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.

TABLE OF CONTENTS

CHAPTER 2

MATERIALS AND GENERAL METHODS.

CHAPTER 4

KERATIN mRNA SEQUENCES.

CHAPTER 5

NON-KERATIN mRNA SEQUENCES.

CHAPTER 6

IDENTIFICATION OF FAST PROTEIN GENES.

CHAPTER 7

CONCLUDING DISCUSSION.

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.