represent two unsequenced, mirror image fragments of 212 bp. Note the presence of the altered Pst I site at the 3′-end of the insert (CGCAG not CTGCAG).

pCFK26

```
        10         20         30         40         50         60
CTGCAGGGGG GGGGGGGGGG GGGGGAGGGA AAGTGGGAGA AGAAGAAGAC ATCAGAGGCC

        70         80         90        100        110        120
---------- ---------- ---------- ---------- ---------- ----------

       130        140        150        160        170        180
---------- ---------- ---------- ---------- ---------- ----------

       190        200        210        220        230        240
---------- ---------- ---------- ---------- ---------- ----------

       250        260        270        280        290        300
---------- ---------- ---------- --GGCCAAAG CCACCGGAGG AGATGGGCAC

       310        320        330        340        350        360
TCCCTCCTCA CTCAGGATGC TTCCAACAGC AGCAGAGGTG CTGGAGCCCA CAGCAGTGTT

       370        380        390        400        410        420
CTGGGGGAAG GAGCTGAGGA TGGGTCCAGG CAGGGTCATC CACTTCTCTT GCCTTCTCCT
                                    Cap Site

       430        440        450        460        470        480
CCTTGGTGCA CAAGGTCGAC CTCCATCCCA CCACCATGTC CTGCTTCGAT CTGTGCCGTC
     Intron Position             Initiation Codon

       490        500        510        520        530        540
CTTGTGGCCC GACCCCACTG GCCAACAGCT GCAACGAGCC CTGTGTGCGC CAGTGCCAGG

       550        560        570        580        590        600
ACTCCCGGGT GGTGATTGAA CCATCTCCTG TGGTGGTGAC CCTGCCTGGA CCCATCCTCA

       610        620        630        640        650        660
GCTCCTTCCC CCAGAACACT GCTGTGGGCT CCAGCACCTC TGCTGCTGTT GGAAGCATCC

       670        680        690        700        710        720
TGAGTGAGGA GGGAGTGCCC ATCTCCTCCG GTGGCTTTGG CC-------- ----------

       730        740        750        760        770        780
---------- ---------- ---------- ---------- ---------- ----------

       790        800        810        820        830        840
---------- ---------- ---------- ---------- ---------- ----------

       850        860        870        880        890        900
---------- ---------- ---------- ---------- ---------- ----------

       910        920        930        940        950        960
---------- ----GGCCTC TGATGTCTTC TTCTTCTCCC ACTTTCCCTC ATCATCATTA

       970        980        990       1000       1010       1020
GGAGTCCATC TTGTCTCCTG CTGTTCTGTG CACTGGGTTC ATCCTGAAGC AAGTTTAGAT

      1030       1040       1050       1060       1070       1080
GGTCCTGCTT TTTCTCCCTC TCACTATGTG AGGGAGGAAG ATGTGCATCC CATTCTGGTG

      1090       1100       1110       1120       1130       1140
AAGGGATGTC TCGAAATAAC CAGCTGCTCT TGTTTGCAGC CTGGACCAGA TCCCTTTCCA

      1150       1160       1170       1180       1190       1200
CATGCATGCT TATTTACCTA GACTCAATAA AGTTTATGCT GCACCCCCCC CCCCCCCCCC
                     Polyadenylation Signal

GCAG
```

contained all of the 5′-untranslated region and extensive portions of sequence from both the coding and 3′-untranslated regions, which were well represented by sequence data already obtained for pCFK23, pCFK17 and several genomic genes.

Conceptual translation of the coding region of pCFK26 demonstrated that this clone was derived from a unique embryonic feather keratin mRNA not yet isolated as a cDNA or genomic sequence. The coding and 3′-untranslated regions of pCFK26 showed similar patterns of homology with the corresponding regions of pCFK23 and the genomic genes, which confirms and illustrates the idea that the keratins are a family of sequence related genes. The 5′-untranslated region of pCFK26 is shown in Figure 4.18, compared to the 5′-untranslated region of gene C. pCFK26 also followed the pattern within the genomic sequences in that the first 37 bases up to the position of the intron displayed a strong degree of homology with gene C, while the degree of homology decreased to about 70% for the 22 bases from the intron to the initiation codon. Although pCFK26 proved valuable in confirming conclusions of mRNA and gene structure indicated by the genomic sequencing data of Molloy *et al.* (1982) and Gregg *et al.* (1983), it also proved an interesting and perplexing exercise in the interpretation of DNA sequencing data, as described below.

4.3.13 pCFK26 - A TERATOGENY.

The length of the pCFK26 insert could not be determined by Pst I excision of the insert because one terminal Pst I site was not regenerated and, for reasons which will become clear later, attempts to size the insert by summation of its small component fragments invariably led to underestimates of insert length. The estimate of 1,200 bp for the length of the pCFK26 insert finally obtained from the sequence data was anomalous. A full length keratin cDNA clone should be about 870 bp, allowing for 20 bp of poly(A) and 30 bp of G+C-tails at each end.

An attempt to explain the structure and mode of origin of pCFK26 is presented in Figure 4.19. The sequence can be broken up into two sections, one of 807 bp,

FIGURE 4.18

COMPARISON BETWEEN THE 5′-UNTRANSLATED REGIONS OF pCFK26 and GENE C.

The nucleotide sequence from the 5′-end of pCFK26 is shown compared with the corresponding region in gene C. In order to demonstrate the presence of sequence blocks differing in homology, the sequence has been displayed in three blocks, from the 5′-end of the mRNA to the position of the intron (5′-leader sequence), from the intron to the protein initiation codon and the first 30 bases of the coding region. Matches between the two sequences are indicated by asterisks.

HOMOLOGY IN THE 5'-UNTRANSLATED

REGION OF THE KERATIN GENES

```
                       10        20        30
pCFK26     ATCCACTTCTCTTGCCTTCTCCTCCTTGGTGCACAAG         5'-LEADER
           ******************************* *****          SEQUENCE
Gene C     ATCCACTTCTCTTGCCTTCTCCTCCTTGGTGAACAAG
           ↑           10        20        30    ↑

       cap site                              5'-splice
```

```
             40        50        60
pCFK26     GTCGACCTCCATCCCACCACC             SEQUENCE FROM
           *** **  ****** **  **               INTRON TO
Gene C     GTCTACTCCCATCCTACAGCC             INITIATION CODON
          ↑  40        50     ↑60

       5'-splice            coding start
```

```
                        70        80
pCFK26     ATGTCCTGCTTCGATCTGTGCCGTCCTTGT         CODING REGION
           ************************** ***           SEQUENCE
Gene C     ATGTCCTGCTTCGATCTGTGCCGTCCCTGT
           ↑            70        80

       coding start
```

FIGURE 4.19

STRUCTURE AND PROBABLE ORIGIN OF THE cDNA CLONE pCFK26.

a. Schematic representation of the structure of the cDNA clone pCFK26 showing (by the divergent arrows) the regions which are mirror image complementary copies.

b. Proposed mechanism for the origin of pCFK26

(1). Synthesis of the first cDNA strand, which proceeded normally to the last base at the 5′-end of the mRNA.

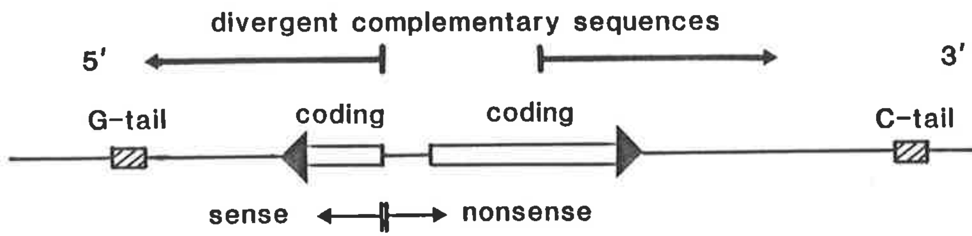
(2). The mRNA template is removed by boiling and the 5′-end of the cDNA snaps back upon itself to base pair in the middle of the coding region. Here the mRNA strand is shown still in close proximity to the cDNA strand so that the relationship between the snapped back cDNA and the mRNA coding region remains clear.

(3). The snapped back cDNA acts as a template for synthesis of a second strand which stops prematurely in the 3′-untranslated region.

(4). The double-stranded cDNA opens out to form a single-stranded cDNA containing mirror image complementary sequences. This acts as a template for second strand synthesis after normal AMV reverse transcriptase hair pin loopback, which proceeds to the end of the first strand.

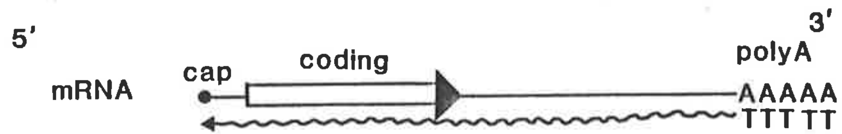
(5). After S_1 nuclease trimming, dC homopolymer tailing, annealing to dG-tailed pBR322 and transfection into *E.coli*, the cDNA clone pCFK26 is produced.

a.

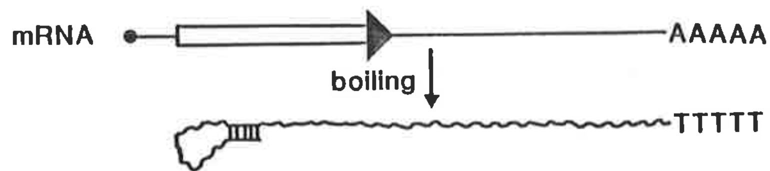


b.

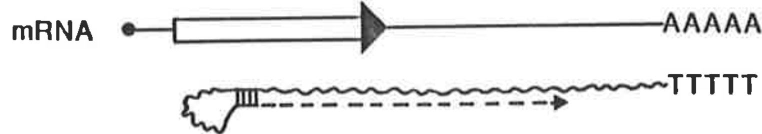
1.



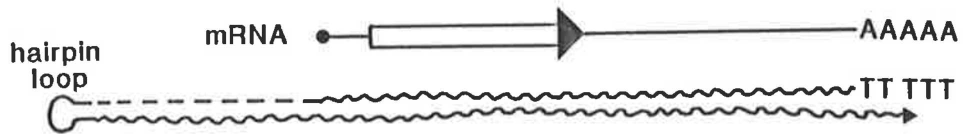
2.



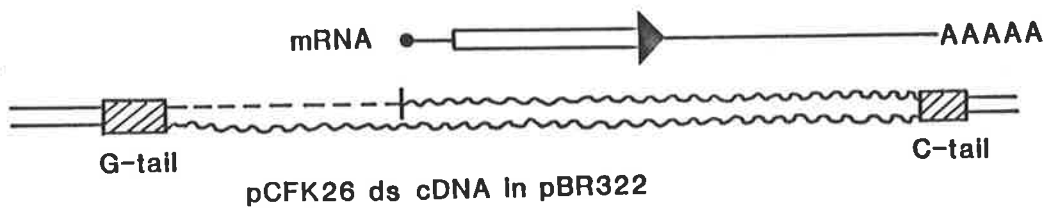
3.



4.



5.



which almost represents a full length copy of a keratin mRNA, being just 10 bases short at the 3′-end. The other shorter section of 392 bp which begins about half way through the coding region and ends about 240 bases short of the 3′-end of the mRNA, is intimately joined to the first section and represents a "sense" copy of the gene while the first section represents a "nonsense" copy, as does the mRNA. This results in two 372 bp sections of sequence being exact complementary copies of one another. This explains the low estimates of insert length obtained by the summation of small restriction fragments since fragments of identical length were produced from these complementary sections of sequence. It also explains the observation (R.B. Saint personal communication) that a large stem-loop structure was seen in the electron microscope when pCFK26 DNA was heteroduplexed to pBR322 DNA, since these regions would be able to base-pair under these conditions producing a structure with a 372 bp stem and a 176 nucleotide loop at the top and a 245 nucleotide at the base of the stem.

It is likely that pCFK26 arose, as shown in Figure 4.19 by the looping back of the 5′-end of a full length first strand cDNA copy of the mRNA which base-paired with a complementary sequence in the middle of the coding region of the first cDNA strand, using this as a primer the AMV reverse transcriptase elongated this new template to about 240 bases short of the full length cDNA, then this structure probably opened out to form a 1,200 residue single-stranded structure which was made double-stranded and trimmed to size by the normal S_1 procedure, to produce the double-stranded molecule which was tailed and became pCFK26, a most nefarious clone.

4.4 DISCUSSION.

4.4.1 GROUP I SEQUENCES CODE FOR EMBRYONIC FEATHER β-KERATIN.

The sequences of three cDNA clones, pCFK17, 23 and 26, which were previously determined to be Group I sequences by their Southern hybridization

characteristics (Saint, 1979) have been shown here to code for keratin proteins similar to short peptides from embryonic feather keratins and keratins from emu and silver gull feathers. This finding has confirmed the hypothesis that the Group I sequences were derived from keratin coding mRNAs (Saint, 1979). The finding of five genomic keratin genes (Molloy *et al.*, 1982; Gregg *et al.*, 1983) and the three cDNA clones described above, which are all different, further indicates that the number of genes coding for embryonic feather keratins is large (probably greater than 20), with the minimum number being eight. Some restriction data not presented in this thesis from six other putative Group I cDNA clones, pCFK32, 46, 69, 72, 145 and 151, which suggested that all of these clones were different from one another and from the three sequenced cDNA clones, shows that there are considerably more than eight keratin genes.

There is no doubt that the Group I cDNA clones code for keratin and since they were derived from embryonic feather tissue, they code for embryonic feather keratins. Despite having similar qualitative and quantitative patterns of sequence homology to the cDNA clones, the genomic sequences cannot be identified unequivocally as embryonic feather keratin genes, since lack of protein, or nucleotide sequence data for adult feather keratins does not eliminate the possibility that they are expressed in adult feathers. Indeed, the recent report of an adult chick feather keratin protein sequence revealed that this protein was nearly the same as the proteins encoded by the genomic and cDNA sequences (Arai *et al.*, 1983).

4.4.2 THE NEW MODEL FOR KERATIN mRNA STRUCTURE.

A rudimentary model for the organization of embryonic feather keratin mRNA was presented by Lockett *et al.* (1979), in which keratin mRNA was described as being about 760 nucleotides long including a poly(A) tract of 60 nucleotides, with at least 150 bases adjacent to the poly(A) tract being unique 3′-untranslated sequence. Further to the 5′-end of the mRNA and covalently linked to the 3′ unique sequence it was proposed that there existed, a repetitive sequence (short

and faithfully conserved or longer and mismatching), a G+C rich coding sequence of about 300 residues and presumably a 5′-untranslated sequence, although the arrangement of these latter mentioned three features within the remaining 550 bases was uncertain.

With the benefit of nucleotide sequencing data, a new model of embryonic feather keratin mRNA structure can be presented. Embryonic feather keratin mRNA can be described as a large family of mRNAs with a conserved length of about 850 nucleotides, including a poly(A) tract of 65 nucleotides (Morris and Rogers, 1979; this thesis, Chapter 3), containing a conserved protein coding sequence of 297 bp including initiation and termination codons, which specifies the production of a large family of homologous proteins containing 97 amino acid residues. The mRNAs contain a 3′-untranslated sequence which is consistently about 435 nucleotides long and contains two regions of between 60-70 nucleotides which are strongly conserved, while the remainder of the 3′-untranslated sequence is only poorly conserved (it is conceivable that the 3′ 150 nucleotides would appear unique by hybridization analysis, since more than half of this region is only poorly conserved). Keratin mRNA also has a 5′-untranslated region of 58 or 59 nucleotides which is divided, in the gene by the presence of an intron, into a rigidly conserved 37 nucleotide block of sequence at the 5′-terminus of the mRNA and a poorly conserved block of 21 or 22 nucleotides which abuts the coding sequence.

4.4.3 THE CODING SEQUENCE OF KERATIN mRNA.

It is obvious that the selection for the retention of protein coding regions is dictated by the requirement for that sequence to specify a functional protein, however this functional selection generally does not extend to the third base of a codon. Why then do keratin mRNAs exhibit a strong bias toward the use of certain codons? This bias resembles that seen for the chicken α and β-globin genes (Richards, 1979, 1980) but differs dramatically from chicken ovalbumin gene usage, in which all four bases are used with equal frequency in the third position of the codon (Grantham *et al.*,

UNIVERSITY OF ADELAIDE LIBRARY

1980). Observations of numerous mRNAs of different origins, have shown broad variations in codon usage, which is apparently species or tissue specific (Grantham *et al.*, 1980). It is probable that the differences in codon usage for these various chicken genes reflect tissue specific differences in tRNA populations or requirements for mRNA secondary structure, either of which may be important in the regulation of mRNA translation.

4.4.4 THE 3′-UNTRANSLATED SEQUENCE OF KERATIN mRNA.

The 3′-untranslated region of keratin mRNAs show areas of sequence which exhibit very low levels of homology. Presumably these mRNAs are derived from a common ancestral gene, and the degree to which portions of the 3′-untranslated regions, or any sequence, have diverged gives an indication of the time of separation of these genes. One would also expect that functionally important sequences would diverge at a slower rate than sequences that did not encode any important functions because of natural selective pressures. In light of this, it is interesting that two large blocks of sequence in the 3′-untranslated region of keratin mRNA appear to be as strongly conserved as the coding region, indicating a functional selection for their retention equivalent to the selective pressure to produce feather keratin proteins.

It is possible that these conserved regions are sites of secondary structure or for the binding of effector molecules which are involved in mRNA stability or perhaps the regulation of translation. It is easy to conceptually visualize secondary structures (as shown in Figure 4.15) in any region one might choose, but usually these are not conserved in other mRNA species and, strangely, structures which did seem to be conserved were found in areas of relatively poor sequence conservation, possibly indicating that these areas of sequence conservation are not required for, or involved in, secondary structures. It is possible that these sequences have a role at the level of the gene rather than the mRNA, perhaps by "editing" existing keratin genes in

the cluster, or producing new ones by initiating gene conversion and duplication events.

4.4.5 THE 5′-UNTRANSLATED REGION OF KERATIN mRNA.

An analysis of the 5′-untranslated region of keratin mRNA (Figure 4.18) revealed that the first 37 bases were strongly conserved, more so than even the coding region, while the remainder of the 5′-untranslated region was only poorly conserved. The junction of these two sequences marks the position of the only intron found in embryonic feather keratin genes. Why should this be so, when it has not been reported for any other gene system? It is unusual to have an intron in an untranslated section of a gene, particularly when there is only one intron (Heilig *et al.*, 1980).

The occurence of the intron is not related to the separation of functional protein domains brought together during evolution to form a single protein, as suggested by Gilbert (1979) as a rationale for the positions of a number of introns in several eukaryotic genes. The position of the intron led Molloy *et al.* (1982) to propose that the keratin intron is present to allow the expression of the same gene in different tissues (e.g., scale or claw) or at a different development stage (adult feather) using a different promoter and 5′-leader sequence. An analogous situation has been described for mouse α-amylase, where different primary transcripts in salivary glands and liver result in mRNAs in those tissues differing only in their 5′-untranslated regions (Young *et al.*, 1981). Many examples of viral RNAs which are differentially spliced to produce a range of viral mRNAs are known (Reddy *et al.*, 1978), and it is not unlikely that similar mechanisms should exist in the control of higher eukaryote gene expression.

If keratin genes are differentially expressed by this mechanism, at different developmental stages or in different tissues then it is possible to propose an explanation for the pattern of sequence homology which exists in the 5′-untranslated

region of keratin mRNA. In such a situation the 22 bases of 5′-untranslated region 3′ to the splice site of a particular mRNA would be the same in either tissue or developmental stage and could not be the site of differential control, there would therefore not be a selective pressure to maintain this sequence. On the other hand the 37 bases at the 5′-end, and any alternative 5′-leader sequence would be amenable to control mechanisms acting both at the level of the gene and the mRNA, and therefore be conserved because of its role in the control of gene expression. Developmental stage or tissue specific effector molecules could act on the 5′-leader sequences at the level of the gene, to enhance the promotion of the correct transcript for the particular tissue or stage, i.e. to "switch" the correct set of genes on. In this simple hypothetical case any set of keratin genes to be expressed in a particular tissue or stage, would have the same 5′-leader sequence and would all be correctly switched on by the effector molecule(s). At the level of the mRNA, stage or tissue specific effector molecules could act to regulate the amount of each protein made by potentiating or slowing the translation of these mRNAs, thus establishing quantitative differences in the levels of keratins expressed in these tissues or stages.

If such mechanisms were to exist one would expect to find alternative promoters and 5′-leader sequences in the 5′-flanking sequences of these genes. No such sequences have yet been found (K. Gregg, personal communication) since sequence data in the intergenic regions of the keratin genes is as yet quite limited.

It is clear from the results presented in this chapter that the Group I cDNA clones represent embryonic feather keratin mRNAs, the question regarding the nature of the Group II clones, however, remains unanswered.

CHAPTER 5.

NON-KERATIN mRNA SEQUENCES.

5.1 INTRODUCTION.

The results presented in Chapter 4 dealt with Group I cDNA clones, all of which proved to be derived from mRNA coding for embryonic feather β-keratins. The derived sequences of these mRNAs resulted in a vast improvement in the understanding of the chick feather β-keratin genes and the keratin proteins themselves. Furthermore, early reports of protein synthetic activity in the embryonic chick feather indicated that these keratins were the only major product (Kemp *et al.*, 1974a, b, 1975). A subsequent study (Walker and Rogers, 1976a) recognized the presence in the proteins extracted from embryonic feathers of a family of non-keratin proteins, which would not incorporate radioactivity when S-carboxymethylated using ^{14}C-iodoacetic acid, indicating the absence of cysteine residues (Dhouailly *et al.*, 1978). These proteins which were named "fast" proteins because of their high electrophoretic mobility on pH 2.7 gels, were found by Walker and Rogers (1976a) to comprise about 10% of total feather protein. It is interesting then, that the embryonic feather cDNA library, described in Chapter 4, appeared to contain two different classes of sequences (Group I and II), which occurred in approximately equal numbers. The hybridization characteristics of the Group II sequences, described in 4.1, were quite different from the Group I sequences, which were shown in Chapter 4 to code for typical feather β-keratins. So the question that emerges is, are these Group II sequences also keratins, or do they code for these fast proteins, or for some other as yet unidentified group of proteins, such as an unrelated group of "housekeeping" functions? A further question is, if these Group II clones do not code for keratins why are they so numerous amongst the cDNA library when 90% of the proteins synthesized in the differentiating embryonic feather are reputedly keratins?

The results presented in this chapter describe an analysis of these Group II cDNA clones. As part of this analysis, the complete nucleotide sequences of four Group II cDNA clones and studies on the occurrence of these sequences, both in

the cDNA library and in embryonic feather mRNA, are presented. The data indicate that a small number of Group II cDNA clones were derived from mRNAs which code for β-keratin proteins previously only identified in chick scales (Walker and Bridgen, 1976). The proteins were typical of both adult (Walker, 1974; Walker and Bridgen, 1976), and embryonic scale β-keratins (Wilton, 1983: Gregg *et al.*, 1984). A large number of the remaining Group II cDNA sequences were shown to be of the same type, but the data presented is insufficient to decisively identify the nature of the proteins for which they code, however it is clear that they code neither for embryonic scale keratins nor typical embryonic feather keratins.

5.2 SPECIFIC METHODS.

5.2.1 ELECTROPHORESIS OF DENATURED RNA.

RNA was denatured with glyoxal and electrophoresed on agarose gels essentially as described by McMaster and Charmichael (1977).

Several μg of poly(A)-plus feather RNA was incubated at 50°C for 60 minutes in a tightly closed tube containing an incubation mix consisting of 1M glyoxal, 10 mM NaH_2PO_4 pH 7.0 and 50% DMSO to totally glyoxylate and permanently denature the RNA. The denatured RNA sample was cooled to 20°C and 5 μl of loading buffer (50% glycerol, 10 mM NaH_2PO_4 pH 7.0, 0.4% bromophenol blue) was added and the sample immediately applied to a 1% agarose submarine gel in 10 mM NaH_2PO_4 pH 7.0 along with glyoxylated 18S and 28S ribosomal RNAs as molecular weight markers. The gel was run at less than 5 volts/cm with constant recirculation of the buffer in order to maintain the running pH below 8.0, above which the glyoxal can dissociate from RNA. When the bromophenol blue had migrated the required distance, the gel was stained with 0.5 μg/ml ethidium bromide and photographed to detect the mobility of the marker RNAs.

5.2.2 TRANSFER OF GLYOXYLATED RNA TO NITROCELLULOSE.

Glyoxylated RNA was transferred to nitrocellulose basically as described by Thomas (1980).

The gel was immediately placed in contact with a nitrocellulose sheet presoaked in 20 x SSCE and blotted essentially as described for the Southern transfer (2.2.16). The transfer was allowed to proceed for 24 hours, or at least overnight (14hr), and the RNA was permanently immobilized by baking the filter at 80°C under vacuum for 2 hours. Hybridization of probes to the immobilized RNA was carried out as described in 2.2.17.

5.2.3 HYBRID SELECTION.

5.2.3a BINDING DNA TO NITROCELLULOSE.

DNA was bound to nitrocellulose for the selection of mRNAs by hybridization using the procedure of Parnes *et al.*, (1981).

Recombinant plasmid DNA was dissolved in water to a concentration of 0.5 μg/ml and heated to 100°C for 10 minutes, chilled quickly on ice and incubated at room temperature for 20 minutes after the addition of an equal volume of 1 M NaOH. The DNA sample was neutralized by the addition of 0.5 volumes of 1 M HCl, 1 M NaCl, 0.3 M sodium citrate and 0.5 M Tris-HCl pH 8.0 and immediately chilled to 0°C. The denatured DNA was then applied as 5 μl aliquots to 3 mm squares of nitrocellulose, allowed to evaporate, then re-applied until about 20 μg of plasmid DNA was bound to the filter. The filters were air dried for 1 hour and gently washed twice in 6 x SSCE at room temperature, then blotted and allowed to air dry for 1 hour. DNA was permanently bound to the nitrocellulose filters by baking at 80°C for 2 hours.

5.2.3b HYBRIDIZATION AND ELUTION OF RNA.

Plasmid DNA only loosely bound to the filter was eluted by washing the filters in boiling H_2O prior to use for hybridization selection of RNA. The poly(A)-plus mRNA was added to hybridization solution (20 mM 1,4-piperazinediethanesulphonic acid (PIPES) pH 6.4, 0.2% SDS, 0.4 M NaCl, 100 μg/ml *E. coli* tRNA and 65% v/v freshly deionized formamide) and heated to 70°C for 10 minutes prior to being added to the filters containing bound plasmid DNA, which were then incubated at 50°C for 3 hours to allow mRNA to hybridize to the immobilized DNA.

Unbound RNA was removed from the filters by ten washes at 65°C using 10 mM Tris-HCl pH 7.6, 0.15 M NaCl, 1 mM EDTA and 0.5% SDS and two washes at 65°C using the same solution without SDS. Bound mRNA was eluted by boiling the individual filters in 200 μl of H_2O containing 20 μl of *E. coli* tRNA and snap chilling. Released mRNA was recovered by ethanol precipitation, ready for translation in the wheat germ cell-free system as described in 2.2.5.

5.3 RESULTS.

5.3.1 SELECTION OF A GROUP II CLONE FOR SEQUENCING.

It was obvious from the results presented in Chapter 4 that the elucidation of the nucleotide sequence of one or more Group II cDNA clones would be necessary to achieve the same level of understanding of the mRNAs from which these clones were derived, as was achieved for the keratin mRNAs. The results of Chapter 4 also indicated that the selected clone shoud have the longest possible insert, since many of the Group I cDNA clones were not full length copies of mRNA and often did not contain sufficient coding sequence to identify the encoded protein. This factor was even more important in the case of the Group II clones since of all the feather proteins, keratins were the only proteins for which sequence data was available, so a

complete coding sequence could be required to establish the protein coding capacity of the clone.

Of all the Group II cDNA clones examined in 4.3.1, i.e. pCFK 3, 10, 15, 19, 22, 5, 25, and 27, pCFK22 had the longest insert, which was composed of two Pst I fragments (370 and 515 bp) totalling 885 bp. A clone of this length, in theory, was long enough to contain a full length copy of a keratin mRNA including long poly(A) and G+C tails. It was therefore chosen for sequencing.

5.3.2 THE NUCLEOTIDE SEQUENCE OF pCFK22.

Figure 5.1 shows the complete nucleotide sequence of pCFK22 which was determined using the strategies described in Chapter 4 by the methods described in Chapter 2. The sequence neither displays a poly(A) tract nor the presence of a recognizable polyadenylation signal at either terminus, therefore, provided that it was derived from a polyadenylated mRNA, it must be missing at least 30 bases from the 3′-untranslated region of the mRNA.

Computer assisted comparisons of the 837 bases (excluding G+C tails) of the pCFK22 sequence with all available keratin sequences, including those of the sheep follicle (Powell *et al.*, 1983), did not reveal the presence of even the shortest keratin-specific nucleotide sequences. If pCFK22 was derived from a keratin-coding mRNA, it was expected that, considering the length of the pCFK22 insert, the comparative analysis described above should have revealed a positive result.

Conceptual translations of the pCFK22 nucleotide sequence into the possible protein sequences also did not reveal the presence of any identifiable protein sequence. The conceptual translation of the six possible reading frames of both strands of pCFK22 are shown in Figure 5.2. There was only one open reading frame from the six possible reading frames, which was larger than about 40 amino acids. This reading frame appeared at the 5′-end of the positive strand of pCFK22 and was 110 amino acids long, but was interrupted by the G+C tail and therefore did

FIGURE 5.1

THE COMPLETE NUCLEOTIDE SEQUENCE OF THE GROUP II cDNA CLONE pCFK22.

The complete nucleotide sequence of the pCFK22 insert was determined using the techniques described in Chapter 4 (4.3.5 to 8). The only recognizable features in the 884 bp of sequence are the G and C homopolymer tails and the terminal Pst I restriction sites (there is also an internal Pst I site starting at base position 510).

pCFK22

```
        10         20         30         40         50         60
CTGCAGGGGG GGGGGGGGAG CACTACTCAC CCTTCTGCCA GGAAGACCTG CATGGTCTCT

        70         80         90        100        110        120
GGGGCCTGAA TGACCACCGA TTCAAGCACC TATATGGCCT CCACCGAGAC CACCACCATG

       130        140        150        160        170        180
ACTACAACCA ACACTGGAGC CCTTATGGCT ACAACAGAAG CTTTGGGAGC CTGTATGGGA

       190        200        210        220        230        240
ACAGGAGCTT GAGCTCTCAT GGAGGCTATT ATGGGCATGG GGACTTCTTT GGTTTTGGCA

       250        260        270        280        290        300
CCGCCACCCC TACTTTTCTC AGATTGGCCA CAGATACTGG TATTGAAGCT GCTATCCATG

       310        320        330        340        350        360
CTAAACCACG CAGACACGGA AGTCAAGAAA TGAGGCAGAT CCTTTGCTAA GACCATGGTG

       370        380        390        400        410        420
CTGAGAAACG CTGGGTACCC ATACCTCTGG CTGTGTTCAG TGAGATGTGT GGGTGCCAGG

       430        440        450        460        470        480
CTGCATTAAC ACTACCTTAT ATGTATAACT ACACTAATAT ACACTAATAT TTCTATCTCT

       490        500        510        520        530        540
TCTGCTATCG CTCTAGAGGG CAATTTAAAC TGCAGAAGTA TTGAGGCACC CCTACACCCT

       550        560        570        580        590        600
GCTGAAAGTC CTGCTTTACT CCACTAGAGT GGGTTAGACA GTTTCCTTCT AAATTAAATA

       610        620        630        640        650        660
ATAGCCACTT TGTGTTGCTA TAGTGATATA TATGGTCCCG TGTCAATGAG AATCAGACTT

       670        680        690        700        710        720
TAACTGCCAT GAAGAAGTCC ACTTTGCTCA ACTATGAACT TCTCCAAAGA CAACACATAG

       730        740        750        760        770        780
AGAACATTAC ATATCTATGC TGTGCCTCAT ATGGTAGTGA TCAGAATTTA TGAAGAACTC

       790        800        810        820        830        840
GTGAGCAAGA AACGTGCCTG GAGTGATGAA TTCCATGAAG AGAGATCATC ATCTGGTATC

       850        860        870        880
AGACCCACTC TACTTGCTGA AGACCCCCCC CCCCCCCCCT GCAG
```

FIGURE 5.2

CONCEPTUAL TRANSLATION OF THE SIX POSSIBLE READING FRAMES OF pCFK22.

The nucleotide sequences of the positive and negative strands of pCFK22 are presented, numbered from top to bottom and bottom to top, in an antiparallel manner. The corresponding conceptual translations of these strands is presented, in single letter code, above and below the respective strands. The longest open reading frame is shown boxed at the 5′-end of the positive strand.

CONCEPTUAL TRANSLATIONS OF pCFK22

```
 L  Q  G  G  G  G  S  T  T  H  P  S  A  R  K  T  C  M  V  S  G  A  *  M  T  T  D  S  S  T  Y  M  A  S  T  E  T  T  T  M
 C  R  G  G  G  G  A  L  L  T  L  L  P  G  R  P  A  W  S  L  G  P  E  *  P  P  I  Q  A  P  I  W  P  P  P  R  P  P  P
  A  G  G  G  G  E  H  Y  S  P  F  C  Q  E  D  L  H  G  L  W  G  L  N  D  H  R  F  K  H  L  Y  G  L  H  R  D  H  H  H
CTGCAGGGGGGGGGGGGGAGCACTACTCACCCTTCTGCCAGGAAGACCTGCATGGTCTCTGGGGCCTGAATGACCACCGATTCAAGCACCTATATGGCCTCCACCGAGACCACCACCATG
        10        20        30        40        50        60        70        80        90       100       110       120
       875       865       855       845       835       825       815       805       795       785       775       765
GACGTCCCCCCCCCCCCCTCGTGATGAGTGGGAAGACGGTCCTTCTGGACGTACCAGAGACCCCGGACTTACTGGTGGCTAAGTTCGTGGATATACCGGAGGTGGCTCTGGTGGTGGTAC
 Q  L  P  P  P  P  A  S  S  V  R  R  G  P  L  G  A  H  D  R  P  G  S  H  G  G  I  *  A  G  I  H  G  G  G  L  G  G  G  H
  A  P  P  P  P  S  C  *  E  G  K  Q  W  S  S  R  C  P  R  Q  P  R  F  S  W  R  N  L  C  R  Y  P  R  W  R  S  W  W  W
 C  P  P  P  P  L  V  V  *  G  E  A  L  F  V  Q  H  T  E  P  A  Q  I  V  V  S  E  L  V  *  I  A  E  V  S  V  V  V  H

 T  T  T  N  T  G  A  L  M  A  T  T  E  A  L  G  A  C  M  G  T  G  A  *  A  L  M  E  A  I  M  G  M  G  T  S  L  V  L  A
*  L  Q  P  T  L  E  P  L  W  L  Q  Q  K  L  W  E  P  V  W  E  Q  E  L  E  L  S  W  R  L  L  W  A  W  G  L  L  W  F  W
 D  Y  N  Q  H  W  S  P  Y  G  Y  N  R  S  F  G  S  L  Y  G  N  R  S  L  S  S  H  G  G  Y  Y  G  H  G  D  F  F  G  F  G
ACTACAACCAACACTGGAGCCCTTATGGCTACAACAGAAGCTTTGGGAGCCTGTATGGGAACAGGAGCTTGAGCTCTCATGGAGGCTATTATGGGCATGGGGACTTCTTTGGTTTTGGCA
       130       140       150       160       170       180       190       200       210       220       230       240
       755       745       735       725       715       705       695       685       675       665       655       645
TGATGTTGGTTGTGACCTCGGGAATACCGATGTTGTCTTCGAAACCCTCGGACATACCCTTGTCCTCGAACTCGAGAGTACCTCCGATAATACCCGTACCCCTGAAGAAACCAAAACCGT
 S  C  G  V  S  S  G  K  H  S  C  C  F  S  Q  S  G  T  H  S  C  S  S  S  S  E  H  L  S  N  H  A  H  P  S  R  Q  N  Q  C
S  *  L  W  C  Q  L  G  *  P  *  L  L  L  K  P  L  R  Y  P  F  L  L  K  L  E  *  P  P  *  *  P  C  P  S  K  K  P  K  P
 V  V  V  L  V  P  A  R  I  A  V  V  S  A  K  P  A  Q  I  P  V  P  A  Q  A  R  H  S  A  I  I  P  H  P  V  E  K  T  K  A

 P  P  P  L  L  F  S  D  W  P  Q  I  L  V  L  K  L  L  S  H  L  N  H  A  D  T  E  V  K  K  *  G  R  S  F  A  K  T  M  V
H  R  H  P  Y  F  S  Q  I  G  H  R  Y  W  Y  *  S  C  Y  P  C  *  T  T  Q  T  R  K  S  R  N  E  A  D  P  L  L  R  P  W
 T  A  T  P  T  F  L  R  L  A  T  D  T  G  I  E  A  A  I  H  A  K  P  R  R  H  G  S  Q  E  H  R  Q  I  L  C  *  D  H  G
CCGCCACCCCTACTTTTCTCAGATTGGCCACAGATACTGGTATTGAAGCTGCTATCCATGCTAAACCACGCAGACACGGAAGTCAAGAAATGAGGCAGATCCTTTGCTAAGACCATGGTG
       250       260       270       280       290       300       310       320       330       340       350       360
       635       625       615       605       595       585       575       565       555       545       535       525
GGCGGTGGGGATGAAAAGAGTCTAACCGGTGTCTATGACCATAACTTCGACGATAGGTACGATTTGGTGCGTCTGTGCCTTCAGTTCTTTACTCCGTCTAGGAAACGATTCTGGTACCAC
 R  W  G  *  K  E  *  I  P  W  L  Y  Q  Y  Q  L  Q  *  G  H  *  V  V  C  V  R  F  D  L  F  S  A  S  G  K  S  L  G  H  H
V  A  V  G  V  K  R  L  N  A  V  S  V  P  I  S  A  A  I  W  A  L  G  R  L  C  P  L  *  S  I  L  C  I  R  Q  *  S  W  P
 G  G  G  R  S  K  E  S  Q  G  C  I  S  T  N  F  S  S  D  H  S  F  W  A  S  V  S  T  L  F  H  P  L  D  K  A  L  V  H  T

 L  R  N  A  G  Y  P  Y  L  W  L  C  S  V  R  C  V  G  A  R  L  H  *  H  Y  L  I  C  I  T  T  L  I  Y  T  N  I  S  I  S
C  *  E  T  L  G  T  H  T  S  G  C  V  Q  *  D  V  W  V  P  G  C  I  N  T  T  L  Y  V  *  L  H  *  Y  T  L  I  F  L  S
 A  E  K  R  W  V  P  I  P  L  A  V  F  S  E  H  C  G  C  Q  A  A  L  T  L  P  Y  H  Y  N  Y  T  N  I  H  *  Y  F  Y  L
CTGAGAAACGCTGGGTACCCATACCTCTGGCTGTGTTCAGTGAGATGTGTGGGTGCCAGGCTGCATTAACACTACCTTATATGTATAACTACACTAATATACACTAATATTTCTATCTCT
       370       380       390       400       410       420       430       440       450       460       470       480
       515       505       495       485       475       465       455       445       435       425       415       405
GACTCTTTGCGACCCATGGGTATGGAGACCGACACAAGTCACTCTACACACCCACGGTCCGACGTAATTGTGATGGAATATACATATTGATGTGATTATATGTGATTATAAAGATAGAGA
 Q  S  V  S  P  V  W  V  E  P  Q  T  *  H  S  T  H  T  G  P  Q  H  L  V  V  K  Y  T  Y  S  C  *  Y  V  S  I  N  R  D  R
 A  S  F  R  Q  T  G  H  G  R  A  T  N  L  S  I  H  P  H  W  A  A  N  V  S  G  *  I  Y  L  *  V  L  I  C  *  Y  K  *  R
 S  L  F  A  P  Y  G  Y  R  Q  S  H  E  T  L  H  T  P  A  L  S  C  *  C  *  R  I  H  I  V  V  S  I  Y  V  L  I  E  I  E

 S  A  I  A  L  E  G  N  L  N  C  R  S  I  E  A  P  L  H  P  A  E  S  P  A  L  L  H  *  S  G  L  D  S  F  L  L  N  *  I
L  L  L  S  L  *  R  A  I  *  T  A  E  V  L  R  H  P  Y  T  L  L  K  V  L  L  Y  S  T  R  V  G  *  T  V  S  F  *  I  K
 F  C  Y  R  S  R  G  Q  F  K  L  Q  K  Y  *  G  T  P  T  P  C  *  K  S  C  F  T  P  L  E  W  V  R  Q  F  P  S  K  L  N
TCTGCTATCGCTCTAGAGGGCAATTTAAACTGCAGAAGTATTGAGGCACCCCTACACCCTGCTGAAAGTCCTGCTTTACTCCACTAGAGTGGGTTAGACAGTTTCCTTCTAAATTAAATA
       490       500       510       520       530       540       550       560       570       580       590       600
       395       385       375       365       355       345       335       325       315       305       295       285
AGACGATAGCGAGATCTCCCGTTAAATTTGACGTCTTCATAACTCCGTGGGGATGTGGGACGACTTTCAGGACGAAATGAGGTGATCTCACCCAATCTGTCAAAGGAAGATTTAATTTAT
 R  S  D  S  *  L  A  I  *  V  A  S  T  N  L  C  G  *  V  R  S  F  T  R  S  *  E  V  L  T  P  *  V  T  E  K  *  I  L  Y
K  Q  *  R  E  L  P  C  N  L  S  C  F  Y  Q  P  V  G  V  G  Q  Q  F  D  Q  K  V  G  S  S  H  T  L  C  N  G  E  L  N  F
 E  A  I  A  R  S  P  L  K  F  Q  L  L  I  S  A  G  R  C  G  A  S  L  G  A  K  S  W  *  L  P  N  S  L  K  R  R  F  *  I

 I  A  T  L  C  C  Y  S  D  I  Y  G  P  V  S  M  R  I  R  L  *  L  P  *  R  S  P  L  C  S  T  M  N  F  S  K  D  N  T  *
*  *  P  L  C  V  A  I  V  I  Y  M  V  P  C  Q  *  E  S  D  F  N  C  H  E  E  V  H  F  A  Q  L  *  T  S  P  K  T  T  H
 N  S  H  F  V  L  L  *  *  Y  I  W  S  R  V  N  E  N  Q  T  L  T  A  M  K  K  S  T  L  L  N  Y  E  L  L  Q  R  Q  H  I
ATAGCCACTTTGTGTTGCTATAGTGATATATATGGTCCCGTGTCAATGAGAATCAGACTTTAACTGCCATGAAGAAGTCCACTTTGCTCAACTATGAACTTCTCCAAAGACAACACATAG
       610       620       630       640       650       660       670       680       690       700       710       720
       275       265       255       245       235       225       215       205       195       185       175       165
TATCGGTGAAACACAACGATATCACTATATATACCAGGGCACAGTTACTCTTAGTCTGAAATTGACGGTACTTCTTCAGGTGAAACGAGTTGATACTTGAAGAGGTTTCTGTTGTGTATC
 Y  G  S  Q  T  A  I  T  I  Y  I  T  G  H  *  H  S  D  S  K  L  Q  W  S  S  T  W  K  A  *  S  H  V  E  G  F  V  V  C  L
L  L  W  K  T  N  S  Y  H  Y  I  H  D  R  T  L  S  F  *  V  K  V  A  M  F  F  D  V  K  S  L  *  S  S  R  W  L  C  C  M
 I  A  V  K  H  Q  *  L  S  I  Y  P  G  T  D  I  L  I  L  S  *  S  G  H  L  L  G  S  Q  E  V  I  F  K  E  L  S  L  V  Y

 R  T  L  H  I  Y  A  V  P  H  M  V  V  I  R  I  Y  E  E  L  V  S  K  K  R  A  W  S  D  E  F  H  E  E  R  S  S  S  G  I
R  E  H  Y  I  S  M  L  C  L  I  W  *  *  S  E  F  M  K  N  S  *  A  R  N  V  P  G  V  M  N  S  M  K  R  D  H  H  L  V
 E  N  I  T  Y  L  C  C  A  S  Y  G  S  D  Q  N  L  *  R  T  R  E  Q  E  T  C  L  E  *  *  I  P  *  R  E  I  I  I  W  Y
AGAACATTACATATCTATGCTGTGCCTCATATGGTAGTGATCAGAATTTATGAAGAACTCGTGAGCAAGAAACGTGCCTGGAGTGATGAATTCCATGAAGAGAGATCATCATCTGGTATC
       730       740       750       760       770       780       790       800       810       820       830       840
       155       145       135       125       115       105        95        85        75        65        55        45
TCTTGTAATGTATAGATACGACACGGAGTATACCATCACTAGTCTTAAATACTTCTTGAGCACTCGTTCTTTGCACGGACCTCACTACTTAAGGTACTTCTCTCTAGTAGTAGACCATAG
 S  C  *  M  D  I  S  H  R  M  H  Y  H  D  S  N  I  F  F  E  H  A  L  F  T  G  P  T  I  F  E  M  F  L  S  *  *  R  T  D
S  F  M  V  Y  R  H  Q  A  E  Y  P  L  S  *  F  K  H  L  V  R  S  C  S  V  H  R  S  H  H  I  G  H  L  S  I  M  M  Q  Y
 L  V  N  C  I  *  A  T  G  *  I  T  T  I  L  I  *  S  S  S  T  L  L  F  R  A  Q  L  S  S  N  W  S  S  L  D  D  D  P  I

 R  P  T  L  L  A  E  D  P  P  P  P  P  C
S  D  P  L  Y  L  L  K  T  P  P  P  P  P  A
 Q  T  H  S  T  C  *  R  P  P  P  P  P  L  Q
AGACCCACTCTACTTGCTGAAGACCCCCCCCCCCCCCCTGCAG
       850       860       870       880
        35        25        15         5
TCTGGGTGAGATGAACGACTTCTGGGGGGGGGGGGGGGACGTC
 S  G  S  *  K  S  F  V  G  G  G  G  G  A
*  V  W  E  V  Q  Q  L  G  G  G  G  G  R  C
 L  G  V  R  S  A  S  S  G  G  G  G  G  Q  L
```

not contain the correct methionine initiation codon. This protein sequence was far too long to be either a feather keratin, known to contain 97 amino acids, or a fast protein, which at that time had been estimated by electrophoresis on SDS gels to be 80 amino acids long (Walker and Rogers, 1976a). The amino acid composition was different too, for it contained lysine, methionine and histidine residues all of which are almost completely absent from feather keratins. The clone was also unusual since, if the reading frame described above was the correct one, the 3'-untranslated region must be well in excess of 500 bases, which is unusual for small proteins like those found in the feather.

Since the sequence of pCFK22 offered nothing to positively identify it as having been derived from a mRNA, the possibility was considered that it was derived from a fragment of ribosomal RNA, the only other possible source of RNA in polysomes. This idea was abandoned because no sequence homology between pCFK22 and all available ribosomal sequences was found by computer comparisons and the pattern of bands detected in Hind III cut chick genomic DNA by pCFK22 sequences (Figure 4.1) was different from that detected with ribosomal cDNA (S. Wilton personal communication). Saint (1979) also found that the cDNA library contained no ribosomal sequences.

The identity of pCFK22 was left as an open question, and remained an enigma.

5.3.3 pCFK22 SEQUENCES IN THE cDNA LIBRARY.

If pCFK22 was derived from a feather mRNA, it was likely that other such sequences would be present in the feather cDNA library, which could allow extension of the pCFK22 sequence and thereby aid in the identification of the mRNA from which pCFK22 was derived. Figure 5.3 shows about 200 clones from the cDNA library screened with a 421 bp Sau 3A fragment from near the middle of pCFK22, subcloned into M13 and labelled as described in 2.2.19d. The pCFK22 colony was not detected by this probe, as indicated by the arrow, because the inoculum for this

FIGURE 5.3

IDENTIFICATION OF SEQUENCES HOMOLOGOUS TO pCFK22 IN THE cDNA LIBRARY.

Approximately 200 cDNA clones from the library were grown on two nitrocellulose filters exactly as described in the legend to Figure 4.16. These filters were screened using a 421 bp Sau 3A fragment from near the centre of the pCFK22 insert, subcloned into M13 and labelled as described in 2.2.19d. The filters were washed at high stringency (0.2 x SSC) and six clones were detected, pCFK10, 71, 120, 121, 164 and 195. pCFK22 was not detected since it did not grow on this filter. The position of pCFK22 is indicated by the arrow.

clone did not grow (note the absence of background radioactivity), but the probe effectively detected the presence of six other clones, pCFK10, 71, 120, 121, 164 and 195, which all gave a very positive response to the probe. These clones were likely to be identical or very similar to pCFK22 since the filters were washed at high stringency (0.2 x SSC), which suggests that pCFK22 is indeed derived from a feather mRNA and that this mRNA or similar mRNAs are fairly abundant in feathers.

pCFK10 was chosen for further study because it contained a fairly large insert (about 690 bp) and was the only positive clone to be previously identified as a Group II cDNA clone.

5.3.4 THE NUCLEOTIDE SEQUENCE OF pCFK10.

The complete nucleotide sequence of pCFK10 (Figure 5.4) was determined using the methods described in Chapter 2 following the strategies described in Chapter 4. The insert sequence contained both G and C tails and a poly(A) tract of 23 As and was therefore presumed to be derived from a polyadenylated mRNA of embryonic feathers. The sequence ATTAAA appears 15 residues from the poly(A) tract of pCFK10 and is presumed to be the sequence required for polyadenylation of the mRNA. The insert of 690 bp was larger than that of 470 bp determined by Pst I restriction of the insert (Figure 4.2a) because of the loss of the Pst I site at the 5′-end of pCFK10, presumably due to a deletion of an adenine during restriction and tailing of the pBR322 vector and the presence of an internal Pst I site 467 bp from the intact Pst I site at the 3′-end of pCFK10. No coding region was apparent from the conceptual translation of this sequence (data not shown), but comparison of the pCFK10 sequence with that of pCFK22 revealed that the two sequences were derived from the same mRNA (Figure 5.5). It was obvious from Figure 5.5 that pCFK10 and pCFK22 overlapped by 562 bp and that pCFK10 extended the sequence of pCFK22 by 93 bp at the 3′-end, including the poly(A) tract.

FIGURE 5.4

THE COMPLETE NUCLEOTIDE SEQUENCE OF THE GROUP II cDNA CLONE pCFK10.

The complete nucleotide sequence of the pCFK10 insert was determined using the techniques described in Chapter 4 (4.3.5 to 8). The sequence of the antisense (mRNA) strand is presented. Note the presence of the poly(A) tract, the unusual polyadenylation sequence (AUUAAA) and the altered Pst I site at the 5′-end of the insert (CTGCG not CTGCAG).

pCFK10

```
        10         20         30         40         50         60
CTGCGGGGGG GGGGGGGGGG CCACGCAGAC ACGGAAGTCA AGAAATGAGG CAGATCCTTT

        70         80         90        100        110        120
GCTAAGACCA TGGTGCTGAG AAACGCTGGG TACCCATACC TCTGGCTGTG TTCAGTGAGA

       130        140        150        160        170        180
TGTGTGGGTG CCAGGCTGCA TTAACACTAC CTTATATGTA TAACTACACT AATATACACT

       190        200        210        220        230        240
AATATTTCTA TCTCTTCTGC TATCGCTCTA GAGGGCAATT TAAACTGCAG AAGTATTGAG

       250        260        270        280        290        300
GCACCCCTAC ACCCTGCTGA AAGTCCTGCT TTACTCCACT AGAGTGGGTT AGACAGTTTC

       310        320        330        340        350        360
CTTCTAAATT AAATAATAGC CACTTTGTGT TGCTATAGTG ATATATATGG TCCCGTGTCA

       370        380        390        400        410        420
ATGAGAATCA GACTTTAACT GCCATGAAGA AGTCCACTTT GCTCAACTAT GAACTTCTCC

       430        440        450        460        470        480
AAAGACAACA CATAGAGAAC ATTACATATC TATGCTGTGC CTCATATGGT AGTGATCAGA

       490        500        510        520        530        540
ATTTATGAAG AACTCGTGAG CAAGAAACGT GCCTGGAGTG ATGAATTCCA TGAAGAGAGA

       550        560        570        580        590        600
TCATCATCTG GTATCAGACC CACTCTACTT GCTGAAGACC CTTTTGTAAA GCTGCAAACT

       610        620        630        640        650        660
TATTCTTTCT GTAACTTCTC ACTCTTTATT CATTAAAATG ATTCTGCATA GTAAAAAAAA
                                   Polyadenylation Sequence

       670        680        690
AAAAAAAAAA AAAAACCCCC CCCCCCCTGC AG
```

FIGURE 5.5

COMPARISON BETWEEN THE NUCLEOTIDE SEQUENCES OF pCFK10 AND pCFK22.

The nucleotide sequences of the two Group II cDNA clones pCFK10 and pCFK22 are shown aligned for comparative purposes. Matches between the two sequences are indicated by an asterisk. The pCFK10 sequence is not shown beyond the poly(A) tract, i.e. the C tail and a small portion of the poly(A) tract have not been included. It is obvious that the two sequences are derived from the same gene.

pCFK10/pCFK22 SEQUENCE COMPARISON

```
            10        20        30        40        50        60        70        80        90       100       110       120
22  CTGCAGGGGGGGGGGGGGGAGCACTACTCACCCTTCTGCCAGGAAGACCTGCATGGTCTCTGGGGCCTGAATGACCACCGATTCAAGCACCTATATGGCCTCCACCGAGACCACCACCATG

           130       140       150       160       170       180       190       200       210       220       230       240
22  ACTACAACCAACACTGGAGCCCTTATGGCTACAACAGAAGCTTTGGGAGCCTGTATGGGAACAGGAGCTTGAGCTCTCATGGAGGCTATTATGGGCATGGGGACTTCTTTGGTTTTGGCA

           250       260       270       280       290       300       310       320       330       340       350       360
22  CCGCCACCCCTACTTTTCTCAGATTGGCCACAGATACTGGTATTGAAGCTGCTATCCATGCTAAACCACGCAGACACGGAAGTCAAGAAATGAGGCAGATCCTTTGCTAAGACCATGGTG
                                                     ** *        *   *******************************************************
10                                                  CTGCGGGGGGGGGGGGGCCACGCAGACACGGAAGTCAAGAAATGAGGCAGATCCTTTGCTAAGACCATGGTG
            10        20        30        40        50        60        70        80        90       100       110       120

           370       380       390       400       410       420       430       440       450       460       470       480
22  CTGAGAAACGCTGGGTACCCATACCTCTGGCTGTGTTCAGTGAGATGTGTGGGTGCCAGGCTGCATTAACACTACCTTATATGTATAACTACACTAATATACACTAATATTTCTATCTCT
    ************************************************************************************************************************
10  CTGAGAAACGCTGGGTACCCATACCTCTGGCTGTGTTCAGTGAGATGTGTGGGTGCCAGGCTGCATTAACACTACCTTATATGTATAACTACACTAATATACACTAATATTTCTATCTCT
           130       140       150       160       170       180       190       200       210       220       230       240

           490       500       510       520       530       540       550       560       570       580       590       600
22  TCTGCTATCGCTCTAGAGGGCAATTTAAACTGCAGAAGTATTGAGGCACCCCTACACCCTGCTGAAAGTCCTGCTTTACTCCACTAGAGTGGGTTAGACAGTTTCCTTCTAAATTAAATA
    ************************************************************************************************************************
10  TCTGCTATCGCTCTAGAGGGCAATTTAAACTGCAGAAGTATTGAGGCACCCCTACACCCTGCTGAAAGTCCTGCTTTACTCCACTAGAGTGGGTTAGACAGTTTCCTTCTAAATTAAATA
           250       260       270       280       290       300       310       320       330       340       350       360

           610       620       630       640       650       660       670       680       690       700       710       720
22  ATAGCCACTTTGTGTTGCTATAGTGATATATATGGTCCCGTGTCAATGAGAATCAGACTTTAACTGCCATGAAGAAGTCCACTTTGCTCAACTATGAACTTCTCCAAAGACAACACATAG
    ************************************************************************************************************************
10  ATAGCCACTTTGTGTTGCTATAGTGATATATATGGTCCCGTGTCAATGAGAATCAGACTTTAACTGCCATGAAGAAGTCCACTTTGCTCAACTATGAACTTCTCCAAAGACAACACATAG
           370       380       390       400       410       420       430       440       450       460       470       480

           730       740       750       760       770       780       790       800       810       820       830       840
22  AGAACATTACATATCTATGCTGTGCCTCATATGGTAGTGATCAGAATTTATGAAGAACTCGTGAGCAAGAAACGTGCCTGGAGTGATGAATTCCATGAAGAGAGATCATCATCTGGTATC
    ************************************************************************************************************************
10  AGAACATTACATATCTATGCTGTGCCTCATATGGTAGTGATCAGAATTTATGAAGAACTCGTGAGCAAGAAACGTGCCTGGAGTGATGAATTCCATGAAGAGAGATCATCATCTGGTATC
           490       500       510       520       530       540       550       560       570       580       590       600

           850       860       870       880
22  AGACCCACTCTACTTGCTGAAGACCCCCCCCCCCCCCCCTGCAG
    **************************           *     *
10  AGACCCACTCTACTTGCTGAAGACCCTTTTGTAAAGCTGCAAACTTATTCTTTCTGTAACTTCTCACTCTTTATTCATTAAAATGATTCTGCATAGTAAAAAAAAAAAAAAAAAAAAAAA
           610       620       630       640       650       660       670       680       690       700       710       720
```

Since it was now obvious that pCFK22 was derived from a real mRNA, the only open reading frame from pCFK22 of reasonable size was more closely examined (Figure 5.2). The region of 330 bp coding for 110 amino acids was tested using the Testcode analysis described by Fickett (1982), and was found to produce a very positive result, having a probability of greater than 98% of being a coding region, while, as a control, a 200 bp sequence from the 3′-end of the clone had a probability of less than 4% of being a coding region. It was also evident that, although the conceptual translation of this reading frame produced a protein much larger (110 amino acids) than fast protein (80 amino acids), which is a histidine rich protein, the pCFK22 protein contained 13 histidines out of 110 amino acids. Table 5.1 shows the conceptually derived amino acid composition of this protein compared with the amino acid compositions of partially purified fast protein and feather keratin. It is obvious from this table that the pCFK22 protein amino acid composition is remarkably similar to that of fast protein in that both these proteins have high levels of histidine, tyrosine and phenylalanine and low levels of cysteine and valine, while keratin does not. These data suggested that pCFK22 coded for fast protein yet the open reading frame in pCFK22 was larger than the estimated size of fast protein. It was possible that a sequencing error had been made in the coding region of pCFK22 such that an initiation codon early in the protein coding region, or another termination codon late in the protein coding region had been lost, but repeated nucleotide sequencing in the protein coding region failed to reveal any sequencing errors.

5.3.5 THE LENGTH OF THE pCFK22 mRNA.

The coding sequence in pCFK22 did not contain an initiation codon and was therefore incomplete and coded for a protein which was longer than 110 amino acids, but how much longer? If the protein coding region and therefore the mRNA were considerably longer than in pCFK22 it was possible that it coded for one of the minor high molecular weight (about 50,000 Mr) proteins of the feather (Figure 3.1),

TABLE 5.1

COMPARISON BETWEEN THE AMINO ACID COMPOSITIONS OF A pCFK22 PROTEIN, EMBRYONIC FEATHER KERATIN AND EMBRYONIC FEATHER FAST PROTEIN.

The theoretical amino acid composition of the protein coded by the open reading frame (ORF) in pCFK22 (see Figure 5.2) was determined by conceptual translation of the open reading frame and calculation of the content of the component amino acids expressed as residues per 100 residues. The amino acid compositions of embryonic feather keratin and fast protein (data of Powell, 1979) are included for comparison.

TABLE 5.1

COMPARISON BETWEEN THE AMINO ACID COMPOSITIONS OF A pCFK22 PROTEIN, EMBRYONIC FEATHER KERATIN AND EMBRYONIC FEATHER FAST PROTEIN.

amino acid	pCFK22 ORF.	fast§ protein	embryonic§ feather keratin
Cysteine	1.80	1.77	6.18
Asparagine	3.60	—*	—*
Aspartic acid	3.60	9.51*	6.79*
Threonine	4.50	1.74	5.37
Serine	7.21	8.46	12.68
Glutamine	3.60	—*	—*
Glutamic acid	5.41	4.73*	8.93*
Proline	3.60	4.45	10.93
Glycine	12.61	17.64	11.98
Alanine	4.50	1.90	4.99
Valine	0.00	2.86	8.01
Isoleucine	2.70	1.56	4.85
Leucine	9.01	7.43	7.36
Tyrosine	7.21	10.87	1.84
Phenylalanine	6.31	8.83	3.85
Lysine	1.80	1.43	0.16
Histidine	11.71	9.34	0.81
Arginine	7.21	7.42	5.04
Methionine	0.90	0.00	0.00
Tryptophan	1.80	0.00	0.00

§ Data from Powell, (1979).

* Values for Asp and Glu include Asn and Gln content, respectively.

but if it was only slightly longer it was most likely that it coded for fast protein with the discrepancy being in the estimated size of the protein.

To resolve the question, embryonic feather poly(A)-plus mRNA was glyoxylated, fractionated on a 1% agarose gel and transferred to nitrocellulose as described by Thomas (1980), with ribosomal RNAs in a parallel track as molecular weight markers. These filters were probed with the same pCFK22 specific probe as was used in Figure 5.3 and a pCFK23 (keratin specific) probe, then washed at high stringency and autoradiographed (Figure 5.6). The "Northern" blot showed the pCFK22 mRNA to be about 200 bases longer than keratin mRNA (about 1,050 bases compared to about 850 bases).

Since the combined sequences of pCFK22 and 10 accounted for about 985 bases of mRNA, allowing about 65 bases for the poly(A) tract, this left only 65 bases at the 5′-end of the mRNA unaccounted for by pCFK22, and allowing about 50 bases for the 5′-untranslated region of the mRNA, only a small portion of the coding sequences of the mRNA could have been missing from pCFK22. This suggested that the protein coded for by pCFK22 was at least 110 amino acids long but less than about 130 amino acids, and therefore could not be one of the 50,000 Mr minor proteins of feathers. It seemed likely then that pCFK22 coded for fast protein and the difference between the length of the pCFK22 protein and the molecular weight estimate of fast protein might be explained either by the presence of a leader sequence on fast protein which is post-translationally removed, or the molecular weight estimate of fast protein may have been a gross underestimate, despite the fact that fast proteins migrated much faster than the 10,000 Mr keratin proteins on SDS gels (Figure 3.1).

5.3.6 HYBRID RELEASE TRANSLATION.

In an effort to identify the protein encoded by pCFK22 this and other cDNA clones, the Group II clones pCFK3, 10,and 15 and the Group I cDNA clones pCFK23

FIGURE 5.6

DETECTION OF SEQUENCES COMPLEMENTARY TO pCFK22 IN EMBRYONIC FEATHER mRNA (NORTHERN BLOT).

Embryonic feather poly(A)-plus mRNA prepared as described in 2.2.1d, was glyoxylated, fractionated on a 1% agarose gel and transferred to nitrocellulose as described in 5.2.1 and 5.2.2. These filters were probed with the same pCFK22 probe described in the legend to Figure 5.3 and a pCFK23 (keratin specific) probe. The filters were washed at high stringency (0.2 x SSC) and autoradiographed. The position of 28 and 18S ribosomal RNAs were determined in a parallel track.

2 3 2 2

28—

18—

and 26, were analysed by translation of the mRNA which these clones selected from total feather mRNA as described in 5.2.3. Plasmid DNA from these clones was denatured and bound to nitrocellulose filters which were hybridized to total embryonic feather poly(A)-plus mRNA, then washed to remove unbound mRNA. mRNA was eluted by heating the mRNA/cDNA hybrids in water to release bound mRNA which was collected and translated in a wheat germ cell-free system as described in 2.2.5.

Figure 5.7 shows the products of these translations fractionated on an SDS slab gel. pBR322 was included in the experiment as a hybrid release control and poly(A)-plus mRNA was included in the translations to act as a marker of translational activity in the embryonic feather. The translation products were not immunoprecipitated since fast protein or keratin antibodies were not available at the time of the experiment, which resulted in a generally high background of feather proteins in all tracks. However one clone, pCFK3, gave an unambiguously positive result, in that it appeared to select a mRNA coding for a protein which co-migrated with fast protein. Although the result was not completely specific it also appeared that pCFK22 selected a mRNA coding for fast protein, yet pCFK10, which was derived from the same gene as pCFK22 did not show enrichment for any of the proteins of the feather. pCFK23 and 26, appeared to bind slightly less fast protein mRNA than was present in the poly(A)-plus track. pBR322 did not significantly bind any mRNA and the radioactivity in the pBR322 track probably represents wheat germ endogenous mRNA activity. pCFK15 was interesting in that apart from the background of feather proteins, there was an extra band just above the higher molecular weight keratins, at about 15,000 Mr, possibly an indication that pCFK15 coded for a scale like β-keratin.

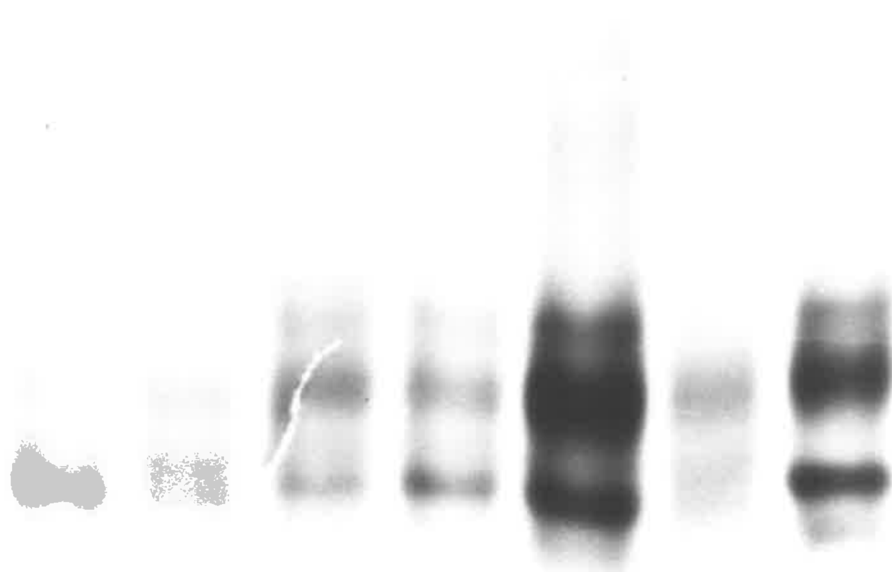
5.3.7 NUCLEOTIDE SEQUENCE OF pCFK3.

The result of the pCFK22 hybrid release experiment, although not completely unambiguous, added weight to the idea that pCFK22 coded for fast protein, however

FIGURE 5.7

HYBRID RELEASE TRANSLATION OF GROUP I AND GROUP II cDNA CLONES.

The Group II cDNA clones pCFK3, 10, 15 and 22 and the Group I cDNA clones pCFK23 and 26 were analysed by translation of the mRNAs which these clones selected from total feather poly(A)-plus mRNA, exactly as described in 5.2.3. Plasmid DNA from these clones was denatured and bound to nitrocellulose filters which were hybridized to total embryonic feather poly(A)-plus mRNA, then washed to remove unbound RNA. mRNA was eluted by heating the mRNA/cDNA hybrids in water to release bound RNA which was collected and translated in a wheat germ cell-free system as described in 2.2.5. These translation products were fractionated on a 10% polyacrylamide SDS slab gel and detected by fluorography. pBR322 was included in the experiment as a hybrid release control and poly(A)-plus mRNA was included in the translations to act as a marker of translational activity in the embryonic feather.


3
10
15
22
23
26
polyA+
pBR

pCFK3 clearly selected a mRNA coding for fast protein, or a protein of similar electrophoretic mobility. The complete nucleotide sequence of pCFK3 (Figure 5.8) was therefore determined to ascertain whether any similarity existed between the sequences of pCFK22 and 3, or whether the nucleotide sequence of pCFK3 gave a better indication of its coding potential than did pCFK22. The pCFK3 insert sequence of 648 bp contained G and C tails, and a poly(A) tract of 21 bases. It was interesting that the sequence ATTAAA appeared 17 bases from the start of the poly(A), and also 15 bases from the poly(A) of pCFK10, as this sequence is not the usual AAUAAA sequence found associated with the poly(A) sequences of eukaryotic mRNAs (Proudfoot and Brownlee, 1976), although the sequence AUUAAA has been reported in chick (Jung *et al.*, 1980), mouse (Hagenbuchle *et al.*, 1980) and human (Goeddel *et al.*, 1981) genes.

Conceptual translation of the pCFK3 sequence did not reveal the presence of any likely coding region (data not shown), although it may not have been present since only 583 bp of mRNA (excluding poly(A) and G+C tails) was represented in this clone, and the 3′-untranslated region deduced from the combined pCFK22 and 10 sequences was 589 bp in length. To gain some indication of whether pCFK3 was all 3′-untranslated region from a fast protein mRNA or at least the same type of mRNA as that from which pCFK22 was derived, the pCFK3 sequence was compared to the combined sequences of pCFK22 and 10 (data not shown). It was obvious that very little homology between the sequences existed, except at the extreme 3′-end where 17 out of the 21 3′-terminal bases of pCFK10 matched those in pCFK3, and this homology dropped to about 60% over the next 40 bases.

The homology at the 3′-end of pCFK3 and pCFK10 mRNAs indicated that the pCFK3 mRNA was probably from the same family of genes as pCFK10 and 22, i.e. fast protein genes. Although keratin genes exhibit more homology in the 3′-untranslated region than did pCFK3 and 10, other keratin genes such as the scale keratins do not possess any significant homology, even at their extreme 3′-ends

FIGURE 5.8

THE COMPLETE NUCLEOTIDE SEQUENCE OF THE GROUP II cDNA CLONE pCFK3.

The complete nucleotide sequence of the pCFK3 insert was determined using the techniques described in Chapter 4 (4.3.5 to 8). The sequence of the antisense (mRNA) strand is presented. Note the presence of the poly(A) tract and the unusual polyadenylation sequence (AUUAAA) also found in pCFK10.

pCFK3

```
        10         20         30         40         50         60
CTGCAGGGGG GGGGGGGGGG GGGGGGGGGG CTTCTGTGCT CTAGGGATCA CTATGGCTTT

        70         80         90        100        110        120
GAGGTCTGAA TAGTGCCATA GGTGGCTGTA CGGGGACTGG TATGGCTACC AAGCTGGTAT

       130        140        150        160        170        180
GGCAGCCGCC ACGGTCACCA CTTCGGTTCA CGATACGGTC AACGCTACGG CTACTGGGGT

       190        200        210        220        230        240
TGGTAAACCC AGCAGGAACG GCACTGACCT GTGAGGATCA ACCCCAGTCT GGACTGACGG

       250        260        270        280        290        300
CGAATCTCAG CAGCATCTTC CTAATGCACA GAGCCCAGTG ACAGCAGCTT CCCTCTGCGG

       310        320        330        340        350        360
AGGCATTGAG GCTCTGCTCT GCTGCCCAAG TGCTCAAAAT GATTGACCTT CTCCTGCTCT

       370        380        390        400        410        420
GTTTCCTCTC TTCCCTCCCA AACTCTCCTC TTCATCCACC GCTGGAGCAA AAGCTTGCTT

       430        440        450        460        470        480
CTCCTCAACA CCCATTCTGT GCAAACTCCA GAAATGCTGT TGGAGTGTCT GGCTTTTGTG

       490        500        510        520        530        540
TTTGTTGCAG TGATGTACAA AAGATATTCT GAATGGATGG TACAGGCACA GTGTACCATG

       550        560        570        580        590        600
CCTTGGGGAA TGTAGTCTTT CTGTTCTGGA ATATATCTGC CTTTCATATT AAAATCTACG
                                            Polyadenylation Sequence

       610        620        630        640        650        660
CTTCATCAGT AAAAAAAAAA AAAAAAAAAA ACCCCCCCCC CCCCCCCCCC CCCCCCCCCT

GCAG
```

(Wilton, 1983).

No firm conclusion as to the origin or coding potential of pCFK3, 22 or 10 could be made but the weight of circumstantial evidence suggested that all three were derived from mRNAs which coded for fast protein, the pCFK3 mRNA being distinct from the pCFK22/10 mRNA.

5.3.8 THE NUCLEOTIDE SEQUENCE OF pCFK15.

In contrast to pCFK3, 10 and 22, pCFK15 did not appear to select a fast protein mRNA when subjected to a hybrid release experiment (Figure 5.7), instead a larger protein of about 15,000 Mr was produced from the pCFK15 selected mRNA. The nature of this protein was determined by analysing the nucleotide sequence of pCFK15 (Figure 5.9). The pCFK15 insert sequence of 587 bp had both terminal Pst I sites regenerated and also contained two internal Pst I sites. pCFK15 also contained G and C tails and the sequence ATTAAA which appeared 13 bases from the short (7 bases long) poly(A) sequence at the 3′-end of the clone.

Conceptual translation of the coding region of the clone, shown above the nucleotide sequence in Figure 5.9 revealed that pCFK15 coded for a scale keratin protein. The 13 amino acid repeat sequence (repeated here $2\frac{1}{2}$ times), which is also reflected in the nucleotide sequence by a perfect 39 base repeat, distinguishes this protein from embryonic feather keratins. Feather keratins appear to have evolved from scale keratins by the deletion of these repeat sequences from the scale genes resulting in shorter proteins with similar sequences (Gregg *et al.*, 1984). It was interesting that the hybrid release experiment did correctly predict the size of the protein encoded by the pCFK15 mRNA. Also of interest was that this scale keratin mRNA, which codes for a protein larger than the feather keratins or fast protein had a 3′-untranslated region of only 353 bases, compared to 435 bases for feather keratin mRNA and 588 bases for the pCFK10/22 mRNA which was thought to code for fast protein.

FIGURE 5.9

THE COMPLETE NUCLEOTIDE SEQUENCE OF THE GROUP II cDNA CLONE pCFK15.

The complete nucleotide sequence of the pCFK15 insert was determined using the techniques described in Chapter 4 (4.3.5 to 8). The sequence of the antisense (mRNA) strand is presented. The conceptual translation of the coding region is presented in three letter code above the nucleotide sequence and the 13 amino acid repeat (present $2\frac{1}{2}$ times) has been boxed. Note the short poly(A) tract and the unusual polyadenylation sequence (AUUAAA) also found in pCFK10 and 3.

pCFK15

```
                                   Gly Gly Leu Tyr Gly Tyr|Gly Gly Ser Ser Leu
CT GCA GGG GGG GGG GGG GGG GGG GAC GGG GGC CTG TAT GGC TAT GGA GGC TCC TCC CTG G
             10          20          30          40          50          60

Gly Tyr Gly Gly Leu Tyr Gly Tyr|Gly Gly Ser Ser Leu Gly Tyr Gly Gly Leu Tyr Gly
 GC TAC GGG GGC CTG TAT GGC TAT GGA GGC TCC TCC CTG GGC TAC GGG GGC CTG TAT GGC T
             70          80          90          100         110         120

Tyr|Gly Arg Ser Tyr Gly Ser Gly Tyr Cys Ser Pro Tyr Ser Tyr Arg Tyr Asn Arg Tyr
 AT GGT AGA TCC TAT GGT TCT GGC TAC TGC AGC CCT TAC TCC TAC CGG TAC AAC AGG TAC C
             130         140         150         160         170         180

Arg Arg Gly Ser Cys Gly Pro Cys ***
 GC CGT GGC AGC TGC GGG CCC TGC TAA GCC AAG CAG AAA TAT TCC CCT CAT GGA AGA GAA T
             190         200         210         220         230         240

 CA CCA ATG GGT TCC CAA CAG AAG ATC TCC ATG TTG CTC TGA TTC AAG ACT ACT GAG CTG T
             250         260         270         280         290         300

 TT CTC TTC AGC CCC ATC AAA TTT ATC CTC TCA ATT CTA CTT TCA GTT TCT ACT ATA ATG C
             310         320         330         340         350         360

 TT CTC CCT TCA TGT TTC CAT CAT ACC TTG TGT AAA TGT AAC AAA ACT GCA GAA ATA ATT C
             370         380         390         400         410         420

 CT GGT TGT TTA GAT AAA CTA AAG CAC AAG AAG GAA AAT GAA TCA TGT GAA TGA ATT CCT T
             430         440         450         460         470         480

 GG TGA AGA TCA GTT GCT TAT TTT CTT GGA GAC TCT GAA TGT GTA TTA ATT TCT CAG TAT C
             490         500         510         520         530         540
```

Polyadenylation Sequence

```
 AT TAA AAG TTT ATT GCA TCA AAA AAA CCC CCC CCC CCC CCC TGC AG
             550         560         570         580
```

A more detailed analysis of the pCFK15 sequence was not carried out since at the time of the present study, the scale keratin genes were the subject of an investigation by S.D. Wilton as part of his Ph.D. studies. Thus, it was possible to recognize that the sequence of pCFK15 was identical to two cDNA clones isolated from a scale cDNA library except that in both cases the scale cDNA clones had longer poly(A) sequences and different 5′-termination points (Wilton, 1983). It was however, very interesting that a scale keratin sequence should be found in a feather cDNA library.

5.3.9 FREQUENCY OF SCALE SEQUENCES IN THE FEATHER cDNA LIBRARY.

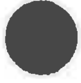
About 200 clones from the feather cDNA library were screened for the presence of scale sequences (Figure 5.10), since it seemed likely that other cDNA clones in the library were derived from scale keratin mRNAs. Only one colony was positive for the pCFK15 probe, but it was not easy to assign it to a particular clone because the nonspecific background binding of radioactivity to the other colonies evident for example in Figure 5.3 was undetectable, even at long exposures using this probe. However, it would be expected that this colony should correspond to pCFK15, since the probe was made from pCFK15. This was confirmed when radioactive ink was used to exactly orientate the filters.

Although only one scale keratin cDNA was detected in the cDNA library it was possible that others existed but were not detected by the pCFK15 probe used, which was derived from the 39 bp repeat sequence in the middle of the coding region. This probe was used because the repeat sequence is common and unique to the majority of scale keratins (Wilton, 1983) and should therefore have cross-hybridized to all scale cDNA clones, not just those identical to the pCFK15 sequence. Unfortunately this repeat region contains substantial secondary structure, as evidenced by sequence specific stoppages of the Klenow Pol I catalysed dideoxy elongation sequencing reactions (data not shown). Secondary structure was likely

FIGURE 5.10

FREQUENCY OF SEQUENCES COMPLEMENTARY TO pCFK15 IN THE cDNA LIBRARY.

Approximately 200 cDNA clones from the cDNA library were grown on two nitrocellulose filters exactly as described in the legend to Figure 4.16. These filters were screened using an end-labelled fragment from the 39 bp (13 amino acid) repeat region of pCFK15. This repeat is characteristic of and unique to scale keratins. The filters were washed at high stringency (0.2 x SSC) and only one clone (pCFK15) was detected. No background radioactivity was bound to the colonies so spots of radioactive ink were used to orientate the filters and hence identify the positive clone.



to also interfere with AMV reverse transcriptase elongation of cDNA transcripts in the repeat region, which would result in short cDNA clones that did not contain this repeat region. This question was not resolved, although, had time permitted, the use of a probe derived from the whole cDNA clone, or at least a region (preferably coding) which did not contain a high degree of secondary structure or sequence homology to the coding regions of feather keratin clones, should have answered this question.

5.3.10 THE PRESENCE OF pCFK15 SEQUENCES IN FEATHER mRNA.

To clearly and directly demonstrate that scale keratin sequences were present in embryonic feather mRNA, poly(A)-plus mRNA from embryonic feathers, which had been glyoxylated then fractionated on a 1% agarose gel and transferred to nitrocellulose, as described in 5.2.1 and 5.2.2, was probed with the same scale specific pCFK15 coding sequence probe described in 5.3.9. Figure 5.11 shows an autoradiograph of this filter next to a similar filter which had been probed with a sequence which would specifically detect feather keratin sequences. The scale specific probe easily detected a mRNA which was marginally longer than the embryonic feather keratin mRNA of about 850 bp, a result consistent with the expected size of scale keratin mRNA (Wilton, 1983). This result confirmed that scale keratin mRNAs were expressed in feather tissues, and from the intensity of the scale mRNA band it was probably present in feather mRNA at levels greater than 1% of total mRNA.

5.4 DISCUSSION.

5.4.1 THE PRESENCE OF SCALE KERATIN mRNA IN EMBRYONIC FEATHERS.

The demonstration that the feather cDNA clone pCFK15 coded for a scale keratin (Figure 5.9), and the observation that this sequence was present in embryonic

FIGURE 5.11

DETECTION OF SEQUENCES COMPLEMENTARY TO pCFK15 IN EMBRYONIC FEATHER mRNA (NORTHERN BLOT).

The Northern blot was performed exactly as described in the legend to Figure 5.6 except that the two filters were probed with a pCFK15 probe from the repeat region (track A)(see the legend to Figure 5.10) and a probe which specifically detects embryonic feather keratin coding sequences (track B).

A B

feather mRNA, presented several interesting questions. Firstly is this mRNA translated in the feather, and if so what function does the protein have?

It seems difficult to imagine that a mRNA is not being translated in a tissue when it is capable of being translated in a cell-free system, unless the mRNA is inactivated, perhaps by some protein packaging process. This seems unlikely since embryonic feathers form a terminally differentiating tissue and it is inconceivable that a mechanism for the expression and inactivation of the mRNAs would have evolved, only to have the tissue die before the mRNA was translated. Also, proteins of the same mobility as scale keratins appear to be present in embryonic feather tissue when the proteins of the feather are fractionated on SDS gels (Figure 3.1), although positive identification would require the use of scale keratin specific antibodies, which could be used to probe a "Western transfer" of the proteins shown in Figure 3.1.

Assuming that scale proteins are expressed in the feathers, one can ask whether they are positively expressed, i.e. for a functional reason, or whether they are there because scale and feather keratins are switched on by similar, but different, signals and the expression of scale keratins in feathers is due to "leakage" of these scale genes which poorly recognize the similar switching mechanisms of the feather. If these scale genes are positively expressed and the hypothesis, which was made in 4.4.5, that all genes expressed in the feather should have the same 5′-leader sequence, is correct, then one would expect that these scale genes should have the same 5′-leader sequence as the feather mRNAs and these scale mRNAs should have a different 5′-leader sequence in mRNA isolated from scale tissue. Although this is speculation, the scale genes, like the feather genes, have only one intron which is situated in the middle of the 5′-untranslated region of the mRNA, which lends the hypothesis to experimental testing.

It is often difficult to obtain full length cDNA clones for a messenger RNA and none have been obtained either from the feather (this thesis) or scale (Wilton, 1983)

cDNA libraries, so the best approach to testing this hypothesis would be to extend a primer which came from the 3′-side of the intron of a scale gene, in the presence of dideoxy nucleotides (see Figure 6.10) to the end of the mRNA, using mRNA derived from either scale or feather tissue. If the above hypothesis is correct, the mRNA sequence before the position of the intron should be identical for mRNA derived from both tissues, and on the 5′-side of the intron the scale mRNA should have a scale specific 5′-leader sequence, and the feather mRNA should have the typical feather 5′-leader sequence.

5.4.2 GROUP II SEQUENCES CODE FOR FAST PROTEIN.

It appeared that only a small number of Group II cDNA clones were derived from mRNAs coding for scale keratins (Figure 5.10), which raises the question, what do the other Group II cDNA clones code for?

Three out of the four Group II cDNA clones sequenced appeared to code for fast proteins (Figure 5.7. and Table 5.1), and two of these, pCFK10 and 22, which were different isolates of the same mRNA sequence, were the same, or nearly the same as five other cDNA clones in the library, indicating that these sequences must represent a significant percentage of the cDNA library. Add to this the fact that pCFK3 was not one of these five cDNA clones detected, presumably because it represents a different fast protein mRNA, and we are drawn to the conclusion that a large number of the remaining Group II cDNA clones were derived from mRNAs which coded for fast protein.

This conclusion is supported by the observation that only 10,000 and 15,000 Mr keratins and fast protein are present, at significant levels, in the proteins of embryonic feathers. The only obstacle to the conclusion that these cDNA clones, in particular pCFK22, code for fast proteins, is the observed molecular weight of 8,400 Mr (about 84 amino acids) for fast protein, when pCFK22 encodes a protein larger than 110 amino acids. Walker and Rogers (1976a) were hampered in their

efforts to obtain a molecular weight estimate for fast protein by the artefactually slow migration of S-carboxymethylated proteins of the embryonic feather on SDS gels. The estimate of 8,400 Mr was therefore obtained by the less accurate method of gel filtration (Walker and Rogers, 1976a). This estimate was also consistent with the mobility of fast protein relative to keratins on SDS gels (Figure 3.1), unfortunately fast proteins would not be expected to be greatly affected by S-carboxymethylation because of the low cysteine content in these proteins, while the cysteine rich keratins would. It is therefore likely that keratins would migrate artefactually slowly on SDS gels but fast proteins would not. Since fast protein molecular weight estimates on SDS gels have only been related to the mobilities of S-carboxymethylated keratins, it is probable that fast proteins could be as large as the protein encoded by pCFK22 (110 to 130 amino acids from mRNA length, Figure 5.6).

The precise identification of pCFK22 and the other Group II cDNA clones and the exact size and nature of fast protein awaited further investigation.

CHAPTER 6.

IDENTIFICATION OF FAST PROTEIN GENES.

6.1 INTRODUCTION.

Although the results presented in Chapter 5 did not positively identify the origin of the cDNA clone pCFK22, they did indicate that this clone was probably derived from a mRNA which coded for fast protein. Positive identification was not possible because pCFK22 was not a full-length cDNA clone, i.e. it did not extend to include the start of the mRNA coding region. Without a full amino acid sequence the protein encoded by pCFK22 could not be identified. In addition to this problem, although the amino acid composition of crude fast protein had been determined (Walker and Rogers, 1976a; Powell, 1979), no amino acid sequence data was available for fast proteins nor had a precise measurement of its molecular weight been achieved, making it impossible to definitively ascribe any conceptually derived protein sequence to fast protein.

It seemed obvious that the complete nucleotide sequence of the gene from which pCFK22 was derived would be necessary to provide a complete protein sequence which would define the encoded protein's size and amino-terminus and allow an examination of the 5′-end of the gene for features in common with other genes expressed in embryonic feathers. Identification of the encoded protein as fast protein however, would require amino acid sequence data from a fast protein molecular species purified away from the proteins of the embryonic feather and also from the other fast protein species.

The techniques for the construction of banks of randomly cleaved genomic DNA fragments cloned into λ-Charon 4A, or other vectors, have effectively enabled the representation of the whole genome of any organism in a library of clones which may be used to select any particular gene derived from the organism using the screening and selection procedures which have evolved along with the cloning techniques (Wensink *et al.*, 1974; Sternberg *et al.*, 1977; Maniatis *et al.*, 1978). It was decided that these procedures would be applied to the selection of the pCFK22 gene from a chick genomic library.

The aim of purifying a fast protein species for the purpose of protein sequencing presents more of a problem than does the isolation and sequencing of the pCFK22 gene. Very little is known about the molecular nature of fast proteins as they have not previously been extensively studied. This is probably in part because attempts to do so have been hampered by problems of protein aggregation, solubility and the observation that the component molecular species of the fast protein family exhibit similar physico-chemical properties.

Electrophoretic techniques are not very effective for separating fast proteins since they migrate as a single band on SDS and acid pH polyacrylamide gels, presumably because they form a family of identical molecular weight proteins. Alkaline pH gels appear to be the most promising electrophoretic system for separating fast proteins, as on these gels they form a ladder of 5 to 7 bands of approximately equal spacing, although they are generally only poorly resolved from one another. Fast proteins are also difficult to solubilize except by using strong denaturants like SDS and urea, and once soluble they tend to aggregate with themselves and other proteins, even in the presence of 8 M urea.

Any attempt to sequence fast proteins would require a single pure fast protein species and the traditional methods of liquid chromatography and gel electrophoresis have proven inadequate for this purpose (Walker and Rogers, 1976a; Powell, 1979), however the new techniques of FPLC (fast protein liquid chromatography) (see Pharmacia technical bulletins) and HPLC (high performance/pressure liquid chromatography) (Rubinstein, 1979) have proven very effective in the separation of difficult to handle proteins and peptides.

Having achieved the purification of a fast protein species, the best approach toward determining its amino acid sequence would be to use one of the newly developed range of automated gas-liquid solid phase protein sequenators (Hewick *et al.*, 1981) which can be used for the determination of up to 60 amino acid residues from the amino-terminus of proteins and large peptides. This technique has the

added advantage that only small amounts of protein (in the range of 10 to 2,000 pmoles) are required for sequence determinations.

The results presented in this chapter describe the isolation of a genomic clone selected from a λ-Charon 4A chicken genomic library (prepared and generously donated by J. Dodgson, J. Engel and R. Axel) using the cDNA clone pCFK22 as a hybridization probe. The pCFK22 gene was localized in the genomic clone and its DNA sequence determined. It was confirmed that this gene coded for a fast protein by comparing the conceptual translation of the gene with the amino acid sequence from the amino-terminal 20 residues of a fast protein component purified using both FPLC and HPLC techniques.

6.2 SPECIFIC METHODS.

6.2.1 SCREENING THE GENOMIC LIBRARY.

The λ-Charon 4A chicken genomic DNA library was screened by plating approximately 7 x 10^4 pfu on each of six pre-dried NZCYM plates cast in 150 mm petri dishes and incubated at 37°C for at least 8 hours or until all plaques were visible and just touching (the plaques were approximately 0.5 mm in diameter at this stage).

Duplicate nitrocellulose filters were prepared from each plate, essentially as described by Benton and Davis (1977). The phage were adsorbed to nitrocellulose filters (30 seconds 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, blot dried and baked at 80°C under vacuum for 2 hours.

The hybridization of radioactive probes to these filters and detection of complementary nucleotide sequences was carried out under conditions of high stringency as described in 2.2.17. Plaques which bound the radiolabelled pCFK22

probe were picked into 1 ml PSB and purified through at least two further platings at low plaque density (usually several hundred plaques on an 80 mm NZCYM plate).

High titre phage stocks of purified recombinants were obtained by flooding a wet 80 mm NZCYM plate, on which the recombinant had been plated at high density (about 10^5 per plate), with two 1 ml washes of PSB. The recovered PSB washes usually contained about 10^{10} pfu/ml and were stored at 4°C over chloroform.

6.2.2 PREPARATION OF RECOMBINANT PHAGE DNA.

DNA was harvested from approximately 10^5 pfu which had been grown overnight at 37°C on two (150 mm) petri dishes. It was essential that these phage were propagated on agarose-base media plates as agar was found to contain potent inhibitors for a wide range of enzymes. Each plate was flooded with 3 ml PSB and left for 5 minutes before the PSB and soft-agarose overlay were transferred to a fresh tube. The plate was again flooded with PSB which was then combined with the first wash and overlay mix and centrifuged at 10,000 rpm for 10 minutes at 4°C. The supernatant was carefully decanted into a fresh tube and the pellet of agarose and cellular debris was washed with 10 ml PSB and the supernatants were combined. The phage were collected after centrifugation at 19,000 rpm for 3.5 hours at 4°C and the pellet was resuspended in 3 ml PSB to which 1.5 g caesium chloride was dissolved. This material was then layered on top of a two step caesium chloride block gradient (specific gravity of the top and bottom phases was 1.4 and 1.6, respectively), and centrifuged at 45,000 rpm for 2 hours at 4°C. The phage band was removed, diluted into three volumes of SET buffer and gently extracted twice with an equal volume of SET-buffered phenol. The aqueous phase was extracted twice with two volumes of ether, ethanol precipitated twice and resuspended in 200 μl of H_2O. This procedure usually yielded 50 to 100 μg of recombinant phage DNA per plate.

6.2.3 EXTENSION OF DNA PRIMERS ON FEATHER mRNA.

End-labelled DNA fragments, prepared as described in 2.2.19b and 2.2.20b, were hybridized to embryonic feather poly(A)-plus mRNA by boiling 0.75 μg of mRNA with an excess of labelled DNA fragment in 10 μl of hybridization buffer (200 mM KCl, 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA), followed by incubation at 65°C for 2 hours. To this hybridization mix was added 4 μl of 10x extension buffer (250 mM Tris-HCl pH 8.3, 100 mM $MgCl_2$, 10 mM DTT and 4 mM of each deoxynucleotide triphosphate) and the volume increased to 40 μl with water. Polymerization was carried out by the addition of 5 units of AMV reverse transcriptase followed by incubation at 42°C for 30 minutes. Elongation products were then ethanol precipitated, redissolved in formamide loading buffer and denatured prior to electrophoresis on 200 x 400 x 0.5 mm sequencing gels (2.2.23c).

6.2.4 PRIMER EXTENDED SEQUENCING.

The nucleotide sequence of mRNAs 5′ to the hybridization of end-labelled DNA primers was determined by extending the primers on mRNA in the presence of dideoxynucleotide triphosphates.

For sequencing, end labelled primer DNA (greater than 200,000 dpm) was annealed to 12.5 μg of poly(A)-plus mRNA in 50 μl of hybridization buffer as described in 6.2.3. To the annealing mix was added 20 μl of 10x extension buffer (250 mM Tris-HCl pH 8.3, 100 mM $MgCl_2$, 10 mM DTT and 200 μM of each deoxynucleotide triphosphate) and the volume increased to 185 μl with water. Of this mix 45 μl was added to each dideoxynucleotide triphosphate, contained in four incubation mixes, to give final concentrations of 50 μM ddATP, 10 μM ddGTP, 10 μM ddCTP and 30 μM ddTTP. Polymerization was initiated by the addition of 10 units of AMV reverse transcriptase followed by incubation at 42°C for 30 minutes.

Elongation products were ethanol precipitated and fractionated by electrophoresis on 6% polyacrylamide gels in 8 M urea, as described in 6.2.3.

6.2.5 FPLC ANION EXCHANGE CHROMATOGRAPHY.

FPLC anion exchange chromatography was carried out using a Pharmacia FPLC system consisting of two P-500 titanium/glass reciprocating piston low pressure pumps, an LCC-500 liquid chromatography controller with plotter/integrator, a UV-1 fixed wavelength (280 nm) single path UV monitor, one MV7 (seven position) and two MV8 (eight position) servo-actuated valves, a Frac-100 fraction collector, a Pharmacia dual pen chart recorder and sundry injection loops. This system was used in conjunction with a 5 ml Mono Q strong anion exchange column.

Crude fast protein samples were dissolved at concentrations ranging from 0.5 mg/ml to 40 mg/ml in buffer A (20 mM Tris-HCl pH 7.6, 8 M urea) and applied to the column which had been pre-equilibrated with buffer A, using a 500 μl injection loop. The protein components were eluted from the column with optimized gradients, as indicated in the figures, ranging from 0 to 100% buffer B (Tris-HCl pH 7.6, 8 M urea, 1 M NaCl) at flow rates ranging from 0.5 to 1.5 ml/min and protein fractions were collected as 0.5 ml fractions using the integrated fraction collector. Chromatographic separations usually ranged in duration from 15 to 25 minutes.

6.2.6 FPLC CATION EXCHANGE CHROMATOGRAPHY.

Cation exchange chromatography of crude fast protein was performed using the same Pharmacia FPLC system and basic procedure as described in 6.2.5, except that a 5 ml Mono S strong cation exchange column was used and buffer A consisted of 50 mM sodium acetate pH 4.8, 8 M urea, while buffer B consisted of 50 mM sodium acetate pH 4.8, 8 M urea, 1 M NaCl.

6.2.7 HPLC REVERSE PHASE C18 CHROMATOGRAPHY.

Reverse phase C18 chromatography was performed using a Varian 5,000 Liquid Chromatograph linked to a Varian UV-50 variable wavelength detector, a Varian CDS-401 controller integrator and a Varian SP-C18 3 μ 15 cm x 4.6 mm reverse phase column.

Initial studies were carried out using 20 mM sodium acetate pH 4.5, which contained 0.1% SDS to solubilize the protein samples which were then applied to the column and eluted by using various gradients from 0 to 50% propan-2-ol. The column was generally run at 40°C and eluted protein detected by monitoring optical absorbance at 280 nm. Subsequent separations were performed in aqueous 0.1% trifluoroacetic acid using elution gradients generated in acetonitrile containing 0.1% trifluoroacetic acid. The column was run at room temperature (23°C) and eluted protein detected at 222 nm.

Fast protein samples derived from FPLC separations were first dialysed against chromatographically pure water to remove buffer and urea, and lyophilized prior to being dissolved in column buffer. Samples from HPLC separations run in 0.1% trifluoroacetic acid and acetonitrile were simply lyophilized prior to being dissolved in column buffer.

6.2.8 HPLC REVERSE PHASE DIPHENYL CHROMATOGRAPHY.

Reverse phase diphenyl chromatography was performed using the same HPLC system described in 6.2.7 except that the column used was a Whatman Protesil-300 10 μ 20 cm x 4.6 mm diphenyl reverse phase column. The buffer system used was the trifluoroacetic acid/acetonitrile system described in 6.2.7, all other conditions including protein sample preparation procedures and chromatography conditions were as described in 6.2.7.

6.2.9 AUTOMATED PROTEIN SEQUENCING.

Protein samples were sequenced with the assistance of M. Snoswell of this department as a departmental service.

Protein samples ranging from 50 to 500 pmoles were applied to an Applied Biosystems 470A Protein Sequencer (using the gas-liquid solid phase automated peptide and protein sequencing principle described by Hewick *et al.*, 1981). The automated Edman degradation released phenylthiohydantoin (PTH) amino acid derivatives which were collected and identified by HPLC reverse phase chromatography using a Waters Liquid Chromatograph linked to a Nova C18 column and a Waters Wisp 710B automated sample injector. The complex elution system used acetate, trifluoroacetic acid and acetonitrile, and is currently being prepared for publication by M. Snoswell.

6.3 RESULTS.

6.3.1 CHARACTERIZATION OF THE pCFK22 GENE.

6.3.1a SELECTION OF RECOMBINANTS FROM THE GENOMIC LIBRARY.

Screening of 4 x 10^5 λ-recombinants (2 to 3 chick genome equivalents) from the chick genomic library with a probe prepared from the cDNA clone pCFK22 resulted in the detection of 12 recombinant phage under low stringency washing conditions (Figure 6.1a) which was reduced to 2 positive phage when washed at high stringency (data not shown). The low number of recombinants (approximately one per genome equivalent) detected at high stringency indicated that the pCFK22 probe was specifically detecting a unique gene present in the chick genome, while the larger number of recombinants detected at low stringency indicated that this gene may form part of a family of sequence related genes.

FIGURE 6.1

DETECTON OF CHICK GENOMIC CLONES CONTAINING SEQUENCES COMPLEMENTARY TO THE cDNA CLONE pCFK22.

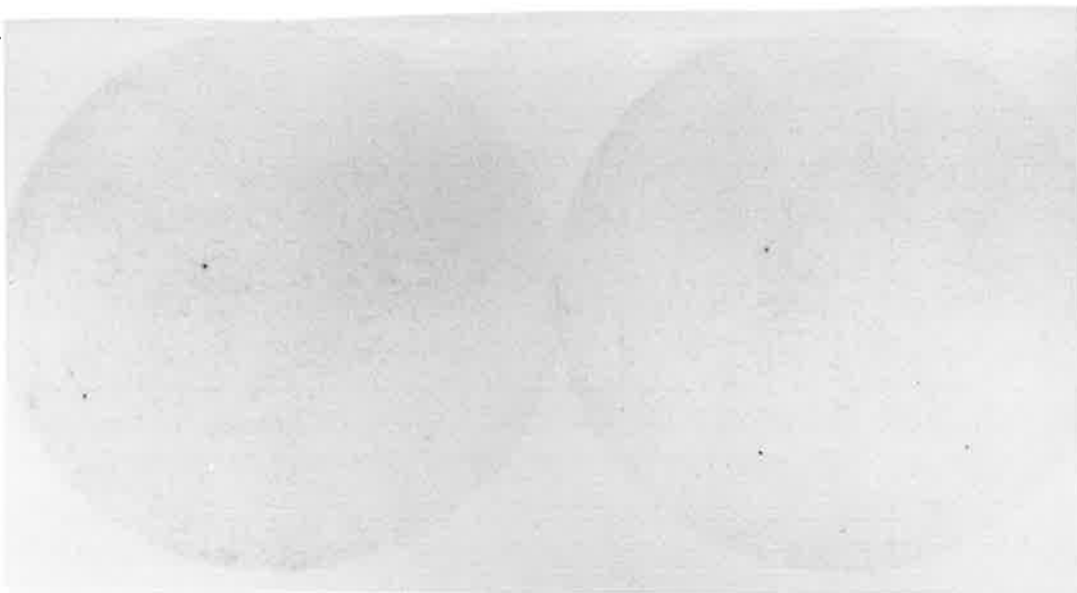
Recombinant phage from the chick genomic library were plated at the densities indicated, the plaques were transferred to nitrocellulose filters and their DNA was immobilized by baking for 2 hr at 80°C under vacuum, as described in 6.2.1. These filters were then annealed to a probe made by labelling an HgiA I fragment from the extreme 5′-end of pCFK22 which had been subcloned into M13. The filters were then washed in 1.0 x SSCE, 0.1% SDS at 65°C and autoradiographed.

a. First round screening; duplicate filters containing the DNA from approximately 7 x 10^4 plaques (150 mm plates)

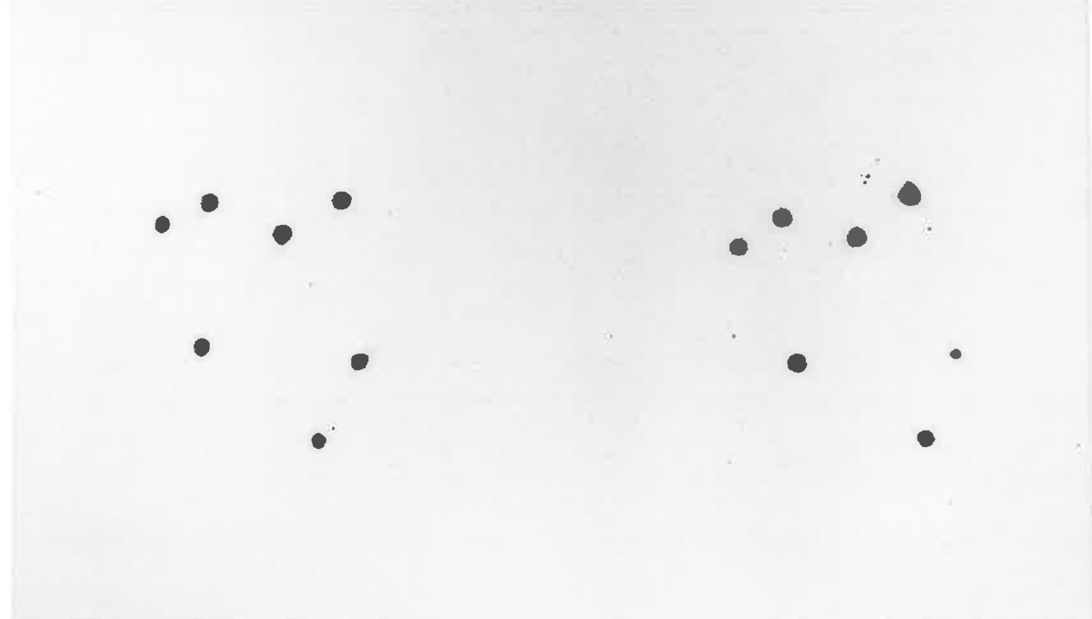
b. Second round screening; duplicate filters containing the DNA from approximately 1 x 10^3 plaques (150 mm plates)

c. Third round screening; purified recombinant phage plated at high and low density on 80 mm plates. All phage present are positive with the pCFK22 probe.

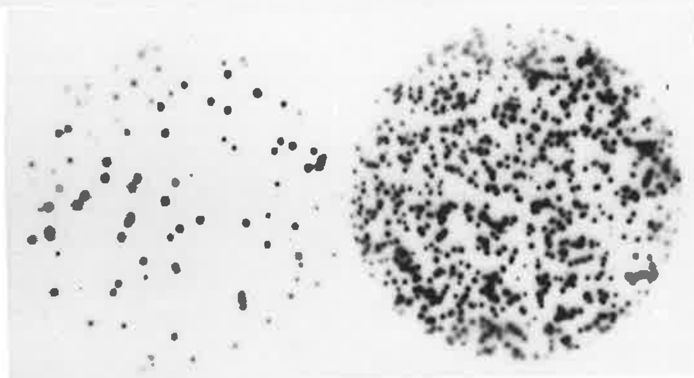
a.



b.



c.



Since the gene which corresponded exactly to the sequence of pCFK22 was needed for the positive identification of the mRNA from which pCFK22 was derived, only the two clones detected at high stringency were further purified by a second round of plating (Figure 6.1b). Phage from plaques isolated from this second screening were further checked for the presence of contaminating λ-recombinants by a third screening, plated at high and low plaque densities, in which all phage present were positive with the pCFK22 probe (Figure 6.1c).

DNAs prepared from these two λ-recombinants, as described in 6.2.2, (designated λCFP1 and λCFP2) were subjected to both single and double digests with restriction endonucleases Eco RI, Hind III and Bam HI followed by fractionation of the digestion products on a 1% agarose gel (Figure 6.2). Although some of the digestions still contained partial digestion products, it was clear from the patterns of bands produced by λCFP1 and 2 that these clones were independant isolates of exactly the same λ-recombinant. This observation indicates that the genomic library was of such a limited size when first constructed, or following subsequent amplification, that only one clone in the whole library contained a pCFK22 gene. The length of genomic DNA inserted into the clone λCFP1 was estimated to be about 16 Kb by the summation of the length of insert restriction fragments generated from this clone (Figure 6.2).

6.3.1b LOCALIZATION OF THE pCFK22 GENE IN λCFP1.

The nucleotide sequence of the cDNA clone pCFK10 extended the sequence of pCFK22 to the poly(A) tract at the 3′-end of the mRNA from which these clones were derived, it was only at the 5′-end of the mRNA that sequence data was incomplete. For this reason the major interest in the gene for pCFK22 lay at the 5′-end of the gene, the sequence of which would allow the complete protein coding region to be determined. The probe chosen to detect the pCFK22 gene in λCFP1 was therefore derived from the extreme 5′-end of the pCFK22 insert and was totally contained within the putative protein coding region.

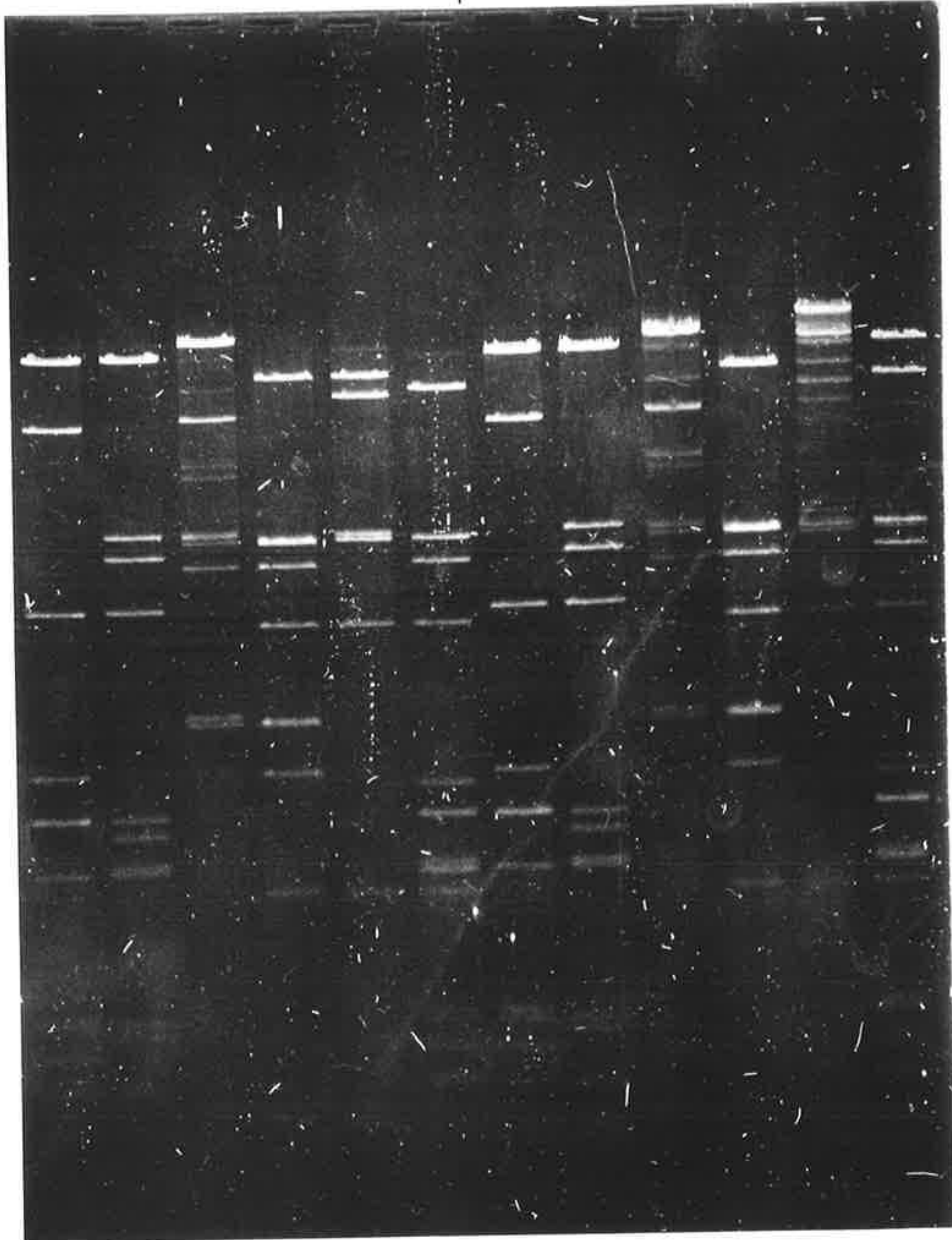
FIGURE 6.2

ELECTROPHORESIS OF DNA FROM λCFP1 AND λCFP2 AFTER CLEAVAGE WITH RESTRICTION ENZYMES.

DNA from two recombinant phage, detected as shown in Figure 6.1, was subjected to both single and double digests with restriction endonucleases Eco RI (E), Hind III (H) and Bam HI (B) followed by fractionation of the digestion products on a 1% agarose gel. DNA fragments were visualized under UV light after staining with ethidium bromide. Despite the presence of partial digestion products, it is clear that λCFP1 (the first 6 tracks) and λCFP2 (last 6 tracks) are identical.

λ CFP1 | λ CFP2

E E/H H H/B B E/B | E E/H H H/B B E/B



Digestion of λCFP1 with restriction endonucleases Eco RI, Hind III, Bam HI and Sal I, singly and in paired combinations resulted in the banding patterns shown in Figure 6.3a when these digests were fractionated on a 1% agarose gel. The DNA from this gel was transferred and immobilized onto a nitrocellulose filter which was probed with the 5′-end of pCFK22. The resulting autoradiograph is shown in Figure 6.3b next to the gel from which the DNA was transferred. Sal I did not cleave the insert DNA or the λ arms so the probe detected the full length clone in this digest, but the probe strongly detected a 0.8 Kb Eco RI, a 2.7 Kb Hind III and a 13.2 Kb Bam HI fragment. A 5.0 Kb Hind III fragment was also weakly detected by the probe. It was later surmised that the probe, which contained a Hind III site 33 bp from its 3′-end, spanned the 2.7 and 5.0 Kb Hind III fragments with the 5.0 Kb fragment being detected by the 33 bp at the 3′-end of the probe, which explained why this fragment was only weakly detected.

As expected, since the probe contained a Hind III site, the 0.8 Kb Eco RI fragment was shortened to 0.65 Kb and 0.15 Kb fragments, both of which were detected with the probe, when digested with Hind III. The 0.8 Kb Eco RI fragment was not cleaved by Bam HI or Sal I. The 2.7 Kb Hind III fragment was not cleaved by Sal I but it was shortened to 2.65 Kb when digested with Bam HI. The 13.2 Kb Bam HI fragment was not cleaved by Sal I.

These data indicated that the sequence detected by the probe was totally contained within the 0.8 Kb Eco RI fragment which contained a Hind III site dividing the fragment into 0.15 and 0.65 Kb fragments. The cDNA clone pCFK22 contained a 0.65 Kb Eco RI/Hind III fragment with the Hind III site being the one at the 5′-end of pCFK22 which was contained in the probe used in the above experiment. It was therefore probable that the gene was oriented with the 0.15 Kb Eco RI/Hind III fragment at the 5′-end of the gene and the 0.65 Kb fragment at the 3′-end. The 0.8 Kb Eco RI fragment was considered the ideal place to start DNA sequencing in order to identify the gene, but the identity and position of flanking DNA fragments

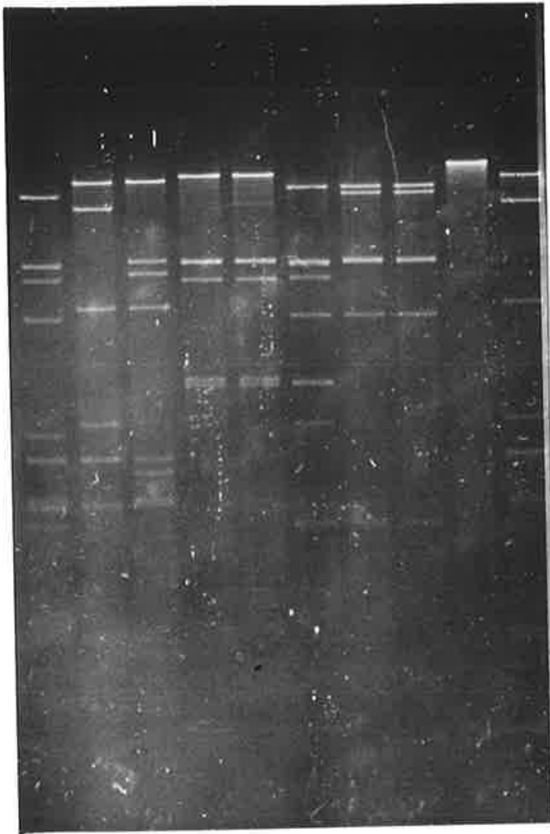
FIGURE 6.3

LOCALIZATION OF THE pCFK22 GENE IN λCFP1.

DNA from λCFP1 was digested with Eco RI (E), Hind III (H), Bam HI (B) and Sal I (S), singly and in paired combinations. The digestion products were fractionated on a 1% agarose gel and the fragments detected as described in the legend to Figure 6.2. The DNA on this gel was transferred and immobilized onto a nitrocellulose filter which was annealed to the same pCFK22 probe described in the legend to Figure 6.1.

a. Agarose gel showing restriction fragments of λCFP1.

b. Autoradiograph of Southern transfer from the gel shown in a. (exposure time was 30 min).


E•B
E
E•H
H•S
H
H•B
B
B•S
S
E•S

E•B
E
E•H
H•S
H
H•B
B
B•S
S
E•S

would be required if the gene were to be completely sequenced.

Longer exposure of the Southern transfer shown in Figure 6.3 revealed the presence of 0.92 Kb, 1.9 Kb and 2.2 Kb Eco RI fragments which also hybridized very weakly to the probe derived from the presumptive coding region of pCFK22 (Figure 6.4), possibly indicating the presence of sequence related genes which are clustered in the genome. Figure 6.4 also revealed a decrease in length of the weakly positive 2.2 Kb Eco RI fragment to 1.64 Kb when digested with Hind III and 1.55 Kb when digested with Bam HI, indicating that the sequence which hybridized to the probe in the 2.2 Kb Eco RI fragment was contained within a 1.55 Kb Eco RI/Bam HI fragment. It was subsequently discovered that these weakly positive fragments were located on either side of the pCFK22 gene, which suggested that there must be at least two homologous genes flanking the pCFK22 gene and that this number is probably less than four since the gene cluster only spans about 5 Kb including the pCFK22 gene.

6.3.1c POSITIVE IDENTIFICATION OF THE pCFK22 GENE.

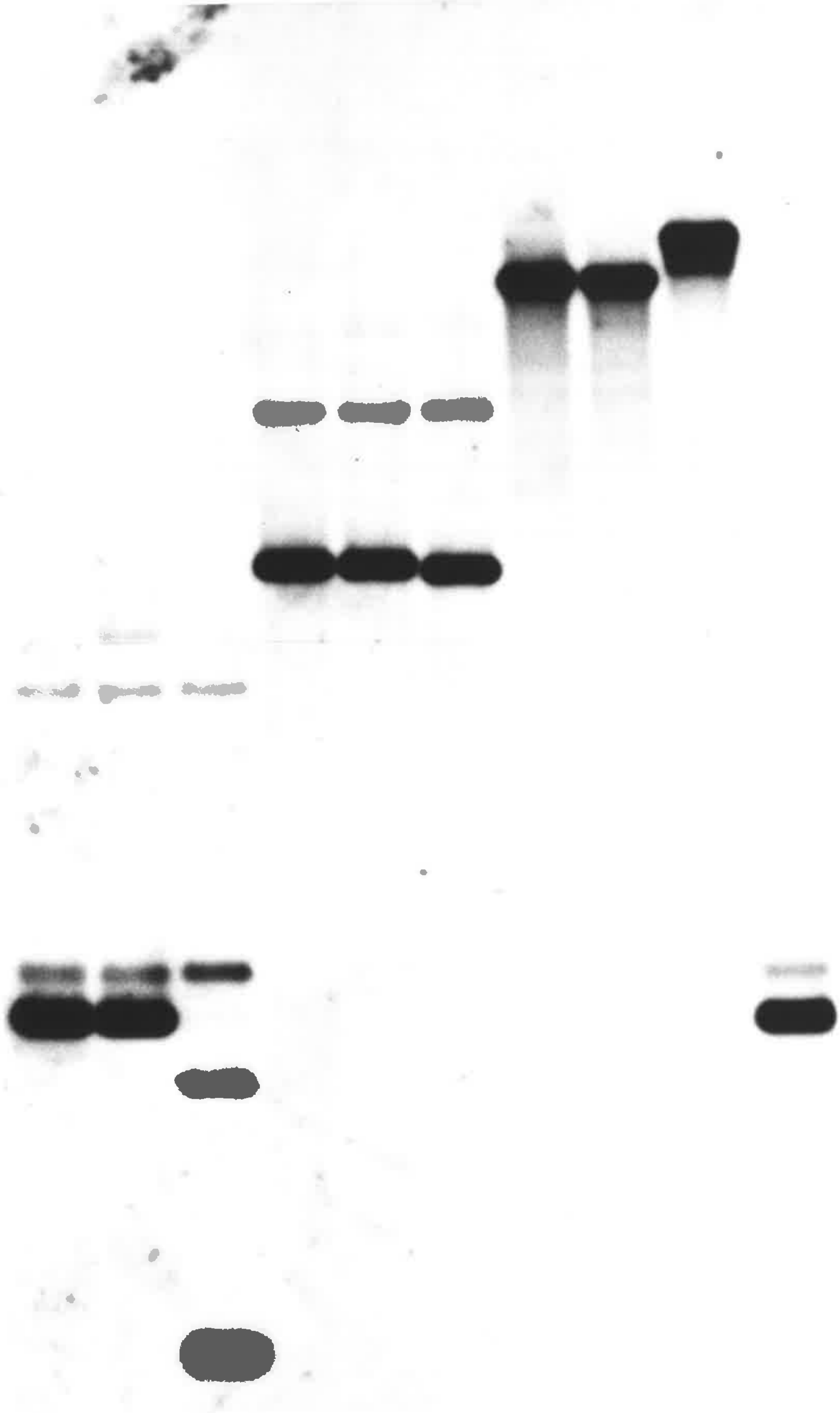
Due to the difficulty in estimating the length of small DNA fragments on agarose gels, which contained few low molecular weight markers, in early experiments the length of the 0.8 Kb Eco RI fragment referred to above (6.3.1b), was incorrectly estimated to be 0.9 Kb. The cDNA clone pCFK22 extended 0.14 Kb beyond to the Hind III site in the 5′ direction (Figure 5.1), so it was incorrectly estimated that the 0.9 Kb Eco RI fragment should extend 0.25 Kb 5′ to the Hind III site and hence extend the pCFK22 sequence by 0.11 Kb at the 5′-end where extra sequence data was required for completion of the protein coding region.

The presumptive 0.9 Kb Eco RI fragment was therefore excised from an agarose gel, ligated directly into M13 and, after transformation, both orientations of the fragment were obtained as M13 subclones to be used for DNA sequencing. Sequence from one orientation definitively identified the gene to be exactly the one from which

FIGURE 6.4

PRESENCE OF OTHER SEQUENCES IN λCFP1 WHICH ARE HOMOLOGOUS TO THE pCFK22 SEQUENCES.

The Southern transfer shown in Figure 6.3b was exposed for a longer time (24 hr as opposed to 30 min). All conditions are the same as described in the legend to Figure 6.3. Note the presence of the weakly hybridizing 0.92, 1.9 and 2.2 Kb Eco RI fragments and the 1.64 Kb Eco RI/Hind III and 1.55 Kb Eco RI/Bam HI fragments.


E/B
E
E/H
H/S
H
H/B
B
B/S
S
E/S

the cDNA clone pCFK22 was derived, since the sequence was identical to the DNA sequence 5′ to the Eco RI site in pCFK22. Sequence from the other orientation did not extend the pCFK22 sequence by 0.11 Kb as erroneously expected, instead the sequence was extended by only 15 bp which increased the coding region potential from 110 amino acids to 115 amino acids (data not shown, but refer to gene sequence in Figure 6.8).

It was obvious that more sequence data was required from the fragments on the 5′-side of the 0.8 Kb Eco RI fragment and although a crude restriction map could be constructed from the data presented in Figure 6.3 it was considered that an accurate map of the clone should be constructed since small fragments could have been missed on the gel shown in Figure 6.3.

6.3.1d RESTRICTION MAPPING OF λCFP1.

When λCFP1 was digested with a range of restriction enzymes (Eco RI, Hind III, Bam HI, Sal I and Xho I) many small fragments less than 0.5 Kb were discovered. It was fortunate that it was decided to map these fragments as it was found that an Eco RI fragment of only 0.07 Kb (70 bp) was adjacent to the 5′-end of the 0.8 Kb Eco RI fragment already sequenced and this very small fragment would almost certainly have been overlooked if a high resolution restriction map had not been determined. Also, a 28 bp Xho I fragment was found which spanned the site for the initiation of mRNA transcription.

Figure 6.5 shows the products of single and double digests of λCFP1 DNA digested with combinations of Eco RI, Bam HI, Hind III and Xho I which were end-filled to enable the detection of small fragments and fractionated on 0.6% and 1.8% agarose gels (Figure 6.5a and b respectively) in order to resolve both large and small restriction fragments. The restriction data revealed that the insert of about 15.5 Kb contained 3 Xho I sites, 3 Bam HI sites, 7 Hind III sites and 10 Eco RI sites excluding the Eco RI linker sites at the ends of the insert.

FIGURE 6.5

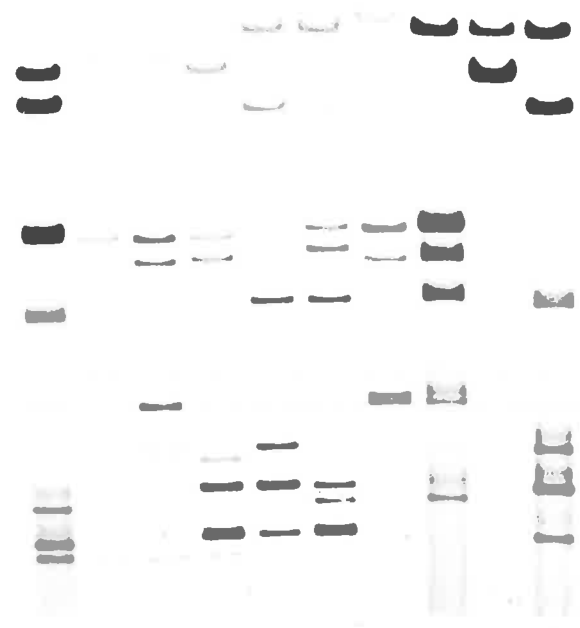
FINE MAPPING OF RESTRICTION SITES IN λCFP1.

DNA from λCFP1 was subjected to both single and double digests with restriction endonucleases Eco RI (E), Hind III (H), Bam HI (B) and Xho I (X) and the resulting fragments were end-labelled, as described in 2.2. Labelled fragments were fractionated by electrophoresis on 0.6% and 1.8% agarose gels and detected by autoradiography. Prior to autoradiography the gels were fixed in 10% acetic acid and extensively washed to remove unincorporated nucleotides which were either still on the gel at the ion front or which ran off the bottom of the gel and diffused back into the gel from the submarine buffer. The fixed and washed gels were then blotted to reduce their thickness to less than 1 mm and air dried before autoradiography. This procedure resulted in low gel backgrounds, sharper bands and reduced exposure times.

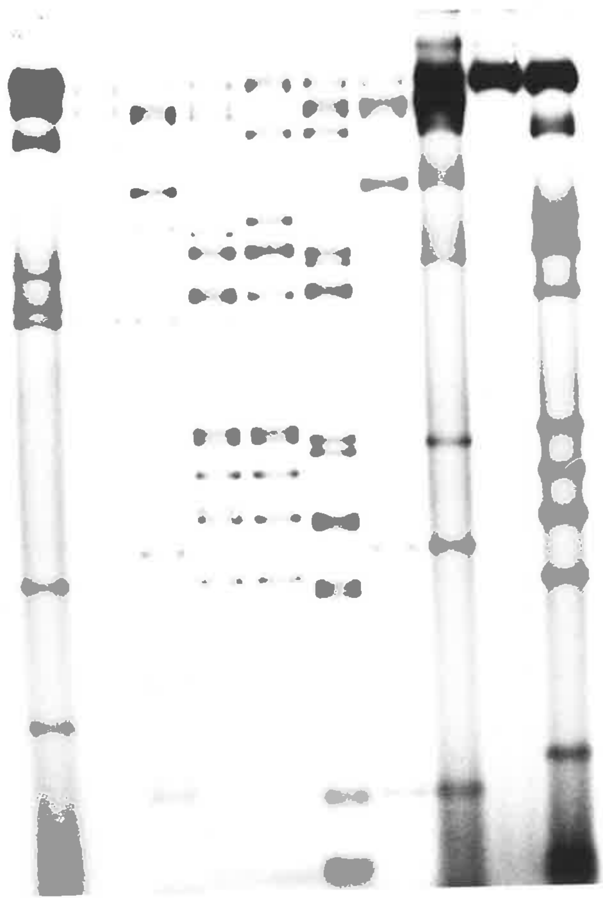
a. 0.6% agarose gel.

b. 1.8% agarose gel.

a. B/X B B/H E/B E E/H H H/X X E/X



b.



In order to localize the large number of Eco RI fragments, the 5 major Hind III fragments were excised from a low melting point agarose gel and digested with Eco RI and fractionated on an agarose gel in parallel with Eco RI cut λCFP1 DNA as size markers (Figure 6.6). The largest Hind III fragment of 24.1 Kb was cleaved by Eco RI to release the entire left vector arm of 20 Kb and a 4.1 Kb fragment which was present in the λCFP1 Eco RI digest. The second excised Hind III band contained two fragments, a 5.7 Kb vector fragment which was not cut with Eco RI, and a 5.8 Kb insert fragment which was cleaved by Eco RI to produce an insert fragment of 0.47 Kb and a portion of the right vector arm of 5.3 Kb. The third Hind III fragment of 4.8 Kb was cleaved by Eco RI to release 0.65, 0.66, 1.57 and 1.92 Kb fragments. The fourth Hind III fragment of 2.7 Kb was cleaved by Eco RI to release 0.92 and 1.64 Kb fragments and the fifth Hind III fragment of 2.65 Kb was cleaved by Eco RI to release 0.88 and 1.8 Kb fragments. A very small amount of cross-contamination of the 2.7 and 2.65 Kb fragments could be seen, but this was to be expected for fragments which migrated so close to one another.

The mapping data presented in Figures 6.5 and 6.6 combined with the restriction digest and Southern transfer data of Figure 6.3 and 6.4 culminated in the production of the restriction map for λCFP1 shown in Figure 6.7. The limit of regions which hybridized to the 5′-end of pCFK22 either weakly or strongly are delineated by the thin and thick bars respectively, and an exploded view is shown depicting the alignment of the sequence derived from the cDNA clones pCFK10 and 22.

6.3.1e THE NUCLEOTIDE SEQUENCE OF THE pCFK22 GENE.

The restriction map described above identified fragments located 5′ and 3′ to the existing pCFK22 gene sequence which was previously described in 6.3.1c. These fragments were sub-cloned directly into M13 and sequenced as described in Chapter 2. The complete sequence of the pCFK22 gene, including 230 bp of 5′ and 110 bp of 3′-flanking sequences is shown in Figure 6.8.

FIGURE 6.6

LOCALIZATION OF THE Eco RI FRAGMENTS OF λCFP1.

In order to localize the large number of Eco RI fragments found in λCFP1, the 5 largest Hind III fragments were fractionated by electrophoresis on a low melting point agarose gel (data not shown) and purified as described in 2.2.18b. The isolated fragments were then digested with Eco RI, end-labelled and fractionated on a 0.8% agarose gel in parallel with labelled Eco RI cut λCFP1 DNA used as markers (tracks labelled M). All conditions for electrophoresis, end-labelling and autoradiography are as described in the legend to Figure 6.5. The tracks have been labelled in order of decreasing Hind III size, ie:

1. 24 Kb (containing a 20 Kb λ-arm);
2. a doublet of 5.7 and 5.8 Kb (the 5.7 Kb fragment is from λ);
3. 4.8 Kb;
4. 2.7 Kb;
5. 2.65 Kb.

A small amount of cross-contamination of fragments 4 and 5 is evidenced by the faint bands in those tracks.

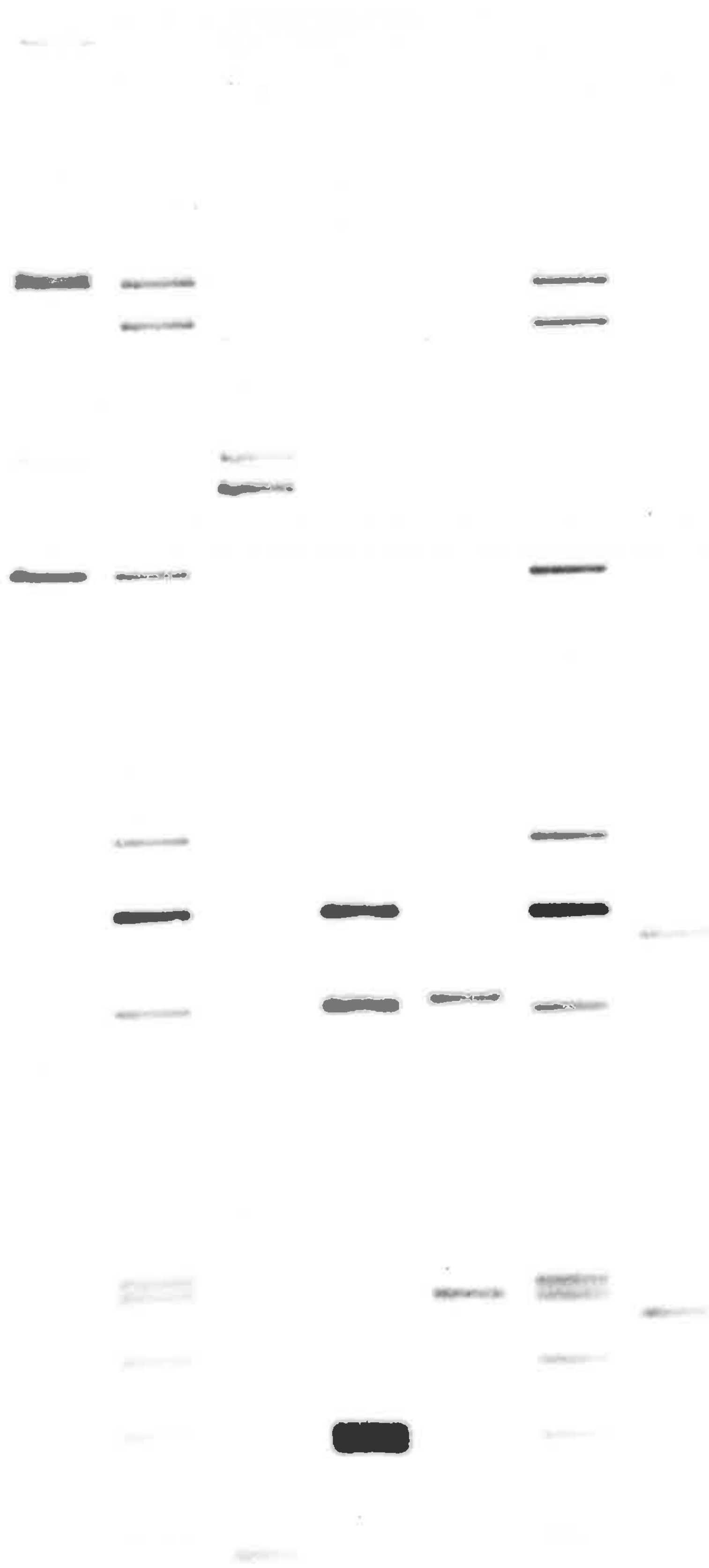

1
M
2
3
4
M
5

FIGURE 6.7

RESTRICTION MAP FOR λCFP1.

The restriction map for λCFP1 was derived from data presented in Figures 6.3, 4, 5 and 6. The limit of regions which hybridized to the pCFK22 probe (see Figure 6.1), either weakly or strongly, are delineated by the thin and thick bars respectively. An exploded view is shown depicting the alignment of the sequence derived from the cDNA clones pCFK10 and 22 and the position of the probe used to select λCFP1 from the genomic library.

Restriction sites are :

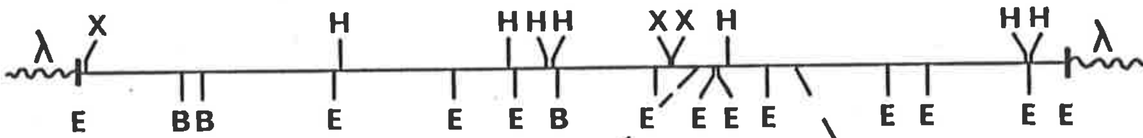
B......Bam HI.

E......Eco RI.

H......Hind III.

X......Xho I.

scale in Kb



scale in bp

FIGURE 6.8

THE COMPLETE NUCLEOTIDE SEQUENCE OF THE pCFK22 GENE.

The complete nucleotide sequence of the pCFK22 gene and its 5′ and 3′-flanking sequences was determined by sequencing several fragments from λCFP1 using the techniques described in Chapter 4. The DNA sequence of 2,017 bp shows the entire gene sequence including the 672 bp intron, the 230 bp of 5′-flanking sequence and 107 bp of 3′-flanking sequence. The TATA box, cap site, 5′-splice site, 3′-splice site, initiation codon, termination codon, polyadenylation sequence and position of poly(A) tract have been boxed using a solid line to indicate their position. Putative alternative TATA box, cap and 5′-splice sites have been boxed using a dashed line.

pCFK22 GENE SEQUENCE

```
         10         20         30         40         50         60
 GAATTCCATG CCATTTTACT AAAAATGTGC TGAGCCTATT TCTTCCTGTT TTAAGATGAA

         70         80         90        100        110        120
 AGAAATTCTG GAGGTATTAC AGGTGTGTCA TCACATTGAA TTTGAAAACT ATGAAAATCA

        130        140        150        160        170        180
 CTGAGTCTGA TTACATCAGT TGGAGGAACA CTGCACGCAA CAAGAGCCAT CGTGCAGCGA

        190        200        210        220        230        240
 ACACTGGAGC TTGAGAGGGT ATATATACAG ACCTCGAGCT CCGGAGCTTC ATTCGCTTCT
                     TATA Box                         Cap Site

        250        260        270        280        290        300
 CTCGAGTTCC TCTCCTCGGT GAACCGGGTA AGTCCACACT GCCTATACTC TACTCTTCTT
                        5'-splice Site

        310        320        330        340        350        360
 GACATGTCTT TGTGTGTATA TATCTGTATG CATGCTGTAA AAATAATATG TGCGTGTCAG
          Alternative TATA Box          Alternative Cap Site

        370        380        390        400        410        420
 TTTTAGATGG TGAATAGAGT GTTTTCTAAG TACTGCCAGG CACTATGTTT TGTCCAGAGA
Alternative 5'-splice Site

        430        440        450        460        470        480
 GGCTGTGGAT GCCCCATCCC TGGAGATGCT CAAGACCAGG TTGGATGGAG CACTGGCCAA

        490        500        510        520        530        540
 CCTGGGCTAG TAGCATATCT GGATGTTGGT GGTCCTGCCT GTGGCAGGGG TTGGAATGGC

        550        560        570        580        590        600
 TTCAAGCTGC ACCAGGGGAG GTTCAAGGCC AGACGTTAGG AAATACTACT TCTCTGAAAG

        610        620        630        640        650        660
 GGTGGTCAGG CACTGGAATG GGCTGCCCAG GGAGGTGGTG GAATCACCAA CCCTGGAGGT

        670        680        690        700        710        720
 GTTCAAGGAA CGTTTGGATG TTGTGTTGAG GGACATGGTT TAATGAGAAC TACTGGTGAT

        730        740        750        760        770        780
 AGGTGGATGG TTGGACTGGG TGATCCTGTG GGCCTTTTCC AACCCTGGTG ATTCTATGGA

        790        800        810        820        830        840
 TGATCCTTGG GGCCCCTTCC AACCCAAGCC ATTCTATGAT TCTATATGAA AAAAGTTTTG

        850        860        870        880        890        900
 GGGGATTCAG AAGGCTAAAC AGACAAATGA GGTGGCTAAG AGGTGAAAAG CCTTCTGGTG

        910        920        930        940        950        960
 AGGACATGAA TTCACTCTGT TCCCTGCTGT GTGTTTCAGG TTTCCCTCCA ACAACCAGCA
                                     3'-splice Site

        970        980        990       1000       1010       1020
 ATGACTTTCC ACAGGGAATT CTACAATGAC GAGCACTACT CACCCTTCTG CCAGGAAGAC
Initiation Codon

       1030       1040       1050       1060       1070       1080
 CTGCATGGTC TCTGGGGCCT GAATGACCAC CGATTCAAGC ACCTATATGG CCTCCACCGA

       1090       1100       1110       1120       1130       1140
 GACCACCACC ATGACTACAA CCAACACTGG AGCCCTTATG GCTACAACAG AAGCTTTGGG

       1150       1160       1170       1180       1190       1200
 AGCCTGTATG GGAACAGGAG CTTGAGCTCT CATGGAGGCT ATTATGGGCA TGGGGACTTC

       1210       1220       1230       1240       1250       1260
 TTTGGTTTTG GCACCGCCAC CCCTACTTTT CTCAGATTGG CCACAGATAC TGGTATTGAA

       1270       1280       1290       1300       1310       1320
 GCTGCTATCC ATGCTAAACC ACGCAGACAC GGAAGTCAAG AAATGAGGCA GATCCTTTGC

       1330       1340       1350       1360       1370       1380
 TAAGACCATG GTGCTGAGAA ACGCTGGGTA CCCATACCTC TGGCTGTGTT CAGTGAGATG
Termination Codon

       1390       1400       1410       1420       1430       1440
 TGTGGGTGCC AGGCTGCATT AACACTACCT TATATGTATA ACTACACTAA TATACACTAA

       1450       1460       1470       1480       1490       1500
 TATTTCTATC TCTTCTGCTA TCGCTCTAGA GGGCAATTTA AACTGCAGAA GTATTGAGGC

       1510       1520       1530       1540       1550       1560
 ACCCCTACAC CCTGCTGAAA GTCCTGCTTT ACTCCACTAG AGTGGGTTAG ACAGTTTCCT

       1570       1580       1590       1600       1610       1620
 TCTAAATTAA ATAATAGCCA CTTTGTGTTG CTATAGTGAT ATATATGGTC CCGTGTCAAT

       1630       1640       1650       1660       1670       1680
 GAGAATCAGA CTTTAACTGC CATGAAGAAG TCCACTTTGC TCAACTATGA ACTTCTCCAA

       1690       1700       1710       1720       1730       1740
 AGACAACACA TAGAGAACAT TACATATCTA TGCTGTGCCT CATATGGTAG TGATCAGAAT

       1750       1760       1770       1780       1790       1800
 TTATGAAGAA CTCGTGAGCA AGAAACGTGC CTGGAGTGAT GAATTCCATG AAGAGAGATC

       1810       1820       1830       1840       1850       1860
 ATCATCTGGT ATCAGACCCA CTCTACTTGC TGAAGACCCT TTTGTAAAGC TGCAAACTTA

       1870       1880       1890       1900       1910       1920
 TTCTTTCTGT AACTTCTCAC TCTTTATTCA TTAAAATGAT TCTGCATAGT AACATCAGCT
                        Polyadenylation Sequence   Polyadenylation Site

       1930       1940       1950       1960       1970       1980
 TGTCTTCTCC TTTTGTCTTT CGTGGCACAC TCCAGAAAGG GACCAGCAGC AGTGCCACTG

       1990       2000       2010
 ACAGACATAC ACTAAGATGG AGGTTAGACT CCAGGTG
```

The conceptual translation of the likely coding region which consists of a protein of 120 amino acids including the N-terminal methionine, is shown in Figure 6.9. The start of the protein sequence was assumed to be the first initiation codon (AUG) 3′ to the last termination codon (the UGA which forms part of the Eco RI recognition sequence at nucleotide sequence position 908 in Figure 6.8). If this gene was typical of the majority of eukaryotic genes it would contain an uninterrupted 5′-untranslated region of 50 to 100 bp leading to the mRNA "cap" site which would be preceded by consensus TATA and CAAT boxes in the 100 or so bases 5′ to the "cap" site. Unfortunately none of these features were evident in the corresponding region of the pCFK22 gene. Two possibilities existed to explain this phenomenon, either the 5′-untranslated region was much larger than most of the eukaryotic genes examined to date, or an intron existed, either in the 5′-untranslated region or the coding region.

To examine the first of these possibilities, several fragments were prepared from the region 3′ to the putative protein initiation site for use as primers on feather mRNA in order to determine the distance from the 5′-end of those fragments to the 5′-end of the mRNA derived from this gene. Two fragments were used, one a 59 bp Eco RI/Hae III fragment which primed about 20 bp before the putative initiation codon, and the other a 170 bp Hae III fragment which primed 110 bp before the putative initiation codon. These fragments were labelled by end-filling and elongated after hybridization to embryonic feather poly(A)-plus mRNA using AMV reverse transcriptase in the presence of cold nucleotide triphosphates, as described in 6.2.3. These transcription products were separated on an 8 M urea polyacrylamide gel along with end-filled pBR322 DNA digested with Msp I used as molecular weight markers. The autoradiograph of this gel is shown in Figure 6.10. The 59 bp Eco RI/Hae III fragment was elongated to 135 to 136 bp indicating that the 5′-untranslated region extends 58 to 59 bp past the putative initiation codon. This result was confirmed by the elongation of the 170 bp Hae III fragment by about 340 bp indicating a 5′-untranslated region of approximately 60 bp.

FIGURE 6.9

CONCEPTUAL TRANSLATION OF THE pCFK22 GENE.

Conceptual translation of the probable coding region of the pCFK22 gene found in λCFP1. The protein sequence is presented in three letter amino acid code aligned with the corresponding nucleotide sequence. Only the nucleotide sequences of the putative coding region and flanking sequences are shown.

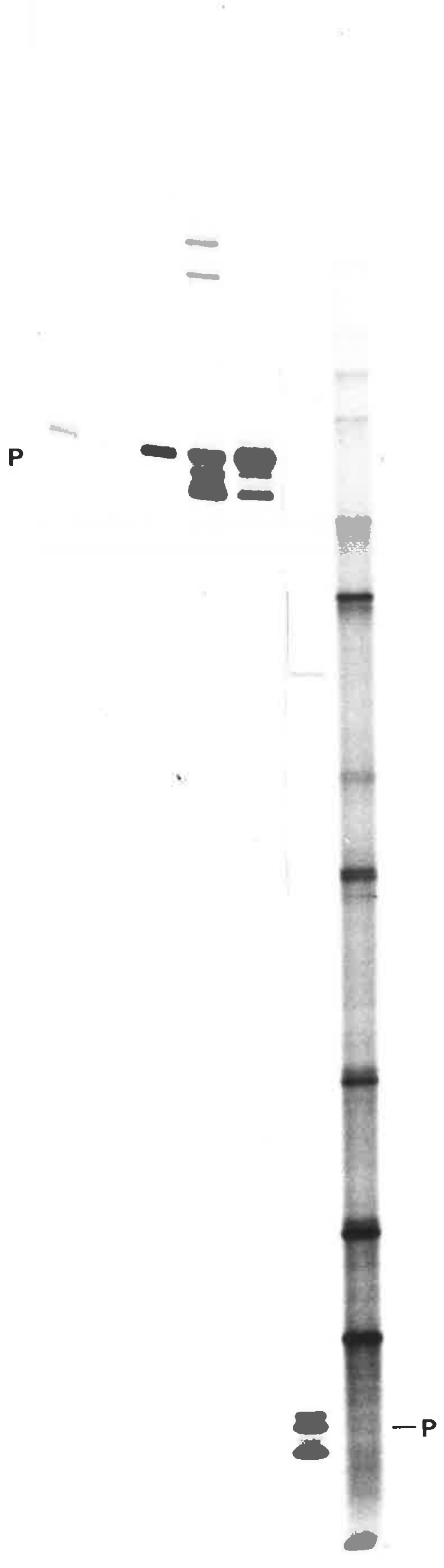
Translation of the pCFK22 gene

```
                                        Met Thr Phe His Arg Glu Phe Tyr Asn Asp
GTG TTT CAG GTT TCC CTC CAA CAA CCA GCA ATG ACT TTC CAC AGG GAA TTC TAC AAT GAC
          940          950          960          970          980          990

Glu His Tyr Ser Pro Phe Cys Gln Glu Asp Leu His Gly Leu Trp Gly Leu Asn Asp His
GAG CAC TAC TCA CCC TTC TGC CAG GAA GAC CTG CAT GGT CTC TGG GGC CTG AAT GAC CAC
         1000         1010         1020         1030         1040         1050

Arg Phe Lys His Leu Tyr Gly Leu His Arg Asp His His His Asp Tyr Asn Gln His Trp
CGA TTC AAG CAC CTA TAT GGC CTC CAC CGA GAC CAC CAC CAT GAC TAC AAC CAA CAC TGG
         1060         1070         1080         1090         1100         1110

Ser Pro Tyr Gly Tyr Asn Arg Ser Phe Gly Ser Leu Tyr Gly Asn Arg Ser Leu Ser Ser
AGC CCT TAT GGC TAC AAC AGA AGC TTT GGG AGC CTG TAT GGG AAC AGG AGC TTG AGC TCT
         1120         1130         1140         1150         1160         1170

His Gly Gly Tyr Tyr Gly His Gly Asp Phe Phe Gly Phe Gly Thr Ala Thr Pro Thr Phe
CAT GGA GGC TAT TAT GGG CAT GGG GAC TTC TTT GGT TTT GGC ACC GCC ACC CCT ACT TTT
         1180         1190         1200         1210         1220         1230

Leu Arg Leu Ala Thr Asp Thr Gly Ile Glu Ala Ala Ile His Ala Lys Pro Arg Arg His
CTC AGA TTG GCC ACA GAT ACT GGT ATT GAA GCT GCT ATC CAT GCT AAA CCA CGC AGA CAC
         1240         1250         1260         1270         1280         1290

Gly Ser Gln Glu Met Arg Gln Ile Leu Cys ***
GGA AGT CAA GAA ATG AGG CAG ATC CTT TGC TAA GAC CAT GGT GCT GAG AAA CGC TGG GTA
         1300         1310         1320         1330         1340         1350
```

FIGURE 6.10

LOCALIZATION OF THE pCFK22 GENE mRNA CAP SITE BY EXTENSION OF A SPECIFIC PRIMER ON mRNA.

Two specific fragments, one a 59 bp Eco RI/Hae III fragment which primed about 20 bp before the putative initiation codon and the other a 170 bp Hae III fragment which primed 110 bp before the putative initiation codon, were end-labelled and used as primers on poly(A)-plus feather mRNA as described in 6.2.3. The elongated primers were fractionated on a 6% 8 M urea polyacrylamide gel along with end-filled Msp I digested pBR322 DNA used as markers. The position of the primers and the full length elongation products are indicated at the side of the gel. Increasing concentrations of mRNA were used with the 170 bp fragment, as indicated above the 170 bp primer tracks, to optimize the proportion of primer radioactivity which was elongated. The marker track is labelled M, while the 170 and 59 bp tracks are labelled appropriately.



170
μg ·5 1 3 6 10 60 M
P
— P

Having established that the site for transcription initiation was approximately 60 bp 5′ to the putative protein initiation codon, this region of the pCFK22 gene sequence was examined for the presence of possible cap sites and consensus TATA and CAAT sequences. None were found, suggesting that an intron might exist between the 5′-end of the primer fragments used, i.e. 20 bp 3′ to the putative protein initiation codon, and the 5′-end of the gene.

The 80 or so bases 5′ to the end of the Hae III primer were therefore examined for the presence of intron 3′-splice sites. Two were found, one (CAGGT) with the splice point 21 bases 5′ to the putative initiation codon (the same consensus splice sequence and distance which exists between the 3′-splice site and initiation codon in the keratin gene system, see 4.3.12), and the other (CAGGG) which occurred about 15 bases 3′ to the putative initiation codon, i.e. within the putative coding region. Although the second mentioned of these splice sites was a better fit to the consensus sequence than the first (Brethnack and Chambon, 1981) when the ten bases on either side were considered, it was felt that it was probably not the correct one because only one base existed between this splice point and the Eco RI site which was known to be in the mRNA (the Eco RI/Hae III fragment described earlier acted as a primer on feather mRNA, the last G of the sequence CAGGG is the first base of the Eco RI site (GAATTC), i.e. the sequence runs, CAGGGAATTC).

In order to resolve which of these 3′-splice sites was the correct one, the end-labelled 59 bp Eco RI/Hae III primer fragment described earlier was elongated on poly(A)-plus embryonic feather mRNA in the presence of the appropriate dideoxynucleotide triphosphates, as described in 6.2.4 and the products separated on an 8 M urea sequencing gel (Figure 6.11). Although the dideoxy adenosine and guanine concentrations were too high in these reactions, resulting in very little radioactivity in sequencing products near the end of the mRNA, it was possible, with difficulty, to read the 74 bases from the primer to the last residue, which was assumed to be an adenosine. The sequence of the mRNA was identical to that of

FIGURE 6.11

NUCLEOTIDE SEQUENCING OF THE 5'-END OF THE pCFK22 GENE mRNA BY EXTENSION OF A SPECIFIC PRIMER IN THE PRESENCE OF DIDEOXYNUCLEOTIDES.

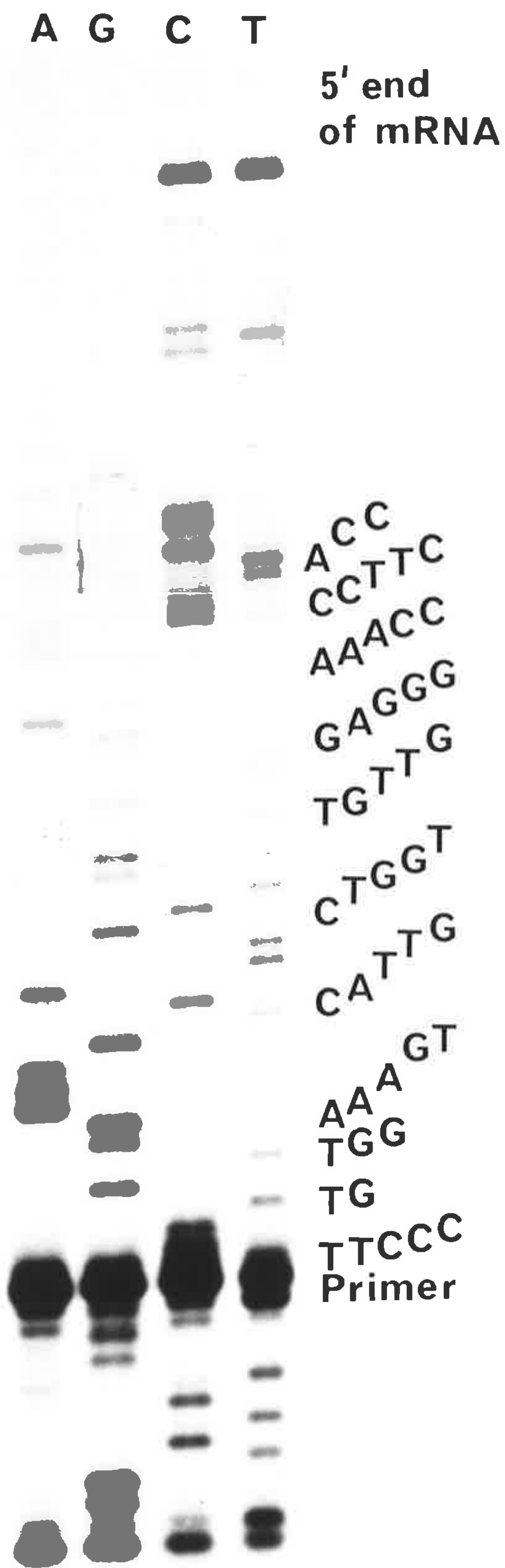
The 59 bp Eco RI/Hae III fragment described in the legend to Figure 6.10 was elongated on poly(A)-plus embryonic feather mRNA in the presence of the appropriate dideoxynucleotide triphosphates, as described in 6.2.4 and the products were separated on an 8 M urea polyacrylamide sequencing gel. The figure is a composite derived from two different autoradiographic exposures of the gel in order that the longer transcription products could be visualized. The determined sequence is shown at the side of the gel approximately aligned with the appropriate bands. The tracks are:

A. dideoxyadenosine reaction.

G. dideoxyguanosine reaction.

C. dideoxycytosine reaction.

T. dideoxythymidine reaction.



A G C T
5′ end
of mRNA
ACC
CCTTC
AAACC
GAGGG
TGTTG
CTGGT
CATTG
AAAGT
TGG
TG
TTCCC
Primer

the pCFK22 gene beyond the first putative splice site adjacent to the Eco RI site but diverged from the gene sequence beyond the second putative splice site, indicating that this was the correct splice site. This result confirmed that the protein was initiated at the methionine codon referred to earlier, since this was the first AUG present in the mRNA sequence.

The 5′-untranslated region sequence derived from priming on feather mRNA in the presence of dideoxynucleotides (Figure 6.11) confirmed that the pCFK22 gene sequence, from the initiation codon to the 3′-splice site, was the same as the mRNA sequence. It also disclosed the 37 nucleotides of mRNA sequence from the position of the intron to the 5′-end of the mRNA (5′-leader sequence) which should be located on the 5′-side of the intron in the pCFK22 gene. This sequence data was used to search the pCFK22 gene DNA sequence 5′ to the initiation codon (about 1 Kb) for a matching sequence, in order to find the transcription initiation site and to reveal the length of the intron. Thus, the 37 bp 5′-leader sequence was found to be separated from the 3′-splice site of the pCFK22 gene by an intron of 672 bp (Figure 6.8). The 5′-leader sequence was found to be identical to that determined by primer extended sequencing on feather mRNA, confirming the residues which were difficult to read in that experiment (Figure 6.11).

A "TATA" or "Goldberg-Hogness" box (Goldberg, 1979), reportedly involved in the efficient initiation of transcription (Wasylyk *et al.*, 1980a, b), was found 31 bases from the transcription initiation site, which fits the predicted spacing for these sequences (Gannon *et al.*, 1979). The TATA box sequence for the pCFK22 gene (TATATAT), is in close agreement with the eukaryotic gene consensus sequence for this structure, as is the sequence around the mRNA start site and the 5′-splice site sequence (CGGGTAA) which was found at the junction of the 5′-leader sequence and the intron (Breathnack and Chambon, 1981).

The "CAAT" box sequence which has been observed about 70 to 80 bases 5′ to the point of transcription initiation in several cellular and viral genes (Efstratiadis

et al., 1980; Benoist *et al.*, 1980) could not be found in the 230 bases of sequence flanking the pCFK22 gene, however, a CAAT box has not been shown to be essential for initiation of transcription and it is absent from many genes, including at least one embryonic feather keratin gene (gene C; Molloy *et al.*, 1982).

A feature of particular interest in the intron sequence was the presence of an alternative TATA box (TATATAT), identical to the first, found 29 bp from an alternative cap site and an alternative 5′-splice site (Figure 6.8). These three features, which concur with the consensus sequences for each structure, together form an alternative promotor and 5′-leader sequence. If the sequence of the 5′-leader in embryonic feathers was not known from the primer extended sequencing experiment (Figure 6.11), the "coincidental" presence of this alternative promotor and 5′-leader sequence would have made it impossible to decide which of these mRNA start points was used in the feather. Although not used in the feather, it is possible that the alternative 5′-leader is used in another tissue or developmental stage.

The sequence of the cDNA clone, pCFK22, revealed that it coded for a protein of at least 110 amino acids. The complete nucleotide sequence of the pCFK22 gene now made it possible to determine the total length of the protein which it encoded to be 120 amino acids, just 10 amino acids longer than that determined for the cDNA clone. Table 6.1 shows that the 10 extra amino acids derived from the pCKF22 gene sequence, combined with the 110 amino acids derived from the cDNA clone sequence, did not significantly change the amino acid composition of the proteins, which still strongly resembled the amino acid composition of fast protein. However, amino acid sequencing data from fast protein was still required to positively identify the pCFK22 gene as a fast protein gene.

6.3.1f COMPARISON OF THE pCFK22 GENE WITH THE KERATIN GENES.

The pattern of structural features which are exhibited by all of the keratin

TABLE 6.1

COMPARISON BETWEEN THE AMINO ACID COMPOSITIONS OF FAST PROTEIN AND THOSE DETERMINED FROM THE CODING REGIONS OF THE GENOMIC AND cDNA pCFK22 SEQUENCES.

The theoretical amino acid compositions of the proteins encoded by the coding regions in the genomic and cDNA pCFK22 nucleotide sequences were determined by conceptual translation and calculation of the content of the component amino acids, expressed as residues per 100 residues. The amino acid compositions of fast protein and feather keratin are included for comparison.

§ Data from Powell, (1979).

* Values for Asp and Glu include Asn and Gln content, respectively.

TABLE 6.1

COMPARISON BETWEEN THE AMINO ACID COMPOSITIONS OF FAST PROTEIN AND THOSE DETERMINED FROM THE CODING REGIONS OF THE GENOMIC AND cDNA pCFK22 SEQUENCES.

amino acid	Genomic pCFK22	cDNA pCFK22	fast§ protein	embryonic§ feather keratin
Cysteine	1.65	1.80	1.77	6.18
Asparagine	4.13	3.60	—*	—*
Aspartic acid	4.13	3.60	9.51*	6.79*
Threonine	4.96	4.50	1.74	5.37
Serine	6.61	7.21	8.46	12.68
Glutamine	3.31	3.60	—*	—*
Glutamic acid	5.79	5.41	4.73*	8.93*
Proline	3.31	3.60	4.45	10.93
Glycine	11.57	12.61	17.64	11.98
Alanine	4.13	4.50	1.90	4.99
Valine	0.00	0.00	2.86	8.01
Isoleucine	2.48	2.70	1.56	4.85
Leucine	8.26	9.01	7.43	7.36
Tyrosine	7.44	7.21	10.87	1.84
Phenylalanine	7.44	6.31	8.83	3.85
Lysine	1.65	1.80	1.43	0.16
Histidine	11.57	11.71	9.34	0.81
Arginine	7.44	7.21	7.42	5.04
Methionine	1.65	0.90	0.00	0.00
Tryptophan	1.65	1.80	0.00	0.00

genes as exemplified by gene C (Molloy *et al.*, 1982) is summarized in Figure 6.12 where they are compared to the features found in the pCFK22 gene. Dramatic structural similarities can be seen between these two gene types, yet they code for proteins which are evolutionarily totally unrelated. No similarities can be seen in the coding or 3′-untranslated sequences except that neither the keratins nor the pCFK22 gene contain any introns. However, remarkable similarities can be seen in the 5′-untranslated regions of these genes: they both have 5′-untranslated regions of the same length (58 nucleotides); they both contain the only intron which exists in each gene; this intron is found in the same position, splitting the 5′-untranslated regions of the genes into a 37 base 5′-leader sequence and a 21 base sequence from the intron to the initiation codon; they both contain TATA boxes but neither has a CAAT box. The genes are structurally dissimilar in this region in that the pCFK22 gene has an intron about twice the length of the gene C intron (pCFK22 gene, 672 bp; gene C, 324 bp).

Comparison of the nucleotide sequences of the pCFK22 gene and the keratin genes revealed that very little sequence homology existed in the 3′-untranslated region. Indeed, even the polyadenylation sequence (AAUAAA) present near the end of all the keratin genes and found in the majority of eukaryotic mRNAs was different in the pCFK22 gene, being AUUAAA (see Figure 6.8), which is not the polyadenylation signal commonly seen in eukaryotic mRNAs (Goeddel *et al.*, 1981). When examining introns from the keratin genes and the pCFK22 gene, it was obvious that, in addition to having introns of markedly different length, there were no regions of sequence homology between the introns of these two gene classes.

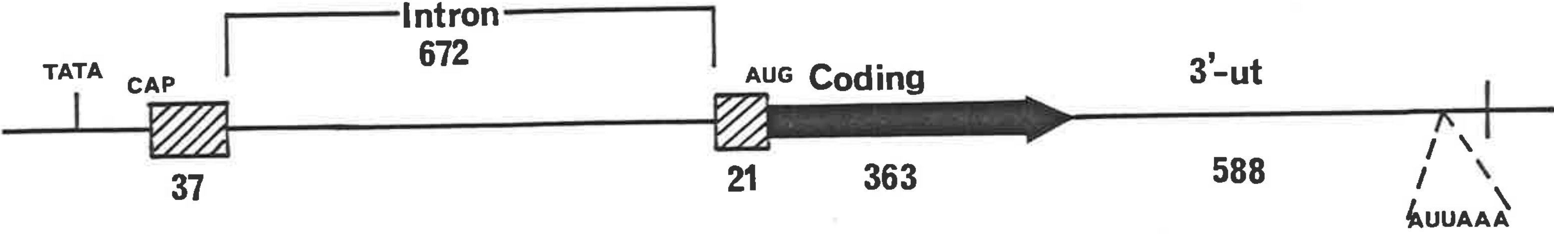
It is of interest however, that the keratin genes and the pCFK22 gene show a significant degree of sequence homology in their 5′-untranslated regions. It can be seen, from Figure 6.13, where the 5′-untranslated sequences of the pCFK22 gene and five keratin genes (a cDNA clone, pCFK26 and four genomic sequences, genes A, B, C and D) are aligned to maximize the homology of the keratin genes with the

FIGURE 6.12

COMPARISON BETWEEN THE STRUCTURAL FEATURES OF THE pCFK22 GENE AND THE KERATIN GENES.

The pattern of structural features displayed by the keratin genes, as exemplified by gene C (Molloy *et al.*, 1982), is displayed here compared to the structural features displayed by the pCFK22 gene. Note the similarity at their 5′-ends, ie. both contain an intron which interrupts the 5′-untranslated sequence at the same point to produce a 37 bp sequence from the 58 bp 5′-untranslated region.

pCFK22 Gene



Gene C

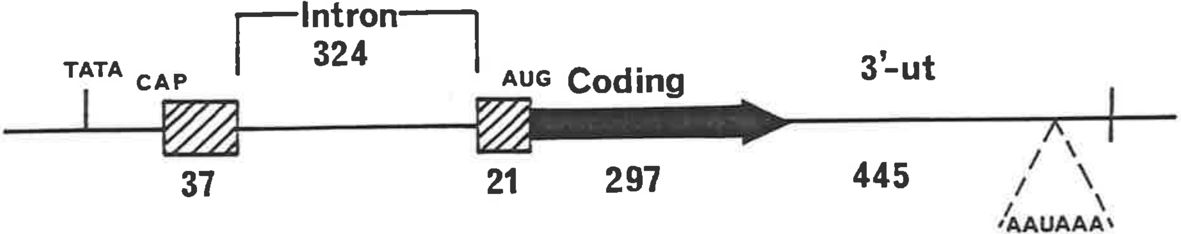


FIGURE 6.13

COMPARISON BETWEEN THE 5′-UNTRANSLATED REGIONS OF THE pCFK22 GENE AND THE KERATIN GENES.

The nucleotide sequences from the 5′-untranslated regions of the pCFK22 gene and five keratin genes (a cDNA clone, pCFK26 and four genomic sequences, genes A, B, C and D) are aligned to maximize the homology of the keratin genes with the pCFK22 gene. Matches between the pCFK22 sequence and the other sequences are indicated by an asterisk. Note that the sequence homology between these sequences is higher 5′ to the intron than 3′ to the intron.

```
                              5'-LEADER SEQUENCE                        3' TO INTRON          CODING
pCFK22          |                                        |                         |
Gene      CTTC|ATTCGCTTCTCTCGAGTTCCTCTCCTCGGTGAACCGG|GTTTCCCTCCAA--CAACCAGCA|ATGAC
          * **|** * ******* *   ***   ***** *******   *|**   ******    ****** * |*** *
Gene A    CCTC|ATCCACTTCTCTTGCCTTCTCCTCCTTGGTGAACAAG|GTCCACCTCCATCCCAACCA-CC|ATGTC
          * **|** * * ***** *   ***   ***** *******   *|**   ******    *   **** |*** *
Gene B    CCTC|ATCCACGTCTCTTGCCTTCTCCTCCTTGGTGAACAAG|GTCGACCTCCAT--CCTACAGCC|ATGTC
          * **|** * ******* *   ***   ***** *******   *|** * *  ***    *   **** |*** *
Gene C    CCTC|ATCCACTTCTCTTGCCTTCTCCTCCTTGGTGAACAAG|GTCTACTCCCAT--CCTACAGCC|ATGTC
              |                                        |**   ******    * **** * |*** *
Gene D        |                                        |GTCCACCTCCAT-CCCACCA-CC|ATGTC
              |** * ******* *   ***   ***** **** **   *|**   ******    * **** * |*** *
pCFK26        |ATCCACTTCTCTTGCCTTCTCCTCCTTGGTGCACAAG|GTCGACCTCCAT-CCCACCA-CC|ATGTC
              |                                        |                       |
              ^                                        ^                          ^
          CAP SITE                           INTRON SPLICE SITE            START CODON
```

pCFKK22 gene, that this sequence homology is measurably higher in the 5′-leader sequence, i.e. 5′ to the intron, than 3′ to the intron. This result is more obvious from Table 6.2 where the percentage of matching bases in these two sections of the 5′-untranslated region are compared with the percentage homology for these sequences when the appropriate regions of the keratin genes are compared with one another.

Table 6.2 shows that although the pCFK22 gene 5′-leader sequence is homologous to the 5′-leader sequence of the keratin genes, the homology exhibited between the keratin genes is substantially higher for this region of sequence (approaching 100%). Also, the homology of the 5′-untranslated sequence 3′ to the intron in the keratin genes is much lower (about 80% when account is taken of manipulations required to align these sequences) than that of the keratin 5′-leader sequences, which is also the situation when these two sections of sequence from the pCFK22 gene are compared with the corresponding keratin gene sequences, i.e. the homology between the pCFK22 gene 5′-leader sequence and the keratin 5′-leader sequences is higher than that observed when the remainder of the pCFK22 gene 5′-untranslated sequence is compared to the keratin genes.

6.3.2 PURIFICATION AND SEQUENCING OF FAST PROTEIN.

6.3.2a SEPARATION OF FAST PROTEINS BY ION EXCHANGE FPLC.

Recently Pharmacia developed a new liquid chromatography system (FPLC) for the fast separation of proteins, consisting of a simple low pressure (less than 5 MPa) dual pump solvent delivery system and gradient programmer linked to a range of columns (Mono Q, strong anion exchange; Mono S, strong cation exchange; Polyanion Si, weak anion exchange; Mono P, chromatofocusing) using a revolutionary column support consisting of spherical beads which have very little

TABLE 6.2

PERCENTAGE SEQUENCE HOMOLOGY OF THE 5′-LEADER SEQUENCE AND THE REMAINDER OF THE 5′-UNTRANSLATED REGION FOR THE pCFK22 GENE AND THE KERATIN GENES.

The two sections of sequence, ie. 5′ and 3′ to the intron, in the 5′-untranslated region of the pCFK22 gene and the 5 keratin genes were aligned to maximize homology, as shown in Figure 6.13 and the percentage of matching bases for each region was calculated. Comparisons between the keratin genes and between the pCFK22 gene and the keratin genes are shown.

TABLE 6.2

PERCENTAGE SEQUENCE HOMOLOGY OF THE 5′-LEADER SEQUENCE AND THE REMAINDER OF THE 5′-UNTRANSLATED REGION FOR THE pCFK22 GENE AND THE KERATIN GENES.

Gene comparison	Percentage sequence homology for the 5′-leader sequence	21 bp 3′ to the intron
Comparison of the pCFK22 gene with the keratin genes.		
pCFK22/Gene A	73	71
pCFK22/Gene B	70	62
pCFK22/Gene C	73	57
pCFK22/Gene D	—*	67
pCFK22/pCFK26	70	67
Comparison of the keratin genes with themselves.		
Gene A/Gene B	97	82
Gene A/Gene C	100	77
Gene A/Gene D	—*	95
Gene A/pCFK26	97	91
Gene B/Gene C	97	90
Gene B/Gene D	—*	82
Gene B/pCFK26	95	86
Gene C/Gene D	—*	77
Gene C/pCFK26	97	73
Gene D/pCFK26	—*	95

* The 5′-leader of Gene D has not been sequenced.

variation in shape or diameter, resulting in high resolution column chromatography with low operational back pressures which allows high flow rates even at quite low operating pressures (1 to 2 MPa). This system was chosen for the initial attempts to purify fast proteins.

The starting material for these investigations was a fraction prepared from total unfractionated S-carboxymethylated feather proteins by the addition of $MgCl_2$ which results in the preferential precipitation of fast proteins (Powell, 1979). This protein fraction, although essentially free of keratins, appeared to be substantially contaminated by higher molecular weight proteins (see Figure 6.17 track Cfp).

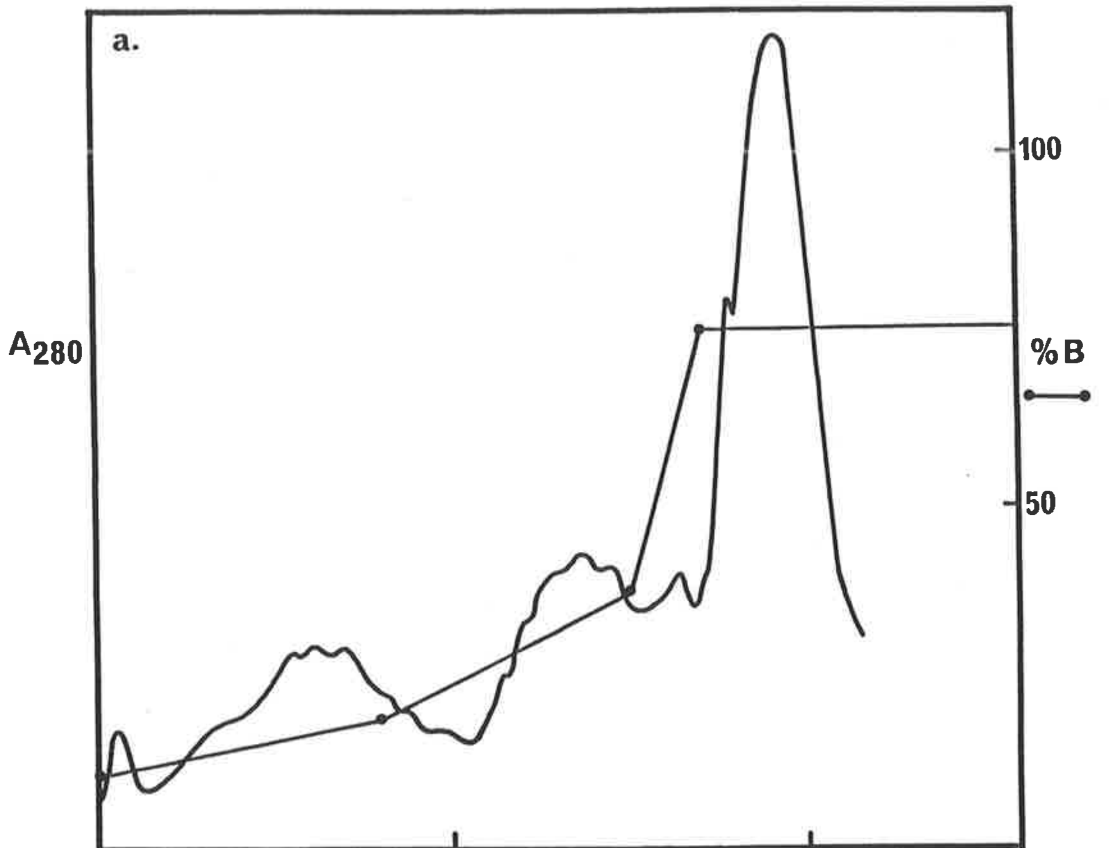
The amino acid composition of fast proteins (Table 6.1) illustrated the presence of large proportions of both the acidic (aspartic acid and glutamic acid) and basic (lysine, arginine and histidine) amino acids. It was thought that the charged nature of these amino acids could result in a purification of fast proteins by subjecting the crude protein to chromatography on either anion or cation exchange chromatography supports.

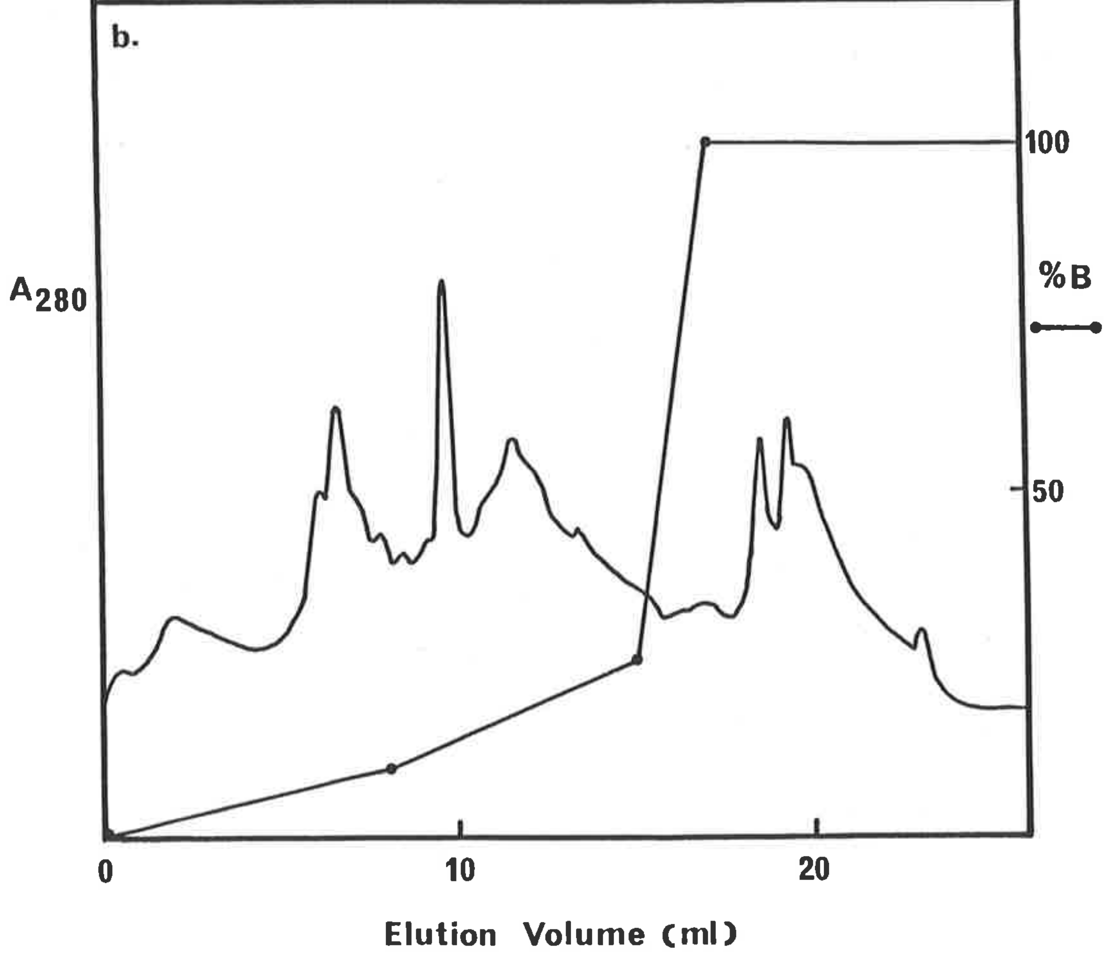
The crude protein fraction was dissolved in 2 M urea since it was only poorly soluble in aqueous solution and 200 μg was then applied to a Mono Q (strong anion exchange) column buffered with Tris-HCl at pH 7.6, thus ensuring that all acidic amino acids would be charged, and eluted with a NaCl gradient which was crudely optimized to maximize the separation of the peaks obtained (Figure 6.14a). It was obvious that although some separation of protein components was occurring each broad peak contained several components which were not resolved, probably due to aggregation. In an attempt to resolve these protein components the urea concentration was increased to 8 M and 200 μg of protein was applied and eluted under similar conditions (Figure 6.14b). In this case the increased urea concentration reduced protein aggregation and resulted in sharper, better resolved peaks, although many protein components obviously still eluted as groups of peaks rather than as well separated homogeneous peaks.

FIGURE 6.14

SEPARATION OF FAST PROTEINS BY FPLC ANION EXCHANGE CHROMATOGRAPHY IN THE PRESENCE OF UREA.

200 μg of crude fast protein, prepared by $MgCl_2$ precipitation from total SCM-feather protein (Powell, 1979), was chromatographed on a Pharmacia Mono Q strong anion exchange column exactly as described in 6.2.5, except that in a. the column was run in 2 M urea and in b. the column was run in 8 M urea. In both cases buffer A was 20 mM Tris-HCl pH 7.6 and buffer B was 20 mM Tris-HCl pH 7.6, 1 M NaCl. The elution gradient used is indicated on the figure.


a.
A280
100
%B
50


b.
A280
100
%B
50
0
10
20
Elution Volume (ml)

Samples of crude fast protein were loaded onto a Mono S (strong cation exchange) column in 8 M urea at pH 4.8 which was chosen in order to ensure that all basic amino acids, including histidine (pI=6.0), were fully charged. The profile shown in figure 6.15a was obtained from 400 μg of crude protein applied to the Mono S column and eluted with the illustrated salt gradient which had previously been optimized for maximum separation of crude fast protein components. At least 20 protein components were distinguishable using this column but when a larger amount of crude fast protein (3 mg) was applied to the column under the same conditions (Figure 6.15b), less peaks appeared to be present and they generally seemed to elute at lower salt concentrations. This effect was probably due to increased interactions and aggregation of the protein components when loaded in higher concentration. It is probable that the many peaks achieved using this column merely reflected different degrees of aggregation during the elution of proteins from the column since fractions taken across the profile shown in Figure 6.15b appeared to contain a similar distribution of proteins when analysed by chromatography on the Mono Q column and by electrophoresis on SDS polyacrylamide gels (data not shown). Cation exchange chromatography was therefore abandoned in preference to the anion exchange column (Mono Q) because there seemed to be less inter-protein interactions and aggregation at the higher pH.

When the salt gradient used to elute 400 μg of the crude fast protein fraction from the Mono Q column, in 8 M urea at pH 7.6, was further tuned to maximize the efficient separation of proteins, a larger number of better resolved peaks was observed (Figure 6.16a). However, when a preparative amount of protein (8 mg, Figure 6.16b) was applied to the column a group of peaks (designated fraction C in Figure 6.16c) decreased in abundance relative to the other peaks, and a new peak (fraction A) appeared which did not bind to the column. This peak (fraction A) decreased markedly in size and was accompanied by a corresponding increase in the size of the fraction C peaks when a similar amount of protein (10 mg) was applied to the column immediately after being heated for 2 minutes at 100°C and eluted

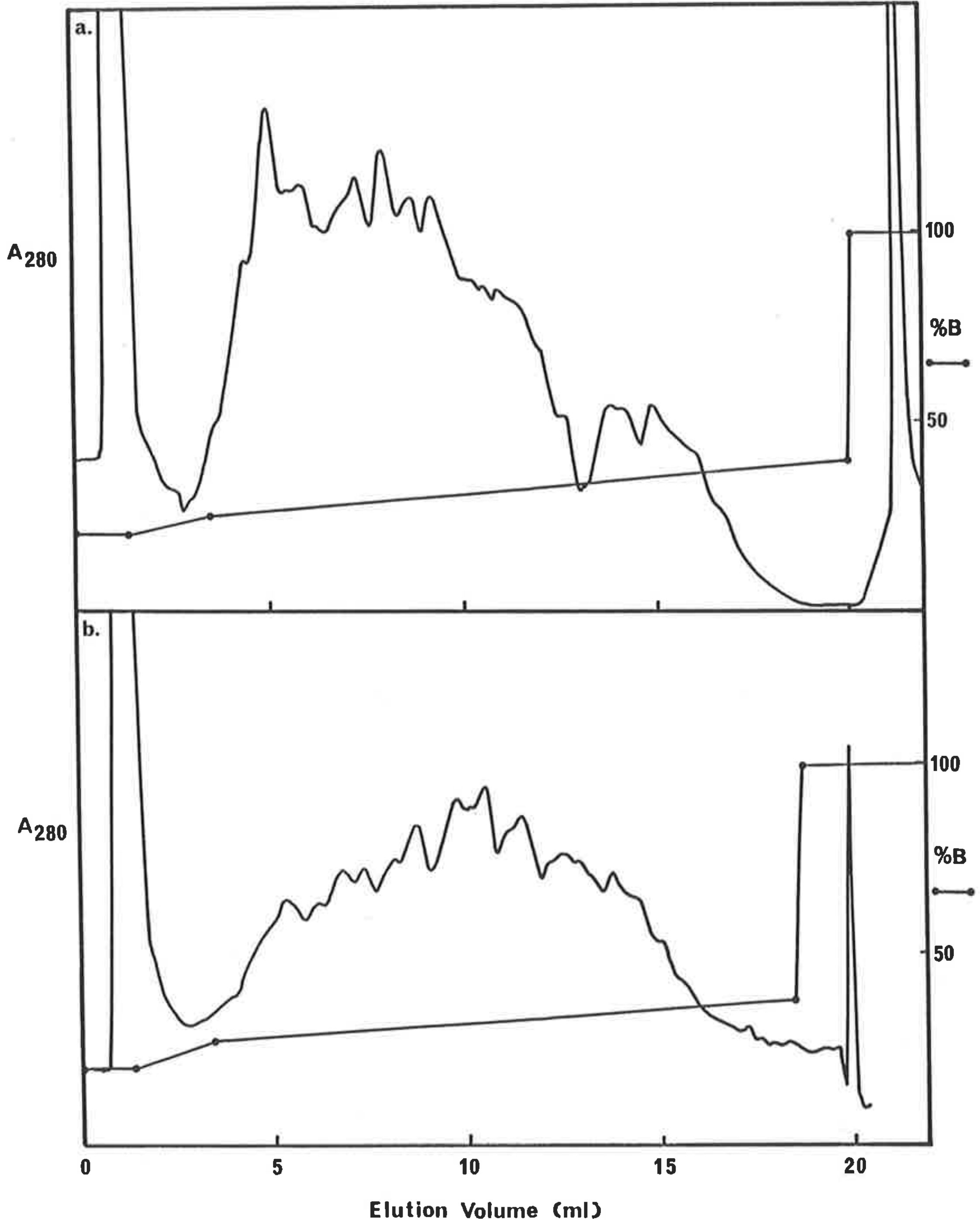
FIGURE 6.15

SEPARATION OF FAST PROTEINS BY FPLC CATION EXCHANGE CHROMATOGRAPHY IN THE PRESENCE OF 8 M UREA.

Samples of crude fast protein, prepared by $MgCl_2$ precipitation from total SCM-feather protein (Powell, 1979) were chromatographed on a Pharmacia Mono S strong cation exchange column, exactly as described in 6.2.6. Buffer A was 50 mM sodium acetate pH 4.8, 8 M urea while buffer B was 50 mM sodium acetate pH 4.8, 1 M NaCl, 8 M urea. All other conditions are as described in the legend to Figure 6.14.

a. 400 μg of fast protein loaded.

b. 3 mg of fast protein loaded.



a.
A280
100
%B
50
b.
A280
100
%B
50
0
5
10
15
20
Elution Volume (ml)

FIGURE 6.16

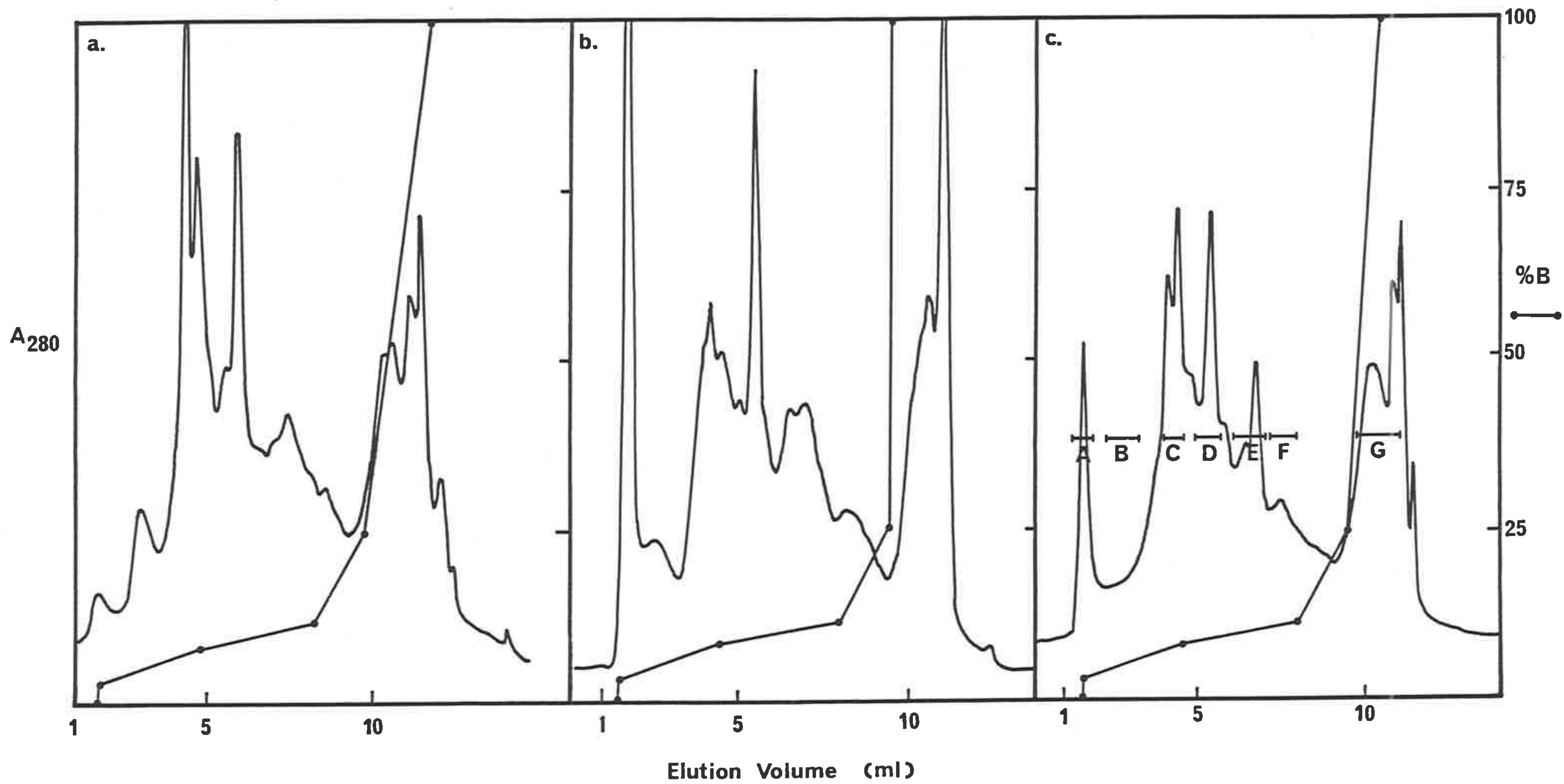
PREPARATION OF FAST PROTEINS BY FPLC ANION EXCHANGE CHROMATOGRAPHY IN 8 M UREA.

Samples of crude fast protein were chromatographed on a Pharmacia Mono Q strong anion exchange column, exactly as described in the legend to Figure 6.14, except that 8 M urea was included in all separations. The fractions, A to G, indicated by bars in c. were collected for further analysis.

a. 400 μg of fast protein.

b. 8 mg of fast protein.

c. 10 mg of fast protein (boiled for 2 min prior to loading).


a.
b.
c.
A280
100
75
%B
50
25
A
B
C
D
E
F
G
1
5
10
1
5
10
1
5
10
Elution Volume (ml)

using the same buffers and salt gradient (Figure 6.16c). This result suggested that when the loaded protein concentration was high the amount of aggregation of the components of this crude fast protein fraction was increased and, since the peak (fraction A in Figure 6.16c), did not completely disappear, this aggregation was not fully disrupted even after boiling in 8 M urea.

The fractions A to G, indicated by bars in Figure 6.16c, were collected and analysed for protein content by electrophoresis on an SDS polyacrylamide gel in the presence of several protein standards (Figure 6.17). Fractions A to G all contained significant amounts of fast protein. This was not expected since the salt needed to elute these fractions varied from 0.0 to 1.0 M NaCl, yet proteins other than fast protein, e.g. the protein in fraction D which co-migrated with scale-like β-keratins, were well resolved under these conditions, even when a very shallow gradient was used.

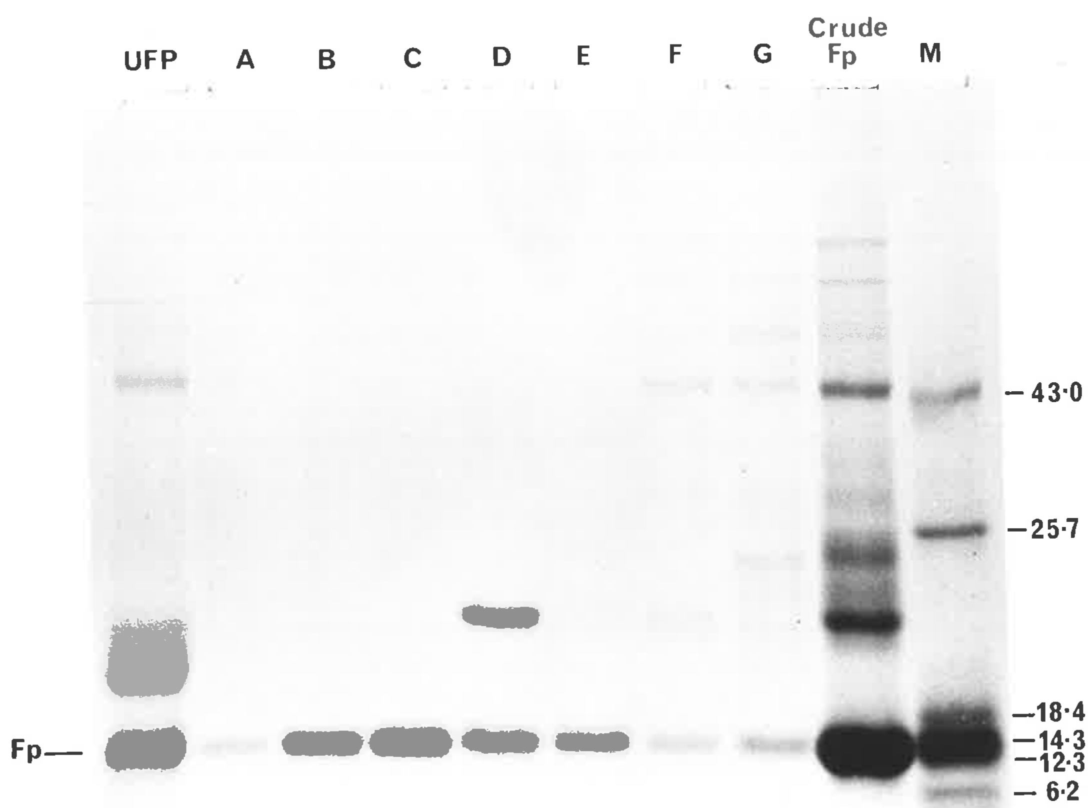
Many of the poorly resolved peaks shown in Figure 6.16 may have been fast proteins aggregated to greater or lesser extents, with the least aggregated protein fraction able to strongly interact with the column, consequently requiring a high salt concentration to elute it from the column, while the most aggregated fraction (fraction A) could not bind at all to the column, presumably because most of the sites on the protein were involved in protein-protein interactions and were not available for protein-column interactions.

Despite these observations, the aim of purifying fast proteins away from the contaminating proteins in the crude fast protein fraction appeared to have been achieved, since fractions A, B, C and probably E contained only proteins which co-migrated with fast proteins. However, the separation of individual protein components of the fast protein family would require the application of other chromatography techniques and preferably those which did not involve ionic interactions, since anion and cation exchange chromatography had already been extensively used. It was not possible to use techniques which involved separation

FIGURE 6.17

SDS GEL ELECTROPHORESIS OF MONO Q FAST PROTEIN FRACTIONS.

The fractions, A to G, indicated by bars in Figure 6.16c, were collected, dialysed to remove urea and acetone precipitated. Aliquots were then electrophoresed on a 10% polyacrylamide SDS slab gel as described in 2.2.8. Samples of crude fast protein (track, "crude Fp"), unfractionated embryonic feather proteins (track, "UFP") and molecular weight standards (BRL low molecular weight range) (track "M") were included as mobility standards. The fractions are indicated by the letters A to G above each track. The position of fast protein has been indicated at the side of the gel and marker protein molecular weights are shown in kilodaltons.


UFP
A
B
C
D
E
F
G
Crude Fp
M
Fp
43·0
25·7
18·4
14·3
12·3
6·2

on protein size, since fast proteins migrated as a single band on SDS gels and were probably all of uniform size. Indeed, Figure 6.17 showed that fast proteins migrated with marker proteins of 12,300 and 14,300 Mr which is in disagreement with the earlier estimate of 8,400 Mr (Walker and Rogers, 1976a), but close to the molecular weight of the protein of 120 amino acids encoded by the pCFK22 gene (see 6.3.1c and Figure 6.9).

6.3.2b PURIFICATION OF FAST PROTEINS BY REVERSE PHASE HPLC.

Since it was not possible to achieve any further separation of the fast protein components by using the properties of protein size or ionic interaction, it was decided to try to separate them by taking advantage of their high content of hydrophobic amino acids (about 30%) which should cause them to bind quite strongly to the hydrophobic reverse phase HPLC columns such as the C-18, diphenyl and cyano columns.

Since solvents containing 8 M urea cannot be used with HPLC column materials, problems were initially experienced when trying to select a solvent system in which these proteins would be soluble. Invariably, buffered solvent systems would not solubilize fast proteins unless 0.1 to 0.2% SDS was added. A great deal of work was undertaken using a 10% acetonitrile, 0.2% SDS solvent buffered with sodium acetate at pH 4.5, with both C-18 and diphenyl columns. Although some encouraging results were obtained, generally problems were encountered with artefactual peaks casued by the solvent system, presumably due to elution of the SDS which was interacting with the column.

The solvent system described above was therefore abandoned in favour of a system recently described by Mahoney and Hermodson (1980), using 0.1% unbuffered trifluoroacetic acid (TFA) as a peptide solvent, which was reported to be an excellent denaturing solvent for large peptides or small proteins such as fast

protein, and had the added advantages of being sufficiently transparent to UV light to allow very sensitive, non-destructive detection of proteins or peptides at low wavelengths (210 to 230nm) and was sufficiently volatile to enable sample recovery by lyophilization.

This solvent readily dissolved both crude fast protein and the fractions purified by Mono Q FPLC (Figure 6.16c). One of these fractions, fraction C, was applied to a high performance C-18 column (Varian 3 μ SP-C18) and eluted using a flow rate of 0.8 ml/min, with an acetonitrile gradient which had been previously optimized (Figure 6.18a). Although fraction C appeared, from SDS gels, to contain only fast proteins, two peaks were observed suggesting that individual components of fast protein were being resolved. The peak which eluted first from the column was not completely symmetrical indicating that these peaks may have contained several protein components.

Attempts to resolve these components by using a shallower elution gradient merely tended to broaden these peaks without increasing resolution (data not shown). However, when the column flow rate was decreased from 0.8 ml/min to 0.5 and 0.2 ml/min (Figure 6.18b and c respectively) the resolution of these components was increased, resulting in the detection of 3 components in the first eluting peak and 2 in the second eluting peak, but these components were not sufficiently resolved to allow the collection of pure samples. The increased resolution observed with decreased flow rates indicated that the protein was physically too large to easily enter the column matrix, and decreasing the flow rate increased the time that the proteins were on the column and therefore allowed greater diffusion into and out of the column matrix, i.e. the pore size of the column (60 Å) was too small to give high resolution for these proteins.

In order to overcome the problem encountered with the C-18 column the fast protein fraction C was applied to a diphenyl reverse phase hydrophobic HPLC column which had a substantially larger pore size (300 Å) and which consequently resulted

FIGURE 6.18

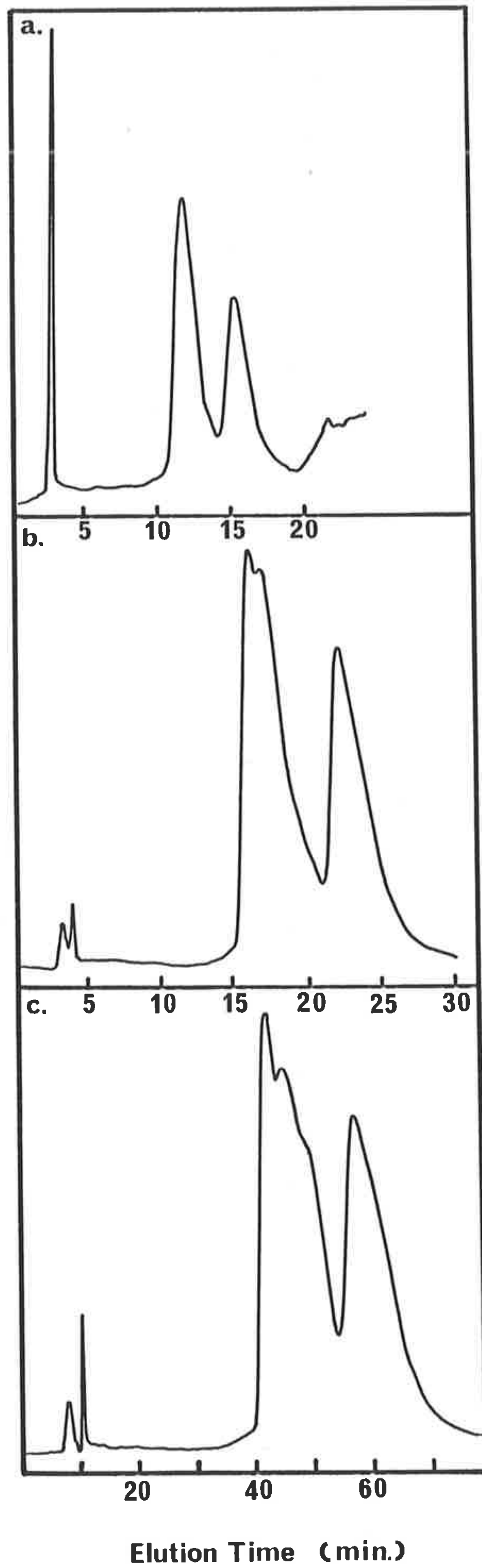
FRACTIONATION OF PURE FAST PROTEINS BY HPLC C-18 REVERSE PHASE CHROMATOGRAPHY.

Aliquots of purified fast protein, fraction C (Figure 6.16c and Figure 6.17, track C) were chromatographed on a Varian SP-C18 reverse phase HPLC column, as described in 6.2.1, under conditions of varying flow rate. The Mono Q fraction was dialysed against chromatographically pure water to remove urea and buffer and lyophilized prior to being dissolved in column buffer. The column was run in 0.1% trifluoroacetic acid and the sample was eluted using a linear gradient of 35% to 55% acetonitrile.

a. 0.8 ml/min flow rate.

b. 0.5 ml/min flow rate.

c. 0.2 ml/min flow rate.


a.
b.
c.
A222
5
10
15
20
5
10
15
20
25
30
20
40
60
Elution Time (min.)

in better resolution of the components of this fraction (Figure 6.19). Although the chromatography of fraction C on the C-18 column resulted in the apparent detection of 5 protein components, only 4 peaks were resolved on the diphenyl column, presumably one of these 4 peaks contained two or more components. The 4 peaks obtained from the diphenyl column were collected, as shown by the bars in Figure 6.19 and designated as peaks I, II, III and IV. These peaks were separately applied to the diphenyl column and eluted using the acetonitrile gradients shown in Figure 6.20 in order to purify these protein components to homogeneity.

Peak I (Figure 6.20a) contained a small amount of a contaminant component present as an early eluting shoulder but was otherwise symmetrical with very little contamination by Peak II. Peaks II, III and IV (Figure 6.20b, c and d respectively) were not completely symmetrical, indicating the presence of several components, some of which were due to contamination by neighbouring peaks. Despite the apparent impurity of some of these peaks, fractions were collected, as indicated by the bars in Figure 6.20, by taking only the centre of each peak in order to avoid contamination of the major peak component by less abundant components which appeared as shoulders and inflexions on these peaks. Since these fractions (designated peak I, II, III and IV) were originally derived from a fraction (fraction C, Figure 6.16c), which appeared to be pure fast protein, it was assumed that they represented pure fast protein, containing one or only a very few fast protein molecular species, and as such they were suitable for protein sequencing.

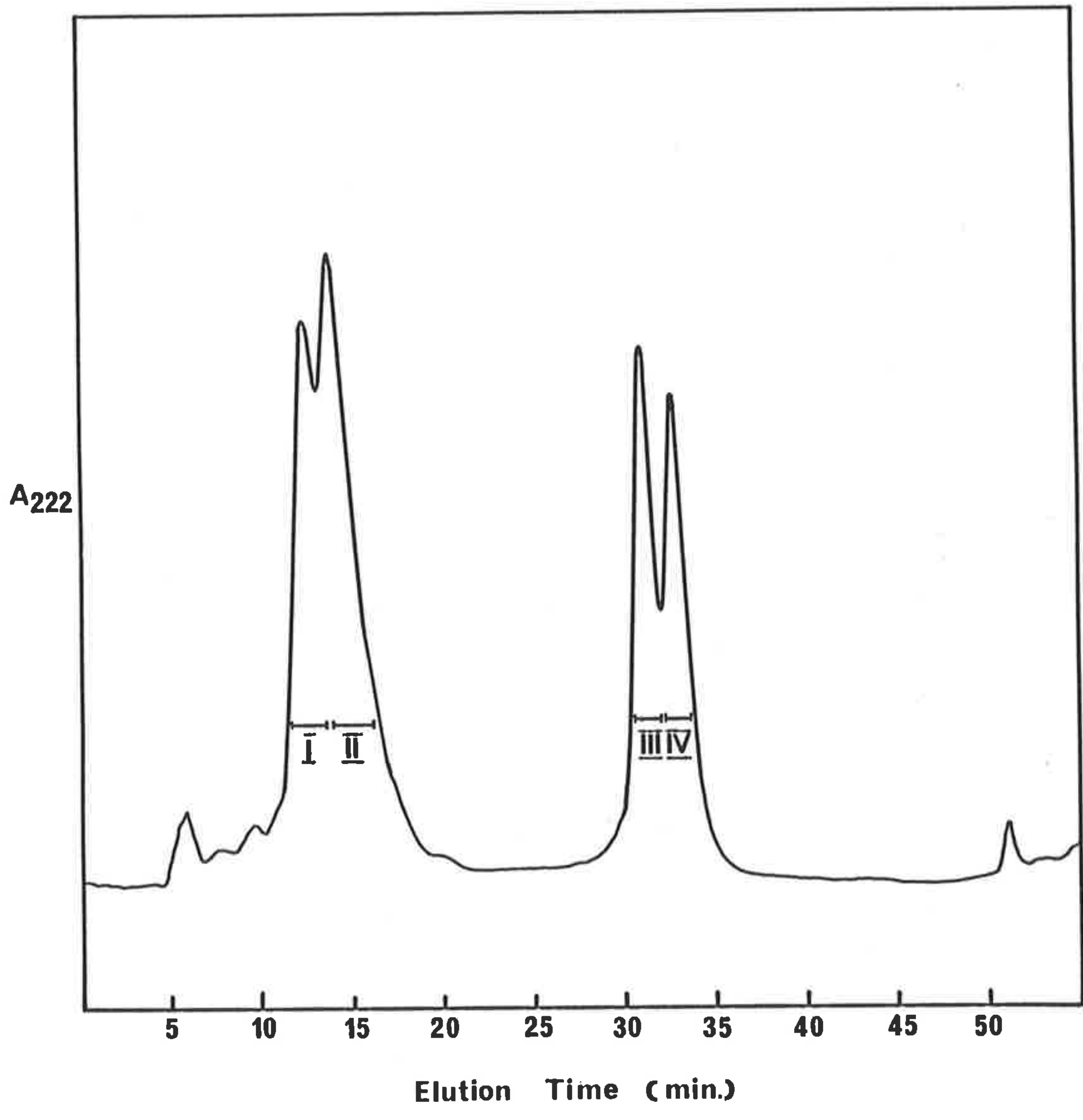
6.3.2c PARTIAL AMINO ACID SEQUENCING OF FAST PROTEIN COMPONENTS.

The advent of newly developed automated gas-liquid solid phase protein sequencing techniques (Hewick *et al.*, 1981) has enabled the rapid determination of the amino-terminal amino acid sequences of many peptides and proteins. The advantage of this procedure over other protein sequencing methods, either manual

FIGURE 6.19

FRACTIONATION OF PURE FAST PROTEINS BY HPLC DIPHENYL REVERSE PHASE CHROMATOGRAPHY.

Purified fast protein, fraction C (Figure 6.16c and Figure 6.17, track C) was chromatographed on a Whatman Protesil-300 diphenyl reverse phase HPLC column, as described in 6.2.8. The buffer system used was the trifluoroacetic acid/acetonitrile system described in the legend to Figure 6.18. A linear gradient varying from 25% to 40% acetonitrile was used to elute the proteins. The column was run at 27°C and 0.8 ml/min flow rate and the proteins were detected at 222 nm. The fractions, I to IV, as indicated by the bars were collected for further analysis.



A222
I
II
III
IV
5
10
15
20
25
30
35
40
45
50
Elution Time (min.)

FIGURE 6.20

PURIFICATION OF FAST PROTEIN COMPONENTS BY HPLC DIPHENYL REVERSE PHASE CHROMATOGRAPHY.

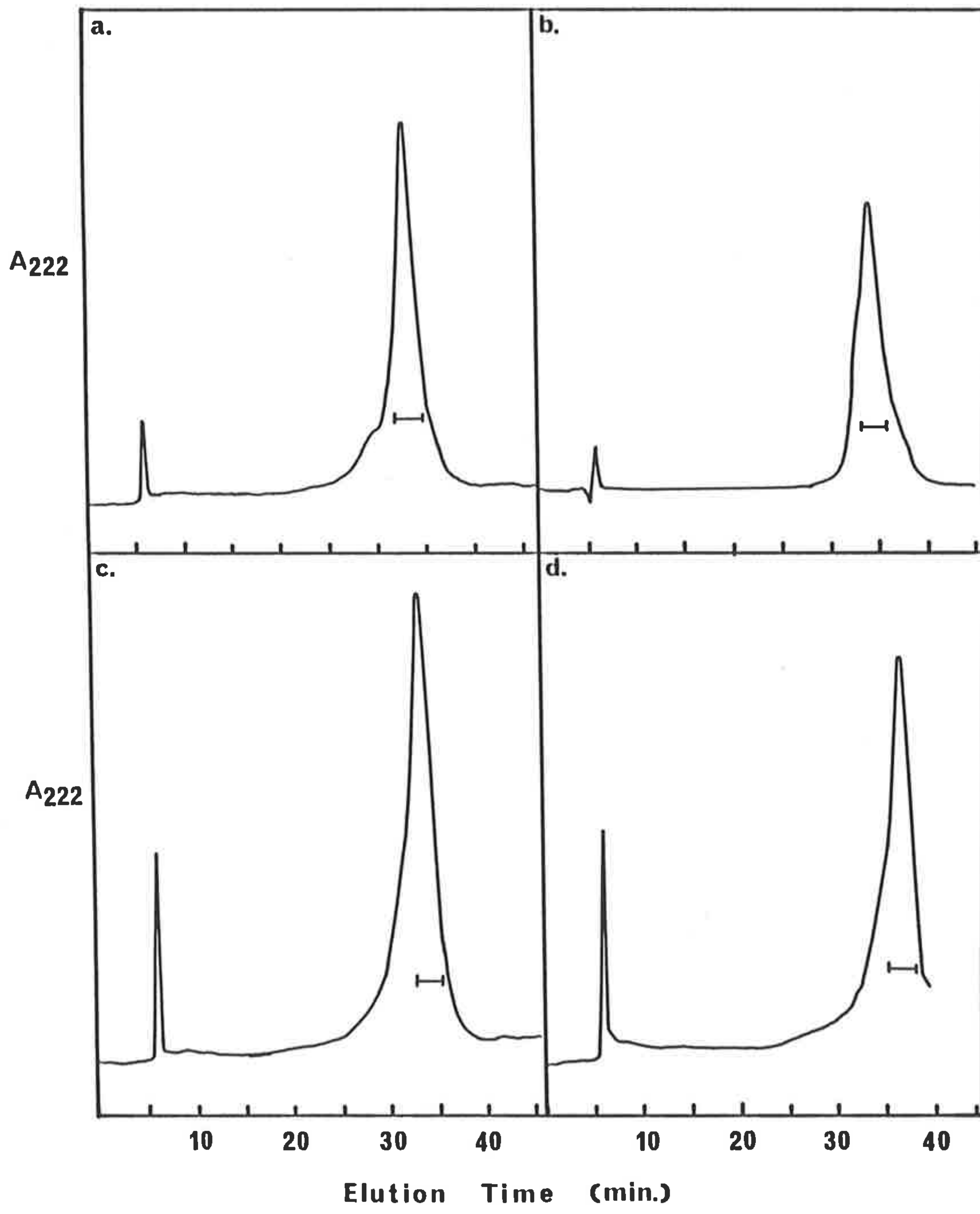
The peaks, I to IV, as indicated by the bars in Figure 6.19, were collected, lyophilized and reapplied to the diphenyl column under the same conditions as those described in the legend to Figure 6.19, except that the elution gradients were varied slightly in order to reduce running times. The bars indicate the portion of each peak collected for further analysis.

a. Peak I.

b. Peak II.

c. Peak III.

d. Peak IV.



a.
b.
c.
d.
A222
A222
10
20
30
40
10
20
30
40
Elution Time (min.)

or automated, is the extreme sensitivity which can be achieved (down to 10 pmoles of protein or peptide).

The amino-terminal amino acid sequences of peaks I, II and III were all determined as described in 6.2.9 by application of these protein samples to a filter mounted in such an automated gas-liquid solid phase protein sequencer (recently acquired from Applied Biosystems Pty. Ltd.). The amino acid sequences so obtained are shown in Figure 6.21 compared with the amino acid sequence derived from the conceptual translation of the pCFK22 gene. It is obvious that the fast protein components, peak I and II possessed a high degree of homology with one another and with the pCFK22 protein.

Twenty amino acids of sequence were determined from the amino-terminus of peak I which appeared to be a single fast protein component very similar if not identical to the pCFK22 protein, since all twenty amino acids exactly matched those of the pCFK22 protein, except at position 16 where a cysteine appeared in the pCFK22 protein sequence but no amino acid appeared to be released during the automated cleavage of the peak I protein. It is possible that this amino acid was a cysteine but no amino acid was released because this cysteine was linked via a disulphide bridge to the cysteine at the carboxyl-terminus of the pCFK22 protein, either of the same protein chain or another protein chain bound to the filter. This would have resulted in the retention of this cysteine residue even though the peptide bond was cleaved. If this explanation is correct, it is more likely that the cysteine was linked to another cysteine via an intra-protein rather than an inter-protein linkage because the fraction (fraction C, Figure 6.16) from which this protein was prepared migrated on SDS gels (Figure 6.17) at approximately the correct molecular weight (14,000 Mr), with no sign of dimer or trimer molecules which would have been present if the protein chains were covalently linked.

Eighteen amino acids were determined from the amino-terminus of peak II with two amino acid residues being released in approximately equal quantities at positions

FIGURE 6.21

THE AMINO-TERMINAL AMINO ACID SEQUENCES OF SEVERAL PURIFIED FAST PROTEIN COMPONENTS.

The amino-terminal amino acid sequences of Peaks I, II and III, purified as shown in Figure 6.20, were determined using an automated gas-liquid solid phase protein sequencer, as described in 6.2.9. The amino acid sequences of the three components are shown, in one letter code, compared to the amino acid sequence derived from the conceptual translation of the pCFK22 gene.

Amino Acid Sequences of C-18 Fractions

pCFK22 gene	MTFHREFYNDEHYSPFCQEDLHGLW
Peak I	TFHREFYNDEHYSPF?QEDL
Peak II	S (above) / T (below) ; Q (above) / W (below) /FHREFYNDEHYSPF?/E
Peak III	SFLHDD?Y

1 and 17, indicating that this peak consisted of at least two molecular species, possibly due to contamination by peak I. The amino acid sequence again was very similar to that of the pCFK22 protein with the same missing amino acid at position 16 which was assumed to be a cysteine. Figure 6.22 shows the derivatized amino acids released during the first seven cycles of the automated sequencing of peak II, as analysed by HPLC C-18 reverse phase chromatography, in order to illustrate the type of data produced by this sequencing procedure and to demonstrate the presence of two amino acids (serine and threonine) at the amino-terminus of the protein.

Peak III sequenced poorly in that the repetitive yield from one amino acid cleavage cycle to the next was very low which resulted in a rapid drop in amino acid "signal", allowing only the first 6 residues to be determined. This rapid loss of signal is recognized in our laboratory to be associated with the sequencing of short peptides, i.e. less than 10 to 12 amino acids (M. Snoswell, personal communication). The sequence obtained contained the correct type of amino acids to be derived from a fast protein but there was no real homology match between this short sequence and the amino-terminus or any region of the pCFK22 protein sequence. The peak III sequence may have been derived from a member of the fast protein family which differed markedly from the other fast proteins which had been sequenced.

6.4. DISCUSSION.

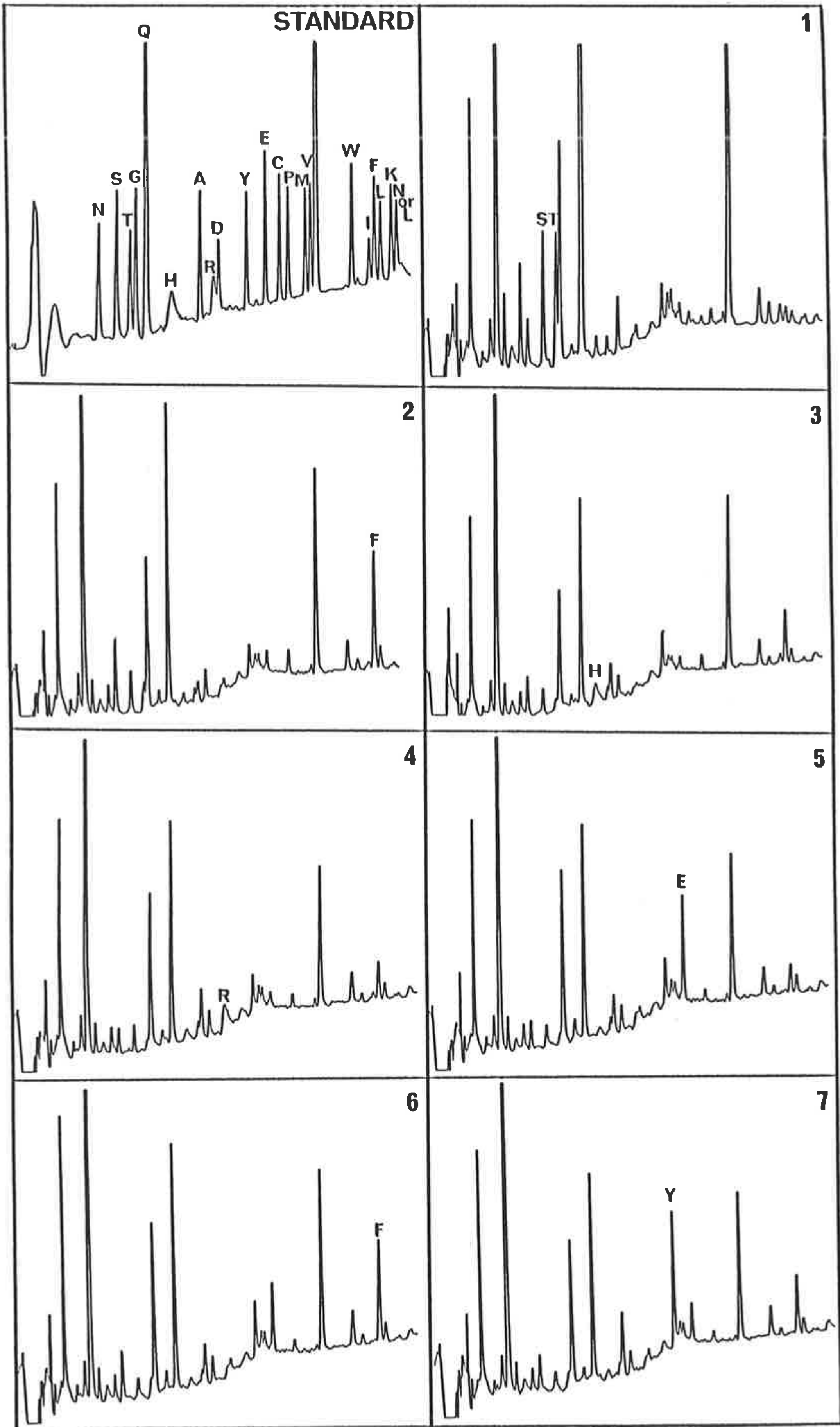
6.4.1 ISOLATION OF SEQUENCES CODING FOR FAST PROTEINS, A FAMILY OF SEQUENCE RELATED GENES.

The isolation and DNA sequencing of the gene from which the cDNA clone pCFK22 was derived demonstrated that this gene coded for a protein of 120 amino acids including the initiating methionine (see Figure 6.8). The amino acid composition of this protein closely matched that of fast protein isolated from

FIGURE 6.22

IDENTIFICATION OF AMINO ACID RESIDUES RELEASED DURING PROTEIN SEQUENCING.

Purified fast protein, Peak II (Figure 6.20), was sequenced using an automated gas-liquid solid phase protein sequencer which released phenylthiohydantoin (PTH) amino acid derivatives which were collected and identified by HPLC C18 reverse phase chromatography. The complex elution system used acetate, trifluoroacetic acid and acetonitrile (M. Snoswell, manuscript in preparation). The C18 HPLC chromatograms for the identification of the derivatized amino acids released in the first seven cleavage cycles are shown (1 to 7, cycles one to seven respectively) along with the fractionation of derivatized amino acid standards ("standard"). The released amino acid identified in each cycle is indicated by the one letter amino acid code above the appropriate peak. Each amino acid peak is similarly identified in the fractionation of derivatized amino acid standards.



STANDARD
Q
S G
N
T
H
A
R
D
Y
E
C
P M
V
W
F
I
L
K
N
or
L
1
S T
2
F
3
H
A222
4
R
5
E
6
F
7
Y

embryonic feathers (Table 6.1) but this protein was calculated to have a molecular weight of 14,070 Mr whereas fast proteins were estimated to have a molecular weight of 8,400 Mr (Walker and Rogers, 1976a).

The protein encoded by the pCFK22 gene was however demonstrated to be fast protein since the amino-terminal 20 amino acids of this protein exactly matched (except for a cysteine residue at position 16) the first 20 amino acids determined for a purified fast protein molecular species (Peak I, Figure 6.20a). Also, analysis of embryonic feather proteins by SDS polyacrylamide gel electrophoresis (Figure 6.17), demonstrated that fast proteins ran as a single band of betweem 12,000 and 14,000 Mr, which is much closer to the calculated weight of the protein encoded by the pCFK22 gene than the original estimate of Walker and Rogers (1976a).

It is clear from these data that the cDNA clones pCFK22, 10 and probably 3 all code for embryonic feather fast proteins. It was also clear from the hybridization of a fast protein coding region probe to the restriction fragments from the λ-Charon 4A clone λCFP1 (Figures 6.3b and 6.4) that sequences flanking the sequenced fast protein gene (previously called the pCFK22 gene) are capable of annealing to the probe, although not as well as the sequenced fast protein gene. Also, protein sequencing data (Figure 6.21) demonstrated the existence of at least two sequence related fast protein species.

These observations suggest that fast proteins exist as a family of sequence related genes found clustered in the genome. This idea was first advanced by Walker and Rogers (1976a) who showed that fast proteins ran as a single band on all polyacrylamide gel systems tested except when electrophoresed on pH 9.5 gels where they separated into about five poorly resolved bands. This observation led them to suggest that fast proteins were a family of at least five sequence related proteins. Fractionation of proteins at pH 9.5 by electrophoresis is analogous to anion exchange chromatography in that in both systems the mobility of the proteins or the time of elution is dependant on the amount of negative charge carried by the protein, highly

charged proteins elute last on an anion exchange column and migrate fastest on a polyacrylamide gel, while the converse is true of poorly charged proteins. Therefore the electrophoretic fractionation of fast proteins at pH 9.5, reported by Walker and Rogers (1976a), can be compared to the fractionation of fast proteins by FPLC anion exchange chromatography using the Mono Q column (Figure 6.16), where the large number of fast protein peaks observed were shown to be due to aggregation. It is probable that the five or so fast protein bands seen on pH 9.5 gels by Walker and Rogers (1976a) were due to different aggregation states of these proteins and not separate fast protein molecular species. Therefore, the results presented in this and the previous chapter represent the first unequivocal demonstration that fast proteins exist as a family of sequence related genes.

It should be pointed out that, apart from a preliminary characterization of fast proteins (Walker and Rogers, 1976a) and the localization of small amounts of mRNA coding predominantly for fast proteins in fractionated total feather mRNA (Powell and Rogers, 1979), the results presented in this and the previous chapter represent the first extensive study of fast proteins, culminating in DNA sequences from fast protein cDNA clones, the complete nucleotide sequence of a fast protein gene and its flanking regions, the purification of several fast protein components and the direct determination of partial amino acid sequences of several fast protein species.

Although Walker and Rogers (1976a) provided the groundwork for this study in that they were the first to recognize the presence of fast proteins in embryonic feathers, it could be noted that their results on the molecular weight of fast proteins were misleading and complicated the indentification of DNA sequences which coded for fast proteins.

6.4.2 THE NATURE OF FAST PROTEINS.

The most outstanding feature of the amino acid sequence derived from the sequenced fast protein gene is the high content of histidine residues (about 12%)

which is unusually high compared to most proteins. An analysis of the primary structure of fast protein does not allow protein structure predictions to be made with great confidence, however some features of the amino acid sequence are quite interesting.

The sequence can be subdivided into three sections, i.e. residues 1 to 48, 49 to 98 and 99 to 119. The first section contains a high proportion of charged amino acids and most of the histidines present in the protein (8 acidic, 4 basic and 10 histidine residues). These charged residues appear to be clustered within this first section to such an extent that in one place there are seven consecutive residues and in another five out of six consecutive residues are charged. The second segment contains few charged residues, 2 acidic, 3 basic and 2 histidines, but is rich in glycines (10) and uncharged polar residues, (18). This segment probably contains alternating β and β-turn structures with some random coil. The terminal segment is basic (2 acidic, 4 basic and 2 histidines) and probably contains β-turn, random coil and some β structure, as would the N-terminal segment. Overall, the protein probably contains less the 10% α-helix and more than 50% β-structure and is likely to be roughly spherical or elipsoidal.

6.4.3 THE BIOLOGICAL ROLE OF FAST PROTEINS.

The biological role of fast proteins is unclear, although it is probable that they are structural proteins. However, it is possible that fast protein, like haemoglobin, which also contains a high proportion of histidine (the only amino acid capable of buffering in the physiological pH range), functions as an intracellular buffer in the cells of the growing embryonic feather. A more plausible suggestion is that fast proteins act as an interfilamentous matrix protein or "glue" which binds the arrays of keratin filaments into a solid inflexible structure. Such proteins exist in wool, i.e. high-sulphur proteins (Crewther, 1976; Powell *et al.*, 1983) and high glycine/tyrosine proteins (Dopheide, 1973), which together make up the wool interfilamentous matrix proteins. Fast proteins show some similarity to the high glycine/tyrosine proteins

of wool in that they have high levels of glycine (12%) and tyrosine (8%) but this is less than the amounts of glycine and tyrosine found in the wool proteins which do not contain histidine.

Histidine rich proteins have been found in mammalian epidermal tissues (for a review see Bernstein, 1983) and one of these, filaggrin (Dale, 1977; Steinert, 1980; Steinert, 1984),which is rich in histidine, arginine and glutamic acid or glutamine is thought to be an interfilamentous matrix protein. Different filaggrins have been isolated from rat and mouse epidermis and have molecular weights of 48,000 Mr (Dale, 1977) and 30,000 Mr (Balmain *et al.*, 1979; Steinert *et al.*, 1981a) respectively, which are both larger than the molecular weight of fast proteins (14,000 Mr).

One difference between fast protein and filaggrin is that they probably have radically different pIs. Fast proteins are essentially neutral molecules with approximately equal amounts of acidic and basic amino acids (10.1% glutamic and aspartic acid, 9.2% arginine and lysine) but with a large amount of histidine (11.8%) which has a pI of about 6.0 and should not affect the pI of the protein significantly. Column chromatography studies suggest that the pI of fast proteins is below about 7.0 since very little fast protein binds to a cation exchange resin (Mono S) at any pH above 6.0. Filaggrins, on the other hand, are reported to have pIs of greater than 10.0, despite being rich in either glutamic acid or glutamine.

6.4.4 CONTROL OF FAST PROTEIN GENE EXPRESSION.

Fast proteins and embryonic feather keratins are separate families of sequence related proteins which are expressed at the same time during embryological development, i.e. keratins and fast proteins form a super-family of co-ordinately expressed embryonic feather proteins. Fast proteins and keratins are expressed at the same time during the development of the embryonic feather but is their expression controlled by the same factor(s)?

The control of keratin gene expression is not yet well understood but it can be seen from the sequences of these genes that they follow a similar pattern at their 5′-ends, the region likely to be involved in the control of gene expression. The 5′-untranslated regions of these genes is well conserved in that it is interrupted by the only intron found in these genes, thus forming a 5′-leader sequence of 37 bp (from the cap site to the intron), which is almost totally conserved between the genes, and a 21 bp sequence from the intron to the initiation codon which is conserved but not as strongly as the 5′-leader sequence. This observation led to the hypothesis (presented in Chapter 4, 4.4.5) that alternative 5′-leader sequences are recognized by stage or tissue specific effector molecules which potentiate the initiation of transcription of the correct set of proteins for that particular stage or tissue. Unfortunately, no alternative 5′-leader sequences or promotors (TATA box) have been found in the embryonic feather keratin sequences.

The fast proteins are evolutionarily unrelated to the keratins in that no amino acid sequence homology exists between them, in fact they have totally different amino acid compositions and polypeptide chain lengths, yet they both possess the same structural features at the 5′-ends of their genes. Remarkably, for evolutionarily unrelated genes, fast proteins also demonstrate a strong degree of sequence homology with the keratin genes at their 5′-ends. These observations suggest that the conserved sequences are involved in the co-ordinate expression of these genes and the presence of a potential alternative TATA box, cap site and 5′-splice site add weight to the hypothesis that alternative 5′-leader sequences are used in different stages or tissues for the expression of the super-family of keratins and fast proteins.

CHAPTER 7.

CONCLUDING DISCUSSION.

7.1 FEATURES OF THE EMBRYONIC FEATHER SYSTEM.

The work reported in this thesis was aimed toward the eventual understanding of the molecular processes and underlying mechanisms involved during the terminal differentiation of the embryonic chick feather. This study was foreshadowed by an extensive examination of keratin protein and mRNA synthesis and a preliminary study of fast protein synthesis in this tissue (as previously discussed in Chapter 1). However, at its commencement, nothing was known, at the molecular level, of the control of embryonic feather gene expression and none of the mRNAs, genes or proteins expressed in the embryonic feather had been fully characterized.

The results presented in this thesis have significantly added to the understanding of embryonic feather gene expression and will unquestionably facilitate the continuation of studies in this direction. The following is a resume of these results discussed in relation to an understanding of gene expression in embryonic chick feathers.

7.1.1 PROTEINS EXPRESSED DURING DIFFERENTIATION.

Since embryonic feathers contain a complex group of proteins, the individual components of which are difficult to unequivocally separate and identify by physico-chemical means, the approach employed to characterize these proteins was via double-stranded cDNA clones made from feather mRNA. This approach resulted in the purification of individual species from the complex mixture and DNA sequencing revealed the primary structure of these mRNAs, thereby identifying the encoded proteins.

As expected a large number of cDNA clones were found to code for typical feather keratins. The result obtained for these clones confirmed earlier observations of various workers (see Chapter 1, 1.3.1) who studied keratin proteins and found them to be a heterogeneous family of proteins. Although they possessed minor differences in amino acid sequence the keratins were basically very similar and could

not be further divided into sub-classes. Indeed, embryonic chick feather keratin protein sequences showed an amazing similarity to adult feather keratins from gull and emu, although they did differ slightly in length.

The observation that embryonic feather keratins form a single class of proteins is consistent with the results of Kemp and Rogers (1972) who found that the proteins of embryonic feathers were similar to those of adult feather barbs but different to those of adult feather rachis and calamus, i.e. adult feathers contain several classes of proteins (barb, rachis and calamus) but embryonic feathers have only one protein class as they are composed only of barb cells.

The nucleotide sequences of the keratin cDNA clones revealed that, like the proteins, keratin mRNAs form a family of structurally similar, sequence-related molecules. All keratin mRNA species were of about the same length (790 bp) and contained a coding region of 297 bp, a 3′-untranslated region of about 435 bp and a 5′-untranslated region of 58 bp. These mRNA sequences resolved the question of keratin mRNA structure and consequently confirmed many of the mRNA structure postulates made by Kemp (1975) and Lockett *et al.* (1979). An estimate of the exact number of keratin species was not possible although of the eight separate keratins sequenced to date (3 cDNA clones and 5 genomic clones), all were different, indicating that the estimate of 19 to 25 different polypeptide chains (Walker and Rogers, 1976a) is correct, or at least a lower limit of protein complexity. These results also indicated that the protein heterogeneity does not arise via post-translational modification but from the presence of multiple homologous keratin genes.

The biological significance of this large number of similar keratin genes is unclear, but it is probably unrelated to the structural requirements of the proteins since their amino acid sequences are nearly identical and is consistent with them having the same secondary structure. The requirement to rapidly produce large amounts of protein, although facilitated by having many genes, can equally be

fulfilled by a single gene or several genes, e.g. globin production in reticulocytes. It is possible that the selection pressures which resulted in a large family of keratin genes were acting at the level of the gene and not the protein, i.e. the large family of clustered genes may be required for some aspect of the control of gene expression or possibly the stabilization of a particular keratin type (feather as opposed to scale or claw) by exchange of genetic material within the gene family which could provide a pool of genetic information capable of "editing" out any gene changes, particularly variations in gene length. With this in mind, it is interesting to note that the feather keratins probably evolved from scale keratins via a deletion event and having evolved, the keratin gene family has exhibited a remarkable conservation of gene and protein length.

The only other proteins known to be expressed in embryonic feather are the fast proteins and until this study very little was known about them. Because of the poor understanding of fast proteins it was not possible to positively identify cDNA clones which coded for them until the gene for a fast protein was isolated, sequenced and compared to the N-terminal amino acid sequences of several fast protein components which were purified using powerful HPLC and FPLC protein separation techniques. These results revealed that fast proteins, like keratins, form a family (of unknown size) of sequence-related genes.

It was found, contrary to the results of Walker and Rogers (1976a) who estimated that fast proteins were about 80 amino acids long, that the true length of these proteins was 119 amino acids. It was not possible to estimate the number of fast protein components found in embryonic feathers although HPLC separations revealed four peaks each consisting of at least one and probably two or three components as indicated by amino acid sequencing data. DNA sequencing and hybridization data also indicated that several fast protein genes existed and were probably clustered in the genome. Further, screening of the feather cDNA library revealed that the proportion of fast protein mRNAs in feather mRNA was greater

than the estimate of 10% by Walker and Rogers (1976a) and Powell (1979). In support of this finding, it is interesting that the proportion of fast proteins in embryonic feathers appears to be between 20 and 40% when feather proteins are fractionated on SDS gels (see Figure 3.1).

It was also observed for the first time that typical scale β-keratins are expressed in embryonic feathers. This was an unexpected result since all previous work had indicated that the proteins expressed in different tissues were found exclusively in those tissues (see Kemp, 1972; Ph.D thesis). It was not simply that a scale-like gene was exclusively expressed in feathers and was specific to feathers, since the identical gene was found in a scale cDNA library (Wilton, 1983). This finding raises the question of whether all the β-keratin genes are expressed in epidermal appendages, with variations in protein components being quantitative rather than qualitative.

7.1.2 CONTROL OF GENE EXPRESSION.

It was thought that some understanding of the control of gene expression in embryonic feathers might be facilitated by a comparison of the nucleotide sequences of the various feather mRNAs and their parent genes. The embryonic feather system is ideally suited to this purpose since several co-ordinately expressed families of genes are found in this tissue enabling comparisons within and between these families. The two major families expressed are the keratins and fast proteins, although comparisons with the scale keratin genes might also be useful since at least one scale gene is expressed in embryonic feathers.

Sequence comparisons within the keratin gene family revealed that they all code for similar proteins of 97 amino acids and the functional regions of the mRNA are conserved in length. Some sequence conservation was observed in the 3′-untranslated regions while the coding regions and 5′-untranslated regions were more strongly conserved. Each gene contained one intron located in the

5′-untranslated region 37bp from the transcription initiation point and this intron divided the 5′-untranslated region into two sequences which differed in homology when the appropriate sequences from the keratin genes were compared.

It was found that the 5′-leader sequence (from the "cap" site to the intron) was more strongly conserved than the remainder of the 5′-untranslated region. This observation led to the hypothesis, presented and discussed in Chapter 4 (4.4.5), that the rigidly conserved 5′-leader sequence is the site of interaction of an effector molecule involved in the initiation of the transcription of stage or tissue specific sets of genes in chick epidermal tissues. Since the epidermis is "determined" by the underlying dermis (see 1.2.2) it may be that the transcriptional switch (stage or tissue specific effector molecule) is derived directly from the dermis, or it may be a secondary messenger stimulated by the primary interaction of the dermis.

An alternative hypothesis is that this conserved 5′-leader sequence functions as a translational control sequence and the tissue or stage specific effectors may regulate the amount of any family(s) of proteins to be synthesized in that tissue. This would appear to be an unacceptable hypothesis since the levels of protein synthesis in the feather appear to be paralleled by mRNA levels (Powell *et al.*, 1976).

The possibility of expression of the same protein in different tissues, e.g. scale keratin in feathers, can be accommodated by the postulation of alternative 5′-leader sequences which would bind different effector molecules in different tissues and thus result in the stimulation of transcription of the same mRNA in different tissues, although these mRNAs would contain different 5′-leader sequences. Support for this idea was found when the fast protein gene sequence was examined.

Fast proteins are evolutionarily unrelated to the keratins, yet it was found that the sequenced fast protein gene was remarkably similar to the keratin genes in the

5′-untranslated region, in that the gene's only intron was found to divide the 5′-untranslated region into two sequences of the same length as those found in the keratin genes and contained good homology (about 70%) to the keratin sequences. Further, in support of the hypothesis, an alternative promotor (TATA box) and 5′-leader sequence were found within the intron of the fast protein gene.

7.1.3 EVOLUTION OF FEATHER GENES.

The β-keratins of embryonic feather almost certainly evolved from the β-keratins of scales since the feather is generally believed to have evolved from the reptilian scale (Kerbert, 1876; Parkes, 1966). Further, DNA sequence analysis of embryonic chick scale β-keratins has revealed strong regions of sequence homology to the feather keratin genes (Gregg *et al.*, 1983) and indicated that the feather β-keratins probably arose via the deletion of a region from one of the scale β-keratin genes. The evolution of the feather keratins as a large family of clustered, sequence-related genes presumably occurred by gene duplication and subsequent mutational divergence.

The coding regions of the keratin genes and important sequences in the 5′-untranslated region appear to have been conserved to a greater extent than the 3′-untranslated region of the genes, perhaps suggesting that the amount to which the 3′-untranslated regions have diverged reflects the amount of evolutionary drift to which these genes have been subjected. This idea is supported by sequence data from the 5′-flanking regions of two keratin genes (L. Crocker and K. Gregg, unpublished observations), in an area away from structures likely to be conserved, e.g. TATA boxes, which show similar levels of homology as is observed in the 3′-untranslated regions.

The coding regions of fast proteins are totally different to the coding regions of the keratin genes at the amino acid and nucleotide sequence level, yet the genes show strong sequence and structural homology in their 5′-untranslated regions.

This observation is diffucult to explain. Is it possible that fast proteins and keratins shared an extremely ancient common ancestral gene and these features have been preserved over that long time span, or is this an example of convergent evolution at the level of the DNA sequence? This question remains unanswered, though the former explanation holds more appeal.

7.2 THE EMBRYONIC FEATHER SYSTEM — FUTURE PROSPECTS.

Although the results presented in this thesis have gone a long way toward the characterization of fast proteins and their genes there are several areas where this work could be extended. The obvious place to start would be to continue the protein and DNA sequencing work in order to further characterize the fast protein family. The protein work could take the form of further HPLC separations of fast protein mixtures followed by protease cleavage and peptide purification in order to provide protein sequence data from the whole protein chain, not just the N-terminus. It would also be a simple matter to sequence the fast protein genes in the genomic clone λCFP1 which flank the fast protein gene already sequenced. Such DNA sequencing work might be extended to cosmid clones in order to establish the number of fast protein genes and examine the possibility that they are linked in the genome to the keratin genes.

An experiment which is immediately approachable to examine the hypothesis that alternative 5′-leader sequences are used when the same gene is expressed in different tissues has already been described in Chapter 5 (5.4.1). Briefly, a primer derived from the 5′-end of the pCFK15 gene (a scale β-keratin gene known to be expressed in both feathers and scales) could be used to sequence the 5′-leaders of scale and feather derived mRNAs. The hypothesis predicts that the sequences prior to the position of the intron would be the same for both scale and feather mRNA, but they should differ 5′ to the position of the intron, i.e. they should have different 5′-leader sequences.

The number of keratin genes expressed in the embryonic feather is at present unclear. This question may be approached by an extensive and time consuming examination of mRNAs expressed in feathers by sequencing multiple cDNA clones from an extensive cDNA library. Alternatively, the number of genes could be defined by an examination of cosmid clones containing feather keratin genes. This approach would have the added advantages of providing the entire gene, including any introns, and would define the extent of keratin gene clustering (such an approach is currently being undertaken in this laboratory).

The most interesting question posed by the embryonic feather system, or any epidermal system, is the elucidation of the mechanisms involved in the switching of groups of genes during differentiation and how this gene control is mediated. In the short term, structural features of importance for efficient gene expression, i.e. DNA sequences important for the initiation of transcription could be examined by *in vitro* mutagenesis of isolated keratin genes followed by transcription *in vitro*, e.g. using HeLa cell extracts, or *in vivo* using *Xenopus* oocytes or cultured epidermal keratinocytes. Preliminary studies of this nature have been started by A. Koltonow, of this laboratory, using micro-injected *Xenopus* oocytes. No expression of keratin genes has yet been observed (A. Koltonow, unpublished observations). This is perhaps not suprising since, if the previously presented hypothesis regarding initiation of stage specific transcription is correct, transcription would not occur, at least efficiently, unless the required tissue or stage specific effector molecules were present. This problem might be overcome if an homologous system was used, e.g. primary cultures from 10-day chick feather keratinocytes.

Studies on the nature of the dermal signal or the proposed stage specific effector molecules could take the form of reconstitution experiments where extracts from the correct stage of dermal tissue are applied to cultured whole epidermis or epidermal cells to look for factors which stimulate the synthesis of keratins. Also, extracts from tissue expressing keratins may be found to stimulate keratin gene expression in

heterologous *in vitro* or *in vivo* transcription systems, thus allowing identification of the components in these tissues responsible for regulation of keratin gene expression.

In summary, the embryonic chick feather system affords an excellent opportunity to examine the control of gene expression in the differentiating epidermis. The system is at a sufficiently advanced stage of characterization that pertinent questions regarding gene control may be approached and answered in the near future. It is believed that the work described in this thesis has significantly advanced the understanding of gene expression in the embryonic chick feather, thus providing an essential background for further progress in this area.

BIBLIOGRAPHY

Adesnik, M. and Darnell, J.E. (1972) *J. Mol. Biol.* **67** : 397 – 406.

Ando, T. (1966) *Biochim. Biophys. Acta.* **114** : 158 – 168.

Arai, K.M., Takahashi, R., Yokote, Y. and Akahane, K. (1983) *Eur. J. Biochem.* **132** : 501 – 507.

Balmain, A., Loehren, D., Alonso, A. and Goerttler, K. (1979) *Devel. Biol.* **73** : 338 – 344.

Barrell, B.G. (1971) *Procedures Nucl. Acid Res.* **2** : 751 – 768.

Bartels, E.A. (1943) *Arch. f. Entwicklmech.* **142** : 763 – 816.

Beckingham-Smith, K. (1973a) *Dev. Biol.* **30** : 249 – 262.

Beckingham-Smith, K. (1973b) *Dev. Biol.* **30** : 263 – 278.

Bell, E. and Thathachari, Y.T. (1963) *J. Cell. Biol.* **16** : 215 – 223.

Benton, W.D. and Davis, R.W. (1976) *Science.* **196** : 180 – 182.

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

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

Birnboim, H.C. and Doly, J. (1979) *Nuc. Acids Res.* **7** : 1513 – 1523.

Birnstein, I.A. (1983) *Biochemistry and Physiology of the skin. vol I. ed. L.A. Goldsmith,* **pp 170 – 183**, Oxford Uni. Press.

Bishop, J.O., Morton, J.G., Rosbash, M. and Richardson, M. (1974) *Nature.* **250** : 199 – 204.

Bollum, F.J. (1968) *In "Methods in Enzymology" Vol XII Part B. Grossman, L. and Moldave, K. (Eds).* **pp. 169 – 173**, Academic Press, New York.

Bonner, W.M. and Laskey, R.A. (1974) *Eur. J. Biochem.* **46.** : 83 – 88.

Both, G.W., Banergee, A.K. and Shatkin, A.J. (1975a) *Proc. Natl. Acad. Sci. U.S.A.* **72** : 1189 – 1193.

Both, G.W., Furuichi, Y., Muthukrishnan, S. and Shatkin, A.J. (1975b) *Cell.* **6** : 185 – 195.

Brawerman, G. (1973) *Methods Cell Biol.* **7** : 1 – 23.

Brawerman, G. (1974) *Ann. Rev. Biochem.* **43** : 621 – 642.

Brawerman, G. (1976) *Prog. Nucl. Acid Res. Mol. Biol.* **16** : 117 – 148.

Brawerman, G., Mendecki, J. and Lee, S.Y. (1972) *Biochemistry.* **11** : 637 – 641.

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

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

Breindl, M. and Gallwitz, D. (1973) *Eur. J. Biochem.* **32** : 381 – 391.

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

Brownlee, G.G. and Cartwright, E.M. (1977) *J. Mol. Biol.* **114** : 93 – 118.

Brush, A.H. (1983) *J. Protein Chem.* **2** : 63 – 75.

Brush, A.H. and Wyld, J.A. (1982) *Comp. Biochem. Physiol.* **73** : 313 – 325.

Brysk, M.M., Gray, R.M. and Bernstein, I.A. (1977) *J. Biol. Chem.* **252** : 2123 – 2127.

Cain, D.F. and Pitney, R.E. (1968) *Anal. Biochem.* **22** : 11 – 20.

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

Chou, P.Y. and Fasman, G.D. (1978) *Advances in Enzymol.* **47** : 45 – 148.

Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) *Cell.* **12** : 1 – 8.

Chrambach, A., Reisfeld, R.A., Wyckoff, M. and Zaccari, J. (1967) *Anal. Biochem.* **20** : 150 – 154.

Cohen, S. (1962) *J. Biol. Chem.* **237** : 1555 – 1562.

Cohen, S. (1965) *Dev. Biol.* **12** : 394 – 407.

Cohen, S. and Carpenter, G. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72** : 1317 – 1321.

Cohen, S., Carpenter, G. and Lembach, K.J. (1975) *In "Adv. Metabolic Disorders" Lust, R. and Hall, K. (Eds.),* **8** : 265 – 284. Academic Press, New York.

Cohen, S. and Stastny, M. (1968) *Biochim. Biophys. Acta.* **166** : 427 – 437.

Crewther, W.G. (1976) *Proc. 5th Int. Wool Test. Res. Conf. Aachen, 1975.* **1** : 1 – 101.

Crewther, W.G., Dowling, L.M., Steinbert, P.M. and Parry, D.A.D. (1983) *Internal. J. Biol. Macromol.* **5** : 267 – 274.

Dale, B.A. (1977) *Biochim. Biophys. Acta.* **491** : 193 – 204.

Darnell, J.E., Wall, R. and Tushinski, R.J. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68** : 1321 – 1325.

Davies, H.R. (1889) *Morph. Jahrb.* **15** : 560 – 645.

Dhouailly, D. (1967) *J. Embryol. Exptl. Morphol.* **18** : 389 – 400.

Dhouailly, D. (1973) *J. Embryol. Exptl. Morphol.* **30** : 587 – 603.

Dhouailly, D., Rogers, G.E. and Sengel, P. (1978) *Dev. Biol.* **65** : 58 – 68.

Dodson, J.W. (1963) *Exptl. Cell Res.* **31** : 233 – 235.

Doel, M.T. and Carey, N.H. (1976) *Cell.* **8** : 51 – 58.

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

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

Efstratiadis, A., Maniatis, T., Kafatos, F.C., Jeffrey, A. and Vournakis, J.N. (1975) *Cell.* **4** : 367 – 378.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.W., O'Connell, C., Spritz, R.A., De Riel, J.B., Forget, B.G., Weissman, S.M., Slightom, J., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) *Cell.* **21** : 653 – 668.

Englund, P.T. (1972) *J. Mol. Biol.* **66** : 209 – 224.

Espinasse, P.G. (1964) *Feather. in A New Dictionary of Birds, ed. A.L. Thomson,* **pp. 272 – 276**, Nelson, London.

Fell, H.B. (1962) *J. Embryol. Exptl. Morphol.* **10** : 389 – 409.

Fickett, J.W. (1982) *Nuc. Acids Res.* **10** : 5303 – 5318.

Fisher, C. and Sawyer, R.H. (1979) *J. Exp. Zool.* **207** : 505 – 512.

Fitton-Jackson, F.L. and Fell, H.B. (1963) *Dev. Biol.* **7** : 394 – 419.

Franke, W.W., Schmid, E., Weber, K. and Osborn, M. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75** : 5034 – 5038.

Fraser, R.D.B., MacRae, T.P., Parry, D.A.D. and Suzuki, E. (1971) *Polymer.* **12** : 35 – 36.

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

Fuchs, E. and Green, H. (1978) *Cell.* **15** : 887 – 897.

Fuchs, E. and Hanukoglu, I. (1983) *Cell.* **34** : 332 – 334.

Fuchs, E. and Marchuk, D. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80** : 5857 – 5861.

Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Crochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) *Nature.* **278** : 428 – 434.

Gedamu, L., Iatrou, K. and Dixon, G.H. (1977) *Cell.* **10** : 443 – 451.

Gilbert, W. (1979) *In Eukaryote Gene Regulation, ICN-UCLA Symposium on Molecular and Cellular Biology, XIV.* **pp. 1 – 12**, R. Axel, T. Maniatis and C.F. Fox, eds. New York, Academic Press.

Gillespie, D., Marshall, S. and Callo, R.G. (1972) *Nature (London) New Biol.* **236** : 227 – 229.

Goeddel, D.V., Leuing, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandiss, R., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W. (1981) *Nature.* **290** : 20 – 26.

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

Gould, H.J. and Hamlyn, P.H. (1973) *FEBS Letters.* **30** : 301 – 305.

Grantham, R., Gautier, C. and Guoy, M. (1980) *Nuc. Acids Res.* **8** : 1893 – 1912.

Green, K.J. and Goldman, R.D. (1983) *J. Cell Biol.* **97** : 224a.

Greenberg, J.R. and Perry, R.P. (1972) *Biochim. Biophys. Acta.* **287** : 361 – 368.

Gregg, K., Wilton, S.D., Parry, D.A.D. and Rogers, G.E. (1984) *EMBO Journal.* **3** : 175 – 178.

Gregg, K.G., Wilton, S.D., Rogers, G.E. and Molloy, P.L. (1983) *in Manipulation and Expression of Genes in Eukaryotes. eds. P. Nagley, A.W. Linane, W.J. Peacock and J.A. Pateman.* **pp. 65 – 72**, Academic Press, Sydney.

Gross, K., Probst, E., Schaffner, W. and Birnstiel, M. (1976) *Cell.* **8** : 455 – 469.

Grunstein, M.L. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72** : 3961 – 3965.

Hagenbuchle, O., Bovey, R. and Young, R.A. (1980) *Cell.* **21** : 179 – 187.

Hamilton (1952) *Lillie's Development of the Chick.* **Ed 3** Henry Holt and Co., New York.

Hamlyn, P.H., Brownlee, G.G., Cheng, C.C., Gait, M.J. and Milstein, C. (1978) *Cell.* **15** : 1067 – 1075.

Hanukoglu, I. and Fuchs, E.V. (1982) *Cell.* **41** : 243 – 252.

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.

Harrap, B.S. and Woods, E.F. (1967) *Comp. Biochem. Physiol.* **20** : 449 – 460.

Hayward, G.S. (1972) *Virology.* **49** : 342 – 344.

Heilig, R., Perrin, F., Gannon, F., Mandel, J.L. and Chambon, P. (1980) *Cell.* **20** : 625 – 637.

Herwick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) *J. Biol. Chem.* **256** : 7990 – 7997.

Hoober, J.K. and Cohen, S. (1967) *Biochim. Biophys. Acta.* **138** : 347 – 356.

Horiuchi, K. and Zinder, N.D. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72** : 2555 – 2558.

Iatrou, K. and Dixon, G.H. (1977) *Cell.* **10** : 433 – 441.

Jacobs-Lorena, M., Baglioni, C. and Borun, T.W. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69** : 2095 – 2099.

Jung, A., Seppel, A.E., Grez, M. and Schutz, G. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77** : 5759 – 5763.

Kato, Y. (1969) *J. Exp. Zool.* **170** : 229 – 252.

Kemp, D.J. (1972) *Ph.D Thesis, University of Adelaide.*

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. (1970) *J. Cell. Sci.* **7** : 273 – 283.

Kemp, D.J. and Rogers, G.E. (1972) *Biochem.* **11** : 969 – 975.

Kemp, D.J., Schwinghamer, M.W. and Rogers, G.E. (1974c) *Mol. Biol. Reports.* **1** : 441 – 446.

Kemp, D.J., Walker, I.D., Partington, G.A. and Rogers, G.E. (1975) *in The Eukaryote Chromosome, eds. W.J. Peacock* and *R.D. Brock,* **pp. 205 – 216** Aust. Nat. Uni. Press, Canberra.

Kemper, B. (1976) *Nature.* **262** : 321 – 323.

Kerbert, C. (1876) *Arch. Mikros. Anat.* **13** : 205 – 262.

Kim, K.H., Rheinwald, J.G. and Fuchs, E.V. (1983) *Mol. Cellular Biol.* **3** : 495 – 502.

Kischer, C. (1963) *Jour. Ultrastructure Res.* **8** : 305 – 321.

Kitano, Y. and Kuroda, Y. (1967) *Exptl. Cell Res.* **48** : 350 – 360.

Kollar, E.J. (1970) *J. Invest. Dermatol.* **55** : 374 – 378.

Krystosek, A., Cawthon, M.L. and Kabat, D. (1975) *J. Biol. Chem.* **250** : 6077 – 6084.

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

Lazarides, E. (1982) *Ann. Rev. Biochem.* **51** : 219 – 250.

Lazarides, E. and Hubbard, B.D. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73** : 4344 – 4348.

Liem, R.K.H. and Hutchinson, S.B. (1982) *Biochemistry.* **21** : 3221 – 3226.

Lillie, F.R. (1965) *"Lillie's Development of the Chick" ed. H.L. Hamilton,.* **pp. 546 – 606**, Holt, Rinehart and Winston, New York.

Lockett, T.J. (1979) *Ph.D. Thesis, Adelaide University.*

Lockett, T.J., Kemp, D.J. and Rogers, G.E. (1979) *Biochemistry.* **18** : 5654 – 5663.

Lucas, A.M. and Stettenheim, P.R. (1972) *Avian Anatomy – Integument. Part II. Agriculture Handbook.* **362** U.S. Dept. of Agriculture, Washington.

Mandel, J.L., Breathnach, R., Gerlinger, P., Le Meur, M., Gannon, F. and Chambon, P. (1978) *Cell.* **14** : 641 – 653.

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.

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72** : 1184 – 1188.

Marbaix, G., Huez, G., Burny, A., Cleuter, Y., Hubert, E., Leclercq, M., Chantrenne, H., Soreq, H., Nudel, U. and Littauer, U.Z. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72** : 3065 – 3067.

Marcu, K. and Dudock, B. (1974) *Nucleic Acids Res.* **1** : 1385 – 1397.

Matulionis, D.H. (1970) *Zeitschrift Fur Anatomie Und Entwicklungs Geschichte.* **132** : 107 – 157.

Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74** : 560 – 564.

Maxam, A.M. and Gilbert, W. (1980) *In "Methods in Enzymology" Grossman, L. and Moldave, K. (Eds). Academic Press, New York.* **65** : 499 – 560.

McMaster, G.K. and Carmichael, G.G. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74** : 4835 – 4839.

Mendecki, J., Lee, S.Y. and Brawerman, G. (1972) *Biochemistry.* **11** : 792 – 801.

Messing, J. (1979) *Recombinant DNA technical Bulletin.* **2** : 43 – 48.

Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nuc. Acids Res.* **9** : 309 – 321.

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74** : 3642 – 3646.

Messing, J. and Vieira, J. (1982) *Gene.* **19** : 269 – 276.

Moll, R., Franke, W.W., Schiller, D.L., Geiger, B. and Krepler, R. (1982) *Cell.* **31** : 11 – 24.

Molloy, P.L., Powell, B.C., Gregg, K.G., Barone, E.D. and Rogers, G.E. (1982) *Nuc. Acids Res.* **10** : 6007 – 6021.

Morris, C.P. (1976) *Hons. Thesis, University of Adelaide.*

Morris, C.P. and Rogers, G.E. (1979) *Mol. Biol. Reports.* **3** : 145 – 149.

Murray, N.E., Brammar, W.J. and Murray, K. (1977) *Mol. Gen. Genet.* **150** : 53 – 61.

Nudel, U., Soreq, H., Littauer, U.Z., Marbaix, G., Huez, G., Leclercq, M., Hubert, E. and Chantrenne, H. (1976) *Eur. J. Biochem.* **64** : 115 – 121.

O'Donnell, I.J. (1973a) *Aust. J. Biol. Sci.* **26** : 401 – 414.

O'Donnell, I.J. (1973b) *Aust. J. Biol. Sci.* **26** : 415 – 437.

O'Donnell, I.J. and Inglis, A.S. (1974) *Aust. J. Biol. Sci.* **27** : 369 – 382.

Panyim, S. and Chalkley, R.C. (1969) *Arch. Biochem. Biophys.* **130** : 337 – 346.

Parkes, K.C. (1966) *Living Bird.* **5** : 77 – 86.

Parnes, J.R., Velan, B., Felsenfeld, A., Ramanathan, L., Ferrini, U., Appella, E. and Sidman, J.G. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78** : 2253 – 2257.

Partington, G.A. (1974) *Ph.D. Thesis, Adelaide University.*

Partington, G.A., Kemp, D.J. and Rogers, G.E. (1973) *Nature New Biology.* **246** : 33 – 36.

Peacock, A.C. and Dingman, C.W. (1968) *Biochemistry.* **7** : 668 – 674.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) *Cell.* **20** : 555 – 566.

Perry, R.P., La Torre, J., Kelley, D.E. and Greenberg, J.R. (1972) *Biochim. Biophys. Acta.* **262** : 220 – 226.

Pinder, J.C., Staynov, D.Z. and Gratzer, W.B. (1974) *Biochemistry.* **13** : 5373 – 5378.

Poppa, L.M. and Bosch, L. (1969) *Fed. Eur. Biochem. Soc. Lett.* **4** : 143 – 149.

Powell, B.C. (1979) *Ph.D. Thesis, University of Adelaide.*

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) *Nuc. Acids Res.* **7** : 2165 – 2176.

Powell, B.C., Sleigh, M.J., Ward, K.A. and Rogers, G.E. (1983) *Nuc. Acids Res.* **11** : 5327 – 5346.

Price, M.G. and Lazarides, E. (1983) *J. Cell Biol.* **97** : 1860 – 1874.

Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature.* **263** : 211 – 214.

Raj Bahandary, U.L. (1968) *J. Biol. Chem.* **243** : 556 – 563.

Raj Bahandary, U.L., Stuart, A., Faulkner, R.D., Chang, S.H. and Khorana, H.G. (1966) *Cold Spring Harbour Symp. Quant. Biol.* **31** : 425 – 437.

Raj Bahandary, U.L. Stuart, A., Hoskinson, R.M. and Khorana, H.G. (1968) *J. Biol. Chem.* **243** : 565 – 571.

Randerath, K. and Randerath, E. (1973) *Methods in Cancer Research.* **9** : 3 – 15.

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.

Reddy, U.B., Gosh, P.K., Lebowitz. P. and Weismann, S.M. (1978a) *Nuc. Acids Res.* **5** : 4195 – 4213.

Reddy, U.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978b) *Science.* **200** : 494

– 502.

Rheinwald, J.G. and Green, H. (1975a) *Cell.* **6** : 317 – 330.

Rheinwald, J.G. and Green, H. (1975b) *Cell.* **6** : 331 – 343.

Rheinwald, J.G. and Green, H. (1977) *Nature.* **265** : 421 – 424.

Richards, R.I., Shine, J., Ullrich, A., Wells, J.R.E. and Goodman, H.M. (1979) *Nuc. Acids Res.* **7** : 1137 – 1146.

Richards, R.I. and Wells, J.R.E. (1980) *J. Biol. Chem.* **255** : 9306 – 9311.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* **113** : 237 – 251.

Roberts, B.E. and Paterson, B.M. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70** : 2330 – 2334.

Rogers, G.E. (1984) *Biochem. Soc. Symp.* **49** (In press).

Roman, R., Brooker, J.D., Seal, S.N. and Marcus, A. (1976) *Nature.* **260** : 359 – 360.

Romanoff, A.L. (1960) *"The Avian Embryo."* MacMillan, New York.

Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M. and O'Malley, B.W. (1978) *Cell.* **15** : 671 – 685.

Roop, D.R., Hawley-Nelson, P., Cheng, C.K. and Yuspa, S.H. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80** : 716 – 720.

Roop, D.R., Kronenberg, M.S. and Clark, J.H. (1984) *UCLA Symp.* (In press).

Rosbash, M. and Ford, P.J. (1974) *J. Mol. Biol.* **85** : 87 – 95.

Rose. J.K. (1975) *J. Biol. Chem.* **250** : 8098 – 8102.

Rottman, F., Shatkin, A.J. and Perry. R.P. (1974) *Cell.* **3** : 197 – 199.

Rougeon, F. and Mach, B. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73** : 3418 – 3422.

Rubenstein. M. (1979) *Anal. Biochem.* **98** : 1 – 7.

Ruderman, J.V. and Pardue, M.L. (1978) *J. Biol. Chem.* **253** : 2018 – 2025.

Rueger, D.C., Huston, J.S., Dahl, D. and Bignami, A. (1979) *J. Mol. Biol.* **135** : 53 – 68.

Saint, R.B. (1979) Ph.D. Thesis, University of Adelaide.

Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979a) *Nature.* **280** : 228 – 294.

Sakano, H., Rogers, J.H. Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S. (1979b) *Nature.* **277** : 627 – 633.

Sanger, F., Brownlee, G.G. and Barrell, B.G. (1965) *J. Mol. Biol.* **13** : 373 – 385.

Sanger, F. and Coulson, A.R. (1975) *J. Mol. Biol.* **94** : 441 – 448.

Sanger, F., Nicklon, S. and Coulson, A.R. (1977) *Proc. Nat. Acad. Sci. U.S.A.* **74** : 5463 – 5467.

Savage, C.R., Inagami, T. and Cohen, S. (1972) *J. Biol. Chem.* **247** : 7612 – 7621.

Sawyer, R.H. (1979) *Develop. Biol.* **68** : 1 – 15.

Schachner, M., Hedley-White, E.T., Hsu, D.W., Schoonmaker, G. and Bihnami, A. (1977) *J. Cell. Biol.* **75** : 67 – 73.

Schaefer, M.P., Boyd, C.D., Tolstoshev, P. and Crystal, R.G. (1980) *Nuc. Acids Res.* **8** : 2241 – 2253.

Scott, A.C. and Wells, J.R.E. (1976) *Nature.* **259** : 635 – 638.

Scott, A.C. and Wells, J.R.E. (1978) *Methods in Cell Biol.* **19** : 257 – 272.

Sengel, P. (1957) *Experimentia.* **13** : 177 – 182.

Sengel, P. (1958a) *Ann. Sc. Nat. Zool.* **20** : 431 – 514.

Sengel, P. (1958b) *Annee Biologique.* **34** : 29 – 52.

Sengel, P. (1971) *Advances in Morphogenesis.* **9** : 181 – 230.

Sengel, P. (1983) *Biochemistry and Physiology of the skin, vol I, ed. L.A. Goldsmith,* **pp. 102 – 131**, Oxford Uni. Press.

Sengel, P., Dhouailly, D. and Kieny, M. (1969) *Develop. Biol.* **19** : 436 – 446.

Shain, W.G., Hilfer, S.R. and Fonte, V.G. (1972) *Dev. Biol.* **28** : 202 – 218.

Shatkin, A.J. (1976) *Cell.* **9** : 645 – 653.

Slatter, D.W., Slatter, I. and Gillespie, D. (1972) *Nature (London) New Biol.* **240** : 333 – 335.

Slatter, I., Gillespie, D. and Slatter, D.W. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70** : 406 – 410.

Smith, G.E. and Summers M.D. (1980) *Anal. Biochem.* **109** : 123 – 129.

Southern, E.M. (1975) *J. Mol. Biol.* **18** : 503 – 517.

Staden, R. (1977) *Nuc. Acids. Res.* **4** : 4037 – 4051.

Staden, R. (1978) *Nuc. Acids. Res.* **5** : 1013 – 1015.

Staden. R. (1979) *Nuc. Acids. Res.* **6** : 2601 – 2610.

Staden, R. (1980a) *Nuc. Acids. Res.* **8** : 817 – 825.

Staden, R. (1980b) *Nuc. Acids. Res.* **8** : 3673 – 3694.

Starkey, R.H., Cohen, S. and Orth, D.N. (1975) *Science.* **189** : 800 – 802.

Staynov. D.Z., Pinder. J.C. and Gratzer. W.B. (1972) *Nature (London) New Biol.* **235** : 108 – 110.

Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) *Cell.* **21** : 681 – 687.

Steinert, P.M. and Cantieri, J.S. (1983) *Biochemistry and Physiology of the Skin. vol I. ed. L.A. Goldsmith,* **pp. 135 – 169**, Oxford Uni. Press.

Steinert, P.M., Cantieri, J.S., Teller, D.C., Lonsdale-Eccles, J.D. and Dale, B.A. (1981a) *Proc. Natl. Acad. Sci. U.S.A.* **78** : 4097 – 4101.

Steinert, P.M. and Idler, W.W. (1975) *Biochem. J.* **151** : 603 – 614.

Steinert, P.M., Idler, W.W., Cabral, F., Gottesman, M.M. and Goldman, R.D. (1981b) *Proc. Natl. Acad. Sci. U.S.A.* **78** : 3692 – 3696.

Steinert, P.M., Jones, J.C.R. and Goldman, R.D. (1984). *J. Cell Biol.* **99** : 22s – 27s.

Steinert, P.M. and Roop, D.R. (1984) *UCLA Symp.* (In press).

Steinert, P.M., Zackroff, R.V., Anyardi-Whitman, M. and Goldman, R.D. (1982) *Methods in Cell Biology.* **24** : 399 – 419.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977) *Gene.* **1** : 255 – 280.

Stewart, M. (1977) *J. Ultrastruct. Res.* **60** : 27 – 33.

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.

Sun, T.T., Eichner, R., Nelson, W.G., Tseng, S.C.G. and Weiss, R.A. (1983) *J. Invest. Dermatol.* **81** : 109s – 115s.

Sun, T.T. and Green, H. (1976) *Cell.* **9** : 511 – 521.

Sun, T.T. and Green, H. (1978) *Cell.* **14** : 469 – 476.

Sutcliffe, J.G. (1978) *Nuc. Acids Res.* **5** : 2721 – 2728.

Swank, R.T. and Munkres, K.D. (1971) *Anal. Biochem.* **39** : 462 – 477.

Symons, R.H. (1975) *Mol. Biol. Reports.* **2** : 277 – 280.

Taylor, J.M. and Tse, T.P.H. (1976) *J. Biol. Chem.* **251** : 7461 – 7467.

Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77** : 5201 – 5205.

Thompson, E.O.P. and O'Donnell, I.J. (1965) *Aust. J. Biol. Sci.* **18** : 1207 – 1225.

Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V. and Leder, P. (1978a) *Proc. Natl. Acad. Sci. U.S.A.* **75** : 725 – 729.

Tilghman, S.M., Curtis, P.J., Tiemeier, D.C., Leder, P. and Weissman, C. (1978b) *Proc. Natl. Acad. Sci. U.S.A.* **75** : 1309 – 1313.

Vogt, V.M. (1973) *Eur. J. Biochem.* **33** : 192 – 200.

Voitkevich, A.A. (1966) *"The Feathers and Plumage of Birds,"* Sidgwick and Jackson, London.

Wahl, G.M., Stern, M. and Starke, G.R. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76** : 3683 – 3687.

Wahli, W., Dawid, I.B., Wyler, T., Weber, R. and Ryffel, G.U. (1980) *Cell.* **20** : 107 – 117.

Walker, I.D. (1974) *Ph.D. Thesis, Adelaide University.*

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 – 340.

Walker, I.D. and Rogers, G.E. (1976b) *Eur. J. Biochem.* **69** : 341 – 350.

Ward, K.A., Edwards, K.J. and Sleigh, M.J. (1983) *Proc. Aust. Biochem. Soc.* **15** : p 70.

Wasylyk, B., Kedinger, C., Cordon, J., Briston, O. and Chambon, P. (1980a) *Nature.* **285** : 367 – 373.

Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R. and Chambon, P. (1980b) *Proc. Natl. Acad. Sci. U.S.A.* **77** : 7024 – 7028.

Weeds, A.G. (1976) *Eur. J. Biochem.* **66** : 157 – 173.

Wensink, P.C., Finnegan, D.J., Donelson, J.E. and Hogness, D.S. (1974) *Cell.* **3** : 315 – 325.

Wessells, N.K. (1961a) *Devlpmtl. 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. (1963) *Exp. Cell Res.* **30** : 36 – 55.

Wessells, N.K. (1965) *Dev. Biol.* **12** : 131 – 153.

Wiche, G., Krepler, R., Artleb, U., Pytela, R. and Denk, H. (1983) *J. Cell. Biol.* **97** : 887 – 901.

Williams, B. and Blattner, F. (1979) *J. Virol.* **29** : 555 – 575.

Wilton, S.D. (1983) *Ph.D. Thesis, University of Adelaide.*

Wilton, S.D., Crocker, L.A. and Rogers, G.E. (1984) *Nuc. Acids Res.* Submitted.

Wilton, S.D. and Rogers, G.E. (1981) *Proc. Aust. Biochem. Soc.* **14** : p 84.

Winter, G. and Fields (1980) *Nuc. Acids Res.* **8** : 1965 – 1974.

Woodcock-Mitchell, J.R., Eichner, R., Nelson, W.G. and Sun, T.T. (1982) *J. Cell Biol.* **95** : 580 – 588.

Yang, N.S., Manning, R.F. and Gage, L.P. (1976) *Cell.* **7** : 339 – 345.

Young, R.A., Hagenbuchle, O. and Schibler, U. (1981) *Cell.* **23** : 451 – 458.

Zackroff, R.V., Idler, W.W., Steinert, P.M. and Goldman, R.D. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79** : 754 – 757.