Expression of the Guinea Pig $\alpha$-Lactalbumin Gene in Transgenic Mice

Thesis submitted for the degree of
Doctor of Philosophy
at the University of London

23rd February 1990

Antonio Maschio
Medical Molecular Biology Unit
Department of Biochemistry
University College, London
Abstract

In order to identify the mechanisms at work in the gene regulation of milk protein genes, and in particular the \( \alpha \)-lactalbumin gene, the guinea pig \( \alpha \)-lactalbumin gene has been analysed both \textit{In Vitro} using a mobility shift assay and \textit{In Vivo} by the creation of transgenic mice.

Results obtained using the \textit{In Vitro} system demonstrated the formation of specific complexes by nuclear proteins prepared from lactating guinea pig mammary glands on probes derived from the guinea pig \( \alpha \)-lactalbumin gene promoter. These complexes were however also formed by proteins prepared from the liver and kidney of lactating guinea pig, and therefore the \textit{In Vitro} approach was judged to be non-instructive in this instance.

Two lines of transgenic mice were constructed by microinjection of a 3.5kb guinea pig \( \alpha \)-lactalbumin genomic fragment, and both lines were found to express the transgene at levels commensurate with the levels seen in the guinea pig. Guinea pig \( \alpha \)-lactalbumin was secreted into the milk of transgenic mice. The transgene mRNA was correctly initiated and processed, and was found to be expressed at high levels exclusively in the mammary glands, and the skin of the transgenic mice. Both transgenic and non-transgenic lactating mice were observed to express the endogenous mouse \( \beta \)-casein gene in the skin. Guinea pig \( \alpha \)-lactalbumin protein was not detected in the skin of lactating transgenic mice.

\textit{In Situ} hybridisation studies revealed that the expression of the transgene in the mammary tissues of transgenic mice was restricted to the secretory acini of the mammary glands, which is the expected site for milk protein production. In the skin, the expression of guinea pig \( \alpha \)-lactalbumin mRNA was found to be confined to the sebaceous glands, more specifically to the ducts formed between the sebaceous gland itself and the hair follicle.
Acknowledgements

I would like to thank my supervisors, Dr. Roger Craig and Dr. Paul Brickell, for their guidance and support throughout this project.

Thanks are also due to many people both inside the Medical Molecular Biology unit and from other institutions who have been of great assistance with certain aspects of this work. In particular, I wish to thank Dr. Robin Lovell-Badge (MRC mammalian development unit, UCL) and Dr. David Murphy (Royal Veterinary College, London) for their persistent if unsuccessful attempts at the creation of transgenic mice; Dr. Dimitris Kioussis and Dr. Andy Mellor (National Institute for Medical Research, Mill Hill) for carrying out successful microinjections, and producing two different lines of transgenic mice; Dr David Katz and Mr. Neil Rayment (department of histopathology, Middlesex Hospital Medical School) for advice with the histological aspects of this work; and all the members of the Medical Molecular Biology unit for their help and advice during the last 3½ years.
Abbreviations

The abbreviations used in this thesis are described in the "Policy of the Journal and Instructions to Authors" of the Biochemical Journal (1988).

In addition, the following abbreviations are used:

- $\alpha$LA: $\alpha$-lactalbumin
- bp: base-pair
- BPV: Bovine Papilloma Virus
- BSA: Bovine serum albumin
- cDNA: complementary DNA
- DTT: Dithiothreitol
- HEPES: N-2-hydroxyethylpiperazine-N'-2-ethane-sulphonic acid
- PIPES: Piperazine-N-N-bis(2-ethane-sulphonic acid)
- Tris: Tris(hydroxy-methyl) amino-methane
- X-gal: 5'-bromo-4-chloro-3-indolyl-$\beta$-D-galactoside
- TBS: Tris buffered saline
- SDS: Sodium dodecyl sulphate
- OLB: Oligo-labelling buffer
Chapter 1

General Introduction

1.1 Gene expression in eukaryotes .................................................. 14
   1.1.1 The importance of the structure of DNA .......................... 14
   1.1.2 The structure of Chromatin ............................................. 14
   1.1.3 The synthesis of mRNA ................................................. 16
   1.1.4 The synthesis of protein ............................................... 22

1.2 Regulation of transcription by cis-acting DNA sequences ......... 23
   1.2.1 The TATA box ............................................................ 23
   1.2.2 The Upstream Promoter Element ................................. 27
   1.2.3 Trans-activating DNA binding proteins ....................... 27
   1.2.4 Mechanisms of trans-activation ................................... 29
   1.2.5 The steroid responsive element .................................. 33
   1.2.6 Enhancers are modular .............................................. 34
   1.2.7 Tissue-specific action of promoters and enhancers ......... 34

1.3 Post-transcriptional control ...................................................... 37
   1.3.1 Alternative splicing ..................................................... 37
   1.3.2 Transport of mRNA to the cytoplasm ......................... 37
   1.3.3 mRNA stability .......................................................... 38
   1.3.4 Translational Control ................................................. 39
Chapter 2

Materials and Methods

2.1 Materials ................................................................................................................... 62

2.1.1 Routine chemicals and plasmids ........................................................................ 62
2.1.2 Enzymes ............................................................................................................... 62
2.1.3 Radiochemicals .................................................................................................... 62
2.1.4 Kits ...................................................................................................................... 62
2.1.5 Antisera ............................................................................................................... 63
2.1. Other materials ..................................................................................................... 63
2.1.7 Solutions ............................................................................................................. 63

2.2 Methods ....................................................................................................................... 66

2.2.1 General procedures ............................................................................................ 66
2.2.2 Restriction enzyme analysis ................................................................................ 66
2.2.3 Transformation of bacterial cells ........................................................................ 67
2.2.4 Preparation of plasmid DNA .............................................................................. 67
   2.2.4a. Large-scale plasmid preparation ................................................................. 67
   2.2.4b. Small scale preparations ............................................................................ 68
2.2.5 Isolation of restriction fragments ...................................................................... 68
   2.2.5a. From agarose gels .................................................................................... 68
   2.2.5b. From polyacrylamide gels ....................................................................... 68
2.2.6 Polyacrylamide gels ............................................................................................ 68
   2.2.6a. Sequencing gels ....................................................................................... 68
   2.2.6b. Native polyacrylamide gels ..................................................................... 69
2.2.7 Preparation of genomic DNA from transgenic mouse tails ............................. 69
2.2.8 Preparation of total RNA ................................................................................... 70
2.2.9 Radiolabelling of DNA .............................................................. 71
   2.2.9a. Oligo-labelling of DNA probes ........................................ 71
   2.2.9b. MMTV RT end-labelling of DNA ...................................... 71
2.2.10 Southern blot analysis ......................................................... 72
2.2.11 Northern Blot analysis ......................................................... 72
2.2.12 S1 nuclease analysis of total RNA ....................................... 73
2.2.13 Maizel protein gels ............................................................... 74
2.2.14 Western blotting ................................................................. 75
2.2.15 In Vitro transcription of α-lactalbumin complementary RNA ... 75
   2.2.15a. Labelling transcription .................................................. 75
   2.2.15b. Cold transcription ......................................................... 76
2.2.16 Preparation of slides for In situ hybridisation sections .......... 76
2.2.17 In Situ hybridisation protocol ............................................. 76
Chapter 3

The structure of the guinea pig \( \alpha \)-lactalbumin gene: regulatory sequence elements

Introduction

3.1 The genomic organisation of the guinea pig \( \alpha \)-lactalbumin gene. 81

Results

3.2 The position of regulatory consensus sequences in the guinea pig \( \alpha \)-lactalbumin gene ................................................................. 85

3.3 Gel retardation analysis of the \( \alpha \)-lactalbumin gene promoter .... 90

3.3.1 Preparation of mammary gland nuclear extracts .................................. 90

3.3.2 Preparation of end-labelled DNA probes from the guinea pig \( \alpha \)-lactalbumin promoter. ........................................................................ 96

3.3.3 Resolution of protein/DNA complexes by native gel electrophoresis .... 96

3.3.4 The formation of reproducible complexes on a whole-promoter probe.. 97

3.3.5 The formation of complexes on a 55bp probe ...................................... 109

3.3.6 A 90bp probe obtained by partial Fok I digestion forms extra bands .... 109

Discussion

3.4 Regulatory sequences in the \( \alpha \)-lactalbumin gene ...................... 117

3.5 The nature of protein-DNA complexes formed on the \( \alpha \)-lactalbumin gene ................................................................. 117
Chapter 4

Construction of transgenic mice carrying the guinea pig α-lactalbumin gene

Introduction

4.1 Cloned milk protein genes are not expressed in tissue culture.....119

4.2 Transgenic mice: an in vivo expression system...............................121

Results

4.3 Generation of transgenic mice.................................................................124

4.3.1 Preparation of the transgene DNA.........................................................124

4.3.2 Microinjection of Mouse Embryos.......................................................124

4.4 Detection of the transgene in founder mice.................................130

4.4.1 Preparation of an α-Lactalbumin cDNA probe..............................130

4.4.2 Identification of Transgenic Mice by Southern Blot Analysis of Tail DNA. 130

4.4.3 Screening of Transgenic Mouse Offspring........................................131

4.4.4 Deletions and Rearrangements of Transgenes in the Mouse Germ Line 132

4.5 Two unusual lines of transgenic mice.................................................145
Chapter 5
Expression of the α-lactalbumin transgene

Introduction

5.1 The expression of milk protein genes in transgenic mice..............146

Results

5.2 The expression of guinea pig α-lactalbumin in transgenic mice....148

5.2.1 Milking Transgenic Mice..............................................................148

5.2.2 Analysis of transgenic mouse milk for secretion of guinea pig α-lactalbumin.
....................................................................................................................149

5.3 Determination of tissue specificity of gene expression..............152

5.3.1 Northern blot analysis of transgenic mouse tissues..................152

5.3.2 Determination of the transcription start site by S1 nuclease protection
analysis........................................................................................................160

5.4 Hormonal regulation of the guinea pig α-lactalbumin transgene...163

5.5 The expression of milk protein genes in mouse skin..............169

5.5.1 The expression of mouse β-casein in lactating mice.................169

5.5.2 α-lactalbumin expression in guinea pig skin.................................169

5.5.3 Guinea pig α-lactalbumin protein is not detectable in transgenic mouse skin
..................................................................................................................169

5.6 High-level expression of the α-lactalbumin transgene..............179
5.7 Tissue specificity and hormonal regulation of milk protein transgenes

5.8 High level transcription of milk protein mRNA in mouse skin

Chapter 6
The identification of cell-specific α-lactalbumin transgene expression by in situ hybridisation

Introduction

6.1 The relationship between mammary gland and skin

6.1.1 Mammary gland and skin epithelia share a common evolutionary background

6.1.2 The development of the mammary gland

6.1.3 The development of other skin appendages

6.2 The localisation of mRNA by In situ Hybridisation

Results

6.3 In situ hybridisation of mouse sections

6.3.1 Subcloning of the guinea pig α-lactalbumin cDNA into a GEMINI vector

6.3.2 Preparation of tissue sections

6.3.3 Hybridisation of tissue sections with α-lactalbumin cRNA

6.3.4 Autoradiographic analysis of hybridised tissue sections
6.4 Microscopic examination of developed and stained slides .......................... 193
  6.4.1 Mammary sections .............................................................................. 193
  6.4.2 Skin sections ....................................................................................... 203
  6.4.3 Sections from other tissues ................................................................ 203

6.5 Correct transgene expression in mammary cells ........................................ 215

6.6 Transgene expression in the skin ............................................................. 215
  6.6.1 The development of sebaceous glands ................................................. 215
  6.6.2 Localisation of transgene expression to the sebaceous basal cells ....... 216

Chapter 7

Conclusion

7.1 The study of milk protein gene expression ................................................... 217
  7.1.1 The commercial importance of milk-specific transgenes ....................... 217
  7.1.2 Approaches to the study of milk protein gene regulation ....................... 217

7.2 Gene expression of the guinea pig α-lactalbumin gene ................................ 218
  7.2.1 In Vitro analysis of the guinea pig α-lactalbumin gene promoter .............. 218
  7.2.2 The expression of guinea pig α-lactalbumin in transgenic mice ............... 219
  7.2.3 The extracellular matrix and milk protein gene expression ..................... 220

7.3 Further Experiments .................................................................................... 221
  7.3.1 Introduction of manipulated genes into transgenic mice ....................... 221
  7.3.2 The construction of a new transgenic mouse line .................................. 225
  7.3.3 Skin-specific α-lactalbumin expression ................................................. 226

References ........................................................................................................ 227
1.1 Gene expression in eukaryotes

1.1.1 The importance of the structure of DNA

The mechanistic study of eukaryotic gene expression first became a possibility with the isolation of mRNA in the late 1960s, and the subsequent rapid development of gene cloning techniques to allow the isolation and characterisation of single-copy cellular genes. The structure of eukaryotic genes is now known in considerable detail (see Breathnach and Chambon, 1981, for a review; fig. 1.1A); much research is now concerned with how and when these genes are expressed.

It is generally accepted that the regulation of gene expression in higher eukaryotes centres mainly on the control of transcription of cellular genes into mRNA. This control is effected not only through the protein factors and the DNA elements they interact with (see section 1.3), but also at the level of the physical structure of the nuclear chromatin which encodes the gene. Changes in chromatin conformation can either allow or restrict access by the protein factors to particular DNA sequences, and are thus found to be responsible for wide-ranging activation and repression of linked genes (reviewed by Weintraub, 1985). Such changes are largely responsible for the creation of differentiated states among cell types, and the permanent nature of these changes ensures the stability of the cell types within differentiated tissues.

1.1.2 The structure of Chromatin

DNA is packaged into the nucleus as a dense structure organised on three levels- the 10nm (100Å) fibre, the 30nm (300Å) fibre and the chromosomal loops. This organisation serves the primary purpose of allowing the inclusion into a cell nucleus, typically 5μm in diameter, of the entire genome, which, in the human, consists of $2.3 \times 10^9$ bp and would extend to an estimated length of 174cm if stretched out. The 10nm fibre is formed by the coiling of the DNA double helix around an octamer of histone proteins (two each of H2A, H2B, H3, H4) which form the
nucleosome (figure 1.1B; reviewed by Kornberg and Klug, 1981), with histone H1 binding to the
linker DNA to 'seal' the nucleosome. The chromatin present in most cell nuclei is further
compacted into a 30nm fibre, which is composed of nucleosome strings arranged into an ordered
structure (reviewed by Felsenfeld and McGhee, 1986). The most favoured model for the
assembly of nucleosomes into a fibre is as a solenoid (Thoma et al., 1979), with six nucleosomes
per helical turn. Molecules of histone H1 are thought to be involved at this stage, by associating
with each other and causing the tight packaging of the nucleosomes onto each other. This 30nm
fibre becomes organised at a further level, as the solenoid is looped around the nuclear scaffold
(reviewed by Eissenberg et al., 1985). Non-histone scaffolding proteins are thought to arrange
the 30nm fibre into a looped structure in a sequence-specific manner (Mirkovitch et al., 1984).
These chromosome loops are further compacted to form metaphase chromosomes during cell
division, when chromosomes become compacted and readily visible under the microscope.
There is a strong suggestion (Matsumoto et al., 1980), that the phosphorylation of the fifth
histone, histone H1, is important in this condensation step and therefore instrumental for the
progression of the cell cycle into mitosis.

Transcriptionally active genes are also packaged into nucleosomes (Lacy and Axel, 1975; McKnight et al., 1978), but the double helix appears to have a different topological structure, which is demonstrated by increased sensitivity to the pancreatic endonuclease
deoxyribonuclease 1 (DNase 1- see Weisbrod, 1982; Eissenberg et al., 1985). An even greater
level of sensitivity to DNase 1 (DNase 1 hypersensitive sites) is observed in the cis -control
regions of transcriptionally active genes. Sensitivity to DNase 1 is possibly induced by
topoisorerase II induced stress in the double helix (Mirkovitch et al., 1984; Villeponteau
et al., 1984), which may be concentrated into the nucleosome-free areas, in which cis -acting
control regions are often located, by the topologically restrictive interactions of the nucleosome
cores with the DNA (Morse and Simpson, 1988). The modification of histone proteins, as well as
the displacement of nucleosomes, may play a part in releasing negative supercoils constrained
within nucleosomes. Norton et al., 1989, demonstrated that acetylation of nucleosome particles,
restricted to the basic domains of the core histones which are thought to interact with the DNA,
caused the number of DNA supercoils associated with each nucleosome to decrease
significantly. The interactions of the DNA with the non-histone nucleosome-associated proteins
HMG 14 and HMG 17 are also thought to be responsible for the maintenance of the DNase 1 sensitive state, probably by causing further torsional stress in the DNA (Weisbrod et al., 1980), or by displacing histone H1 (Weintraub, 1985).

The active state of chromatin is further affected by the methylation state of the DNA, present as 5-methyl cytosine and concentrated into CpG dinucleotides around the 5' end of the gene (reviewed by Bird, 1987; Cedar, 1988). Transcriptionally active genes are found to be demethylated, and this demethylation would appear to be important only at the 5' end (Busslinger et al., 1983). The experiments of Becker et al. (1987) suggested that methylation prevented the in vivo binding of transcription factors. Recent reports of the isolation of a protein (MeCP) which specifically binds methylated CpGs in vivo and in vitro and may prevent the binding of other proteins to the DNA (Meehan et al., 1989; Antequera et al., 1989) suggest a mechanism by which these phenomena may be explained.

1.1.3 The synthesis of mRNA

Messenger RNA was first isolated from polysomal fractions by precipitation as an 18s ribonucleoprotein (mRNP) fraction released by the chelation of Mg2+ ions by EDTA (covered by Lewin, 1987). Comparison of these transcripts with the genomic DNA from which they were derived revealed the eukaryotic transcription unit to be very different from the prokaryotic genes which were known at the time (see reviews by Breathnach and Chambon, 1981; Nevins, 1983).

Before being translocated from the nucleus to the cytoplasm, mRNA is modified extensively both at the 3' and 5' ends, and in the body of the message by the splicing out of introns (fig. 1.2). The initiation of transcription is directed by the gene promoter, in particular the TATA box (see section 1.3), and it has been shown that the 5' end of the mRNA maps to this initiation site (Ziff and Evans, 1978). The 5' end of the message is modified very early during its synthesis by the addition of a cap structure (Nevins, 1983), consisting of a 7-methyl G residue joined 5' to 5' to the end of the polynucleotide chain by the action of methyl- and guanylyltransferases. The original first residue of this chain is then methylated at position 2, and in a small number of cases may be methylated at the N6 position (Lewin 1987). This cap is involved in the translation of the mRNA, by interacting with the Cap Binding Proteins (CBP) of the translation initiation complex, which interact with the ribosomal subunits and lead to initiation.
(Shatkin, 1985), and may also be important in preventing degradation of the mRNA (Banerjee, 1980). The cap structure is, however, dispensed with by many mRNAs from eukaryotic viruses, and is therefore probably not strictly essential for translation (Banerjee, 1980).

Before the completed transcript is spliced, the 3' end of the mRNA is modified by the addition of a poly-A tail (Nevins and Darnell, 1978). The polyadenylation site occurs 10 to 25bp downstream of the consensus sequence AAUAAA. This sequence is thought to direct polyadenylation, in some mRNAs by interacting with another downstream sequence, the GT cluster (typically YGTGTYY), to form a hairpin loop (Birnstiel et al., 1985). Such an interaction specifies a site at which the transcript, which proceeds beyond the polyadenylation site to terminate in a termination region rather than at a specific termination site (Darnell, 1982), can be cleaved (Nevins, 1983), and the poly-A tail subsequently added at this site. The function of the poly-A tail is thought to be to limit the rate at which mRNAs are degraded in the cytoplasm (Nevins, 1983).

The capped and polyadenylated transcript is then spliced to remove the non-coding introns. Examination of the sequences at splice junctions reveals consensus splice donor and acceptor sequences, \((\text{C}/\text{A})\text{AG}/\text{GT}(\text{A}/\text{G})\text{AGT}\) and \((\text{\overline{T}/\overline{C}})\text{6N}(\text{C}/\text{T})\text{AG}/\text{G}\) respectively (Mount, 1982). The introns are excised in a preferred sequence, but not in any conserved order (Roop et al., 1978). During the splicing reaction, the introns are known to form a branched-circular lariat structure (Padgett et al., 1984), which is released as the 5' and 3' exons are spliced together. A family of small nuclear ribonucleoproteins (snRNPs; U1 to U7) is known to be involved in these processes (reviewed by Sharp, 1987) and their interaction with the target pre-mRNA leads to the formation of a spliceosome, or splicing complex, at the splice site. This complex is essential for the splicing reaction, and is assembled on the splice site by the sequential binding of snRNPs. The processed mRNA is then removed to the cytoplasm for eventual translation into protein.
Figure 1.1

The structure of eukaryotic genes.

A. A model eukaryotic RNA polymerase II transcription unit

The promoter, consisting of one or more upstream promoter elements (UPE) and a TATA box is located at the 5' end of the gene, upstream of the transcription initiation site. There may be a consensus 'initiator' sequence downstream of the TATA box (Smale and Baltimore, 1989), which includes the mRNA start site. The coding sequence of the DNA is arranged into exons, which are present in the processed mRNA product. The introns are spliced out. Polyadenylation occurs downstream of the AAUAAA consensus, and RNA polymerase II terminates in a 'termination region' shortly after (Birnstiel et al., 1989).

B. The structure of DNA in the nucleosome

DNA is coiled twice round the nucleosome core proteins, and 'sealed' by histone H1 interactions with the linker DNA (after Kornberg and Klug, 1981).
A. UPE UPE TATA initiation AAUAAA termination
promoter exon intron exon

polyadenylation

B. nucleosome core
2x H2A, B, H3 & H4.
Messenger RNA is capped virtually instantly as it is synthesised (Darnell 1982). Subsequently polyadenylation takes place, at a polyadenylation site which is 10 to 20bp downstream of the AAUAAA polyadenylation signal. Methylation will then take place, followed by splicing. The introns are spliced out by the spliceosome complex (S), which is formed by the snRNPs, the splice donor (sd) and splice acceptor (sa) being fused together while the intron is released as a lariat structure.
1.1.4 The synthesis of protein

mRNA is translated by the ribosomal apparatus to direct the assembly of proteins from the free amino acid pool (see Lewin, 1987). Proteins are assembled N-terminus to C-terminus via the base-pairing of aminoacyl tRNAs with the individual codon triplets of the mRNA.

Much of recent research has been centred on the mechanism of translation initiation, as it is at this initial step that control of gene expression may be effected (discussed in section 1.3.4). The first step involves the binding of the 40s ribosomal subunit with the cap structure on the mRNA and involves the eukaryotic initiation factors (elF 1 to 3, elF 4A to D and elF 5), crude preparations of which have been shown to contain the cap binding proteins (CBP) (Shatkin, 1985). CBP 1 binds to the cap, and then forms a stable complex called CBP II (also called elF-4F) with two other polypeptides. The initiating tRNA, plus GTP, are bound by elF 2, which binds to the 40s ribosomal subunit. CBP II is thought to be involved in the formation of the 40s initiation complex, by binding the mRNA to the 40s subunit.

The 40s subunit is then thought to migrate along the mRNA in an ATP dependent manner, until it reaches the first AUG of the mRNA (the scanning model for translation: reviewed by Kozak, 1989). Here tRNA^{met} is thought to be responsible for stopping the migration through its base-pairing to the initiator codon. The RNA sequence in the vicinity of the initiating AUG is highly conserved among eukaryotic mRNAs, with a consensus GCCGCC\[^A\]GCAUGG (Kozak, 1989). Mutation of this sequence can significantly impair initiation.

The elongation process then involves the binding of the 60s ribosomal subunit, and the sequential addition of amino acids as specified by codon-anticodon pairing between the mRNA and the tRNAs. The process is dependent upon elongation factors (eEF1 and eEF2) and the hydrolysis of GTP for translocation of the nascent polypeptide chain between the A and P sites on the ribosome, and concomitant movement of the ribosome along the mRNA strand. The elongation process is terminated when the ribosome arrives at a stop codon, and involves the release factor eRF.
1.2 Regulation of transcription by cis-acting DNA sequences.

1.2.1 The TATA box

It is generally understood that transcription from eukaryotic genes is regulated by means of specific DNA sequences known as promoters and enhancers (Maniatis *et al.*, 1987). The promoter is a region of DNA situated proximal to the 5' end of the gene, and is responsible for directing the correct initiation of mRNA transcription. Typically, a promoter will consist of a TATA box (reviewed by Breathnach and Chambon, 1981) at -20 to -30, and one or more upstream promoter elements (UPEs) 20 to 70bp further upstream (Sassone-Corsi and Borelli, 1986).

Eukaryotic genes coding for proteins are transcribed by RNA polymerase II (pol II). The function of the TATA box is to direct correct initiation of the mRNA by RNA polymerase II. Experiments involving mutation of the TATA box lead to heterogeneous initiation of the transcript but do not necessarily alter the rate of transcription, demonstrating that the TATA box functions to define the site of initiation rather than its rate (Benoist and Chambon, 1981). As is the case in prokaryotes, however, the polymerase alone is incapable of directing precise initiation of mRNA transcription; a host of additional factors are required. Five such activities (which may consist of more than one polypeptide each) have been isolated, and named TFIIA, TFIIB, TFIID, TFIIE and TFIIF (Lillie and Green, 1989). It has been shown that a combination of these five factors and pol II will successfully initiate a minimal promoter, consisting of just a TATA box and a transcription initiation site.

By resolving the complexes formed on the adenovirus major late promoter by native gel electrophoresis, Buratowski *et al.* (1989) have recently been able to determine an ordered sequence of events which leads up to the initiation of transcription from the TATA promoter (fig. 1.3). TFIIID was found to be the first protein to bind, was essential for the formation of all complexes, and binds to the TATA motif. TFIIID is found to be widely conserved throughout evolutionarily distant eukaryotes, suggesting a broad similarity in the mechanisms that lead to initiation; yeast TFIIID is functional in vitro with mammalian transcription systems lacking in TFIIID (Buratowski *et al.*, 1988). TFIIA is the next factor to be incorporated into the complex, and to a certain extent stabilises the binding of TFIIID, possibly by altering the interaction of TFIIID with the TATA box. The binding of TFIIB creates the third complex, which, when examined by DNAse I...
footprinting, gives rise to a wide area of protection which extends beyond beyond the TATA box to include the start site. It also allows the binding of the next component, pol II, and is thought to act as a 'bridge' to the polymerase. The binding of pol II is followed by the association of TFIIE to the complex. It is not to be excluded that in vivo pol II will associate with TFIIB/E before binding to the TFIID/A complex positioned on the TATA box, or indeed that the latter complex does not preassemble.

A protein named RAP 30 has been isolated as a protein that binds pol II, and characterised as part of the TFIIE fraction (Buratowski et.al.,1989) or as TFIIF, a subdivision of TFIIE (Lillie and Green, 1989). RAP 30 is bound to another protein, RAP 74, forming the RAP 30/74 heterodimer. This dimer is essential for the formation of the first phosphodiester bond at a promoter (Burton et.al.,1988), has recently been shown to include an ATP dependent DNA helicase activity (Sopta et.al.,1989), although it is not clear whether this helicase is RAP 74 or another polypeptide. The hydrolysis of ATP leads to the formation of an activated transcription complex (Buratowski et.al.,1989).

Some genes, typically housekeeping genes, do not possess a TATA box, but an alternative element near the start site helps to determine the site of transcription initiation (Dynan, 1986). Thus both the hypoxanthine phosphoribosyl transferase (HPRT) gene and the mouse dihydrofolate reductase (DHFR) gene lack TATA consensus sequences in their promoters. They both, however, possess a GC rich 5' flanking region which contains several potential binding sites for the ubiquitous transcription factor Sp1, known as a GC box (typically GGGCGG: Kadonaga et.al.,1986). GC boxes, as well as CAAT boxes (CCAAT), are typically found 40 to 100 nucleotides upstream of the start site in TATA promoters, and are thought to be responsible for the upregulation of transcription, at least 10-fold in the case of Sp1 (Breathnach and Chambon, 1981). If, in non-TATA promoters, these GC boxes are responsible for the initiation of transcription, then there may be a fundamental difference in the initiation process; the details of initiation on these promoters are unknown.
Buratowski *et al.* (1989) were able to demonstrate the sequence of complex formation at the transcription initiation site. TFIID is the TATA recognising factor, and the rest of the initiation complex forms around it.

1. TFIID binds to the TATAAA motif; the transcription initiation site is indicated.
2. This complex is stabilised by TFIIA.
3. TFII B is believed to form a 'bridge' to pol II; this allows the RNA polymerase to bind. TFIIA may be discarded at this stage.
4. TFII E/F binds downstream of pol II. TFII F comprises the RAP 30/74 ATP-requiring activity which is necessary for the formation of the activated transcription complex.

(Adapted from Buratowski *et al.*, 1989)
1. TF II D - TATAAA

2. A TF II D - TATAAA

3. A TF II D - TATAAA
   RNA polymerase II

4. A TF II D - TATAAA
   RNA polymerase II
1.2.2 The Upstream Promoter Element

The function of the UPEs is to modulate the rate of transcription; ubiquitous examples are the CAAT box and the GC box already mentioned (Dynan and Tjian, 1985), while genes controlled in a tissue-specific or temporally regulated manner will often possess more specialised UPEs. UPEs are binding sites for transcription-activating proteins, and may overlap eachother, creating the possibility of competition for site occupancy between two proteins with different effects. For example, the glucocorticoid receptor, a transcriptional activator, has been shown to bind DNA overlapping the UPE responsible for conferring cAMP responsiveness (Akerblom et al., 1988). The proteins which bind to these elements are usually promoter-specific activators, and their role is to accelerate one or more steps leading to either the initiation of transcription or elongation. Mutagenesis studies suggest that the number and type of UPE is instrumental in determining the relative strength of the promoter (Maniatis et al., 1987), that UPEs can be inverted with respect to the TATA box without loss in activity (Serfling et al., 1985) but that deletion of varying amounts of bases from the DNA which lies between the UPE and the TATA box can be detrimental, especially if the stereospecific alignment of the UPE with respect to the TATA box is altered (Takahashi et al., 1986). These data suggest that the activator proteins which bind to the UPEs may interact with the transcription complex which assembles around the TATA box.

1.2.3 Trans-activating DNA binding proteins

Before considering the mechanisms by which transcriptional activation by DNA-binding proteins is thought to occur, it is well to define the different types of trans-activators characterised to date. It has become apparent, especially from studies involving yeast activators, that the domains for DNA binding and transcriptional activation are distinct, and separable (reviewed by Ptashne, 1988). Thus the bacterial LexA DNA binding domain, when exchanged for the DNA binding domain of the yeast activator GAL 4 in a hybrid molecule, directs the activating effects of the GAL 4 activation domain exclusively to yeast genes which have a LexA operator placed upstream of the start site.

Activator proteins can be subdivided into at least four structural groups, reviewed by Mitchell and Tjian (1989), and Johnson and McKnight (1989). The 'zinc finger' domain was first
observed in TFIIA, a pol III transcription factor which is active in the transcription of 5S rRNA genes. There are at least two classes of zinc finger structures (Evans and Hollenberg, 1988). These are the C₂H₂ group, in which a Zn²⁺ ion is chelated by two Cys residues and two His residues, and the Cₓ group, of which GAL 4 and the steroid hormone receptors are members, where one or more Zn²⁺ ions are chelated by four, five, six or more conserved cys residues. The DNA binding sequence specificities of the zinc finger proteins are diverse, and the site of specificity is not believed to lie within the finger.

A second variety of DNA binding domain is the homeodomain, first detected in Drosophila homeotic gene products, but since recognised in a variety of eukaryotic transcriptional activators. Consisting of a region of 60 or so amino acids, displaying a sequence of conserved basic and hydrophobic residues, the homeodomain is found in many mammalian activators such as the octamer binding proteins oct-1/2, the pituitary-specific activator pit-1, and Unc-86, a developmental regulator of Caenorhabditis elegans. Analysis of these proteins reveals the existence of a larger region of homology, known as the POU domain (Pit - Oct - Unc; Herr et al., 1988). Whether this POU domain is involved in other functions apart from DNA binding in these transcriptional activators remains to be seen.

The third binding domain is that first described for the activator C/EBP. Other members of this family include AP1/Jun, Fos, and CREB (see Mitchell and Tjian, 1989), and display a conserved region of 30 amino acids associated with a 'leucine zipper', a region containing four leu residues spaced at intervals of seven amino acids. The leucine zipper is involved in dimerisation reactions between these factors rather than DNA binding.

The fourth type of domain is a basic α-helix found on NF-1, AP-2 and SRF (Mitchell and Tjian, 1989). A number of other activator proteins do not seem to fall into any category, at the moment at least (Johnson and McKnight, 1989).

The activating domains of transcriptional activators are also diverse. GAL 4 possess two regions of negatively charged amino acids, which form amphipathic α-helices. The possession of negative charge appears to be the key characteristic, as point mutations which add negatively charged residues increase activity (Ptashne, 1988). These negatively charged regions are shared by other activators, such as AP-1/Jun and the glucocorticoid receptor. Sp1 has four activating domains, two of which are rich in glutamine residues, and glutamine-rich regions occur in
1.2.4 Mechanisms of trans-activation

Recently (Lillie and Green, 1989; Ptashne, 1988) there has been an accumulation of evidence to suggest that the target of the activating domains is the binding of TFIID. For example, DNA-bound ATF (activating transcription factor; Lee et al., 1987) alters the conformation of TFIID and the nature of its binding to the DNA (Horikoshi et al., 1988a). The same effects are apparent with the yeast activator GAL 4, (Horikoshi et al., 1988b). The yeast homologue of TFIID has been characterised (Buratowski et al., 1988; Cavallini et al., 1988) and cloned by two separate groups (Hahn et al., 1989; Horikoshi et al., 1989). Buratowski et al. and Cavallini et al. (1988) both demonstrated that the yeast TFIID was capable of substituting for mammalian TFIID in a mammalian pol II in vitro transcription system. In addition, the yeast activator protein GAL 4 will trans-activate mammalian, insect and plant genes (Ptashne 1988). This would seem to suggest that the conservation of TFIID and the ability of activators to function in different species are not unrelated. The existence of more than one type of transcription factor activator domain, though, argues against TFIID being the only target, unless there are multiple varieties of TFIID or the 'normal' TFIID has more than one activation site. It seems likely that the acidic-helix type of domain interacts with TFIID (Ptashne, 1988; Mitchell and Tjian, 1989) but the other domains, the proline-rich and glutamine-rich varieties, may interact with other factors (TFII proteins, or RAP30/74), or directly with pol II.
Figure 1.4

Models for the mechanism of gene activation

A: The interaction between the activator protein, which binds UPE or enhancer elements, and the target protein is central to the current models for activation. The DNA is imagined to loop out of the intervening space. Clearly, specificity of activation is possible at several levels with such a model: different activator proteins will bind to different activator sites, and may interact with different target proteins or different forms of the same target. The target protein may well be TFII D (see below; after Lillie and Green, 1989)

B: Factor cooperativity could operate in a scheme such as this. Each factor stabilises the interaction between the activating protein and the target. Such interactions are known to take place, and may not require DNA binding by the auxiliary proteins.

C: Another model for factor cooperativity. The target protein continuously scans the activator sites for the presence of activator proteins, which thereby stabilise the association of the target protein with the target site. The continual scanning of extra factors (3 and 4) keeps the target on site and allows continuous initiation by RNA polymerase II.
1.2.5 Enhancers and silencers

Associated with promoters, and often structurally similar, are enhancers and silencers. Traditionally, enhancers have been regarded as being functionally different from promoters, in that they do not direct initiation of transcription, but serve to regulate its level. They are capable of acting at a distance, irrespective of their orientation with respect to the promoter. Recently, though, the distinction between enhancers and UPEs has become increasingly blurred (Maniatis et al., 1987). In yeast, only one class of activation sequence exists, fulfilling both roles. The mouse metallothionein responsive UPE behaves like an enhancer if separated from its associated TATA box (Serfling et al., 1985) as will the heat shock regulatory element (Bienz and Pelham, 1986) but only if two copies are placed in close association with each other: one copy is insufficient. The method of enhancer action is thought to be similar to that of UPE action (reviewed by Ptashne, 1986; Maniatis et al., 1987; Ptashne, 1988). The fact that UPEs and enhancer modules prove to be similar (see section 1.2.7) supports this line of thought, but leaves the bridging of hundreds or even thousands of base-pairs of DNA by enhancer-binding proteins to be explained. The most favoured model is the looping model, where DNA is imagined to loop out, allowing enhancer-bound proteins to contact the polymerase complex (Ptashne 1986). Other models, such as the causation of a conformational change in the helical structure of the DNA by the activator protein leading to increased transcription, are largely rejected by Ptashne (1986), but may operate in certain situations; for example, the expression of the hsp 70 gene is associated with displacement of a nucleosome; this is likely to cause a conformational change in the DNA structure. There is evidence that transcription itself displaces histones from the nucleosome (Lorch et al., 1988), which is likely to have a similar effect. More importantly, the occupation of a promoter or enhancer region by activator proteins may prevent the formation of a repressive chromatin structure (Wolffe and Brown, 1988; note, however, that this article is concerned with oocyte 5S rRNA genes, which are transcribed by pol III) and promoters and enhancers of transcriptionally active genes are often found in nucleosome-free stretches of DNA (Mitchell and Tjian, 1989).

Silencers are DNA sequences which confer down regulation upon a gene. They can be identical to enhancer sequences, but bound by proteins which repress rather than activate. An example of this situation exists in the β-interferon enhancer (Goodbourn et al., 1986). This
normally binds two cellular proteins which repress transcription; upon viral infection however, these proteins are displaced by an activating factor which binds to an adjacent region of the enhancer (Zinn and Maniatis, 1986). Alternatively, silencers can act to repress gene expression without seemingly interacting with proteins, as is the case with c-myc (Linzer, 1985).

1.2.6 The steroid responsive element

The SRE is a good example of a family of enhancer sequences which can bind steroid receptors to activate or repress gene expression (Yamamoto, 1985), and a pertinent example in milk protein gene expression, as lactogenesis is under the antagonistic regulation of the steroids glucocorticoid and progesterone (see section 1.6). In vitro binding studies have confirmed that the steroid receptors bind to SRE (for example, see Payvar et al., 1983; reviewed by Beato, 1989). The consensus Glucocorticoid Responsive Element (GRE) derived from these studies consists of a pentadecamer, which incorporates a highly conserved sequence TGTTCT, and is able to mediate induction by progesterone (von der Ahe et al., 1985; Strähle et al., 1987), mineralocorticoids and androgens as well as glucocorticoids, indicating that the elements are in general not receptor-specific (Beato 1989). SRE often occur in several copies, and may operate synergistically to favour cooperative binding of receptors (Tsai et al., 1989).

An observation by Adler et al., 1989, suggests that activation and inactivation of gene transcription by steroid receptors can be mediated by two separate mechanisms. They observe that the rat prolactin gene is negatively regulated by estrogen if the estrogen responsive element is removed from the gene promoter, and confirm that the inhibitory effect is mediated by the estrogen receptor in the absence of its DNA binding domain. The only prerequisite for inhibition is the presence of the binding site for the tissue-specific activating transcription factor pit-1. This clearly implies that a protein-protein interaction between the activated estrogen receptor and pit-1 is responsible for the inhibitory effect, possibly by the inhibition of the pit-1 mediated stimulation.
1.2.7 Enhancers are modular

Promoters and enhancers are composed of a number of short DNA sequence elements (or 'Modules': see Dynan, 1989). These elements are the binding sites for one or more transcription regulating proteins, and in the case of enhancer modules are often referred to as 'enhancers'. But, just as a CAAT box is not a promoter but a promoter module, which forms part of a functional promoter, so enhancer modules are not enhancers but cooperate to form a functional enhancer. Thus two SV40 enhancer modules are necessary to form a functional SV40 enhancer. Loss of one module will inactivate the enhancer completely. In terms of binding of regulatory proteins, it is believed that these modules, which were originally defined by genetic means, contain multiple overlapping protein binding sites, called enhansons, which probably bind one regulatory protein each (Dynan, 1989). Enhansons can be active either alone or in pairs with identical or nonidentical partners. An enhancer module, consisting of one or two such enhansons in an active configuration, will be reiterated a number of times at variable distances to form a functional enhancer. Therefore we can regard the heat shock element (HSE) as an enhancer module: a pair of HSEs will function as an enhancer. Alternatively, placed in the proximity of a promoter, a single module will function as an UPE (Bienz and Pelham, 1986). One exception to this generalisation is the CAAT box; multiple CAAT boxes will not function as enhancers if removed from the immediate vicinity of the TATA box (Maniatis et al., 1987). Thus although promoters and enhancers are functionally distinct, they both rely on similar or in some cases identical basic units to achieve their respective goals.

1.2.8 Tissue-specific action of promoters and enhancers

Some promoters and enhancers, such as those for housekeeping genes or the SV40 enhancer, do not show any tissue-specific properties, and are universally active. Other promoters, however, are functional only in a given cell type, and will contain tissue-specific promoter elements in addition to the ubiquitous sequences such as the TATA and CAAT boxes. Examples include the enhancers from insulin, interferon, elastase, α-fetoprotein, and MHC genes (see Voss et al., 1986). Promoters can also direct tissue-specific responses, as exemplified by the prealbumin gene promoter, which is only active in liver cells. It is linked to a
similarly liver-specific enhancer, which will enhance transcription from any promoter, but only in liver cells (Costa et al., 1986).

Not all tissue-specific control is so clear-cut, however. Enhancer elements may interact with different transcription factors in different tissues, and these interactions may be influenced by and may require the concerted action of more than one transcription factor. This can make initial results, especially if obtained in vitro, extremely confusing.

The best characterised tissue specific system is that of the immunoglobulin genes, in which the difficulties mentioned above are clearly demonstrated. The octamer sequence ATGCAAAAT is a motif which functions as a *cis*-regulatory element in the promoters of many ubiquitous genes (e.g. snRNA genes, histone H2B genes; Herr et al., 1988), and also functions as a regulatory element in the B cell-specific immunoglobulin gene promoter. The reason for this apparent disparity of function, that is general activity in some genes and tissue-specific activity in immunoglobulin genes, can be explained by the fact that the same sequence is recognised by several different transcription factors. One of these, oct-1, is generally found in all cells and, in the case of the histone H2B octamer, regulates cell cycle S-phase-specific expression (Fletcher et al., 1987). Furthermore, it can exert this regulatory function independently of the other H2B promoter elements, apart from the TATA box. Another factor, oct-2, is B cell-specific (Scheidereit et al., 1987). *In vitro* the purified oct-2 will selectively activate immunoglobulin genes. Thus the same regulatory element, which is bound by two functionally different proteins, can enhance two separate transcription events. Since both oct-1 and oct-2 are present in B cells, the question arises as to how oct-2 can exert its specific effect on immunoglobulin genes, especially as experiments using gel retardation and methylation interference footprinting have failed to show any selective binding properties. The cooperative action of other, possibly tissue-specific, protein factors and DNA elements is probably the answer, as four separate protein-DNA interactions have been identified in the immunoglobulin heavy chain enhancer (Maniatis et al., 1987). Genomic footprinting of these sites and of the octamer reveals that they are occupied only in B cells, even though the factors which bind to them are ubiquitous.

It is also known that protein-protein interactions can modify the activity of transcriptional regulators. This, of course, occurs in the interaction between pol II and the TF II factors, but also affects more specific transcription factors such as oct-1. It is known that the herpes simplex virus
(HSV) transactivator protein VP16 (or VMW65), which is responsible for the activation of viral immediate-early genes during lytic infection, does not bind DNA. Rather, it forms a complex with oct-1 to redirect the binding specificity from the octamer to a sequence found in the HSV immediate-early region, TAATGARAT. This interaction is mediated through the homeodomain of oct-1 (Stern et al., 1989). Of the other types of protein-protein interaction have been observed, one is 'squelching' (see Ptashne, 1988) which involves the 'off-DNA' interaction between an activator and the activator target protein which effectively down regulates gene expression. One problem with this theory is that if the (universal) activator target protein is the ubiquitous TFIID (see section 1.2.1) then a universal downregulation would take place, and this is not the case. Another variety of interaction is the repression of prolactin synthesis by estrogen (Adler et al., 1989; section 1.2.6), which involves three-way interaction between estrogen (a ligand), its receptor, and a transcription factor.

A further perplexing picture is painted by the properties of the transcription factor NF-κB (reviewed by Lenardo and Baltimore, 1989). This factor was isolated as a 'tissue specific' factor which bound to the κ light chain enhancer in B cells (Sen and Baltimore, 1986). It was soon found, however, to be a ubiquitous factor which is normally present in the cytoplasm of many different cell types, and is complexed to an inhibitor protein, called IκB. A modification of IκB, possibly phosphorylation, by NF-κB inducers allows the translocation of NF-κB to the nucleus. Here it seems to perform a variety of different roles in different cell types, for example stimulating the production of lymphokines in T-cells, but not in B-cells. Other transcription factors have been implicated in the control of NF-κB induced response, either by being required to act in concert with NF-κB or by binding to silencer sequences.
1.3 Post-transcriptional control

The transcription of the DNA template into mRNA is only the first step in the synthesis of protein. In the case of genes encoding mRNA the activity of a gene is expressed only through active protein, and therefore the regulation of mRNA processing, stability and translation is important to the control of gene expression.

1.3.1 Alternative splicing

Unprocessed eukaryotic mRNA comprises both exons, which encode protein, and introns, which do not (reviewed by Breathnach and Chambon, 1981). Selective processing which deletes specified exons from the mature mRNA can allow for one gene to encode more than one protein. An example of this is the calcitonin/α-CGRP gene, which encodes both calcitonin and α-CGRP (Amara et al., 1982). Calcitonin and α-CGRP have common N-terminal regions, but diverge at the C-termini. This is due to common usage of the first three exons in the unprocessed mRNA, and selective use of exons four, five and six. Calcitonin mRNA, expressed in thyroid C cells, incorporates the fourth exon, and uses the adjacent polyadenylation site. CGRP mRNA, on the other hand, is expressed in neuronal cells by the action of a trans-acting factor (Emeson et al., 1989) to cause the splicing together of exons three and five, and the use of the polyadenylation signal adjacent to exon six.

This mechanism is only relevant to complex transcription units, but as intron splicing is essential to the formation of a functional mRNA, we can envisage the possibility of leaving a mRNA unspliced, and therefore inactive, until the relevant signal is given for the cell to produce the gene product. This process is known to occur in Drosophila, for example in P-element transposition (reviewed by Bingham et al., 1988).

1.3.2 Transport of mRNA to the cytoplasm

Transcription takes place in the nucleus, while translation is effected in the cytoplasm. Therefore the transport of mRNA to the cytoplasm is a potential site for regulatory activity, which
has been observed to occur in cells infected with HIV (Malim et al., 1989). The HIV-1 \textit{rev} gene (and its HTLV-1 counterpart \textit{rex}; Green and Zapp, 1989) is found to encode a trans-activator which increases export of RNA from the nucleus. This mechanism is found to operate late in infection, and leads to the export of as yet unspliced RNA, which is essential for packaging into virions as genomic RNA. To date, this mechanism has not been demonstrated to occur in other organisms.

1.3.3 mRNA stability

The rate of mRNA degradation is of vital importance to the levels of gene expression within the cell. Although transcription is the most studied, and probably the most significant level at which gene expression is regulated, transcriptional regulation must act in tandem with the control of mRNA stability to modulate the steady-state levels of mRNA in the cytoplasm. The regulation of mRNA stability (reviewed by Brawerman, 1987) is known to be important in the case of proteins which are expressed at very high levels. Milk proteins, which are expressed at exceptionally high levels exclusively during lactation, are controlled in this manner. The lactogenic hormone prolactin is known to increase the stability of casein mRNA 17 to 25 times, as well as simultaneously stimulating its transcription two to four fold (Guyette et al., 1979). This is also true of the histone protein mRNAs, whose stability increases five-fold during the S (DNA synthesis) phase of the cell cycle, when the demand for histone proteins is very high, and transcription is stimulated three to five fold. In these cases, it would be very uneconomical for the cell to have to transcribe a huge amount of unstable mRNA.

The interdependent modulation of transcription and mRNA stability is obviously of importance, and is found to be a general feature of systems in which mRNA stability control is known to occur. Evidently, with increasing mRNA stability comes a concomitant decrease in the response of gene expression to the rate of transcription; therefore, as well as increasing expression levels, changes in mRNA stability can be employed to reduce expression levels more rapidly than would be possible by reducing the transcription rate of a comparatively stable mRNA.

The mechanisms by which these stabilising (or destabilising) effects are mediated are gradually being elucidated. Brawerman (1987) proposes the importance of stem-loop structures, which are found in the 3' untranslated regions (3'UTRs) of many differentially stabilised mRNA
species, and suggests that they may impede degradation of the mRNA molecule by 3' exonuclease activities. There may be special sequences within the 3' UTR (the bacteriophage λ sib structure; eukaryotic AU rich sequences) which regulate the mRNA half-life by directing a specific endonuclease activity to remove the protecting 3' UTR.

Recently Müllner et al. (1988), working with the human transferrin receptor (hTR; the chief means of iron uptake into the cell), showed that the stability of hTR mRNA was markedly decreased in the presence of iron. It was demonstrated that iron-controlled regulation of mRNA stability was due to the presence of five palindromic sequences, as well as a stem-loop structure in the 3' untranslated region (3'UTR). At least four palindromes, and the stem-loop structure, are required for iron-dependent regulation of mRNA stability. Subsequently (Müllner et al., 1989) it was shown that the stability of the mRNA was mediated by the binding to the palindromic sequences of a specific RNA-binding protein factor, which is in turn inactivated by iron. The binding of this factor probably inhibits a separate RNase activity, which acts selectively on the stem-loop. Therefore the presence of excess Fe ions in the cytoplasm causes rapid loss of RNA stability, and a consequently a rapid reduction in the steady-state mRNA level and concomitant levels of hTR synthesis (fig. 1.5).

The 3' end of the histone H3 mRNA is also involved in stability control, but is dependent on translation of the mRNA, and contact of the ribosome with the 3' end, to induce degradation (Graves et al., 1987). Another kind of mechanism which requires translation of the mRNA is observed in β-tubulin (Yen et al., 1988). The amino-terminus of the β-tubulin peptide is recognised on emergence from the ribosome, and acts to destabilise the mRNA, probably by activating an RNase. It has been suggested that this RNase is carried by the ribosome, which would explain the result of Graves et al., by proposing that the 3' end of the histone H3 mRNA carries an RNase activating sequence.

### 1.3.4 Translational Control

The selective modulation of the rate of translation of mRNA is a further level at which gene expression can be controlled. The iron metabolism of the cell provides another excellent example in this case. Ferritin, which is responsible for the chelation of excess iron in the cytoplasm, is positively translationally regulated by iron. There is a palindromic sequence in the
5'UTR of ferritin mRNA which is identical to the palindromic sequences observed in the 3'UTR of the hTR receptor (section 1.3.3: fig. 1.5). This palindrome has been demonstrated to be essential for this negative regulation, and can be substituted for one of the hTR palindromes (Casey et al., 1988). Casey et al. suggest that a protein binding to these palindromes may be responsible for both retarding translation (by obstructing the ribosome) and preventing degradation. The presence of iron degrades the binding factor, thus increasing the levels of ferritin to chelate excess iron, and decreasing the levels of transferrin, thus reducing iron uptake by the cell. This suggestion is supported by the findings of Mülner et al. (1989).

Translational control is also known to proceed in other ways, which may or may not involve the 5'UTR (reviewed by Fink, 1988). The yeast GCN4 mRNA is regulated through translation of a small peptide at its 5' end, which inhibits initiation at the start site; at times of amino acid depletion, the production of these peptides is suppressed. The 3' ends of interferon and tissue plasminogen activator (tPA) are also involved in translational control, in the case of tPA by affecting the essential polyadenylation of the transcript.
A model for iron regulation in the cell using post-transcriptional methods has been put forward by Müllner et.al.,(1989), and Casey et.al. (1988). The existence of an Iron Responsive Factor (IRF) through which these effects are mediated was suggested by Müllner et.al.

A. In high iron conditions, the stem-loop structure of the hTR mRNA is susceptible to attack by a specific RNase activity, which leads to rapid mRNA degradation and concomitant reduction of iron uptake into the cell.

B. When iron becomes scarce, a protein called IRF binds to the palindromic sequences in the 3'UTR, and in some way inhibits the RNase activity. Higher levels of hTR ensue, and more iron is taken up into the cell.

C. In the case of ferritin, the palindrome is found in the 5'UTR. In conditions where iron is plentiful, ferritin mRNA is translated normally, and high ferritin levels lead to sequestration of iron from the cytoplasm.

D. Under conditions of iron scarcity, IRF binds to the palindrome in the ferritin mRNA 5' UTR; this causes a reduction in translation levels, possibly through physical obstruction of the ribosomal apparatus.

After Müllner et.al. (1989).
Ribosomes

translation product

AAA

AAA
1.4 Milk Proteins

The biosynthesis of milk is the foremost specialisation of the class mammalia, being common to all mammals, even the egg-laying monotremes. Composed primarily of water, milk contains a large array of carbohydrates, fats, salts, organic compounds and proteins (reviewed by Jenness, 1974). The proteins in milk are either synthesised in the mammary gland or directed into milk secretions from the blood. The latter category, which includes proteins such as serum albumin and immunoglobulins, comprises 5-10% of bovine milk proteins (see Larson and Jorgensen, 1974). At least 90% of milk proteins, though, are synthesised in the mammary gland from free amino acids. These proteins are broadly divisible into two groups, the caseins and the whey proteins, originally defined by virtue of their method of purification. The milk proteins of the cow, *Bos taurus*, have been more closely studied than those of any other species, and therefore much of the available data refer to bovine protein, which will be referred to hence unless specified.

1.4.1 Caseins

The original definition of a casein protein is according to the purification method of Hammarsten (quoted in a general review by Weller, 1979), who precipitated a milk protein fraction by acidification to pH 4.6. However, not all proteins which have been identified as belonging to a casein gene family are precipitated at pH 4.6, whereas some noncasein proteins are. The caseins are perhaps best defined as proteins which possess an ester-bound phosphate group (Jenness, 1974) which tend to aggregate, and do not seem to have a clearly defined organised structure. They exist in the form of calcium-dependent micelles, spherical particles with a diameter of 40-280nm, and are largely responsible for the distinctive appearance of milk.

The caseins comprise a heterogeneous group of proteins which were initially subdivided into three major classes. These were the α casein complex, β casein, and the γ casein complex, classified in decreasing order of mobility on electrophoresis at pH 8.6. The α and β-caseins are described as calcium-dependent, because they are precipitated by low Ca²⁺ concentrations, and are maintained in suspension in milk by their interaction with κ-casein. The major components of the predominant α casein complex are the α₅ caseins, further subdivided into αₛ⁰ and αₛ₂
casein groups, and the κ caseins, of which there are two genetic variants (see Larson and Jorgensen, 1974). Each of these groups can be further subdivided according to the number of carbohydrate and phosphate groups bound. These fractions are precisely defined according to their solubility characteristics, especially in the presence of Ca$^{2+}$ ions. There is an additional fraction known as λ casein which is more vaguely defined (see Weller, 1979) and, unlike other caseins, does not form aggregates.

β caseins, the next most abundant, form a group which is composed of several fractions. γ casein has now been defined as a fraction of β casein, derived by proteolysis, and not a separate protein. Guinea pig milk contains over twice as much casein as bovine milk (Jenness, 1974), yet, due to the commercial insignificance of guinea pig milk, little is known about guinea pig caseins.

Guinea pig milk proteins were fractionated by sedimentation at pH 4.6 by Craig et al., 1976, who determined that guinea pig milk was composed of 78% caseins and 22% whey protein. Analysis of the casein fraction by SDS PAGE revealed the presence of three major casein species, which were named casein A (α$\text{s}_2$), B (α$\text{s}_1$) and C in decreasing order of electrophoretic mobility. This result suggests the presence of a similar population of caseins as that seen in bovine milk (Jenness, 1974) and rat milk (Rosen et al., 1975).

1.4.2 Whey proteins

The whey fraction of milk is defined as those proteins which remain soluble at pH 4.6. This therefore includes both the noncasein proteins and those caseins which are soluble at pH 4.6 (Jenness, 1974). The noncasein proteins present in bovine milk whey are β-lactoglobulin, α-lactalbumin, bovine serum albumin, IgG, lactoferrin and an associated protein known as lactollin. β-lactoglobulin is the most abundant whey protein in bovine milk, has a molecular weight of 18300, occurs in four polymorphs A- D, and is commonly found in the milks of ruminants. It does not usually contain carbohydrate, although this is not unknown. β-lactoglobulin is poor in Cys residues, and rich in Pro. The function of β-lactoglobulin remains unknown, although its similarity to retinol binding protein may suggest that it is involved in vitamin A transport (Pervaiz and Brew, 1985; Papiz et al., 1986).
α-lactalbumin, also a major component of bovine milk, has a molecular weight of approximately 15000, and has a well-defined structure containing 8 Cys residues linked together by 4 disulphide bridges (see Ebner and Schanbacher, 1974, and Brew and Hill, 1975, for reviews). The structure and primary sequence of α-lactalbumin (first determined for α-lactalbumin from bovine milk, see Brew et al., 1970; for human α-lactalbumin, see Findlay and Brew, 1972; also for guinea pig α-lactalbumin, see Brew, 1972) as well as the nucleotide sequence of both the cDNA (Hall et al., 1982) and the gene encoding it (Hall et al., 1987) bear a striking resemblance to lysozyme, implying evolution from a common ancestral gene. Although at 123 amino acids α-lactalbumin is slightly smaller than lysozyme (130 amino acids in human), the identical pairing of the Cys residues in the disulphide bridges as well as the amino acid sequence similarity suggest that the protein structures may be similar. α-lactalbumin and lysozyme share related enzymic functions: lysozyme catalyses the hydrolysis of β1-4 glycosidic bonds, while α-lactalbumin is involved in the formation of a β1-4 glycosidic bond. The methods by which these functions are propagated are, though, completely different (Hall and Campbell, 1986), as α-lactalbumin has no enzymic activity but merely functions as a specifier protein. α-lactalbumin is the B protein of the enzyme complex lactose synthase, and has the function of altering the substrate specificity of galactosyl transferase to allow the transfer of a galactosyl moiety onto glucose rather than N-acetyl glucosamine, the more usual substrate. This potentiates the synthesis of lactose, a glucose-galactose disaccharide.

1.4.3 Guinea pig whey proteins

In the guinea pig β-lactoglobulin is absent (Brew and Campbell, 1967), and α-lactalbumin is the most abundant whey protein (Craig et al., 1976). It is present in vast excess of the concentration required to fulfil its biosynthetic role, and it may have the additional role of supply of cysteine to the suckling newborn. The protein has a molecular weight of 14500 and appears to be synthesised in the first instance as larger polypeptide, pre-α-lactalbumin (mw 15500), which is subsequently cleaved at the N-terminus to produce α-lactalbumin (Craig et al., 1976, 1978). Comparison of the amino acid sequence for guinea pig α-lactalbumin (Findlay and Brew, 1972) and the nucleotide sequence of the guinea pig α-lactalbumin genomic clone (Laird et al., 1988) shows the presence of 19 amino acids between the initiating ATG and the N-terminal Lys residue.
This sequence is believed to act as a signal sequence to direct the nascent α-lactalbumin polypeptide to the endoplasmic reticulum, and is subsequently removed (see Craig et al., 1979).

1.4.4 The role of α-lactalbumin in lactose biosynthesis

Lactose is the major milk carbohydrate in the vast majority of mammalian species, with the notable exception of the Californian sea-lion (Pilson and Kelly, 1962). It is also found in very small quantities in the monotremata and the family Otarioidea, i.e. seals, sea lions and walruses (Jenness 1974). It is composed of a disaccharide of glucose and galactose molecules, joined by a β1-4 glycosidic link. This bond is formed in the mammary gland by the milk enzyme complex lactose synthetase, which is created by the interaction of galactosyl transferase and α-lactalbumin.

Galactosyl transferases are universal enzymes present in the golgi apparatus of all cells. They are the generally accepted enzymic markers for golgi, used to detect golgi fractions during subcellular fractionation. They catalyse the transfer of galactose from UDP galactose to various acceptors, and can also catalyse the transfer of galactose to glucose, but only when glucose is present at very high concentrations (for a review, see Ebner and Schanbacher, 1974).

Galactosyl transferases are involved in the glycosylation steps in the synthesis of glycoproteins and glycolipids, as well as the synthesis of polysaccharides and mucin. The galactosyl transferase responsible for the synthesis of lactose transfers galactose moieties onto N-acetylglucosamine, giving rise to products such as:

\[
galactose - \beta(1-4) - \text{GlcNAc} - \beta - \text{glycoprotein}
\]

In the presence of α-lactalbumin, glucose can function as the acceptor in the place of N-acetyl glucosamine. Other galactosyl transferases, which do not effect transfers to N-acetyl glucosamine, are not involved in lactose biosynthesis, but the 'correct' galactosyl transferase has been identified in a wide variety of cell types and tissues. In order to effect a transfer to glucose, however, the inordinately large Km of galactosyl transferase with respect to glucose has to be lowered. α-lactalbumin possesses this activity, and is only synthesised in the mammary gland, making lactose a mammary gland specific disaccharide.
1.5 Milk protein gene expression

1.5.1 Studies on the gene expression of milk proteins: the isolation of mRNA

Measurement of the relative levels of DNA and RNA in mammary tissue during pregnancy and lactation (Nelson et al., 1962) had demonstrated that the production of milk proteins was associated with growth of the mammary gland, and the production of large quantities of RNA. As it became established that hormonal action was exerted through modification of the expression of RNA and proteins, and that lactation was under the control of a multiplicity of hormones (see Banerjee, 1976), the mammary gland presented itself as an ideal system in which to study the hormonal regulation of gene expression. Early studies involved the assay of milk protein production by the estimation of lactose synthetase activity and the Ca$^{2+}$-rennin precipitation of $^{32}$P casein (Rosen et al., 1980). These experiments, based on whole animals or organ slices in culture, demonstrated that the mammary gland is regulated in a variety of different ways during lactogenesis, and that milk protein gene expression is specifically regulated by the synergistic and antagonistic action of different hormones.

These results, however, did not shed any light on the control of RNA by hormones; in order to investigate how RNA synthesis was affected, several groups isolated polysomal RNA from lactating tissue, and demonstrated that it was capable of directing milk protein synthesis in vitro (Fairhurst et al., 1971; Campbell et al., 1973; Gaye et al., 1973). Total lactating mammary gland mRNA was later shown to be capable of the same (Craig et al., 1976; reviewed by Craig and Campbell, 1978). This technique allowed the estimation of mRNA levels at different stages during pregnancy (Rosen et al., 1975; Nakhasi and Qasba, 1978; Takemoto et al., 1980), and demonstrated that prolactin directly induced casein mRNA in organ culture (Houdebine and Gaye, 1975). Analysis of the mRNA sequences thus identified also confirmed the presence of a signal peptide on pre-$\alpha$-lactalbumin, and that, in common with other secreted proteins, milk proteins were mainly synthesised on membrane-bound polysomes (Siekevitz and Palade, 1980; Craig et al., 1978).

The synthesis of partial length cDNA clones derived from the mRNA isolates (Houdebine, 1976) allowed considerable progress in the rapid detection of small quantities of message, and
confirmed the results obtained by *in vitro* message translation (Houdebine and Gaye, 1975). Experiments along the same lines by Rosen and colleagues demonstrated the importance of prolactin and hydrocortisone acting synergistically to effect the maximal stimulation of casein mRNA synthesis (Rosen and Barker, 1975; Matusik and Rosen, 1978), and that prolactin could stimulate casein mRNA transcription and enhance its stability in the cytoplasm (Guyette *et al.*, 1979).

By the use of cDNA hybridisation probes it was possible, therefore, to investigate not only the control of milk protein synthesis, but also to begin to address questions about the molecular mechanisms through which this control might be effected.

1.5.2 The isolation of milk protein genomic clones

Although experiments using cDNA probes to determine mRNA in whole animals and organ culture have begun to demonstrate the intricacies of the mechanisms of regulation of milk protein genes, they cannot answer questions relating to the structure of the gene itself and the importance of this structure in the regulation of the gene. In order to do this, it is necessary to be able to study the fate of manipulated genes in expression systems, and to this end the genes encoding many different species of milk protein have been cloned and characterized. These include three rat casein genes (Yu-Lee *et al.*, 1986), bovine α-casein (Yu-Lee *et al.*, 1986) and bovine β-casein (Gorodetsky *et al.*, 1988), bovine α-lactalbumin (Hurley and Schuler, 1987), bovine β-lactoglobulin (Jamieson *et al.*, 1987), whey acidic protein (WAP) from rat and mouse (Campbell *et al.*, 1984), human α-lactalbumin (Hall *et al.*, 1987) and guinea pig α-lactalbumin (Laird *et al.*, 1988).

The structure of milk protein genes have been found to be remarkably consistent across species barriers; thus the α-caseins of rat and cow (Yu-Lee *et al.*, 1986) show a conserved 5' exon-intron structure, in which the signal peptide and casein kinase phosphorylation sequences are encoded by different exons, suggesting evolution of an ancestral gene by the addition of exons. The coding sequences of the mouse and rat WAP genes are virtually identical, as are the regions of the introns next to the splice junctions (Campbell *et al.*, 1984), and the α-lactalbumin genes of rat, man and guinea pig show extensive conservation (Hall *et al.*, 1987).
1.6 Hormonal control of milk protein gene expression

1.6.1 Mechanism of hormone action

Hormones may be subdivided into two classes: the steroid hormones, which are based on the structure of cholesterol and are relatively water-insoluble, and the peptide hormones, which are water-soluble. Therein lies the key operational difference between the two subclasses - cell membranes are permeable to steroid hormones but not to peptide hormones. In the female, steroid hormones are further subdivisible into the ovarian steroids, such as estrogen and progesterone, and the adrenal steroids such as the glucocorticoids. The expression of milk proteins is regulated by both classes of steroids, and also by peptide hormones, notably prolactin and insulin. Prolactin is a peptide with a molecular weight of 24Kd, and is part of a large gene family of peptide hormones, which includes growth hormone and placental lactogen (see Vonderhaar, 1987). These peptides have been shown to be involved in the stimulation of mammogenesis, with the possibility of substituting for prolactin in experimental systems (Topper and Freeman, 1980).

1.6.1.1 Steroid hormones

The method by which steroids mediate their effects has long been the subject of enquiry, since the discovery of ecdysone-induced chromosome puffs in the giant chromosomes of Drosophila in 1960 (Clever and Karlson, 1960; see Beato, 1989). Today it is clear that the hormone acts through binding to a cellular receptor, which may be cytoplasmic or nuclear in its localisation, and subsequent activation of the transcription of specific genes (recently reviewed by Rories and Spelsberg, 1989, and Burnstein and Cidlowski, 1989). The steroid, being lipid-soluble, passes freely through the cell membrane and either binds to a cytoplasmic form of its steroid- and tissue-specific receptor or is translocated to the nucleus, where the binding event takes place (see fig.1.6).

The structure of steroid receptors has been determined in detail (reviewed by Evans, 1988; Beato, 1989) and shown to consist of distinct domains: the DNA binding domain, which possesses Cx type Zn$^{2+}$-stabilised DNA binding fingers (section 1.2.3), the hormone binding domain, and an immunogenic hypervariable N-terminal region which is required for maximal
efficiency and may also be involved in the transcriptional activation function (Burnstein and Cidlowski, 1989). These domains have proved to be separable and interchangeable: deletion of the steroid-binding domain of the human glucocorticoid receptor (hGR) created a constitutively activating molecule, which, when transfected into cells, increased transcription from glucocorticoid-responsive genes in a glucocorticoid-independent manner (Hollenberg et al., 1987). Furthermore Green and Chambon (1987) showed that replacing the human estrogen receptor DNA binding domain with that from the hGR confers estrogen inducibility onto glucocorticoid-responsive genes.

The binding of the steroid to its receptor results in the formation of a transcriptionally activating hormone/receptor complex, which in turn binds DNA (Yamamoto, 1985). Analysis of glucocorticoid receptor (GR) complexed MMTV DNA by DNase 1 footprinting (Payvar et al., 1983) shows that the binding is sequence specific, identifying five cis -acting elements of about 20bp, in the 5' LTR. Such elements have been shown to confer hormone-inducibility to otherwise non-responsive genes when incorporated 5' or 3' of the gene promoter (Yamamoto, 1985). Known as Hormone Responsive Elements (HRE), these sequences share strong homologies, and are able to bind more than one receptor type (Beato 1989). There is also the suggestion (Rories and Spelsberg, 1989) that there are specific chromatin acceptor sites with which the activated receptor complexes interact. They consist of nuclear proteins associated with the DNA, possibly in a tissue specific manner, and may provide a second level of control of gene activation by steroids.

1.6.1.2 Peptide Hormones

The prolactin receptor was first isolated from rabbit mammary gland, and subsequently shown to be absent from the fat pad cells and specific to the mammary epithelium (Suard et al., 1979). It is also, however, present in liver, kidney, and a variety of other tissues. Purification of the receptor, for example by affinity to ovine prolactin (Necessary et al., 1984), yields a number of peptides with binding activity. It is possible to disaggregate these peptides into 'core binding units', or 40Kd units, which have a molecular weight in the region of 35-42 (Vonderhaar, 1987). A larger unit (85-90Kd) has been isolated more recently from rat ovaries (Mitani and Dufau, 1986) and their presence has been confirmed in mammary tissue (Vonderhaar, 1987). It is known
that the prolactin receptor exists in a cryptic state, in which it does not bind prolactin, but may be rapidly activated by a variety of stimulators; whether this state consists of different aggregations of the core units or whether it is created by the presence or absence of further factors is unknown.

The mechanisms involved in the mediation of peptide hormone activity, in particular in the modulation of gene expression, remain comparatively obscure. It is known that peptide hormones, such as insulin, EGF and prolactin, become internalised after receptor binding. When the receptor has been activated by hormone binding, it clusters with other hormone receptors (Schreiber et al., 1983) and is subsequently internalised via clathrin-coated pits (Harrison and Kirchhausen, 1983; Brown et al., 1983). Receptors may be internalised as a matter of course, in the absence of bound hormone (Brown et al., 1983), and subsequently recycled to the cell surface: the presence of ligand does, however, appear to accelerate the rate of internalisation (or retard the recycling process; Ciechanover et al., 1983) and may cause autophosphorylation of the receptor (Zick et al., 1983). The act of internalisation involves the transformation of the coated pit into a coated vesicle, and subsequently to an endosome vesicle. In this structure it is thought that the ligand is dissociated from its receptor, and the receptor recycled; the hormone may be subsequently degraded in lysosomes (Brown et al., 1983; see fig.1.7). Prolactin is known to be secreted in the milk of several species (Vonderhaar, 1987) At least part of this prolactin reaches the milk by the intracellular route, implying that some of the prolactin is taken up into the golgi rather than into lysosomes.

The binding of the hormone to its receptor is responsible for the perpetration of the hormonal effects (Shiu and Friesen, 1976). How this is achieved, however, remains unclear. It is known that peptide hormones release 'second messengers', such as cAMP, cGMP, Ca\(^{2+}\) ions, \textit{Ras} p21, polyamines and prostaglandins (Korn \textit{et al.}, 1987; Rillema \textit{et al.}, 1988), which are responsible for carrying out the effects ascribed to the hormone. The possible involvement of these mediators with milk protein gene expression has been reported (cGMP - Matusik and Rosen, 1980; spermidine - Rillema \textit{et al.}, 1977; prostaglandin F\(_{2\alpha}\) - Rillema, 1975), and it seems clear that prolactin differs from many peptide hormones and does not act by stimulating adenylate cyclase and raising cAMP levels. Its action may involve polyamines, cGMP and phosphorylation mechanisms (reviewed by Rillema \textit{et al.}, 1988).
Figure 1.6

Mechanism of steroid hormone action

Being lipid-soluble, steroid hormones can permeate through the cell membrane (1) to bind (2) to receptors which are localised in the cytoplasm of the cell. This receptor/hormone complex then translocates (3) to the nucleus, where gene activation by specific interaction with the chromatin (4) is believed to take place. Increased transcription (5) gives rise to the effect attributed to the steroid.
Steroid

1

Receptor

2

activated receptor

3

Nucleus

translation

mRNA

4

5

CHROMATIN

product
Figure 1.7
Mechanism of peptide hormone action

Peptide hormones are not lipid soluble, and thus are unable to cross the cell membrane save by binding to a membrane receptor. Complexed (activated) receptors are believed to cluster, and cause activation of a second messenger (X), possibly by phosphorylation. This second messenger is thought to migrate to the nucleus and mediate hormonal effects on gene expression.

The receptor/hormone complexes are internalised via 'coated pits' to form coated vesicles: the hormone then passes into a vesicle known as the endosome, and is subsequently either degraded in lysosomes or excreted from the cell via the golgi apparatus. The receptor is meanwhile recycled to the cell surface.
1.6.2 The mammary response to lactogenic hormones

Lactation is extensively regulated by the action of hormones; this is as evident in mammogenesis (see Topper and Freeman, 1980) as it is in lactogenesis (Tucker, 1974, 1988; Vonderhaar and Ziska, 1989). There is a consensus that lactation is mainly regulated by the concerted actions of prolactin, glucocorticoids, progesterone and insulin, with thyroid hormones and placental lactogen playing a part in some species (Topper and Freeman, 1980). The activity of each hormone is influenced by the others, and the sensitivity of mammary cells to different hormones varies relative to the advancement of pregnancy or lactation (Quirk et al., 1988). The individual activity of each hormone is thus difficult to differentiate from the integrated activity of the lactogenic hormones; this, paradoxically, makes it impossible to construct a coherent picture of this integrated control.

1.6.2.1 Prolactin

Many early experiments pointed to the main hormonal stimulation during lactation being due to the action of prolactin (Assairi et al., 1974), and in some species placental lactogen (Rosen et al., 1978), with glucocorticoid being required for maximal expression (Devinoy et al., 1978; Ganguly et al., 1979). It is likely that there is in fact an absolute requirement for glucocorticoids at least in mouse casein expression (Ganguly et al., 1980; Mehta et al., 1980), but that in some organ culture experiments a certain amount of glucocorticoid is retained within the tissue for several days (Mehta et al., 1980), thus confusing the results. Furthermore, rabbits, which are frequently used in the study of the effects of prolactin on lactation, are much more sensitive to the effects of prolactin than other mammals (Tucker, 1988). The inhibition of glucocorticoid receptor binding to the nuclear membrane attenuates the synthesis of casein, confirming the importance of the steroid hormone (Majumder et al., 1983). The only species in which prolactin appears to be the only lactogenic hormone is the tammar wallaby Macropus eugenii (Nicholas and Tyndale-Biscoe, 1985).

Prolactin has been shown to act in several ways on the expression of milk proteins. While inactive in duct formation, it is essential for terminal differentiation of the mammary alveoli during mammogenesis (Topper and Freeman, 1980) and promoting the development of the golgi apparatus in mammary cells (Ollivier-Bousquet, 1978). It is directly responsible for the
transcriptional and post-transcriptional regulation of casein mRNA in both rats (Guyette et al., 1979) and rabbits (Houdebine et al., 1978), raising the rate of transcription and stabilising the mRNA, and is essential for the synchronisation of expression of α-lactalbumin and caseins (Vonderhaar et al., 1973). Although the mechanisms by which prolactin acts remain obscure it is known to bind to a membrane receptor (section 1.1.1.2), and that this binding is functionally essential (Shiu and Friesen, 1976). Prolactin is known to stimulate the expression of its own receptors at physiological hormone concentrations, but also to repress receptor expression in response to the injection of large doses of hormone (Djiane et al., 1979). Prolactin receptors are positively modulated by glucocorticoids (Sakai and Banerjee, 1979) and negatively modulated by progesterone (Djiane and Durand, 1977), which may function by attenuating the positive influence of glucocorticoids (Sakai and Banerjee, 1979).

1.6.2.2 Progesterone

The inhibition of lactation by progesterone is well documented (see Topper and Freeman, 1980). High levels of progesterone during pregnancy are thought to attenuate lactation (Assairi et al., 1974) and prevent the accumulation of casein and α-lactalbumin mRNA (Houdebine and Gaye, 1975). The synthesis and secretion of rat α-lactalbumin are known to be inhibited by progesterone and synthetic progestins (Quirk et al., 1985). The mechanism of progesterone inhibition is, though, unclear. It has been suggested that progesterone could effectively compete for glucocorticoid receptors in the cytosol, and thus negate the stimulatory effects of glucocorticoid (Matusik and Rosen, 1978; Ganguly et al., 1982). This has been refuted by Terada et al., (1988), who showed that the inhibitory effects of progesterone on α-lactalbumin and casein expression in explant culture were strongest in the presence of prolactin and insulin alone, and attenuated by the addition of glucocorticoid. The same results were obtained by Sankaran and Topper (1988) and Jahn et al. (1989), who suggest that progesterone exerts its effects by inhibiting the prolactin stimulation of casein synthesis. There is evidence that displacing progesterone from its receptor removes the inhibitory effect (Jahn et al., 1987), and indeed that the binding of the progesterone agonist R5020 to the glucocorticoid receptor, which only takes place at very high R5020 concentrations, stimulates the production of casein (Jahn et al., 1989). As there is a rapid drop in progesterone receptors post-partum (Kuhn, 1977),
which is synchronous with the onset of lactation, a mechanism involving progesterone binding to its own receptor can be envisaged.

1.6.2.3 Glucocorticoids

As discussed in section 1.6.2.1, glucocorticoids are essential for the potentiation of the stimulatory activity of prolactin. They are also known to exhibit a biphasic behaviour with respect to mouse α-lactalbumin gene expression, being stimulatory at low levels but inhibitory at high levels (Ono and Oka, 1980a,b), while maintaining a stimulatory effect on casein gene expression up to a plateau level at high steroid concentration. This phenomenon may be reflected in the asynchronous expression of α-lactalbumin and casein in guinea pig mammary gland explants observed by Burditt et al. (1981). The appearance of α-lactalbumin mRNA two days pre-partum precedes that of casein mRNA, which appears at partuition. By eight hours post-partum, however, the level of α-lactalbumin mRNA has decreased two-fold, while that of casein is constant.

There is a parallel to be drawn here between the action of glucocorticoid, that is to potentiate the response to prolactin, and the effect of progesterone, which inhibits the response to prolactin. Jahn et al. (1987) suggest that the stimulatory activity of glucocorticoid observed in pregnant animals may be due to the blocking of the negative influence of progesterone; after the removal of progesterone with a specific antiprogestin, the stimulatory effect of glucocorticoid is much weaker. Quirk et al. (1988) observed that in the rat, inhibition of α-lactalbumin by both progesterone and high levels of glucocorticoid was lost two days pre-partum, before the disappearance of the progesterone receptors. This suggests that a common mechanism may be involved in mediating the effects of both steroids.

1.6.2.4 How do glucocorticoids and progesterone interact?

A theory which might explain the linked mechanisms of glucocorticoid and progesterone action is suggested by the observations of von der Ahe et al. (1985) and Strähle et al. (1987), that both the PR and the GR can bind to and mediate their effects through the same SRE. The observations of Quirk et al. support this postulation, and they advance a model in which the
activated progesterone and glucocorticoid receptors could compete for nuclear acceptor sites in the chromatin; the GRE/PRE.

Using the synthetic progestin ORG2058, Quirk et al. (1985) observed specific repression of rat α-lactalbumin over general protein synthesis, with half-maximal effects evident at ORG2058 doses one order of magnitude inferior to the dissociation constant (Kd) for the progestin/receptor complex. This manifests itself as a shift to the left of the progesterone dose/response curve with respect to the Kd for receptor occupancy, which, assuming that binding to one PRE is sufficient to inactivate the gene, suggests that there may be several PRE in the α-lactalbumin gene. Occupation of the GR by the progestin is ruled out, as the Kd for this complex is another order of magnitude higher than that for the progestin/PR complex.

In order to try to explain the biphasic response to glucocorticoid presented by α-lactalbumin, Quirk and Funder (1989) carried out similar experiments on the glucocorticoid receptor. They observe that the studies of Ono and Oka (1980a,b) involved the use of cortisol, which is non-physiologic in the mouse; it also displays a higher affinity for the mineralocorticoid receptor (type I) than for the glucocorticoid receptor (Type II), which could suggest a stimulation of α-lactalbumin synthesis by the activated type I receptor, and inhibition by the type II receptor. By using a highly type II specific ligand, they demonstrated that occupancy of type II receptors was initially stimulatory, but become rapidly inhibitory at steroid concentrations well below the receptor Kd. They propose a model for receptor binding to the GRE in the α-lactalbumin gene, whereby the occupancy of one GRE has a 'turn-on' effect, while the occupancy of any of the other sites has an overriding 'turn-off' activity. Occupancy of any of these GRE by a progesterone receptor would lead to a 'turn-off' event. The data they present is consistent with this theory, and suggests that the α-lactalbumin gene should have one positive GRE and four or five negative ones, which would explain the half-maximal inhibitory effects of glucocorticoid being observed at well below Kd for receptor occupancy (see also Quirk et al., 1985). The presence of six potential GRE is confirmed by analysis of the α-lactalbumin genomic sequences of the rat (Quasba and Safaya, 1980).

There are several observations which would suggest other means of interaction between steroid receptors. One is that estrogen receptors can mediate repression without binding DNA (Adler et al., 1989; section 1.2.6), possibly by direct interaction with transcription factors. This is
particularly attractive as both systems would appear to involve the interaction between steroid receptors and proteinaceous activators (pit-1 and the prolactin mediator). One can imagine that progesterone and glucocorticoids could compete to activate or repress the prolactin activation mediator. Another is the apparent competition by steroid receptors not for binding sites, but for factors that mediate their enhancer function (Meyer et al., 1989). It is observed that progesterone stimulation in HeLa cells was inhibited by the activated estrogen receptor (ER), without interaction between the two receptors or the ER and the DNA. The same effect was seen for ER inactivation of GR induced transcription, and ER induced transcription was in turn repressed by activated GR and PR. Any one of these models, plus perhaps the squelching model of Ptashne (see section 1.2.8) could account for the transcriptional interference observed; furthermore it should be borne in mind that prolactin acts in diverse ways, and therefore interaction with its effects may take place at more than one level.

1.6.2.5 Other hormones involved in α-lactalbumin gene expression

Insulin is an essential part of the hormonal mixture required for α-lactalbumin synthesis. It is known to be involved in mammary proliferation, and is believed to be involved in α-lactalbumin gene expression, possibly through regulation of gene transcription (reviewed by Meisler and Howard, 1989). IGF-1 (insulin-like growth factor 1) has been shown to be able to substitute for insulin in mammary cells (Prosser et al., 1987), with a reduction in efficacy.

Thyroid hormone L-T₃ (reviewed by Samuels et al., 1989) is also postulated to be involved in the control of α-lactalbumin synthesis, and has been shown to influence the binding of prolactin to prolactin receptors (Bhattacharya and Vonderhaar, 1979), perhaps by activating cryptic prolactin receptors (Vonderhaar, 1987), and to selectively enhance α-lactalbumin synthesis over casein synthesis (Vonderhaar, 1987).

Estrogen is necessary early in mammogenesis for the lactogenic response to prolactin and glucocorticoid to occur, and may be required during lactation itself (Tucker, 1988). EGF, on the other hand, while stimulating mammary proliferation, is probably inhibitory to the expression of milk proteins (Sankaran and Topper, 1984).

Of the peptide hormones belonging to the prolactin gene family, growth hormone and placental lactogen have also been implicated in the potentiation of lactogenesis, acting to mimic
the action of prolactin; the effects, however, are not consistent between species and may not be physiologically relevant (Bremel and Schuler, 1987).
Chapter 2
Materials and Methods

2.1 Materials

2.1.1 Routine chemicals and plasmids

All routine chemicals were obtained from sources already described (Laird et al., 1988), with the exceptions as listed below. The mouse β-casein cDNA clone pCMβ13 was the gift of Dr. J. Rosen, Baylor College of Medicine, Houston, Texas, USA. The mouse β-actin clone BS-C10 was the gift of Dr. R. Haffner, Chester Beatty Institute for Cancer Research, London. All other plasmids and all bacterial strains were sourced in the Medical Molecular Biology Unit, Middlesex Hospital Medical School, London, and were the gifts of the scientists working therein.

2.1.2 Enzymes

Restriction endonucleases were purchased from Biolabs (CP laboratories ltd., Bishops Stortford, GB), BRL (Bethesda, USA) or from Boehringer Corporation Limited, E. Sussex, GB. T7 RNA polymerase, E. coli DNA polymerase (Klenow fragment) and MMLV reverse transcriptase were purchased from Boehringer Corporation Limited, E. Sussex, GB.

2.1.3 Radiochemicals

5'α^{32}P dCTP (3000μCi/mmol), 5'α^{32}P-ATP (3000 Ci/mmol) and 5'γ^{32}P ATP (3000 Ci/mmol) were purchased from NEN (New England Nuclear), Dreieich, FRG.

2.1.4 Kits

The Geneclean kit was obtained either direct from Bio 101, La Jolla, California, USA, or through Stratech Scientific, GB. The Sequenase sequencing kit, version
2.0, was obtained from United States Biochemical, Cambridge, GB. The Pierce BCA protein assay kit was the gift of Dr. A. Moore, M.H.M.S.

2.1.5 Antisera

Rabbit anti-guinea pig α-lactalbumin antiserum was the gift of Dr. A. Moore, Medical Molecular Biology Unit, M.H.M.S., London. The goat anti-rabbit IgG was sourced from DAKO immunoglobulins, Denmark, and was also the gift of Dr. Moore.

2.1.6 Other materials

Hybond-N and -C hybridisation membranes were obtained from Amersham International, Aylesbury, GB. Schleicher and Schuell nitrocellulose membranes were the gift of Dr. A. Moore, M.H.M.S. Gel running and blotting equipment was obtained from BRL, Bethesda, USA, or from Bio-Rad, Watford, GB. All agarose was sourced from BRL, Bethesda, USA. Autoradiographic film was obtained from Fuji film, London (U.K.) or from Kodak Ltd., Hemel Hempstead, Herts. (UK).

2.1.7 Solutions

40% Acrylamide for sequencing: 38% (w/v) acrylamide, electrophoresis grade (Analar); 2% bisacrylamide, electrophoresis grade (Analar)

100x Denhardt's: 2% BSA; 2% polyvinylpyrrolidone; 2% ficoll, all w/v

SSC 15mM sodium citrate, 150mM sodium chloride, pH7.0

TE 10mM Tris-Cl, pH8.0, 1mM EDTA
<table>
<thead>
<tr>
<th>Buffer/medium</th>
<th>Composition/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBE (agarose gels)</td>
<td>90mM Tris-90mM boric acid, pH 8.35, 1mM EDTA.</td>
</tr>
<tr>
<td>TBE (sequencing gels)</td>
<td>100mM Tris-100mM boric acid, pH 8.3, 20mM EDTA.</td>
</tr>
<tr>
<td>TAE (all gels)</td>
<td>100mM Tris-Ac, pH 7.5, 20mM EDTA.</td>
</tr>
<tr>
<td>TBS</td>
<td>10mM Tris-Cl, pH 7.4; 0.9% NaCl</td>
</tr>
<tr>
<td>MEA</td>
<td>200mM MOPS, pH 7.0, 10mM EDTA, 50mM NaAc.</td>
</tr>
<tr>
<td>ψ-b broth</td>
<td>2% Difco Bacto Tryptone, 0.5% Difco Bacto Yeast Extract, 0.4% MgSO₄, 10mM KCl</td>
</tr>
<tr>
<td>TFB 1</td>
<td>100mM RbCl, 50mM MnCl₂, 30mM KAc, 10mM CaCl₂, 15% v/v glycerol, pH 5.8 with acetic acid, filter sterilise</td>
</tr>
<tr>
<td>TFB 2</td>
<td>10mM MOPS, pH 7.0 with NaOH, 10mM RbCl, 75mM CaCl₂, 15% v/v glycerol, filter sterilise</td>
</tr>
</tbody>
</table>
Oligo-labelling buffer:

Solution O 1.25M Tris-Cl, pH8.0, 0.125M MgCl₂
Solution A 1 ml Solution O, 18μl β-mercaptoethanol, 5μl 0.1M dATP, dGTP, dTTP.
Solution B 2M HEPES, pH6.6
Solution C Random primers (d(NTP))₆, 90 O.D./ml, in TE
OLB A:B:C in ratio 100:250:150

Media for handling mouse oocytes for microinjection:

Stock solutions: 10xA 5.534g NaCl, 0.356g KCl, 0.162g KH₂PO₄, 0.293g MgSO₄.7H₂O, 2.61g Na lactate, 1.0g glucose, 0.06g penicillin, 0.05g steptomycin
Total volume 100ml.

10xB 2.101g NaHCO₃, 0.01g phenol red in 100ml

100xC 0.036g Na pyruvate in 10ml

100xD 0.252g CaCl₂ in 10ml

10xE 5.958g HEPES, pH7.4, 0.252g phenol red, in 100ml
M2 and M16 media:

<table>
<thead>
<tr>
<th>Stock</th>
<th>M2</th>
<th>M16</th>
</tr>
</thead>
<tbody>
<tr>
<td>10xA</td>
<td>10ml</td>
<td>10ml</td>
</tr>
<tr>
<td>10xB</td>
<td>1.6ml</td>
<td>1.6ml</td>
</tr>
<tr>
<td>100xC</td>
<td>1.0ml</td>
<td>1.0ml</td>
</tr>
<tr>
<td>100xD</td>
<td>1.0ml</td>
<td>1.0ml</td>
</tr>
<tr>
<td>10xE</td>
<td>8.4ml</td>
<td>-</td>
</tr>
<tr>
<td>D.I. water</td>
<td>78ml</td>
<td>78ml</td>
</tr>
<tr>
<td>BSA</td>
<td>400mg</td>
<td>400mg</td>
</tr>
</tbody>
</table>

Total volume 100ml

Filter sterilise through 0.45μm millipore filter, and store at 4°C. Replace every week.

2.2 Methods

2.2.1 General procedures

All solutions were made up using deionised (D.I) water, obtained by the treatment of reverse osmosis (R.O.) water supplied from a central Middlesex Hospital source with an ELGA water purification apparatus. Routine procedures were performed as described by Maniatis et al. (1982).

2.2.2 Restriction enzyme analysis

All endonuclease digestions were carried out in the restriction buffer supplied by the enzyme manufacturer, at the suggested temperature. Genomic digests were carried out overnight, while plasmid digests were left for 2 to 3 hours. Agarose gels were made up to the indicated percentages of agarose in either TBE or TAE buffers, and run according to established procedure (Maniatis et al., 1982).
2.2.3 Transformation of bacterial cells

*E. coli* JM83 was transformed according to a protocol provided by M. Scott, UCSF, USA. A single colony of JM83 was picked into 5ml *\varphi*b, and grown at 37°C until OD\textsubscript{550} reached 0.3. This culture was used to inoculate 100ml of prewarmed *\varphi*b, which was shaken at 37°C until OD\textsubscript{550} was 0.48. The bacteria were then chilled on ice, and centrifuged for 5 minutes at 2500rpm in a pre-cooled rotor, 4°C. At this stage all work was transferred to the cold room. Cells were taken up in 30ml ice-cold T\textsubscript{FB} 1, and left on ice for 40 minutes, before being sedimented at 2500rpm, 4°C. The cells were then gently resuspended in 4ml T\textsubscript{FB} 2, dispensed into pre-cooled eppendorf tubes in 100\textmu l aliquots, and snap-frozen in a dry ice-ethanol bath. Frozen cells were stored at -70°C.

Prepared competent cells were transformed by thawing on ice for 1 hour, and then adding 10ng DNA in ligase buffer. After incubation on ice for 20 minutes, cells were heat-shocked at 42°C for 2 minutes, chilled on ice for 2 minutes, and incubated at 37°C with 4 volumes *\varphi*b for 1 hour. The transformed bacteria were then plated out on *\varphi*a (*\varphi*b + 12g/L Difco agar) agar plates, containing 100\mu g/ml ampicillin, and spread with 50\mu l 25\mu g/ml X-gal.

2.2.4 Preparation of plasmid DNA

2.2.4a. Large-scale plasmid preparation

Large scale preparations were carried out according to the alkali lysis method of Maniatis *et al.* (1982), with the following modifications: cells were grown overnight in 1L cultures of *\varphi*b containing 10\mu g/ml ampicillin. After centrifugation for 20 minutes at 3000rpm (4°C) to sediment the bacteria, the medium was poured off and the cells resuspended in 40ml 50mM glucose, 25mM Tris-Cl, pH8.0, 10mM EDTA. The bacteria were lysed by the addition of 80ml 0.2M NaOH, 1% (w/v) SDS and chromosomal DNA was precipitated with 40ml 3M KAc, pH4.8 with acetic acid. The precipitate was sedimented by centrifugation at room temperature in a Sorvall GS-3 rotor at 9000rpm for 40 minutes, and the supernatant was filtered
through nylon gauze into 500ml centrifuge buckets. Plasmid DNA was precipitated by addition of 0.6vol isopropanol and immediate centrifugation at 6000rpm, 10 minutes, room temperature. The DNA was resuspended in 7ml 0.1M Tris-Cl, pH8.0, 2mM EDTA, and purified by isopycnic banding in CsCl (Katz et.al., 1973). DNA preparations were resuspended in 1x SSC, and treated with boiled RNAse A/T1 to remove RNA contamination; the plasmid was then phenol extracted and ethanol precipitated. Concentration were determined by UV absorbance spectrophotometry at 260nm.

2.2.4b. Small scale preparations

Minipreparations of plasmid DNA were carried out by the alkali lysis method described by Maniatis et.al. (1982), with the omission of the lysozyme lysis step.

2.2.5 Isolation of restriction fragments

2.2.5a. From agarose gels

DNA fragments were excised from low melting-point or standard agarose gels by excision of the UV-visualised band, taking care to minimise exposure of the DNA to UV light. DNA was extracted from agarose using the Geneclean kit, exactly as described by the manufacturer.

2.2.5b. From polyacrylamide gels

Polyacrylamide gels were prepared to the desired specification as described by Maniatis et.al. (1982). Radiolabelled DNA fragments were excised from polyacrylamide gels according to the NH₄Ac elution procedure of Maxam and Gilbert (1980).

2.2.6 Polyacrylamide gels

2.2.6a. Sequencing gels

6% polyacrylamide/urea sequencing gels were prepared by mixing 18g urea, 5.4ml 40% 19:1 acrylamide:bisacrylamide solution, 3.6ml 10x sequencing
TBE, 13.8ml water and 220 µl 22.8%(w/v) ammonium persulphate. The urea was dissolved by heating in the microwave oven for 5 to 10 seconds, the gel mix was degassed for 1 minute at a vacuum pump, and 8µl N,N,N,N-Tetramethylethylenediamine (TEMED) was added. The gel was poured into a taped glass mould, 400mm x 200mm x 0.3mm, with a syringe, and allowed to set for upwards of 1 hour. Samples were denatured before loading by dissolution in 80% (v/v) formamide, 1mM EDTA, 1mg/ml bromophenol blue, 1mg/ml xylene cyanol and heating to 95°C for 4 minutes. The samples were then loaded directly onto the gel, and electrophoresed at 30w until the dye markers had reached the desired position.

2.2.6b. Native polyacrylamide gels

Non-denaturing gels were prepared to the required percentage using 40% acrylamide solution as used in sequencing gels, 720µl 50x TAE buffer, water and 300µl 22.8%(w/v) ammonium persulphate. Samples were loaded directly onto the gel in agarose gel loading buffer (5% glycerol, 2mM EDTA pH7.0, 0.5mg/ml bromophenol blue) and electrophoresed for the required time at 20 to 25w.

2.2.7 Preparation of genomic DNA from transgenic mouse tails

10mm of tail was removed under CO₂ anaesthesia from each newborn mouse at 10 days of age. The tail piece was placed into an eppendorf tube containing 700µl of Tail Buffer (50mM Tris.Cl, pH 8.0,100mM EDTA,100mM NaCl,1% SDS). 25µl of 10mg/ml proteinase K was added, and the tubes shaken at 55°C overnight in an orbital incubator. An equal volume of phenol/ chloroform was added and the tubes shaken gently, by hand, for 15 minutes, before being centrifuged in an IEC Centra-M microcentrifuge for 20 minutes. The aqueous phase was saved to a fresh tube, and the extraction repeated, followed by extraction with an equal volume of chloroform. 0.6 vol of isopropanol was then added and the tube shaken violently. The DNA now becomes visible as a clump, and can be hooked out. The DNA was washed in 4 washes of 70% ethanol and
once in 100% ethanol before being air-dried and left to resuspend in 50μl of water for a few minutes.

2.2.8 Preparation of total RNA

Total RNA was prepared from various tissues by a modification of the method of Chomczynski and Sacchi (1987). Frozen tissue pieces, weighing 0.1 to 0.2g, were placed whole into a chilled glass/teflon homogeniser, and 0.5ml guanidinium lysis solution (solution D; 4M guanidinium isothiocyanate, 25mM sodium citrate, pH7.0, 0.5% sarcosyl, 0.1M β-mercaptoethanol) was added immediately. The tissue pieces were gradually encouraged to thaw and simultaneously homogenised by gentle, and then more violent, action of the homogeniser pestle. This technique ensures that the tissue is disrupted into guanidinium buffer as soon as it is thawed, and RNA degrading enzymes are instantly disabled. The homogenate was transferred to a 1.5ml Eppendorf tube, and NaAc pH 4.0 was added to 0.2M, followed by 0.5ml of water-saturated phenol and 100μl of water-saturated 49:1 chloroform:isoamyl alcohol mixture. After incubation on ice for 15 minutes, the tubes were centrifuged in an Eppendorf microfuge in the cold room (4°C). The upper phase was removed, and re-extracted with phenol and chloroform as above. RNA was pelleted by the addition of 1 volume of isopropanol and incubation at -20°C for at least 1 hour, followed by centrifugation for 10 minutes at 4°C in a microfuge. The pellets were vacuum desiccated for 15 minutes and resuspended in 100μl 0.5% SDS. Sample concentrations were determined by spectrophotometry at 260nm.
2.2.9 Radiolabelling of DNA

2.2.9a. Oligo-labelling of DNA probes

Linearised DNA fragments and circular plasmids were labelled essentially according to the oligo-labelling technique of Feinberg and Vogelstein (1984). 50ng of purified DNA, free from agarose, was heated to 95°C for 4 minutes, in the appropriate volume of water in a sealed eppendorf tube. After chilling on ice for 5 minutes, 5μl oligo-labelling buffer (OLB) was added, followed by 2.4μl α32P dCTP (3000 Ci/mmol) and 1μl Klenow polymerase (2u/μl), to a total volume of 25μl. The mixture was incubated for 14-16 hours at room temperature, and stopped by the addition of 75μl 100mM EDTA, pH8.0. Unincorporated radiolabel was removed by centrifugation on a Sephadex G50 spun column, and specific activity of the probe was determined by counting 1μl of the eluate in a liquid scintillation counter, using Ecoscint scintillant. Probes were denatured just before use by heating to 95°C for 5 minutes, and immediately chilling on ice for 5 minutes.

2.2.9b. MMTV RT end-labelling of DNA

Linear DNA fragments with 5' protrusions containing a G residue were end-labelled with α32P dCTP and MMTV reverse transcriptase. 5μl dCTP was added to 100ng purified DNA, 5μl 0.1mM dATP, dGTP, dTTP and 1u enzyme, using the buffer provided by the manufacturer in a total volume of 25μl. The reaction was allowed to proceed for 1 hour at 37°C, and stopped by phenol extraction of the DNA and removal of unincorporated label by the method described in 2.2.5a.
2.2.10 Southern blot analysis

Digested DNA was separated on 0.8% agarose/TAE gels and was denatured by immersing the gels in a gently shaken tray of 0.5M tris, 1.5M NaOH for 2x 20 minutes. Neutralisation was carried out similarly, using a solution of 0.5M Tris-Cl, pH 7.0, 2M NaCl. The neutralisation was continued until the pH of this solution remained neutral after contact with the gel. Capillary transfer was performed essentially as described by Southern (1975) onto Amersham Hybond-N nylon membranes. The DNA was fixed to the membrane by UV cross-linking, and the membrane was washed free of agarose in 2x SSC, 0.1% SDS, at room temperature for 20 minutes. Prehybridisation was carried out for at least three hours, but less than six hours, in freshly prepared 10% dextran sulphate, 6x SSC, 5x denhardt’s solution and 0.1% SDS, containing 250µg/ml freshly denatured herring testis DNA. This solution was discarded before addition of the hybridisation solution, identical except for the addition of approximately 5x10^7 cpm freshly denatured DNA probe (2.2.5a). Hybridisation was allowed to proceed overnight, and the blot was subsequently washed in 500ml 2x SSC, 0.1% SDS at 65°C for 2x 10 minutes. Dextran sulphate clinging to the nylon membrane was removed by gentle rubbing with a gloved hand. Blots were finally washed to the desired stringency, without SDS, air dried and autoradiographed at -70°C between screens.

2.2.11 Northern Blot analysis

Total RNA was precipitated in 0.3M NaAc with 2.5 volumes of cold ethanol, incubating at -20°C for 1 hour. The RNA was resuspended in 30µl Northern dissolving buffer (1x MEA, 50% formamide, 6.8% formaldehyde) by vortexing for 10 seconds and heating to 65°C for 5 minutes. 5µl of agarose gel loading buffer was added, and the samples were mixed thoroughly before being loaded onto an MEA buffered 1.6% agarose/6.8% formaldehyde gel. Gels were run either overnight at 40v or for 3 to 4 hours at 150-200v. The RNA was visualised on the gel by ethidium bromide fluorescence, and photographed. The gel was then
soaked in 20x SSC for 20 minutes before being capillary blotted to Amersham Hybond-N nylon membranes as described by Southern (1975). Blots were fixed by UV irradiation for 5 minutes, and washed in 2x SSC, 0.5% SDS, for 20 minutes at room temperature.

Hybridisation and prehybridisation solution was prepared as follows; 5g of dextran sulphate and 3.5g SDS were weighed into a glass bottle, and dissolved in 12.5ml 20x SSC, 5ml 100x denhardt's and water up to 50ml, by stirring on a magnetic stirrer and heating to 65°C in a water bath. Blots were prehybridised for 3 to 6 hours at 65°C, shaking, in 20ml prehybridisation solution, with the addition of 250μg/ml freshly denatured herring testis DNA. The prehybridisation solution was replaced with hybridisation solution, identical except for the addition of approximately 5x10^7 cpm freshly denatured DNA probe. After overnight incubation in a shaking water bath at 65°C, blots were washed in 500ml 2x SSC, no SDS, at 65°C. Dextran sulphate was removed by rubbing with a gloved hand. The blots were washed once more in 500ml 2x SSC, 0.1% SDS, at 65°C for 10 minutes, and then stringently washed in 0.15x SSC, no SDS, at 65°C for 20 minutes. The blots were subsequently air dried, and autoradiographed between screens at -70°C.

2.2.12 S1 nuclease analysis of total RNA

A modification of the method of Berk and Sharp (1977) was used, as follows: cDNA probes were labelled with T4 polynucleotide kinase and γ-32P ATP as described by Maniatis et.al. (1982). 10^4cpm of DNA probe was coprecipitated with 20μg of total RNA by addition of NaAc to 0.3M, 2.5 volumes of cold ethanol and incubation at 20°C for 1 hour. The nucleic acid was precipitated by centrifugation in a microfuge for 10 minutes, and the pellet was vacuum desiccated before being resuspended in 20μl hybridisation buffer (80%(w/v) deionised formamide, 50mM PIPES, pH6.4, 400mM NaCl, 1mM EDTA). This was pipetted into an eppendorf tube, and sealed using a Maxam and Gilbert apparatus. Samples were heated to 90°C in a water bath for 30 minutes, and then
incubated overnight in a dry-air shaking incubator at 52°C. While still at 52°C, 300μl of S1 digestion buffer (30mM NaAc, pH4.4, 280mM NaCl, 4.5mM ZnSO₄ and 100u S1 nuclease) was added to each tube, the samples were vortexed and incubated at 21°C for 2 hours. The undigested DNA/RNA hybrid molecules were phenol extracted, ethanol precipitated, taken up in sequencing gel loading buffer, and electrophoresed on a 6% polyacrylamide-urea sequencing gel (Maniatis et al., 1982) until the bromophenol blue had reached the end of the gel.

2.2.13 Maizel protein gels

SDS-PAGE was performed according to the method of Maizel (1972). The gel was cast as 2 layers, the resolving gel and the spacer gel. For a 10% resolving gel, 16.6ml solution A (30g acrylamide (Analar), 0.8g bisacrylamide (Analar), water to 100ml) was added to 5ml resolving gel buffer (1M Tris-CI, pH8.9) 400μl 10% SDS, 270μl 1% (w/v) ammonium persulphate, 17.7ml D.I. water and degassed at a vacuum pump. 65μl TEMED was added, and the gel poured into a prepared gel mould (BRL) to within 1cm of the bottom of the comb teeth. Butanol was layered on top of the gel to prevent the formation of air bubbles at the top of the resolving gel.

Spacer gel (1ml solution A, 1.25ml 166mM Tris-CI, pH6.7, 0.1ml 10% SDS, 65μl ammonium persulphate, 7.6ml water, 12μl TEMED) was added on top of the resolving gel once the resolving gel had set. The comb was inserted into the spacer gel, and the gel allowed to set.

Protein samples were prepared by freeze-drying and resuspending in 25μl Maizel sample buffer (1ml β-mercaptoethanol, 1g SDS, 20ml glycerol, 10ml 166M Tris-CI, pH6.7, 0.4g bromophenol blue, water up to 100ml) and heating to 75°C for 5 minutes. After heating, samples were loaded directly onto the gel.

Gels were run in electrode buffer (6g Tris-base, 28.8g glycine, 10ml 10% SDS, water to 1L) at 14mA for 20 minutes, then 40mA for 2 to 3 hours.
2.2.14 Western blotting

Maizel gels were removed from the glass moulds, and soaked in blot buffer (25mM Tris-Cl, pH8.3, 192mM glycine, 20% methanol). Amersham Hybond-C or Schleicher and Schuell nitrocellulose filters were cut to size, and wetted in blot buffer. The filter was placed onto the gel, and filter+gel were sandwiched between sheets of Whatman 3mm paper, also soaked in blot buffer. The whole sandwich was placed between the plexiglas supports of a Bio-Rad Transblot blotting cell filled with blot buffer, with the gel proximal to the negative terminal. A current of 210mA was applied overnight.

Blots were blocked by incubation in 2%BSA in TBS for 1 hour at room temperature. The rabbit anti-guinea pig α-lactalbumin antiserum was then applied at a 100x dilution in TBS, 4°C, overnight. The blots were washed in 100ml volumes of TBS+0.05% NP40, 2x 10 minutes, then TBS, 2x 5 minutes, shaking, at room temperature. The second antibody, a goat anti-rabbit IgG obtained from Daco was applied at a 400x dilution in TBS for 2 hours, shaking, at room temperature. The blot was washed as before, rinsed in D.I. water, and developed in 30mg 4-chloro-1-napthol dissolved in 10ml ice-cold methanol, supplemented with 30μl 30% H_{2}O_{2} in 50ml TBS. Colour takes between 15 minutes and several hours to develop. Once the bands had developed, the blot was placed in D.I. water, and photographed as soon as possible.

2.2.15 In Vitro transcription of α-lactalbumin complementary RNA

2.2.15a. Labelling transcription

Riboprobe plasmid pgpK98 was linearised with Eco R1, purified by phenol extraction and ethanol precipitation, and resuspended in water at a concentration of 1μg/μl. 2μg of linear template was mixed with 5μl commercial T7 RNA polymerase buffer, 5μl 100mM DTT, 25μl RNasin, 5μl 5x nucleotide mix (2.5mM UTP, GTP, CTP), 5μl 35S-ATP and 1u of T7 RNA polymerase in a total volume of 25μl; the components were mixed at room temperature to prevent precipitation
of template DNA by spermidine in the buffer, and incubated at 37°C for 80 minutes. 2u of RNase-free DNase was then added, together with 25u RNasin, and incubation was continued for a further 10 minutes at 37°C. After incubation, the volume was made up to 200μl, 10mM with respect to DTT and 2M with respect to NH₄Ac. At this stage 2μl was counted by TCA precipitation and liquid scintillation, to determine the degree of incorporation. To the remaining mixture, 20μl of E.coli tRNA was added as a carrier and the RNA was precipitated by the addition of 2.5 volumes cold ethanol and incubation at -20°C for 1 hour. The RNA pellet was resuspended in 10mM DTT, 0.3M NaAc, and reprecipitated before being vacuum desiccated to dryness and resuspended in the required volume of 10mM DTT.

2.2.15b. Cold transcription

Cold transcriptions were set up essentially as described for labelling transcriptions in 2.2.12a, except for the incorporation of 2.5mM ATP in the 5x nucleotide mix, and the omission of 32P-ATP. The entire reaction products were examined by agarose gel electrophoresis after incubation for 60 minutes at 37°C.

2.2.16 Preparation of slides for in situ hybridisation sections

Slides to be used in ISH were handled with washed gloves at all times. New, unused microscope slides were placed in a stainless steel slide rack, washed in several changes of Teepol detergent in tap water, and left to rinse under the tap for 1 hour. They were then rinsed thoroughly in D.I. water, and dried by dipping in 100% ethanol and air-drying. Dry slides, still in their racks, were wrapped in aluminium foil and baked at 180°C overnight. Just before use, slides were coated by drawing a thin layer of 5%poly-L-lysine in PBS on one surface of the slide, and allowing it to dry.
2.2.17 *In Situ* hybridisation protocol

The following protocol was adapted from that generously provided by Ms. Helen Parkes, Medical Molecular Biology Unit, M.H.M.S, London, according to the advice of Mr. Neil Rayment, department of Histopathology, M.H.M.S, London.

Slides must be dry, and preferably at least 1 week old, before commencing.

**Step 1 Dewax**

- **Xylene 1**  
  15 Seconds
- **Xylene 2**  
  15 seconds
- **100% ethanol 1**  
  15 seconds
- **100% ethanol 2**  
  15 seconds
- **95% ethanol**  
  15 seconds
- **70% ethanol**  
  15 seconds
- **50% ethanol**  
  15 seconds
- **D.I. water**  
  5 minutes
Step 2. Prehybridisation:

PBS
0.1M glycine in PBS
0.3% Triton X-100 in PBS
PBS
PBS
PBS
1mg/ml proteinase K in
0.1M Tris-Cl, pH 7.5, 50mM EDTA, 37°C
4% paraformaldehyde in PBS
PBS
PBS
0.25% acetic anhydride in triethanolamine
50% formamide, 2x SSC
37°C

Step 3. Hybridisation:

20μl hybridisation solution, plus 2x10^6cpm ^35S-labelled probe, are placed on a coverslip washed as in 2.2.13. The prehybridised slide is wiped dry with a tissue, without disturbing the section, and placed section-down onto the hybridisation solution. Slides covered with hybridisation solution and coverslips are incubated in a humid chamber (a plastic box lined with Whatman 3mm soaked in 2x SSC, 5% formamide) at 43°C overnight.
Step 4. Washing:

immerse slides in 4x SSC, and allow cover slips to float off. Washing was performed as follows:

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x SSC, 37°C</td>
<td>20 minutes</td>
</tr>
<tr>
<td>4x SSC, 37°C</td>
<td>20 minutes</td>
</tr>
<tr>
<td>4x SSC, 37°C</td>
<td>20 minutes</td>
</tr>
<tr>
<td>20mg/ml RNase A in 10mM</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Tris-Cl, pH7.5, 0.5M NaCl, 1mM EDTA 37°C</td>
<td></td>
</tr>
<tr>
<td>2x SSC, shaking, 55°C</td>
<td>30 minutes</td>
</tr>
<tr>
<td>0.1x SSC, 55°C</td>
<td>30 minutes</td>
</tr>
</tbody>
</table>

The sections were then dehydrated, and air dried:

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>70% ethanol, 0.3M NH₄Ac</td>
<td>2 minutes</td>
</tr>
<tr>
<td>95% ethanol, 0.3M NH₄Ac</td>
<td>2 minutes</td>
</tr>
<tr>
<td>100% ethanol, 0.3M NH₄Ac</td>
<td>2 minutes</td>
</tr>
<tr>
<td>100% ethanol, 0.3M NH₄Ac</td>
<td>2 minutes</td>
</tr>
<tr>
<td>air dry</td>
<td>2 hours</td>
</tr>
</tbody>
</table>
Slides were subsequently autoradiographed by dipping in 50% Ilford K5 gel-form emulsion, 2% glycerol, air-drying for 2 hours and exposing at 4°C in a light-proof slide box. After 1 week, slides were developed at room temperature as follows:

- **Kodak D-19 developer** 2.5 minutes
- 1% acetic acid atop bath 30 seconds
- Ilford Amfix fixer > 5 minutes
- D.I. water 15 minutes
  (several changes)

**Step 5. Staining:**

Developed slides were stained with hematoxylin and eosin, as follows:

- Immerse in Haematoxylin, use undiluted
- rinse in running tap water until stain becomes blue
- dip in acid alcohol (1% conc. HCL in 70% ethanol) 3 or 4 times: any stained emulsion should clear
- Immerse in undiluted eosin 2 to 3 minutes
- wash in running tap water 2 to 3 minutes
- 70% ethanol 2 minutes
- 90% ethanol 2 minutes
- 100% ethanol 2 minutes, 2 changes
- Immerse in Histoclear 2 minutes, 2 changes

Slides were subsequently dried with a tissue, and sections were mounted in Histomount mounting medium.
Chapter 3

The structure of the \( \alpha \)-lactalbumin gene: regulatory sequence elements

Introduction

3.1 The genomic organisation of the guinea pig \( \alpha \)-lactalbumin gene

The structure of \( \alpha \)-lactalbumin genes appears to be very well conserved within exon sequences (see Hall and Campbell, 1986; Hurley and Schuler, 1987). The transcription unit spans 1.8kb, and encodes a mRNA of 750 bases split into four exons (fig. 3.1), separated by three introns. The positions of intron-exon boundaries are conserved, at amino acid residues Leu-26 (intron 1), Lys-79 (intron 2), and Trp-104 (intron 3), indicating that the ancestral gene was established prior to the divergence of species (see Hall et al., 1987). The sizes of the introns themselves are consistent, giving due allowance for the insertion of a species-specific Alu repetitive sequence in the first intron of the human gene (Hall et al., 1987; see also Laird et al., 1988). The sequences of the introns are much less conserved than those of the coding regions, as would be expected given the age of the postulated ancestral gene.

The most striking similarities outside the coding region occur in the 5' flanking sequences, where the promoters of the different \( \alpha \)-lactalbumin genes appear to consist of the same promoter elements. The TATA box, present in most eukaryotic genes (see section 1.2.1) is centred at -27 in the guinea pig, and -25 in the human. Both these genes lack the CAAT consensus and the GC box (Benoist and Chambon, 1981; Laird, 1985) but possess a 30bp sequence, RGAAGGRAA(N)TGGACAGAAATCAA(CG)TTTCTA, which is found in the promoters of the three rat casein genes, the bovine \( \alpha \)-casein gene, the guinea-pig \( \alpha \)-casein gene, the human \( \alpha \)-lactalbumin gene and the rat \( \alpha \)-lactalbumin gene (Hall et al., 1987), and the guinea pig \( \alpha \)-lactalbumin gene (Laird et al., 1988). This sequence, called the 'milk box', is a potential mammary-specific control element for milk protein genes.
Additionally, Laird (1985) identified the presence of a consensus TTCCTA/GA in the upstream region, whose function remains unknown; an SV40 enhancer core sequence (G)TGGATA/TAGA(TG) (Weiher et al., 1983) at -962, which, due to the presence of a guinea pig LINE (Long Interspersed repeated sequence; Singer and Skowronski, 1985) at -512 (Laird 1985), is probably inactive in the regulation of the guinea pig α-lactalbumin gene; and the GRE core TGTGA/CT was noted to occur twice in the upstream region.

Protein-DNA interactions are known to be of central importance in the regulation of gene expression (see section 1.2), and it is therefore useful to examine the sequence of the gene, including both the 5' and 3' flanking regions, to assess the likely importance of sequence elements; this would aid the subsequent construction of mutated genes which would allow the functional assay of such sequences.

In recent years many more regulatory sequence elements have been identified in a number of genes; and techniques allowing the detection of specific protein/DNA interactions have been developed (Varshavsky, 1987; Hennighausen and Lubon, 1987). In the following chapter, the sequence of the guinea pig α-lactalbumin gene is reexamined using such techniques, in the light of more recent data concerning the specific interactions between proteins and DNA.
Exon structure of the human, guinea pig and rat α-lactalbumin genes

The boxes, numbered 1 to 4, indicate exons. The exons share a close homology at the sequence level (Hall et al., 1987; Laird et al., 1988), but the introns are of variable length (Hall and Campbell, 1986). This is partly attributable to the insertion of repetitive sequences into the introns: an Alu sequence is found in the human intron 1, and the rat contains a rodent B2 repeat in intron 3. The guinea pig gene has a LINE sequence upstream of the promoter region, at -512.

In addition, the positions of repeating di- or trinucleotides present within the rat genome are indicated: a, (TCC)_{23}; b, (TG)_{24}; c, (TAT)_{18}; d, (TG)_{21}. 

83
3.2 The position of regulatory consensus sequences in the guinea pig α-lactalbumin gene

A list of consensus sequences identified in a variety of genes has been compiled by Wingender (1985). These sequences were collated as a search file for the Beckman Microgenie sequence analysis software, and the guinea pig α-lactalbumin genomic sequence was analysed (fig. 3.2). Two sites for nuclear factor 1 (CTF/NF1; Jones et al., 1984) were identified in the distal upstream sequences, plus three GRE core sequences upstream of the cap site, one in the coding sequence, and two in the first intron. None of these GRE cores are in the context of one of the 15bp homologies observed for other genes, for example in MMTV (mouse mammary tumour virus) (Payvar et al., 1983) and in the human metallothionein gene (Karin et al., 1984), so their possible role in transcriptional regulation is hypothetical. The TTCCT^A/GGA homology identified by Laird (1985) does not identify with any consensus sequences characterised to date, and neither does the milk box.

Analysis of the 3′ untranslated region (3′ UTR) revealed the presence of conserved sequence repeats, associated with possible stem-loop structures, in the guinea pig, rat and human α-lactalbumin genes (Fig. 3.3). Müllner et al. (1988, 1989) have characterised stem-loop structures and repeated palindromic sequences as being involved in the regulation of mRNA stability of the human transferrin receptor; as it is known that milk protein mRNA is regulated through stability modulation (Guyette et al., 1979), such a possibility is attractive.
Figure 3.2

Consensus regulatory sequences present at the 5' end of the guinea pig α-lactalbumin gene

The nucleotide sequence of the guinea pig α-lactalbumin promoter, exon 1 and intron 1 is shown; consensus sequences identified in the promoter and intron 1 of the guinea pig α-lactalbumin gene are boxed.

- GRE glucocorticoid responsive element core
- NF1 nuclear factor 1 binding site
- TATA TATA box consensus
- CAP mRNA initiation and capping site
- Met first methionine of α-lactalbumin peptide
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>AAGCTTTTTTA</td>
<td>ACAAGATGAG</td>
<td>CTTCCTCTTG</td>
<td>TTGATTTTTT</td>
<td>GGAGGTCTTG</td>
<td>CTTAGGTTTA</td>
<td></td>
</tr>
<tr>
<td>GAAGTTTATCT</td>
<td>CATGAAGTCT</td>
<td>TTTGCTTATA</td>
<td>CCAATGTCTT</td>
<td>CAAGAGTCTT</td>
<td>ACGTACTTA</td>
<td></td>
</tr>
<tr>
<td>TTCTCCAGCA</td>
<td>GTTTTAACAT</td>
<td>TTCTGTTTTTA</td>
<td>ATACATTAGGT</td>
<td>TTTGATTCTT</td>
<td>TTCAATTTTTA</td>
<td></td>
</tr>
<tr>
<td>GTTTAGAGTG</td>
<td>TGTGAAGAGA</td>
<td>TGTGCGTTATA</td>
<td>ATTTTTAACT</td>
<td>TCGCATGTTG</td>
<td>GAAAGCCTAT</td>
<td></td>
</tr>
<tr>
<td>TTTTCGAGCA</td>
<td>CCAATTTACTA</td>
<td>AAGGAGGCTTT</td>
<td>TTTCCAGCAA</td>
<td>AGGTGTGTGG</td>
<td>CTGGGTTTGG</td>
<td></td>
</tr>
<tr>
<td>AAAAATAAAG</td>
<td>GGGCTGAGTT</td>
<td>TGTGGAGTTT</td>
<td>GTCTCTGATAT</td>
<td>CTTCTAACCT</td>
<td>GTTCCACTTA</td>
<td></td>
</tr>
<tr>
<td>TCTTGGGGTG</td>
<td>TGTGGTTTTTT</td>
<td>CCAGTACCAT</td>
<td>GCTGTGTTTTA</td>
<td>TCACAGTGGC</td>
<td>TTTATGTTAT</td>
<td></td>
</tr>
<tr>
<td>AATTTCAGCT</td>
<td>CAGGTTGGGT</td>
<td>GATGCCGCTT</td>
<td>TCTTGTCTTT</td>
<td>GTTGCCCAT</td>
<td>AAAATTTCTT</td>
<td></td>
</tr>
<tr>
<td>GTACTATTTAT</td>
<td>AGGTCTCTTT</td>
<td>TGTTTCCATAA</td>
<td>TGAATTTTTT</td>
<td>AATTTCTATA</td>
<td>ATTCGCGAAG</td>
<td></td>
</tr>
<tr>
<td>ATATGCTCTT</td>
<td>GGAATPTTAA</td>
<td>TCGAATTTGC</td>
<td>ATTAACTCTG</td>
<td>TATAATGATT</td>
<td>TTGAAGCAT</td>
<td></td>
</tr>
<tr>
<td>GCCATPPPTC</td>
<td>ACTATPPGTG</td>
<td>TTTCTCTACT</td>
<td>AAGAGCAAGG</td>
<td>GATTTCCTTCC</td>
<td>CATTTCTCTG</td>
<td></td>
</tr>
<tr>
<td>TATCCTATTG</td>
<td>AATTTCTTTT</td>
<td>GCCATGAAGC</td>
<td>ACCATACTAA</td>
<td>GAGATTCTTC</td>
<td>TCGGAATTGT</td>
<td></td>
</tr>
<tr>
<td>TTCCATGAT</td>
<td>GAGTTGCTTA</td>
<td>CATAGGCGCT</td>
<td>TACTCTGATT</td>
<td>CTGGGATGTG</td>
<td>ATGACATACA</td>
<td></td>
</tr>
<tr>
<td>TCTCTCTTTT</td>
<td>ACATCTCTTA</td>
<td>TGTCATTCGG</td>
<td>GGGAAGAGTG</td>
<td>AATTCTCTTT</td>
<td>GGCTTTCPAT</td>
<td></td>
</tr>
<tr>
<td>TTTACTTCTT</td>
<td>GATTTGCCTCA</td>
<td>TTTGATTTTG</td>
<td>CCGTGAAGAT</td>
<td>TCTTTTCCTC</td>
<td>TTAGAACAGG</td>
<td></td>
</tr>
<tr>
<td>TTTGGGACAT</td>
<td>CCTTTTCTCT</td>
<td>AGCACTAGAC</td>
<td>CTGCAAGAAA</td>
<td>CAAAGGTTAT</td>
<td>ATCAAAGCCA</td>
<td></td>
</tr>
<tr>
<td>AGGGGGGAAA</td>
<td>CATTTACAAAT</td>
<td>TCTTGAACCTT</td>
<td>AGCAGGAGTAA</td>
<td>GAGAAATA</td>
<td>TTAGGATTAT</td>
<td></td>
</tr>
<tr>
<td>ACTAGATTTG</td>
<td>GGGAGGGGGG</td>
<td>GAAATAAGGA</td>
<td>TGAATTAATG</td>
<td>AAGAAGCTGC</td>
<td>CAACCTTTCAG</td>
<td></td>
</tr>
<tr>
<td>TCTATCTTTT</td>
<td>CATGACTATA</td>
<td>CTGTCTCTCAT</td>
<td>CTCCTTTTCT</td>
<td>GATGTAAGGC</td>
<td>TGTGCTTATAT</td>
<td></td>
</tr>
<tr>
<td>TGTCTTGGAA</td>
<td>GAGTCTCTCA</td>
<td>ATTAATTTAA</td>
<td>GAGGTGAATC</td>
<td>TGGAGCTGT</td>
<td>CTGCAATTCTA</td>
<td></td>
</tr>
<tr>
<td>GGTTCACATC</td>
<td>AGCGAGCCAAG</td>
<td>ATGGCTCTTC</td>
<td>TTTCCCCTTCT</td>
<td>GTTGCGTTTG</td>
<td>GGCATCCTGT</td>
<td></td>
</tr>
<tr>
<td>TTCTCTCCTT</td>
<td>GCAGCGCAAG</td>
<td>CAACCTACA</td>
<td>AAATCTGCGT</td>
<td>GTCTCATGAG</td>
<td>TTGAAGCACC</td>
<td></td>
</tr>
<tr>
<td>TGCAGCGCTA</td>
<td>CCGACACAC</td>
<td>ACTTTGCGCTG</td>
<td>AAGTGTGATT</td>
<td>CTCTATTCCA</td>
<td>TCAATTTCAT</td>
<td></td>
</tr>
<tr>
<td>TGCCCGCTA</td>
<td>CCGACACATC</td>
<td>ACTTTGCGCTG</td>
<td>AAGTGTGATT</td>
<td>CTCTATTCCA</td>
<td>TCAATTTCAT</td>
<td></td>
</tr>
<tr>
<td>TCGGTCTTGG</td>
<td>TCTGGTCACC</td>
<td>ATTTGTCTCT</td>
<td>CTCCCTCTTT</td>
<td>CCACCTTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCATATCTC</td>
<td>AATAATTTTC</td>
<td>TGATATTTTT</td>
<td>AGCGTTTTGGT</td>
<td>CTACTCTTTC</td>
<td>TTCAATTTTG</td>
<td></td>
</tr>
<tr>
<td>TCGCTATATC</td>
<td>TTGTCTCGAG</td>
<td>CTTTTTTGAA</td>
<td>ACTGTCTCTA</td>
<td>GATCATCAAG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Intron 1**
Figure 3.3

Extensive homologies in the 3' untranslated regions of the α-lactalbumin genes from human, guinea pig and rat.

A. Comparison of the 3' untranslated regions from the α-lactalbumin genes of human, guinea pig and rat reveals extensive conservation within a region of approximately 200bp in human and guinea pig. The rat sequence diverges more significantly, except for a 200bp central 'core', which contains a putative stem loop structure. It is interesting that the 'loop' part of this structure diverges in all three genes, while the inverted repeats, indicated by the facing arrows, are well conserved.

B. The shape of the putative stem-loop. This putative structure may be part of a more complex secondary organisation orchestrated within the 3' UTR of the α-lactalbumin genes.
A. Human α LA 3'UTR
Guinea pig α LA 3'UTR
Rat α LA 3'UTR

1 GTGTCTGCTGTCCCTTTGGACCCCGGCTCCACTCACACTGGAATACCTCTTCCTTCCCTAAAT
1 GCACCTGATGTATCTGCAATCCTGAGCTCTTCCCATACACCCAGAAAACCCCTCTCTCCCTCAT

------------------------
CTGTAGTGACATCGGATACCATGCTCTCTGAGGACTGCTGCTAAGGATGCATGACTGGTGCACT

BGACATTAGACCTCCTTTGAAACTCTTTGATGGGACTTTTGTCAACAGGAGTGCGACTGGCACACT

GGACTGCAAACCCTTGCTTAGTGACGGCGACTGTGCTTGTCTGCTTCTAGTCTAGGAT

B.

G
T
G
T
C
T
A
T
C
G
A
G
G
C
G
C
C
G
T
T
G
C
A
T
T
G
G
T
A
C
G
C
G
T
A

GTCTCTGAGCGCTGGGA CTGTGTCTAAGGAT
3.3 Gel retardation analysis of the α-lactalbumin gene promoter

The control of gene expression is often effected through the tissue-specific binding of proteins to regulatory elements incorporated in the DNA sequence (see chapter 1). One of the most popular methods of detecting such interactions between protein and DNA is the gel retardation or mobility shift assay (Fried and Crothers, 1981), which detects interactions between protein and DNA by observing the alteration of the mobility of end-labelled DNA probes through native polyacrylamide gels when they are complexed by protein(s) (see fig. 3.4). This technique has the advantages of simplicity and sensitivity, and the number and relative positions of the complexes observed can provide information about the nature of the protein/DNA interactions being monitored (see Hennighausen and Lubon, 1987). The chief limitation of gel retardation analysis is that it does not provide any data as to the in vivo importance of the binding reactions observed; this can make results confusing and indeed misleading, as demonstrated in the case of the immunoglobulin enhancer binding proteins (Ephrussi et al., 1985; Sen and Baltimore, 1986).

3.3.1 Preparation of mammary gland nuclear extracts

The binding of protein to DNA in vitro requires the use of high-grade nuclear extracts, which contain undegraded nuclear proteins still capable of exercising their DNA binding activities. The preparation of such extracts from purified cell nuclei has been described by Dignam et al. (1983), involving the removal of nucleic acid and histone proteins through precipitation in high salt, and has been shown to yield nuclear extracts active in binding assays (see Hennighausen and Lubon, 1987). In order to purify nuclei from lactating mammary gland tissue, the method of Bathurst (1979) was adapted, using a gentler detergent treatment and protease inhibitors in order to preserve protein structure: mammary glands, liver and kidneys were dissected from peak lactating and virgin guinea pigs, Duncan-Hartley strain, obtained from Porcellus animal breeding, Heathfield, Sussex (GB), and placed into ice-cold solution 1 (0.3M sucrose, 10mM HEPES, pH7.9, 10mM KCl, 1.5mM MgCl₂, 0.1mM EGTA, 0.5mM PMSF, 0.5mM DTT, 1% Trasylol). All further steps were carried out in the cold room at 4°C, except for centrifugation steps which were performed in the main laboratory, using Sorvall equipment pre-cooled to 4°C. The tissue was
weighed, minced in a tissue mincer, and homogenised in 2.5 volumes of solution 1 plus 1% Triton X-100 by four strokes of a teflon pestle in a glass homogeniser, 0.4mm clearance, rotating at 1000rpm. A further volume of solution 1 plus Triton was added, and the homogenate was filtered through one, then four, layers of nylon mesh, in order to remove membranous cell debris. Nuclei were precipitated by centrifugation in a Sorvall SS34 50ml tube for five minutes, at 2500rpm. Fat collects round the neck of the tube, and this was removed with a paper tissue before resuspending the nuclear pellet in 50ml solution 1 plus 1% Triton by vortexing. After centrifugation at 2500rpm, the washing procedure was repeated twice more. This protocol can be seen to give a clean preparation of intact nuclei, in which both nuclear membranes and the nucleolar structure are clearly visible (fig. 3.5a).

The intact nuclei were immediately lysed in a high salt buffer according to a modification of the protocol of Dignam et al. (1983) as described by Lubon and Hennighausen (1987), with the further substitution of protease inhibitors leupeptin, pepstatin A and antipain by trasylol. All steps were performed in the cold room, at 4°C. The nuclear pellet was resuspended in 5ml solution 2 (10mM HEPES, pH7.9, 420mM NaCl, 1.5mM MgCl2, 0.1mM EGTA, 5% v/v glycerol, 0.5mM PMSF, 0.5mM DTT, 0.1% Trasylol) and the nuclei were lysed with 10 strokes of a dounce homogeniser, pestle B. The lysate was mixed gently on a magnetic stirrer for 30 minutes, and then centrifuged at 17000rpm (20000g) in a Sorvall SS34 rotor, 4°C, for 20 minutes. The supernatant, which contains the nuclear proteins but is free from pelleted nucleic acids and nuclear membranes, was subsequently dialysed for five hours against two changes of solution 3 (20mM HEPES, pH7.9, 75mM NaCl, 0.1mM EDTA, 20% v/v glycerol, 0.5mM DTT, 0.5mM PMSF, 0.1% Trasylol), and recentrifuged in the SS34 rotor at 17000rpm for 20 minutes. The supernatant was aliquoted into eppendorf tubes and snap-frozen in liquid nitrogen. Aliquots were maintained at -70°C.

One aliquot was used for protein quantitation by the Pierce BCA method. Typically, concentrations of 10 to 20μg/μl were obtained (see fig. 3.5b).

1: section 2.1.4
In order to assay for specific binding between DNA sequence elements and protein fractions present in the nuclei of particular cell types, DNA and crude nuclear extracts can be incubated together *in vitro* and the complexes generated subsequently separated by electrophoresis on non-denaturing polyacrylamide gels.

Radiolabelled DNA probes encoding the DNA sequence in question are prepared as described in fig. 3.7. 1ng, or approximately $10^7$ cpm of probe is incubated together with 1-10μg of nuclear extract (section 3.3.1) and an empirically determined amount, typically 1 to 10μg, of unlabelled competitor DNA (1).

The assay relies on the fact that in order to bind DNA *in vivo*, sequence-specific DNA binding proteins attain a more favourable thermodynamic state when bound to their respective sequence elements than when interacting with DNA in a non-specific manner. In addition, as the competitor DNA is in vast excess over both the labelled probe and the binding capacity of the nuclear extract, the number of non-specific binding events involving the labelled probe will be very low. Therefore an equilibrium between probe, nuclear proteins and competitor DNA can be reached where only the thermodynamically advantageous sequence-specific binding events take place on the probe (2). The vast majority of non-specific binding is absorbed by the competitor DNA (3), while some probe will inevitably remain uncomplexed (4).

When such a reaction is separated on a native 4% polyacrylamide gel (5), the mobility of DNA molecules will be reduced if they are complexed by protein. The probe will be retarded by a specific amount, dependent on the protein(s) with which it is complexed; it will therefore form a distinct band on the gel (6). The uncomplexed probe will run normally, and not be retarded (7).
Protein binding site

Radiolabelled DNA probe

Unlabelled competitor DNA

Nuclear extract

specific protein-DNA interactions

Non-specific protein-DNA interactions

Uncomplexed probe

1

2 3 4

5

6 7
Figure 3.5

Preparation of nuclear extract from lactating guinea pig mammary gland

A. Nuclei were prepared from lactating guinea pig mammary glands, liver and kidneys as described in section 3.3.1. The nuclei are free from cellular debris, and nucleolar structure is clearly visible. Both nuclear membranes are maintained intact. Photographed under phase contrast illumination on a Nikon Diaphot inverted microscope, using Kodak Ektachrome EP-50 (Tungsten) film.

Original magnification 400x

B. The nuclear extract protein concentration was measured using the Pierce BCA assay. 0, 5, 50 and 100μg of BSA dissolved in nuclear extract buffer were used to generate a standard curve. 5μl of nuclear extract were tested. Typically, the protein concentration was determined to be between 10 and 20μg/μl.

L Liver
K Kidney
MG Mammary Gland
A

B

\[ y = -4.7636 + 246.97x \quad R^2 = 0.990 \]
protein
\( \mu G \)

A 650

protein

95
3.3.2 Preparation of end-labelled DNA probes from the guinea pig α-lactalbumin promoter.

In order to facilitate the generation of probes from the region of interest of the α-lactalbumin promoter, a 312bp Ssp 1 / Nae 1 fragment was isolated from the genomic clone pgpH68 (Laird et al., 1988) and subcloned into the polylinker of pUC12 (see fig. 3.6) to form a promoter-specific plasmid, pgp312, from which various end-labelled fragments could then be obtained.

In order to generate linear double-stranded probes specific to defined regions of the promoter, 2μg plasmid DNA were restricted with either Xba 1 or Hind III and purified by phenol extraction and ethanol precipitation. The DNA was labelled as described (section 2.2.9b) and purified by precipitation with NH₄Ac/ethanol and NaAc/ethanol as described by Maxam and Gilbert (1980). The purified, labelled DNA was digested with a second enzyme for one hour in order to release the required fragment (fig. 3.7) and phenol extracted, before being loaded on to a 6% native polyacrylamide gel and electrophoresed for 2 hours at 30W. The gel was wrapped in Saran wrap and autoradiographed for 15 minutes between two glass plates, using radioactive ink markers to permit subsequent alignment of the film to the gel. The band corresponding to the required labelled fragment was cut out of the film, and the film placed over the gel: the acrylamide containing the fragment was then excised by cutting through the hole in the film, and the DNA eluted overnight in NH₄Ac elution buffer as described by Maxam and Gilbert (1980).

3.3.3 Resolution of protein/DNA complexes by native gel electrophoresis

The conditions for optimal binding and complex resolution in electrophoresis were determined empirically. Initially, experiments were conducted essentially according to the protocols of Lubon and Hennighausen (1987). 10μg of plasmid pgp312/1 was digested with 10u of Xba 1, labelled with 32P dCTP, and a 321bp α-lactalbumin promoter fragment was released by secondary digestion with Pst 1. This fragment was gel purified, and 1ng (approximately 10⁷ cpm) was incubated with 1 to 10μg of nuclear extract and 1 to 3μg of poly (dl-dC) in 25μl of 0.1% (w/v) Triton X-100, 12% (w/v) glycerol, 1mM EDTA, 5mM MgCl₂, 1mM DTT, 10mM Tris-Cl pH7.5 and 80mM NaCl. The reagents were mixed at room temperature, adding the nuclear extract last, and incubated at room temperature for 30 minutes. Bromophenol blue loading buffer was then
added, and samples were electrophoresed on a 4% 1:30 bisacrylamide:acrylamide TAE-buffered gel at 30mA for 2 hours, at room temperature. The gel was subsequently soaked in 5% glycerol, 5% methanol and 5% acetic acid before being dried and autoradiographed overnight at -70°C.

The use of poly (dl-dC) competitor DNA gave evidence of very weak binding (Fig 3.8), but binding was enhanced by the use of λ DNA as a competitor (Fig. 3.9). A single band is visible. Clearly, however, the binding is not consistent, and the use of poly (dl-dC) appears to reduce rather than enhance the sensitivity of the assay. These problems could be due to a variety of factors: firstly, the addition of variable amounts of nuclear extract alters the reaction concentration of various buffer components which are present both in the nuclear extract buffer and in the binding assay reaction buffer; secondly, incubation at room temperature does not provide a uniform reaction temperature from day to day for all the reaction tubes; and thirdly, pH and temperature variations are evident in the gel during electrophoresis.

3.3.4 The formation of reproducible complexes on a whole-promoter probe

In order to achieve reproducible binding patterns, nuclear extract was diluted before use using nuclear extract buffer (solution 3). 5μl of diluted nuclear extract were added to each reaction, and the amount of nuclear extract was varied by altering the dilution ratio. In addition, the reactions were set up in an ice bath, and the tubes were incubated in a water bath at 25°C. This was found to be the optimum temperature for binding: incubation at 30 or 37°C led to a reduction in binding efficiency. The gel was run in the cold room at 4°C, using buffer recirculation to prevent pH variation. Under these conditions reproducible binding patterns were observed (fig.310).

Three bands are visible, the third only becoming so when an excessive nuclear extract to competitor DNA ratio is used. Under these conditions a smearing effect is apparent, which can be attributed to non-specific binding between proteins and the probe DNA, and possibly to weak protein-protein interactions which are not evident at lower protein concentrations. It did not prove possible to improve the resolution using this 321bp probe, which as well as being physically too large (Hennighausen and Lubon (1987) suggest a maximum probe size of 300bp for this technique) also contains 2 GRE, the milk box, and the TATA box, which would together be expected to lead to an excessive number of different binding events. In order to reduce the
complexity of the binding pattern and to specifically address the formation of complexes on the milk box, a smaller probe was isolated and tested.
pgpH68 DNA was digested with Ssp I and NaeI to release a 312bp fragment spanning the promoter region of the gene, from the middle of the first exon to 30bp upstream of the milk box (MB). This fragment was ligated into the Sma I site in the polylinker of pUC12, and transformed into E.coli JM83 cells. White colonies were grown up in 5ml cultures, and DNA was prepared by the rapid method of Maniatis et.al. (1982). Clones were analysed for orientation of the insert, and one clone for each orientation was grown up by the large-scale method (section 2.2.4a).
The promoter-specific plasmid pgp312 was digested with Xba I. This enzyme linearises the plasmid, leaving a 5' overhang which can be efficiently labelled with MMTV reverse transcriptase and α-32P by 'filling in'. The labelled linear molecule can then be digested with a second enzyme to release a probe of the required length.

1. A 55bp milk box-specific probe can be released by digesting to completion with Fok I.
2. A 90bp probe, containing the entire milk box, is released by partial digestion with Fok I.
3. A 235bp probe, containing the milk box, the TATA box and 2 GRE is released by partial digestion with Fok I.
4. Digestion with Pst I, which cuts at a unique polylinker restriction site, releases the entire 321bp α-lactalbumin fragment.

Probes of the required size were isolated on 6% native polyacrylamide gels as described in section 2.2.6b.
A 321bp probe encoding the whole of the guinea pig α-lactalbumin gene promoter was cut from pgp312 by end-labelling at the polylinker Xba I site, and secondary digestion at the polylinker Pst I site. The probe was purified by gel electrophoresis on a 6% native polyacrylamide gel, and eluted from the polyacrylamide as described (section 2.2.5b). The probe was incubated with mammary nuclear extracts according to the protocol of Lubon and Hennighausen (1987), and the reaction mixture was electrophoresed on a 4% native polyacrylamide gel at 30mA for 2 hours.

The synthetic polymer poly (dl-dC) was used as a non-specific competitor DNA. The origin (O), free probe (F), and bound (retarded) fraction (B) are indicated; binding is extremely weak.
Figure 3.9

Gel retardation of α-lactalbumin gene promoter probe by nuclear proteins from lactating guinea pig mammary gland using λ competitor DNA

The same probe as in fig. 3.8 was incubated with lactating guinea pig mammary gland nuclear extract. Bacteriophage λ DNA, obtained commercially from Pharmacia, was used as a non-specific competitor. The binding reaction is enhanced, but remains inconsistent. Origin (O), bound fraction (B), and free probe (F) are indicated.

<table>
<thead>
<tr>
<th>Lane</th>
<th>nuclear extract</th>
<th>λ DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4μg</td>
<td>2μg</td>
</tr>
<tr>
<td>2</td>
<td>4μg</td>
<td>1μg</td>
</tr>
<tr>
<td>3</td>
<td>3μg</td>
<td>3μg</td>
</tr>
<tr>
<td>4</td>
<td>3μg</td>
<td>2μg</td>
</tr>
<tr>
<td>5</td>
<td>3μg</td>
<td>1μg</td>
</tr>
<tr>
<td>6</td>
<td>2μg</td>
<td>3μg</td>
</tr>
<tr>
<td>7</td>
<td>2μg</td>
<td>2μg</td>
</tr>
<tr>
<td>8</td>
<td>2μg</td>
<td>1μg</td>
</tr>
<tr>
<td>11</td>
<td>1μg</td>
<td>3μg</td>
</tr>
<tr>
<td>12</td>
<td>1μg</td>
<td>2μg</td>
</tr>
<tr>
<td>13</td>
<td>1μg</td>
<td>1μg</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 3.10

An α-lactalbumin promoter probe binds reproducibly to mammary nuclear proteins

A 321bp probe spanning the guinea pig α-lactalbumin promoter from the Ssp I site upstream of the milk box to the Nae I site in the first exon was excised from pgp312 by end-labelling at the polylinker Xba I site, followed by secondary digestion at the polylinker Pst I site to release the entire guinea pig insert. 1ng (10^7 cpm) of probe was incubated with nuclear extract from lactating guinea pig mammary gland according to a revised protocol (section 3.3.3), and the reaction mixture loaded onto a 4% native polyacrylamide gel.

The position of the gel origin (O) and the free probe (F) are indicated. Two stable complexes were formed (B2 and B3), plus a third complex (B1), which was only formed in the presence of a high concentration of nuclear proteins.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Nuclear Extract</th>
<th>λ DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>3.5μg</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>6μg</td>
<td>4μg</td>
</tr>
<tr>
<td>4</td>
<td>5μg</td>
<td>4μg</td>
</tr>
<tr>
<td>5</td>
<td>4μg</td>
<td>4μg</td>
</tr>
<tr>
<td>6</td>
<td>3μg</td>
<td>4μg</td>
</tr>
<tr>
<td>7</td>
<td>2μg</td>
<td>4μg</td>
</tr>
<tr>
<td>8</td>
<td>3.5μg</td>
<td>8μg</td>
</tr>
<tr>
<td>9</td>
<td>3.5μg</td>
<td>6μg</td>
</tr>
<tr>
<td>10</td>
<td>3.5μg</td>
<td>4μg</td>
</tr>
<tr>
<td>11</td>
<td>3.5μg</td>
<td>2μg</td>
</tr>
<tr>
<td>12</td>
<td>3.5μg</td>
<td>1μg</td>
</tr>
</tbody>
</table>
3.3.5 The formation of complexes on a 55bp probe

In order to determine whether increased sensitivity could be obtained from the gel retardation assay, a probe was obtained by the labelling of pgp312/1 at the Xba 1 site, and secondary restriction with Fok 1 (fig. 3.7). This fragment spans 55bp. Bandshift analysis of this fragment using nuclear extracts prepared from the liver, kidney and mammary gland of lactating guinea pig revealed the formation of a single complex. The complex was not, however, observed to be tissue-specific, as an identical band was generated by mammary gland, kidney and liver nuclear extracts (Fig. 3.11a). Competition with unlabelled probe was successful in reducing binding to the labelled probe, showing a specific effect over increased concentrations of λ DNA (Fig.3.11b).

A further test was conducted using nuclear extracts prepared from the mammary gland, liver and kidney of a sexually mature virgin guinea pig (fig. 3.12). The extracts prepared from liver similar and kidney show complex formation to those prepared from lactating guinea pig mammary gland, liver and kidney, but equivalent concentrations of extract prepared from virgin guinea pig mammary gland appear incapable of forming complexes with the 55bp probe. In addition, the gel in fig. 3.12 reveals the presence of a second, fainter band which had not been previously detected; this band is common to all nuclear extracts except virgin guinea pig mammary gland, but may be an artifact of this particular gel. Incubation of the milk-box probe with liver nuclear extract in the absence of λ competitor DNA results in the formation of three additional complexes: these are probably non-specific, as competition with λ DNA prevents their formation.

3.3.6 A 90bp probe obtained by partial Fok 1 digestion forms extra bands

2μg pgp312/1 was linearised at the Xba 1 site and labelled as described (section 2.2.9b). Secondary digestion with Fok 1 was allowed to proceed for 30 minutes at 37°C, using 2u of enzyme. The relative impurity of the DNA which has been subjected to the labelling procedure and phenol extraction favours incomplete digestion under these conditions. The labelled, partially digested DNA was then separated on a 6% native polyacrylamide gel (section 2.2.6b). Three radioactive bands were excised from this gel, corresponding to 55, 90 and 235bp probes.
The 55bp probe is identical to that previously prepared by complete digestion of labelled pgp312/1 with Fok I.

The 90bp probe includes the milk box consensus sequence, while the 235bp probe additionally encodes 2 GRE and the TATA consensus. 1ng of each probe was incubated with varying quantities of lactating guinea pig mammary gland nuclear extract. The reactions were analysed on a single 4% native polyacrylamide gel, which was dried and autoradiographed overnight at -70°C. The result (fig. 3.13) shows that while the 55bp probe gives rise to a single band as expected, the 90bp probe gives rise to 2 additional bands. These bands could be ascribed to the presence of the milk box consensus in this probe. The 235bp probe only forms a single band, presumably due to excessive probe complexity.
A 55bp probe forms a single specific complex with nuclear proteins from mammary gland, liver and kidney.

A. A 55bp probe was cut from pgp312 by end-labelling at the polylinker Xba 1 site and secondary digestion with Fok I. 1ng of labelled probe was incubated with varying amounts of nuclear extracts from lactating guinea pig mammary gland, liver and kidney, using λ DNA as a non-specific competitor.

<table>
<thead>
<tr>
<th>Lane</th>
<th>λ DNA</th>
<th>Nuclear Extract</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3μg</td>
<td>8μg Kidney</td>
</tr>
<tr>
<td>2</td>
<td>2μg</td>
<td>6μg Kidney</td>
</tr>
<tr>
<td>3</td>
<td>1μg</td>
<td>4μg Kidney</td>
</tr>
<tr>
<td>4</td>
<td>3μg</td>
<td>8μg Liver</td>
</tr>
<tr>
<td>5</td>
<td>2μg</td>
<td>6μg Liver</td>
</tr>
<tr>
<td>6</td>
<td>1μg</td>
<td>4μg Liver</td>
</tr>
<tr>
<td>7</td>
<td>4μg</td>
<td>2μg mammary gland</td>
</tr>
<tr>
<td>8</td>
<td>3μg</td>
<td>2μg mammary gland</td>
</tr>
<tr>
<td>9</td>
<td>2μg</td>
<td>2μg mammary gland</td>
</tr>
<tr>
<td>10</td>
<td>1μg</td>
<td>2μg mammary gland</td>
</tr>
<tr>
<td>11</td>
<td>0.6μg</td>
<td>2μg mammary gland</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>10μg mammary gland</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

2ng of labelled probe was added to lane 6 in error. This leads to the greater observed intensity in the bound fraction band.

B. The 55bp probe was incubated with mammary gland nuclear extract and λ DNA as a non-specific competitor, plus decreasing quantities of unlabelled probe.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Nuclear extract</th>
<th>λ DNA</th>
<th>unlabelled 80bp probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2μg</td>
<td>0.6μg</td>
<td>150ng</td>
</tr>
<tr>
<td>2</td>
<td>2μg</td>
<td>0.6μg</td>
<td>100ng</td>
</tr>
<tr>
<td>3</td>
<td>2μg</td>
<td>0.6μg</td>
<td>75ng</td>
</tr>
<tr>
<td>4</td>
<td>2μg</td>
<td>0.6μg</td>
<td>50ng</td>
</tr>
<tr>
<td>5</td>
<td>2μg</td>
<td>0.6μg</td>
<td>25ng</td>
</tr>
<tr>
<td>6</td>
<td>2μg</td>
<td>0.6μg</td>
<td>12.5ng</td>
</tr>
<tr>
<td>7</td>
<td>2μg</td>
<td>5μg</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>2μg</td>
<td>3μg</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>2μg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Origin (O), free probe (F), and bound fractions (B) are indicated.
Nuclear extracts were prepared from the mammary glands, kidneys and liver of a sexually mature virgin guinea pig. Protein concentration was determined by the Pierce BCA method. Approximately 3μg of each extract was incubated with the milk box-specific 55bp probe (fig. 3.7), in the presence or absence of λ DNA competitor.

(V) Virgin guinea pig

(L) Lactating guinea pig

<table>
<thead>
<tr>
<th>Lane</th>
<th>λ DNA (μg)</th>
<th>Nuclear Extract</th>
<th>Amount (μg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Mammary gland (L)</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Kidney (L)</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Liver (L)</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Mammary Gland (V)</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>Liver (V)</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>Kidney (V)</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>Mammary gland (V)</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>Liver (V)</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>Kidney (V)</td>
<td>3</td>
</tr>
</tbody>
</table>
Supplementary bands are detected with a 90bp probe containing the milk box consensus sequence.

Probes were prepared by partial digestion of labelled pgp312/1 with Fok I as described in section 3.3.6.  

1μg of each probe was incubated with nuclear extracts prepared from the mammary gland, kidney and liver of a lactating guinea pig, and the reactions were analysed by separation on a native 4% polyacrylamide gel.

<table>
<thead>
<tr>
<th>Lane</th>
<th>λ DNA (μg)</th>
<th>Nuclear Extract</th>
<th>Amount (μg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>MG</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>MG</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>L</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>K</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>MG</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>MG</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>L</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>K</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>MG</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>MG</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>L</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>K</td>
<td>3</td>
</tr>
</tbody>
</table>

Origin (O), bound fractions (B) and free probes (F1-3) are indicated.
Discussion

3.4 Regulatory sequences in the α-lactalbumin gene

The expression of the guinea pig α-lactalbumin gene is regulated in two key aspects; it is tissue-specific, and it is both induced and repressed by various hormones. Sequence analysis of the promoter has revealed the presence of 6 putative glucocorticoid responsive element core sequences. Quirk and Funder (1988) propose that the presence of multiple GRE is responsible for the observed dose-biphasic response of α-lactalbumin genes to the influences of glucocorticoid (see section 1.6.2.4), and note that the rat α-lactalbumin gene contains 6 GRE cores. The presence of the same number of GRE cores in the guinea pig α-lactalbumin gene would lend weight to these theories, although it is interesting to note that whereas the guinea pig gene contains 3 GRE upstream of the TATA box, one in the 5' untranslated sequence, and 2 in the first intron, in the rat gene 5 core sequences are found in the first 2 introns. It would be interesting to ask whether the position of these GRE could affect their roles as 'turn-on' or 'turn-off' elements, and to relate this to the much higher levels of α-lactalbumin expression seen in the guinea pig.

3.5 The nature of protein:DNA complexes formed on the α-lactalbumin promoter

The primary interest in this analysis was to identify a possible role for the milk box in the regulation of milk protein gene expression, and particularly in the control of tissue specificity. Lubon and Hennighausen (1987) first identified protein binding to mammary promoters using probes derived from the mouse WAP (Campbell et al., 1984). They were able to identify one minor complex which appeared to be mammary-specific, but which was lost when probes were competed with E.coli DNA in place of poly d(IC). The most stable complex formed on the mouse WAP promoter, complex (a), has a similar mobility to the complex observed with the guinea pig α-lactalbumin 55bp probe (fig. 3.10; fig. 3.11a), and maps by footprinting to the same part of the...
promoter (-100 to -200). The sequences of the two genes at this point are, however, not homologous. In both cases the complex is not mammary-specific.

The 90bp guinea pig α-lactalbumin probe gives rise to 2 extra bands when compared to the 55bp probe. These bands may be due to complex formation on the milk box consensus, but even these complexes do not appear to be tissue-specific. That is not to say, however, that they are not involved in tissue specific activation of α-lactalbumin gene expression. Experiments carried out on the immunoglobulin gene enhancer have shown that sequence elements with tissue-specific effects can appear to bind to universal proteins (Maniatis et al., 1987). The 'universal protein' is, however, two different proteins; oct-1 is a universal activator protein (Fletcher et al., 1987) while oct-2 is specific to B cells (Scheidereit et al., 1987). In B cells, the cooperative action of other tissue-specific protein factors may allow preferential binding of oct-2 (Maniatis et al., 1987). This could well be the case in the α-lactalbumin gene, but it was not possible to resolve the complexes formed by large probes in sufficient detail to allow this possibility to be examined.

Extracts prepared from the mammary glands of virgin guinea pig (fig. 3.12) do not bind the 55bp probe. It is possible that this indicates that the undeveloped epithelium of the virgin mammary gland does not express the necessary binding proteins, even though extracts derived from the liver and kidneys of the same virgin animal gave rise to a retarded band with a mobility identical to that seen with extracts prepared from lactating animals. This suggests that the binding event observed, as well as being non-tissue-specific, is not induced by hormones, but may well be dependent on the functional differentiation of the tissue epithelium. Although ductal ingrowth into the mammary fat pad of guinea pigs occurs for 2 months after birth (Anderson 1978), subjective examination of virgin animals reveals the presence of little epithelial material in the mammary fat pads. Nuclear preparations from mammary fat pads of virgin guinea pigs were judged to be poor in nuclei, and the nuclear extracts derived therefrom were of low concentration. For this reason the inability of nuclear proteins from virgin guinea pig mammary gland to form complexes with the milk box probe must be subject to the possibility that the protein detected in these extracts is not nuclear protein, but a combination of cellular protein from adipocytes and other impurities.
Chapter 4

Construction of transgenic mice carrying the guinea pig α-lactalbumin gene

Introduction

4.1 Cloned milk protein genes are not expressed in tissue culture

The functional characterisation of a eukaryotic gene necessitates the expression of manipulated clones in a eukaryotic expression system. This is usually carried out in cell culture systems, utilising immortalised cell lines derived from eukaryotes. These cells, however, are not a perfect simulation of the whole-animal environment. They are immortalised, and are not subject to the morphogenetic stimuli associated with the internal structure of body tissues. In addition, long periods of culture in synthetic media on plastic culture dishes may have potentiated loss of cell functions which would be otherwise required in vivo. Notwithstanding these problems, a large number of cloned genes are expressed when transfected into such cells, and may be artificially regulated by the administration of hormones and other exogenous stimuli; this phenomenon has been and remains extremely important in the study of eukaryotic gene expression.

Milk protein genes, however, are not responsive in cell culture (see Rosen, 1987; Laird et al., 1988; also Lee et al., 1988, and references therein). Laird et al. and Yu-Lee and Rosen (1988) observed that BPV based constructs carrying the guinea pig α-lactalbumin gene or a rat α-casein minigene respectively directed correct expression of guinea pig α-lactalbumin and rat α-casein, but that the gene was not responsive to hormone treatment. In the case of the α-lactalbumin gene it is possible that this gene was in fact driven by the BPV enhancer present in the vector; Yu-Lee and Rosen, however, observed that culture on floating collagen gels (Emerman and Pitelka, 1977) was required for the constitutive expression of the α-casein minigene. Successful regulation of milk protein gene expression in tissue culture has been
achieved using primary cell cultures (Suard et al., 1983) which are not transformed, cultured on floating collagen gels. These cells did not replicate, but synthesised casein and \( \alpha \)-lactalbumin in response to prolactin stimulation. It was observed that the 'rolling' of the collagen gel allowed the cells to group into a duct-like structure, which was enhanced if the cells were grown embedded in the gels; in this case both the growth of the cells and the expression of milk proteins proved to be prolactin-dependent. Wiens et al. (1987) obtained similar results growing mouse primary mammary cells on adipocytes, thus mimicking the \textit{in vivo} situation, where the mammary ductal system extends by branching through the mammary fat pad. The cells underwent the ultrastructural changes associated with lactation, such as enlargement of the golgi, gave rise to ductal structures, and laid down a basement membrane. Differentiation of the epithelial cells and milk protein production were dependent on lactogenic hormones, whereas ductal morphogenesis was not. Exposure of cells to lactogenic hormones resulted in the expression of milk proteins, but not if the cells had been cultured without adipocytes or on fibroblasts. The evidence points to the requirement for the establishment of an extracellular matrix (ECM), which is known to affect morphogenesis in endothelial systems (Ingber and Folkman, 1989), and is important in both morphogenesis and lactogenesis in mammary epithelia (see Bissell and Hall, 1987; also Blum et al., 1987). Bissell and Aggeler (1987) suggest that the ECM interacts with transmembrane proteins and the cytoskeleton to influence gene expression through mRNA processing, stability, translation, and possibly even transcription, mainly through changes in cell shape.

Primary cells, however, are short-lived and demanding to culture; and impossible to stably transfect with exogenous genes. More recently, therefore, an immortalised mouse mammary epithelial cell line named COMMA 1D has been isolated from lactating mice (Danielson et al., 1984) which displays correct hormonally regulated lactogenic and mammogenic responses. The cell line is heterogeneous, and produces a large amount of ECM components (Bissell and Hall, 1987), which may explain its ability to differentiate and synthesise milk proteins. Isolated single clones of this cell line seem to lose the lactogenic potential, with notable exceptions in the BNW-7 clone isolated by Campbell et al. (1988) and the HC11 clone isolated by Ball et al. (1988). These clones express casein in a hormonally regulated manner when
cultured on plastic; however, the fact that they have lost the potential to be influenced by the ECM casts a shadow over the validity of the results obtained from these experiments.

4.2 Transgenic mice: an in vivo expression system

Cloned, manipulated genes may also be studied via the germline transformation of whole animals. Techniques allowing the removal of embryos from female mammals, culturing them briefly \textit{in vitro}, and subsequently returning them to pseudo-pregnant foster mothers have been crucial to the development of germline transformation techniques (Brinster, 1982). Several manipulative techniques are applicable to the transformation of embryos:

1. Combining cells from two embryos to produce chimeric animals.
2. Introducing transformed teratocarcinoma cells into the blastocyst of the embryo to produce mosaic animals (Mintz and Cronmiller, 1981).
3. The possibility of nuclear replacement, which has been suggested by the success of nuclear replacement experiments in amphibians (McGrath and Solter, 1983).
4. The infection of embryos with viral DNA or intact viral particles in culture (Jamieson 1976)
5. The transfection of sperm with recombinant DNA sequences. The sperm then fuse naturally with eggs, and lead to the generation of transgenic animals (Lavitrano \textit{et al.}, 1989a).

All these techniques have their advantages and disadvantages, and are reviewed by Palmiter and Brinster, 1986, except for the method of Lavitrano \textit{et al.}, which, although promising, has proved not to be repeatable to date (Brinster \textit{et al.}, 1989; Lavitrano \textit{et al.}, 1989b). By far the most popular and successful technique, though, is pronuclear microinjection (also reviewed by Palmiter and Brinster, 1985, 1986; Murphy and Hanson, 1988). The technical problems with this technique and the high equipment outlay are handsomely offset by its relative reliability and speed. The microinjection techniques originated in the methods developed for microinjection of mRNA, and subsequently cloned DNA, into \textit{Xenopus} oocytes (Gurdon and Melton, 1981) and into transformed cells in culture. They were adapted for the microinjection of mRNA and DNA into mouse eggs (Brinster \textit{et al.}, 1980, 1981), and in 1980 the first report of transgenic mice
generated from microinjected eggs was published by Gordon et al., followed by four other groups: Brinster et al., 1981, Constantini and Lacy, 1981, Wagner, E.F. et al., 1981, and Wagner, T.E. et al., 1981. The foreign genes were expressed in most cases, and the DNA appeared to be integrated into the germ line as well as into the somatic tissues. The dramatic demonstration of the physiological alteration of a mouse by injection of a cloned rat growth hormone gene under the control of the mouse metallothionein promoter by Palmiter et al., 1982, leading to abnormally rapid and extensive growth of the mouse, fociussed much interest on the transgenic animal.

Microinjection is conceptually simple: about 500 copies of the intended transgene are injected into the male pronucleus (the larger of the two) of a fertilised one-cell mouse ovum. The DNA may integrate into the mouse genome, either in single or multiple copies, which are usually arranged as head to tail tandem repeats (see Palmiter and Brinster, 1985; Murphy and Hanson, 1988). The site of integration would appear to be random, and integration has been observed in all autosomes as well as both X and Y chromosomes (Palmiter and Brinster, 1986). Palmiter and Brinster have also suggested that randomly generated chromosomal breaks are the rate-limiting step in the integration of foreign DNA, and that the integration is initiated by the ends of the injected DNA fragment. This would explain why linear molecules integrate more readily than circular ones, even though it is known that linear DNA is rapidly circularised on microinjection into the nucleus; ligations may occur between the chromosomal breaks and the linear injected DNA, rather than intramolecularly to circularise the injected molecules (Palmiter and Brinster, 1985, 1986). Once one copy of the transgene has inserted, more copies may integrate by homologous recombination with the integrated DNA. Analysis of the junction areas between injected DNA and chromosomal DNA reveals rearrangement, translocation and duplication of chromosomal sequences, resembling the breakpoints of naturally occurring chromosomal translocations.

Most, but not all, microinjected genes are appropriately expressed. The first example of correct tissue specific expression (Brinster et al., 1983) involved the microinjection of an immunoglobulin κ gene on a pBR322-based plasmid vector. Other injected genes, such as the insulin gene, globin genes and the α-crystallin gene have also shown correct expression, suggesting that if a gene is expressed at all it will be correctly expressed (Storb et al., 1984; Swift et al., 1984; Bucchini et al., 1986). The inclusion of plasmid or viral DNA in the injected construct,
however, appears to adversely affect expression. Bucchini et. al. (1986) observed pancreatic expression of the human insulin gene in transgenic mice after microinjection of an 11kb whole-gene fragment while previous experiments using a pBR322 based construct containing the human insulin gene had failed to show expression (Burki and Ullrich, 1982).

Integration of the transgene at a different chromosomal location to the endogenous gene does not seem to affect expression, which confirms theories about the trans-acting nature of tissue-specific transcription factors. The variable levels of expression observed suggest, however, that the accessibility of the DNA to these factors is influenced by chromosomal location (see section 1.1.2). When a transgene is not expressed it is presumably integrated into heterochromatin (see Storb et. al., 1984), with the exception of non-expressed copies of a tandem array; why some copies should be silent is unknown, but expression levels seldom correlate to the copy number in a tandem array (Palmiter and Brinster, 1986).

With these results in mind, the creation of mice transgenic for guinea pig α-lactalbumin holds much promise. The primary aim of this project is to establish transgenic mice carrying the guinea pig α-lactalbumin gene. It is hoped that the cloned gene will respond correctly to tissue specific and hormonal stimuli in the context of a whole-animal system, and that further experiments using manipulated clones of the gene will subsequently allow us to define DNA sequences that are important in these responses. This chapter describes the creation of two different lines of mice carrying the guinea pig α-lactalbumin transgene.
4.3 Generation of transgenic mice

4.3.1 Preparation of the transgene DNA

The guinea pig α-Lactalbumin genomic clone isolated by Laird (1985) in λ L47.1 has been subcloned into pAT 153 and named pgp H68 (Laird, 1985; Laird et al., 1988). The entire coding sequence, plus 1kb of 3' untranslated sequence and 1kb of the 5' flanking region are contained within this subclone. When transfected into mammary cell lines (Jack, 1987; Laird et al., 1988) these sequences proved sufficient to direct correct initiation and polyadenylation of the guinea pig α-lactalbumin gene cloned into a BPV based replicating vector. Guinea pig α-lactalbumin was secreted into the growth medium of the transfected cells. As the BPV construct proved to be constitutively expressed, it remains a possibility that the α-lactalbumin gene was driven from the BPV enhancer, which would imply that the guinea pig sequences were sufficient to direct correct initiation and termination of a translatable transcript, but that the the guinea pig enhancers were either inactive, or over-ridden by the BPV sequences.

The DNA used for microinjection into mice was derived from pgpH68 (see fig. 4.1) and contains the sequenced guinea pig genomic clone in its entirety; more uncharacterised 5' upstream sequence was available in the original λ clone 84G1, but, as the presence of a guinea pig LINE sequence at -512 suggests that all regulatory sequences are likely to be downstream of this point, a microinjection fragment was prepared incorporating the same guinea pig sequences as were present in the BPV vector. No plasmid or viral sequences were included, and the microinjected DNA was in linear form.

4.3.2 Microinjection of Mouse Embryos

The DNA fragments were microinjected into single cell mouse embryos by Dr. Dimitris Kioussis, N.I.M.R., London (GB), as described in fig. 4.2. The resulting offspring were analysed for presence of the transgene as described below.

46 offspring were born from four surrogate mothers. At 10 days of age, 10mm of tail was removed from each mouse and analysed as described in order to determine the presence or
absence of the transgene. At the age of six weeks the transgenic mice were transferred from N.I.M.R. to the M.H.M.S site and mated to CBA/J males to further the transgenic lines.
Isolation of a microinjection fragment from pgp H68.

30μg of pgp H68 DNA was digested with Hind III, Eco RI, and Ava I. The resulting fragments were separated on a 0.8% agarose gel. The 3.8Kb band containing the guinea pig α-lactalbumin gene was cut out of the gel, and the DNA was eluted from the agarose by using the GeneClean method. The DNA was resuspended in D.I. water, and any remaining glass particles were removed by purification on Elutips DNA binding columns, obtained from Schleicher and Scheull. This ensures that the injection needle does not become clogged by microparticulate matter. The DNA was then adjusted to a final concentration of 1μg/ml in 10mM Tris-Cl, pH7.4, 0.2mM EDTA, and the entire 3.8Kb fragment, containing all four exons of the guinea pig α-lactalbumin coding sequence, plus the 5' flanking sequences and the 3' UTR and polyadenylation signal (see chapter 3), was microinjected into fertilised mouse oocytes as described (fig. 4.2).
Fertilised mouse ova were obtained from superovulated (B10xCBA/J)F1 hybrid female mice, and placed into M2 medium (chapter 2.1.7). Cumulus cells, which clump around the eggs, were released by digestion for 5 minutes with 50μg/μl hyaluronidase. The eggs were picked out with a pipette, washed once in fresh M2 medium and twice in M16 medium (chapter 2.1.7) before being transferred to M16 microdrop cultures.

For injection, eggs were transferred to M2 medium, washed twice, and loaded into a depression slide injection chamber filled with M2. The eggs were immobilised with a holding pipette, and 2pl of a 1ng/μl DNA solution was injected into the male pronucleus with a pulled glass needle. Cells which appeared to survive the injection were placed on one side of the depression. These were later collected into M16 medium, and cultured overnight in M16 microdrop cultures in a 5% CO2 incubator. The cells which developed to the 2-cell stage were reimplanted into pseudopregnant female (B10xCBA/J)F1 mice by injection into the infundibulum at the top of the uterus with a pulled glass pipette.

The resulting pregnancies were allowed to develop to term, and the offspring were born naturally. 10mm of tail was removed from each pup at 10 days of age, and DNA prepared from this tissue (section 2.2.7) was analysed for the presence of the transgene by Southern blotting.
Holding pipette

fertilised ovum

nucleus

nucleolus

DNA solution injected.

single-cell embryo

integration into mouse genome

culture

reimplantation

pseudopregnant foster mother

DNA extraction

Litter

transgenic

normal

southern analysis
4.4 Detection of the transgene in founder mice

4.4.1 Preparation of an α-Lactalbumin cDNA probe

In order to detect the presence of the guinea pig α-lactalbumin gene in a transgenic mouse it is essential to be able to differentiate it from the endogenous mouse gene(s) in a southern blot hybridisation assay. A potential difficulty arises due to the high degree of exhibited by the α-lactalbumin genes from various species (Chapter 3, fig. 3.1), which suggests the possibility of an elevated amount of cross-hybridisation between mouse and guinea pig. Southern analysis of guinea pig and mouse genomic DNA with the entire pgpK9 plasmid does result in the detection of multiple bands in both species, suggesting hybridisation with other genes apart from α-lactalbumin (fig. 4.3).

The sequence for mouse α-lactalbumin is not available, but comparisons performed between the α-lactalbumin sequences of guinea pig, rat and human confirm the three genes to possess extensive homology. This homology, while strong in coding sequences, is also very extensive at the 5' end of the gene, which encodes the signal peptide, and in the 3' untranslated region of the mRNA. Here in particular long stretches of 100% homology can be identified in all three genes (see chapter 3). These sequences are present in the guinea pig α-lactalbumin cDNA clone pgpK9 (Hall et al., 1982), so in order to exclude them a 205bp band which contains most of exon 2, exon 3 in its entirety, and the beginning of exon 4 was cut from pgpK9 by digestion with Rsa I (Fig 4.4). This band was isolated on a 1.5% agarose gel, eluted from the agarose using the GeneClean kit, oligo-labelled to a specific activity of 5x10⁸ cpm/µg, and used to probe Southern blots of mouse and guinea pig genomic DNA (section 2.2.10a). A single band of 6.7kb was visualised in Hind III digested guinea pig genomic DNA (fig. 4.5, lane GP). No cross-hybridisation to mouse DNA was detected when washing to a stringency of 0.1x SSC at 65°C (fig. 4.5, lane M), except for very faint bands on overexposure of the blots (fig. 4.7, lane M).

4.4.2 Identification of Transgenic Mice by Southern Blot Analysis of Tail DNA

Mice were screened for presence of the transgene at 10 days of age. High molecular weight DNA was prepared from tail tissue as described in section 2.2.7, 10 mm of tail typically yielding 50 to 100 µg of DNA. The concentration was determined by spectrophotometric
determination of the \( A_{260} \) of a 1:100 dilution, and the purity of the sample checked by
determination of the \( A_{260} \) to \( A_{280} \) ratio. The DNA was then analysed by electrophoresis on a
0.8% agarose gel, followed by Southern transfer to Amersham Hybond-N and hybridisation with
the K9/Rsa probe (fig. 4.4).

After overnight exposure, the presence of two transgenic mice is confirmed (Fig. 4.5). Six bands are detectable in the Hind III digested DNA of one mouse (101.1), of which 2 bands
are extremely faint, and may be due to partial digestion; and two bands in the DNA of the second
(101.8). As Hind III does not cut within the transgene DNA, it can be deduced that each one of
these bands represents a separate insertion of one or more copies of the transgene into the
mouse genome, separated from each other by Hind III sites. The intensity of each band is similar
to that of the single band produced by the single-copy \( \alpha \)-lactalbumin gene in guinea pig DNA;
therefore each insertion into the mouse genome consists of only one or two copies of the
transgene.

From these positive mice, both females, two breeding lines were established: Line 101.1, with a copy number of 4 to 6, and line 101.8, with a copy number of two. The transgene
proved to be transmissible in both cases in a mendelian fashion, and through both sexes (fig.
4.6).

4.4.3 Screening of Transgenic Mouse Offspring

The offspring of transgenic breeding pairs were screened by isolation of tail DNA as
described in section 2.2.7, and transgene detection by Eco RV digestion followed by southern
blot analysis. Eco RV releases a 1.6kb internal fragment from the guinea pig \( \alpha \)-lactalbumin gene,
and as all copies of the transgene contribute a 1.6kb band autoradiographic detection is
enhanced. It is thus possible to read a southern blot after a few hours of exposure to Kodak X-AR
film. In addition, it was observed that Eco RV is considerably more efficient in cutting genomic
DNA derived from tail tissue than other enzymes such as Hind III. This allows digestion times to
be reduced to 3 hours rather than overnight, all of which accelerates the time-consuming
process of mouse screening.

10\( \mu \)g of tail DNA was restricted with Eco RV for three hours and electrophoresed on a
0.8% agarose gel overnight at 40v. The DNA was transferred to Amersham Hybond-N by capillary
transfer, and fixed to the filter by UV irradiation for four minutes on a UVP transilluminator. The blots were screened with the 205 bp K9 probe, and the expected 1.6kb band was apparent in positive samples (Fig 4.7). In addition, several minor bands are generated through hybridisation to the transgene DNA external to the Eco RV sites within the transgene; the presence of 8 or 9 such bands of equal intensity in the offspring of line 101.1 confirms the transgene has integrated as 4 or 5 independent single-copy insertions.

4.4.4 Deletions and Rearrangements of Transgenes in the Mouse Germ Line

During routine screening of newborn mice in order to identify the carriers of the transgene, some positives appeared to give a much fainter hybridisation signal than their siblings (fig. 4.7, lane 1) and were observed to lack external Eco RV bands. As the total amounts of DNA loaded on to the gels was not seen to vary in ethidium bromide stained photographs of the gels or in the spectrophotometric data, the faint samples were further analysed by restriction with Hind III and Bam H1. The difference between the restriction patterns of these samples and others from the same litters demonstrates that copies of the transgene have been deleted from the genomes of the mice in question, and furthermore that transgenes have been rearranged on at least one occasion (fig. 4.8).

The sizes of the 'external' fragments generated by Eco RV digestion can reveal more information about the arrangement of the transgenes in the chromatin. Most of these bands are between 4 and 10kb in size, which implies each insertion is roughly between 4 and 10kb from the nearest Eco RV site. As sites for restriction enzymes with hexanucleotide recognition sequences, such as Eco RV, theoretically occur once every 4kb on average, it can be deduced that the transgenes are at least 4 to 8kb apart, and possibly much more. One of the external bands, however, is smaller than the 1.6kb internal band, and is of the size (1.4kb) which would be generated if transgenes were integrated as a head-to-tail tandem repeat. This raises the possibility that at least two copies of the transgene could be integrated in tandem.
The guinea pig α-lactalbumin cDNA clone pgpK9 hybridises to multiple bands in guinea pig and mouse genomic DNA

Genomic DNA was prepared from the livers of DH strain guinea pig and CBA/J mouse. 10μg of each sample was digested with Hind III overnight, and electrophoresed on a 0.8% agarose gel. The DNA was transferred to Amersham Hybond-N by Southern transfer, and hybridised with the entire guinea pig α-lactalbumin clone pgpK9, oligo-labelled to an activity of 5x10^8 cpm. The blot was subsequently washed at 65°C, in 0.1xSSC.

Multiple bands are detectable in both guinea pig and mouse DNA; as Hind III does not cut the guinea pig α-lactalbumin gene within the transcribed sequence, it can be inferred that the probe cross-hybridises to other genes in guinea pig DNA. One or possibly two of the bands visible in guinea pig DNA are common to mouse DNA.

<table>
<thead>
<tr>
<th>GP</th>
<th>Guinea pig genomic DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Mouse genomic DNA</td>
</tr>
<tr>
<td>C</td>
<td>Common Bands</td>
</tr>
</tbody>
</table>
20μg of pgpK9 was digested with *Rsa I* and the fragments separated on a 1% agarose/TAE gel. The 205bp fragment containing part of exon 2, exon 3 and part of exon 4 was excised from the gel and the DNA eluted with the Geneclean kit. The linear DNA recovered was labelled by the oligo-labelling method, and used to probe Northern and Southern blots.
10 µg of DNA extracted from the tails of each of 46 founder mice, plus DNA prepared from the livers of adult guinea pig and normal mouse, was restricted with Hind III and electrophoresed on a 0.8% agarose/TAE gel. The DNA was transferred to Amersham Hybond-N and hybridised with a 205bp fragment of the guinea pig α-lactalbumin cDNA clone pgpK9.

A λ/Hind III ladder was run alongside the guinea pig sample: the positions of the fragments are indicated.

GP: Guinea pig DNA
M: Mouse DNA

Lane 3, mouse 101.1; lane 11, mouse 101.8.
A. First generation of transgenic mice

46 mice were analysed: two were positive, both females.

<table>
<thead>
<tr>
<th>Mouse number</th>
<th>Sex</th>
<th>Transgene</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.1 to 99.16</td>
<td>14M, 2F</td>
<td>All Negative</td>
</tr>
<tr>
<td>101.1</td>
<td>F</td>
<td>+</td>
</tr>
<tr>
<td>101.2</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>101.3</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>101.4</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>101.5</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>101.6</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>101.7</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>101.8</td>
<td>F</td>
<td>+</td>
</tr>
<tr>
<td>102.1</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.2</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.3</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.4</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.5</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.6</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.7</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.8</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.9</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.10</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.11</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.12</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.13</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.14</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.15</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.16</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>102.17</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>102.18</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>PA.1 to PA.4</td>
<td>2M, 2F</td>
<td>All Negative</td>
</tr>
</tbody>
</table>
B. Transgene transmission from founder 101.1

Mouse 101.6 was mated with a CBA/J stud male, and the offspring were analysed for the presence of the transgene. The transgene proved to be transmissible in a mendelian fashion, and from both sexes.
Figure 4.7

Southern analysis of offspring transgenic mice

10 μg of DNA isolated from the tails of the offspring from line A1 was restricted with *Eco RV* and electrophoresed on a 0.8% agarose gel. *Eco RV* releases an internal 1.6kb fragment from the guinea pig α-lactalbumin gene. Lane number 1 shows evidence of the loss of multiple copies of the transgene, retaining just one copy.

from left:

- λ: λ DNA digested with *Hind III*.
- 1 to 15: Line A1 transgenic mouse offspring genomic DNA, extracted from tail tissue.
- G: Guinea pig genomic DNA.
- M: Normal mouse genomic DNA.
10μg of tail DNA from 10 transgenic mice was restricted with *Bam H1*. An internal fragment of 320bp is excised; the probe K9/Rsa hybridises to this fragment and to sequences upstream of it, which do not contain further *Bam H1* sites. Mouse number two can be seen to have lost multiple copies of the transgene, leaving two insertions. The appearance of a new *Bam H1* band implies the rearrangement of one of the transgenes or of the flanking genomic sequences. Note that 4 out of 5 transgenic animals have not rearranged, suggesting close linkage of the transgene insertions.

<table>
<thead>
<tr>
<th>lane</th>
<th>DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normal CBA/J mouse</td>
</tr>
<tr>
<td>2 to 10</td>
<td>line A1 offspring</td>
</tr>
</tbody>
</table>
Discussion

4.5 Two unusual lines of transgenic mice

Two lines of mice transgenic for the guinea pig \(\alpha\)-lactalbumin gene have been generated. Both lines show independent integration of transgene monomers into the mouse chromatin, all apparently in the same chromosome, with the possibility of the existence of one tandem insertion in line 101.1; line 101.8 was not exhaustively analysed in this aspect, but each insertion would appear to be single-copy (fig. 4.5). This arrangement is unusual, as most transgenes are observed to integrate as head-to-tail tandem repeats of several hundred copies (Palmiter and Brinster, 1985, 1986). It has been proposed (Brinster et al., 1985) that transgene integration is mediated via chromosomal breaks, in the manner of recombinational breakpoints; thus the integration of multiple independent insertions requires the occurrence of multiple chromosomal breakpoints. This does seem a little unlikely, given the tendency for injected DNA to form multimers almost immediately (Palmiter and Brinster, 1986) and the logistics involved in the generation of 6 chromosomal breakpoints within a region of a single chromosome which is small enough to display linkage of all 6 integrants. Nevertheless, it is difficult to propose an alternative scheme.

The efficiencies of gene transfer into the mouse germ lines were low: of 200 eggs injected, 46 pups were born (23%), but only 2 were transgenic (4.3% of live births, or 1% of injected eggs). These results are comparable with the results obtained for other gene transfer experiments involving milk protein genes (Andres et al., 1987; section 5.1), but are significantly lower than the efficiencies of 20% quoted by Palmiter and Brinster (1986). The reason for this is unclear; the quality of the injected DNA, as well as the size and possibly the sequence of the transgene, may all affect the insertion efficiency. Alternatively, the possibility remains that the low efficiency observed is purely due to chance.
Chapter 5
Expression of the α-lactalbumin transgene

Introduction

5.1 The expression of milk protein genes in transgenic mice

It has been established (see chapter 4) that transgenes are often correctly expressed in transgenic mice. The difficulty experienced with hormonal regulation of gene expression in cloned mammary cells has prompted other groups to look at the possibility of expressing milk proteins in transgenic mice. Andres et.al. (1987) and Pittius et.al. (1988) have used the mouse whey acidic protein (WAP) promoter isolated and characterised by Hennighausen's group (Hennighausen and Sippel, 1982) to direct the expression of both Ha-ras and human tissue plasminogen activator (tPA) to the mammary gland.

Andres et.al. obtained a low efficiency of transgene integration (five transgenic animals from 560 microinjected eggs (0.9%) and 71 live births (3.5%)), with a copy number of 1 to 5. The transgenes were autosoally transmitted to the offspring of the founder mice. The WAP-ras construct was found to be expressed in the mammary gland in a manner dependent on lactogenic hormone stimulation during lactation. The levels of expression were, though, about 50 times inferior to the levels of endogenous WAP. In addition, low levels of WAP-ras expression are noticed in the brains from lactating females of two transgenic mouse lines, but not in muscle, kidney, liver, spleen, lung or ovary from lactating females and not at all in male transgenic mice. The endogenous WAP gene is found to be restricted to the mammary glands of lactating females. In one line of mice, constitutive expression of the transgene in salivary gland was observed.

Pittius et.al. generated six lines of transgenic mice carrying a WAP-tPA construct, with integration copy number varying between 3 and 50. Human tPA was detectable in the milk of four of the lines derived from these mice, and WAP-tPA mRNA was detectable in the mammary glands of lactating females, at a level approximately 100 times lower than that of the endogenous WAP gene, and appeared to be correctly initiated. Very low levels of expression (100 times lower than in mammary gland) were also detected in the tongue, salivary gland and kidney of transgenic mice.
animals. Endogenous WAP was detected in the mammary glands of lactating mice, plus at much lower levels (10^4 to 10^6 times lower) in the pituitary gland, pancreas, adrenal gland, tongue, liver, thymus and heart.

Rosen and colleagues established transgenic mice carrying the entire cloned rat β-casein gene (Lee et al., 1988) in a 14kb genomic fragment, containing 3.5kb of 5' sequence and 3kb of 3' flanking DNA. Eight lines of mice were generated, with copy numbers of between 1 and 10, and transgenes were found to be both hormonally regulated during lactation and correctly initiated. Expression levels were again depressed, varying between 1% and 0.01% of the endogenous mouse β-casein gene. Lower levels of 'aberrant' expression were noted in the brain of one mouse. Subsequently, in order to better characterise the sequences responsible for the tissue-specific expression of the rat β-casein gene, transgenic mice carrying a rat β-casein-chloramphenicol acetyl transferase (CAT) chimeric gene were constructed (Lee et al., 1989). The construct consisted of 0.5kb of 5' flanking DNA, plus the first exon and 0.5kb of the first intron of the β-casein gene linked to the CAT gene from pSVOCAT (Gorman et al., 1982), and detection of transgene activity was performed by the assay of tissue extracts for CAT activity. CAT activity was detected in the mammary glands of lactating female transgenic mice as expected, but not in the mammary fat pads of male transgenic mice, or in liver, spleen, pancreas, kidney, brain, heart, lung or salivary gland. Constitutive expression was detected, though in the thymi of all analysed transgenic lines; this expression was not sex-specific, but was surprisingly attenuated during pregnancy and lactation in females, concurrently with the increase in serum prolactin levels.

It is apparent, therefore, that milk protein transgenes can be correctly processed in mice, but also that the patterns of milk protein gene expression are far from clear or consistent even in endogenous mouse genes, and that the gene promoter is not necessarily the only important control region (compare the results of Andres et al., 1987, and Pittius et al., 1988). It may also be that the linking of DNA sequences not naturally found together in the shape of chimeric transgenes can cause an altered pattern of tissue-specific response (Lee et al., 1989).

The following chapter describes the observation of hormonally regulated and tissue-specific expression of guinea pig α-lactalbumin in transgenic mice.
Results

5.2 The expression of guinea pig α-lactalalbumin in transgenic mice

5.2.1 Milking Transgenic Mice.

Expression of RNA polymerase II transcribed genes results in the production of a protein, which in the case of α-lactalalbumin is directed to the golgi and secreted. It was observed that C127 cells transfected with pLJ1, a BPV based plasmid encoding the guinea pig α-lactalalbumin gene (Laird et al., 1988), successfully secreted guinea pig α-lactalalbumin into their culture medium. If the transgene is being correctly expressed in the transgenic mouse, then the presence of guinea pig α-lactalalbumin in the milk of transgenic mice is to be expected. In order to test for the production of guinea pig α-lactalalbumin in transgenic mouse milk, it is first necessary to milk the mice. As the offspring of the transgenic females may also be transgenic, and therefore valuable, it is not desirable to remove the young for an extended period of time or to use lactation inducing drugs such as oxytocin in order to facilitate the milking of the mouse. Attempts to milk lactating transgenic mice after removal of the young for a few hours were unsuccessful, so milk was obtained by sacrificing a single pup in the litter and removing the contents of its stomach, which is clearly visible through the abdominal skin of the infant mouse.

Pups from the litters of both normal and transgenic mothers were killed by decapitation and the stomach was removed. The entire contents, which have the appearance of a thick white paste, were transferred to an eppendorf tube and solubilised in 100μl of 1% SDS. The protein concentrations were determined using the Pierce BCA protein assay.

Milk from normal mice and from guinea pigs was obtained by expression of milk from lactating mice and guinea pigs obtained from Porcellus.
5.2.2 Analysis of transgenic mouse milk for secretion of guinea pig α-lactalbumin.

The volume of milk corresponding to 100μg of protein was determined for each of the milks obtained from guinea pig, CBA/J mouse, transgenic mouse, and the non-transgenic offspring of both non-transgenic and transgenic mouse lines. These quantities of milk, plus 10μg of purified guinea pig α-lactalbumin, were freeze-dried and redissolved in 25μl of Maizel loading buffer (section 2.2.13). After heating to 75°C for 4 minutes, the samples were loaded onto a 12% SDS-polyacrylamide gel (section 2.2.13) and electrophoresed until the bromophenol blue had reached the bottom of the gel.

The proteins were transferred to an Amersham Hybond-C nitrocellulose filter by electroblotting at 210mA overnight in a Bio-Rad blotting cell (section 2.2.14) and subsequently probed as described with the rabbit anti-α-lactalbumin antiserum. The blot was photographed as soon as the image had fully developed (fig 5.1).

The milk of the line 101.1 mice, and to a marginally lesser extent the milk of line 101.8 mice, can be seen to contain a large amount of a protein species the same size as guinea pig α-lactalbumin which reacts with the anti guinea pig α-lactalbumin antiserum (fig. 5.1, lanes 6 and 8). The antiserum used is not purely anti-α-lactalbumin, and detects other components of guinea pig milk, including the caseins, apparent as higher molecular weight bands (fig. 5.1, lane 2). It is not to be ruled out that some of the 'spread' visible in the α-lactalbumin band in guinea pig milk is due to reaction with some lower molecular weight compounds from guinea pig milk to which the antibody may also be sensitive. Mouse milk, however, is clearly free from any such reaction; and while it is possible, and indeed probable, that this antiserum will react with mouse α-lactalbumin, such a reaction was not evident even when 500μg of CBA/J mouse milk was analysed. It is therefore a safe deduction that the 14.5kd band in transgenic mouse milk is due to the expression of guinea pig α-lactalbumin at high levels. The levels of guinea pig α-lactalbumin apparent in the milk of transgenic mice indicates that the transgene is being expressed at levels commensurate with, if not equal to, those of the endogenous gene in the guinea pig.
Figure 5.1

Analysis of transgene expression in milk of transgenic mice

The protein contents of each milk sample were measured by the Pierce BCA protein assay. The volumes of milk corresponding to 100µg of protein from guinea pig, normal mouse and transgenic mouse milks were freeze dried, loaded onto a denaturing 12% SDS/polyacrylamide gel and separated by electrophoresis. The proteins were then blotted to a Schleicher and Schuell nitrocellulose filter and probed with a rabbit anti-guinea pig α-lactalbumin antiserum.

1 10µg purified guinea pig α-lactalbumin
2 100µg whole guinea pig milk proteins
3 100µg normal CBA/J mouse milk proteins
4 500µg normal CBA/J mouse milk proteins
5 100µg milk proteins from non-transgenic offspring of mouse 99.16 (99.16.5)
6 100µg milk proteins from transgenic offspring of mouse 101.1 (101.1.6)
7 100µg milk proteins from non-transgenic offspring of mouse 101.1 (101.1.3)
8 100µg milk proteins from transgenic offspring of mouse 101.8 (101.8.4)
9 100µg milk proteins from non-transgenic offspring of mouse 101.8 (101.8.1)
10 100µg milk proteins from non-transgenic offspring of mouse 101.8 (101.8.2)

The bromophenol blue loading dye front is visible (BPB).
5.3 Determination of tissue specificity of gene expression.

5.3.1 Northern blot analysis of transgenic mouse tissues

It is known that milk protein genes are not expressed exclusively in mammary cells, and
that various researchers have reached differing conclusions about the tissue distribution of both
milk protein transgenes and endogenous milk protein genes in transgenic mice (see section
5.1). Having determined that the transgene is expressed in the mammary glands of transgenic
mice, the RNA present in various tissues of a line 101.1 lactating transgenic mouse was analysed
for the presence of guinea pig α-lactalbumin mRNA.

A lactating transgenic mouse was killed three days post-partum. Mammary gland, liver,
kidney, spleen, heart, lung, salivary gland, thymus and skin tissues were removed from the
carcass immediately and snap frozen in liquid nitrogen. Total cellular RNA was then prepared by a
modification of the guanidinium/ acid phenol method of Chomczynski and Sacchi (1987), as
described in section 2.2.8. The filter was probed with the 205bp pgpK9 Rsa 1 fragment (fig. 4.4),
and autoradiographed overnight at -70°C using Fuji film.

The result (fig 5.2) shows a 750bp band corresponding to guinea pig α-lactalbumin mRNA
in the lactating guinea pig mammary gland control, and in the mammary gland and skin of the
transgenic mouse. The levels of message in skin and mammary tissues are surprisingly similar.
This blot was then re-autoradiographed for 1 week at -70°C, using Kodak X-AR film. The extra
exposure and extra sensitivity of the film allow the visualisation of bands of identical size in salivary
gland, heart and lung of transgenic mouse (fig. 5.3a). The same filter was washed in two changes
of boiling water, for 5 minutes each, followed by shaking at 80°C in 0.1% SDS for 20 minutes to
strip it of the α-lactalbumin probe, and reprobed with mouse actin (Fig.5.3b).

The guinea pig α-lactalbumin cDNA probe K9/Rsa does not hybridise appreciably to
mouse sequences in Southern analysis (see chapter 4) and does not detect any RNA species in
lactating non-transgenic mouse tissues. In lactating mice from both transgenic lines, however, a
mRNA of 750bp is detected. This corresponds to the size of correctly processed guinea pig α-
lactalbumin mRNA, and is expressed specifically in the lactating mammary gland and the skin of
lactating mice. Skin apart, there is negligible expression of this mRNA species in any other tissue
which was assayed. Line 101.8 showed expression of an identical RNA species at approximately
equal levels to line 101.1 (fig. 5.4) in both mammary and skin tissues. As these mice, on the basis of Western and Northern blotting data, appear to be functionally identical with regard to transgene expression, subsequent experiments concentrated on mice from line 101.1.
Figure 5.2

Northern blot analysis of RNA prepared from transgenic mouse tissues

In order to determine the expression pattern of the α-lactalbumin transgene, RNA prepared from various tissues of a lactating transgenic mouse was size-separated by electrophoresis on a 1.6% formaldehyde/agarose gel and transferred to an Amersham Hybond-N nylon membrane. The blot was probed with a 205bp fragment from the guinea pig α-lactalbumin cDNA clone pgpK9, and exposed overnight at -70°C. The positions of the ribosomal RNA markers visualised by ethidium bromide staining are shown.

1 lactating guinea pig mammary gland
2 lactating transgenic mouse mammary gland
3 lactating transgenic mouse kidney
4 lactating transgenic mouse thymus
5 lactating transgenic mouse small intestine
6 lactating transgenic mouse salivary gland
7 lactating transgenic mouse heart
8 lactating transgenic mouse lung
9 lactating transgenic mouse liver
10 lactating transgenic mouse skin
Figure 5.3
Extended exposure allows detection of transgene mRNA heart, lung
and salivary gland

A. The blot used in fig. 5.2 was re-exposed for 1 week at -70°C using Kodak X-AR film.

1 lactating transgenic mouse skin  
2 lactating transgenic mouse liver  
3 lactating transgenic mouse lung  
4 lactating transgenic mouse heart  
5 lactating transgenic mouse salivary gland  
6 lactating transgenic mouse small intestine  
7 lactating transgenic mouse thymus  
8 lactating transgenic mouse kidney

B. The same blot was stripped by washing twice in boiling water for 5 minutes, followed by 20 minutes at 80°C in 0.1% SDS, and subsequently reprobed with mouse β-actin cDNA.

1 lactating transgenic mouse mammary gland  
2 lactating transgenic mouse kidney  
3 lactating transgenic mouse thymus  
4 lactating transgenic mouse small intestine  
5 lactating transgenic mouse salivary gland  
6 lactating transgenic mouse heart  
7 lactating transgenic mouse lung  
8 lactating transgenic mouse liver  
9 lactating transgenic mouse skin

A 1.9kb actin band is visible in all lanes. Guinea pig mammary gland mRNA did not give rise to a signal when probed with mouse actin.
Figure 5.4

Comparison of RNA from both transgenic mouse lines

Total RNA was prepared from the mammary gland, liver and skin of a lactating transgenic mouse from line 101.8. Comparison with RNA from line 101.1 mammary gland shows the presence of an identical band, confirming that the transgene is expressed in both lines of transgenic mice.

1  2μg lactating guinea pig mammary gland RNA (narrow lane)
2  10μg lactating line 101.1 mouse liver RNA
3  10μg lactating line 101.8 mouse liver RNA
4  10μg lactating line 101.8 mouse mammary gland RNA
5  10μg lactating line 101.8 mouse skin RNA
6  10μg lactating line 101.1 mouse mammary gland RNA
5.3.2 Determination of the transcription start site by S1 nuclease protection analysis.

In order to confirm that the observed guinea pig α-lactalbumin message is being correctly initiated in the transgenic mouse, the 5' end of the message was mapped using S1 nuclease. A suitable DNA probe from the 5' end of the guinea pig α-lactalbumin gene was isolated from pgp H68. 20μg of DNA was digested with Hind III and Hpa II, and the 1300bp fragment spanning from the promoter to exon 1 was isolated from a 1% agarose gel with the Geneclean kit. This fragment was digested with Rsa I, and the resulting 900 and 400bp bands were separated on a 1% agarose gel. The 900bp band was eluted with the Geneclean kit, and end labelled with T4 polynucleotide kinase as described by Maxam and Gilbert (1980). The unincorporated nucleotides were removed by centrifugation at 1900 rpm in an IEC Centra-R benchtop centrifuge on a 1ml Sephadex G50 spun column. 10^4 cpm of purified probe was hybridised as described (section 2.2.12) to total RNA prepared from lactating guinea pig mammary gland and skin, to total RNA from lactating transgenic mouse mammary gland, liver, kidney and skin, and to total RNA prepared from lactating non transgenic mouse mammary gland as a negative control. Samples were subsequently digested with 100u of S1 nuclease. In addition, to better determine at which nucleotide transcription was initiating, bacteriophage M13 was sequenced using the Sequenase sequencing kit. The products of the S1 digestion and of the sequencing reactions were electrophoresed on a 6% polyacrylamide/urea sequencing gel and autoradiographed overnight at -70°C.

A doublet of bands is visible with guinea pig mammary gland RNA (fig. 5.5), and the same doublet appears in transgenic mouse mammary gland and transgenic mouse skin. The size of the major protected fragment is 133bp as determined by comparison with the M13 sequence, and corresponds to the protection of an RNA fragment of the size which would be predicted if the previously mapped start site for guinea pig α-lactalbumin is being employed (Laird 1985). A smaller protected fragment is also visible, 127bp in size, which must correspond to a secondary initiation site of guinea pig α-lactalbumin mRNA: this fragment has been previously observed (Laird 1985) but has never been explained.
Figure 5.5

S1 mapping analysis of transgene mRNA

The transcriptional start site employed by the transgene in the mammary gland and skin of transgenic mice was mapped by S1 nuclease analysis. A 900bp probe spanning the promoter and first exon of the guinea pig α-lactalbumin gene was obtained by digestion of pgpH68 with *Hpa* II and *Rsa* I, labelled with γ-32P ATP and T4 polynucleotide kinase (Maxam and Gilbert, 1980). This probe was hybridised to 20μg total RNA from various tissues; S1 nuclease was then used to digest the single-stranded unhybridised material, and the samples were analysed by electrophoresis on a 6% polyacrylamide-urea gel.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>undigested probe</td>
</tr>
<tr>
<td>1</td>
<td>lactating guinea pig liver mRNA</td>
</tr>
<tr>
<td>2</td>
<td>lactating guinea pig mammary gland mRNA</td>
</tr>
<tr>
<td>3</td>
<td>lactating guinea pig skin mRNA</td>
</tr>
<tr>
<td>4</td>
<td>lactating transgenic mouse mammary gland mRNA</td>
</tr>
<tr>
<td>5</td>
<td>lactating transgenic mouse skin mRNA</td>
</tr>
<tr>
<td>6</td>
<td>lactating transgenic mouse liver mRNA</td>
</tr>
<tr>
<td>7</td>
<td>lactating normal mouse mammary gland mRNA</td>
</tr>
</tbody>
</table>
5.4 Hormonal regulation of the guinea pig α-lactalbumin transgene.

Milk protein genes are extensively regulated by changes in the hormonal environment which take place during the development of pregnancy and the onset of lactogenesis (reviewed by Banerjee, 1976; Topper and Freeman, 1980; Tucker, 1988). In the guinea pig, casein mRNA is produced at high levels, but is not detectable before partuition (Burditt et.al., 1982). α-Lactalbumin, on the other hand, is detectable from 2 days pre-partum, and is found to decline after birth. In the mouse, on the other hand, casein mRNA is detectable in mid-pregnant mammary tissues (Takemoto et.al., 1980). The synthesis of casein mRNA reflects the differences in mammary gland development which are evident between the two species; whereas the mouse mammary gland can be seen to develop throughout pregnancy (Anderson, 1978), in the guinea pig mammary proliferation only begins a few days before the birth of the young and peaks 2 to 6 days post-partum (Nelson, 1962). The rat has a similar pattern of mammary development to the mouse (Anderson, 1978), and in rat mammary gland α-lactalbumin mRNA can be seen to increase after the first week of gestation (Nakhasi and Quasba, 1979). Although Takemoto et.al. (1980) were unable to detect mouse α-lactalbumin in pre-partum mouse mammary gland due to its low levels of expression, it is reasonable to postulate that α-lactalbumin mRNA levels should increase gradually in mouse mammary gland during pregnancy. Consequently, in addition to determining whether the transgene contains sufficient sequence to be modulated by hormones in the mouse, it is of interest to determine whether the guinea pig or the rat pattern of expression is followed in the transgenic mouse.

Accordingly, various tissues (see fig. 5.6) were isolated from virgin transgenic mouse, one week pregnant transgenic mouse, two week pregnant transgenic mouse, lactating normal mouse and male transgenic mouse. RNA was prepared from these tissues, analysed by Northern blotting (section 2.2.11) and autoradiographed for 72 hours. α-lactalbumin mRNA is detectable in two week pregnant transgenic mouse mammary gland, as well as the guinea pig control, but not in one week pregnant transgenic mouse mammary gland, or in virgin transgenic mouse (Fig 5.6a). Further analysis of tissues from virgin, male and 7-day pregnant transgenic mice reveals that guinea pig α-lactalbumin is detectable as a faint band in mammary gland after 7 days of
pregnancy (fig. 5.7). The gene shows similar patterns of expression in the skin, being strongly expressed after two weeks of pregnancy, but is not detectable after 1 week of pregnancy at this level of sensitivity. Other transgenic mouse tissues, as expected, show no expression of guinea pig α-lactalbumin in two week pregnant transgenic mouse. This pattern of expression is similar to that seen in mouse WAP, and distinctly different from that of mouse β-casein, which is expressed strongly from early pregnancy (Rosen, 1987). Therefore the guinea pig transgene, while retaining the regulatory characteristics of a whey protein gene as opposed to a casein gene, is found to behave as dictated by the physiology of the mouse, and is expressed strongly from mid-pregnancy, but very weakly in early pregnancy.
The hormonal regulation of transgene mRNA in transgenic mice

A and B: Total RNA was extracted from various tissues of virgin transgenic mouse, one week pregnant transgenic mouse, two week pregnant transgenic mouse and male transgenic mouse. 20μg of each sample was analysed by Northern blotting (section 2.2.11). The blot was autoradiographed at -70°C for 72 hours.

A: GP lactating guinea pig mammary gland RNA
1 7 day pregnant transgenic mouse mammary gland RNA
2 14 day pregnant transgenic mouse mammary gland RNA
3 14 day pregnant transgenic mouse liver RNA
4 14 day pregnant transgenic mouse kidney RNA
5 14 day pregnant transgenic mouse spleen RNA
6 14 day pregnant transgenic mouse small intestine RNA
7 14 day pregnant transgenic mouse heart RNA
8 14 day pregnant transgenic mouse lung RNA
9 14 day pregnant transgenic mouse thymus RNA
10 14 day pregnant transgenic mouse brain RNA
11 14 day pregnant transgenic mouse abdominal skin RNA
12 transgenic male mouse liver RNA
13 transgenic male mouse skin RNA
14 transgenic male mouse mammary fat pad RNA

B: GP lactating guinea pig mammary gland RNA
1-8 virgin transgenic mouse mammary gland, liver, kidney, spleen, heart, lung, thymus and skin RNA
9-16 lactating non-transgenic mouse mammary gland, liver, kidney, spleen, heart, lung, thymus and skin RNA

165
A: RNA from the same samples as those used in fig. 5.4a,b was electrophoresed on a 1.6% formaldehyde/agarose gel using wide lane slots. The gel was analysed by Northern blotting and hybridisation to the guinea pig α-lactalbumin cDNA probe K9/Rsa. The blot was autoradiographed for 1 week at -70°C.

1 lactating transgenic mouse mammary gland RNA
2 virgin transgenic mouse mammary gland RNA
3 virgin transgenic mouse skin RNA
4 7 day pregnant transgenic mouse mammary gland RNA
5 7 day pregnant transgenic mouse skin RNA
6 male transgenic mouse mammary fat pad RNA

B: The same blot was reprobed with a mouse β-actin cDNA probe, and autoradiographed for 72 hours at -70°C.
5.5 The expression of milk protein genes in mouse skin

5.5.1 The expression of mouse β-casein in lactating mice

Having determined the presence of a high level of α-LA mRNA in the skin of transgenic animals (fig. 5.2), which appears to be correctly initiated and processed (fig. 5.4), a mouse β-casein cDNA probe pCMβ13, obtained from Rosen's laboratory, was used to control for endogenous milk protein gene expression in the mammary glands of transgenic mice. The blot described in fig. 5.2 was therefore washed twice for 5 minutes in boiling water, and for 20 minutes in 0.1% SDS at 80°C before being reprobed with oligo-labelled pCMβ13, and washed in 0.1x SSC at 65°C. RNA hybridising to the mouse β-casein cDNA was found to be present not only in the mammary gland of the lactating transgenic mouse, but also in the skin, at levels commensurate with those seen in the mammary gland and with those achieved by the α-lactalbumin transgene (fig. 5.8a). In order to determine whether this unexpected expression in skin is due to a possible alteration in the patterns of milk protein gene expression caused by the presence of the transgene, RNA was prepared from skin, mammary gland and liver of a lactating non-transgenic CBA mouse. 20μg of each RNA sample, plus 20μg control samples from a lactating transgenic mouse and a male transgenic mouse were analysed by Northern blotting and hybridisation with the same mouse β-casein cDNA probe. The presence of β-casein mRNA in non-transgenic lactating mouse skin was confirmed by this experiment (fig. 5.8b), suggesting that the skin-specific expression of milk protein mRNA is not an aberration caused by the presence of the transgene, but is physiologic in CBA mice.

5.5.2 α-lactalbumin expression in guinea pig skin

In order to determine whether the expression of milk proteins in the skin is restricted to CBA mice, the skin of lactating guinea pig was analysed for the presence of guinea pig α-lactalbumin. Total RNA was prepared from lactating guinea pig skin as described in section 2.2.8. and analysed by northern blot analysis (fig. 5.9). Overexposure of the blot reveals the presence of distinct bands in the guinea pig skin track. These are also present in the guinea pig mammary gland track, a plurality of bands being visualised apparently due to the extreme overexposure of this blot. The signal in the skin track would appear to be real, and not an artifact. This suggests
guinea pigs express $\alpha$-lactalbumin in skin tissues during lactation, and that this phenomenon may not be limited to mice.
Figure 5.8

Detection of β-casein in lactating mouse mammary gland

and skin tissues

A. The blot used in fig. 5.2 was stripped of α-lactalbumin probe by washing in boiling water and reprobed with the mouse β-casein cDNA probe pCMβ13 and autoradiographed overnight at -70°C. The presence of β-casein mRNA at very high levels in mammary gland and skin tissues is apparent.

1 lactating guinea pig mammary gland
2 lactating transgenic mouse mammary gland
3 lactating transgenic mouse kidney
4 lactating transgenic mouse thymus
5 lactating transgenic mouse small intestine
6 lactating transgenic mouse salivary gland
7 lactating transgenic mouse heart
8 lactating transgenic mouse lung
9 lactating transgenic mouse liver
10 lactating transgenic mouse skin

B. RNA extracted from normal lactating mouse, as well as lactating transgenic mouse and male transgenic mouse, was analysed by northern blotting using a mouse β-casein cDNA probe. The blot was autoradiographed overnight at -70°C. The positions of the ribosomal RNA bands visualised by ethidium bromide staining of the gel are indicated.

1 lactating normal mouse skin
2 male transgenic mouse liver
3 male transgenic mouse skin
4 lactating normal mouse liver
5 lactating normal mouse mammary gland
6 lactating transgenic mouse liver
7 lactating transgenic mouse mammary gland

C. The same filter was rehybridised with a mouse β-actin cDNA probe, and autoradiographed for 72 hours at -70°C.
Total RNA was prepared from the skin and mammary glands of a lactating guinea pig as described in section 2.2.8. 20μg of RNA were separated on a 1.6% agarose/formaldehyde gel and probed with the 205bp guinea pig α-lactalbumin cDNA probe. The gel was washed in 0.2% SSC at 65°C before being exposed to Kodak X-AR film at -70°C for 1 week. The overexposed autoradiograph clearly shows the presence of signal in guinea pig skin.
5.5.3 Guinea pig α-lactalbumin protein is not detectable in transgenic mouse skin

In order to assay for the presence of guinea pig α-lactalbumin in the skin of transgenic mice, 2 cm\(^2\) of skin from a lactating transgenic mouse was homogenised in 2 ml 1\% SDS using a polytron, and subsequently boiled for 10 minutes. Fur and other debris was removed by centrifugation in a microfuge, and the supernatant assayed for total protein using the Pierce BCA protein assay kit. 100\(\mu\)g of skin extract protein was electrophoresed on a 10\% polyacrylamide/SDS gel, and blotted to Hybond-C nitrocellulose. The filter was probed with a rabbit anti-α-lactalbumin antiserum as described (section 2.2.14).

The result (fig. 5.10) shows that whereas guinea pig α-lactalbumin is readily detectable in subcellular fractions of guinea pig mammary gland, it is not present in an equivalent amount of protein extract made from transgenic mouse skin. It should be borne in mind, however that it is difficult to extract protein from skin (D. Katz, personal communication), and that a definitive result would necessitate the use of more sensitive techniques such as immunocytochemistry.

Protein was extracted from transgenic mouse skin by tissue homogenisation in SDS. The amount of protein present in the extract was determined by the Pierce BCA assay. At the same time, protein was extracted from a golgi enriched fraction of lactating guinea pig mammary gland. Bio-Rad protein size markers were used to estimate protein size.

1 Bio-Rad protein size markers
2 empty lane
3 1μg purified guinea pig α-lactalbumin
4 100μg skin protein extract
5 500μg skin protein extract
6 empty lane
7 100μg protein extract from lactating guinea pig mammary gland golgi-enriched fraction
5.6 High-level expression of the α-lactalbumin transgene

In chapter 4 the creation of a line of transgenic mice carrying the guinea pig α-lactalbumin gene was reported. The behaviour of multiple independent, and yet apparently linked, insertions of a transgene has not been described by other groups; and the importance of position effects, such as those described by Greaves et al. (1989), makes the expression of any transgene an uncertain matter. In addition to this, the previously reported results obtained with milk protein transgenes imply that extremely low levels of expression should be expected (see section 5.1).

This, however, is not the case. Guinea pig α-Lactalbumin is expressed at very high levels in the transgenic mice, a situation which differs from those previously described on several counts. Firstly, the level of expression is one or two orders of magnitude higher than that seen with other milk protein genes (section 5.1) with the exception of the sheep β-lactoglobulin gene expressed in transgenic mice by Simons et al. (1987). In this case transgene expression was evident in mouse milk and in mammary tissues at levels comparable to those seen in the sheep. β-lactoglobulin is the major whey protein of ruminants, but is absent in the mouse; α-lactalbumin is present in mouse milk, but at very low levels; yet both these transgenes are expressed at very high levels in transgenic mice. In contrast, WAP is a major protein of mouse milk (Hennighausen and Sippel, 1982), yet WAP constructs are expressed at very low levels in transgenic mice (Andres et al., 1987; Pittius et al., 1988). Presumably this must either be due to the effects of deleting important sequences in the structural portion or the 3’ untranslated region of the WAP gene, or to the epistatic effects described by Palmiter and Brinster (1986) which are brought about by the interaction of artificially juxtaposed DNA elements.

This theory has also been advanced by Rosen to explain the thymus-specific aberrant expression of rat β-casein-CAT constructs in transgenic mice (Lee et al., 1989). Both the β-casein-CAT constructs and a β-casein minigene were expressed at very low levels in transgenic mice, which again is difficult to reconcile with the high levels of expression seen with guinea pig α-lactalbumin transgene. The only common factor in these data is that when entire milk protein transgenes are studied, such as the guinea pig α-lactalbumin and sheep β-lactoglobulin genes.
(this chapter; Simons et al., 1987), high levels of expression can be recreated in transgenic mice; however if truncated constructs are used, created either by promoter swapping (Andres et al., 1987; Pittius et al., 1988; Lee et al., 1989) or minigene construction (Lee et al., 1988) the levels of expression are substantially reduced and inappropriate expression can occur. It is notable that both the β-casein minigene and the β-casein-CAT construct of Lee et al. (1988, 1989) contained the noncoding first exon and the first intron of the rat β-casein gene, yet this did not prove sufficient to give high-level expression.

5.7 Tissue specificity and hormonal regulation of milk protein transgenes

On the other hand, broadly correct tissue-specific and hormonal regulation have been demonstrated by virtually all milk protein gene constructs injected into transgenic mice. Therefore it would seem that while the sequences present in the promoters of milk protein genes are not sufficient to direct high-level expression, they are capable of responding to hormonal stimulation, and are usually virtually inactive in incorrect tissues or in the absence of hormones. The notable exception is the rat β-casein-CAT construct of Lee et al. (1989), which is expressed in thymus, and is actually down-regulated during pregnancy and lactation. The reason for this remains unclear.

The extremely low levels of 'inappropriate' expression of the guinea pig α-lactalbumin transgene seen in heart, lung and salivary gland, and the similar phenomena reported by other groups, may be a reflection of naturally occurring 'inappropriate' expression. Pittius et al. (1988) observed endogenous mouse WAP expression in a variety of tissues, though perplexingly not in the same tissues as documented by Andres et al. (1987). It is possible that low-level aberrant expression may even be strain-specific in mouse: Andres et al. and Pittius et al. used different strains of mice.

The early onset of transgene mRNA synthesis in transgenic mice is an indication that in mouse, as in rat (Nakhasi and Quasba, 1979), α-lactalbumin is expressed after the first week of gestation. The parallels in mammary development which exist between the 2 rodents would lead us to expect this, and the implication is that the transgene has responded to the controlling
influences of the physiological stimuli present in the mouse, even though these differ markedly from those of the guinea pig. If we accept this, it can be inferred that the transgene control sequences, while successfully mediating the physiological signals dictating the tissue specificity and timing of α-lactalbumin gene expression, do not respond to any such signals which may dictate the level of expression: it is evident that the levels of transgene expression are commensurate with those seen in the guinea pig, rather than those of the endogenous mouse α-lactalbumin gene. Therefore, the limiting factor in high level expression must be presumed to rest with the sequence of the endogenous gene rather than in the physiology of the mouse; the cloning and sequencing of mouse α-lactalbumin gene would clarify this point, and comparison between mouse and guinea pig α-lactalbumin genes may provide important data to those who may wish to maximise the levels of expression from milk protein transgenes.

5.8 High level transcription of milk protein mRNA in mouse skin

Section 5.5 describes the identification of guinea pig α-lactalbumin mRNA in the skin of transgenic mice. While in the first instance this was taken to be an example of inappropriate expression, experiments using a mouse β-casein probe to detect endogenous mouse β-casein mRNA in both transgenic and normal CBA/J mouse skin reveal that skin-specific expression of β-casein mRNA is normal in CBA/J mice. Given the unavailability of a probe for mouse α-lactalbumin, this evidence strongly suggests that the observed transgene expression in skin is not inappropriate, but that CBA/J mice naturally express α-lactalbumin as well as β-casein mRNA in their skin tissues. The fact that this observation has never before been reported is curious. Pittius et al. (1988) did not detect the expression of endogenous mouse WAP mRNA in the skin of ICR/B6C3 F1 hybrid transgenic mice, but apart from this report data on skin do not appear in the literature concerning lactation. In addition, work on the expression of milk protein genes in normal animals was carried out mostly before the availability of RNA detection methods, and thus does not document the extra-mammary production of milk protein mRNA. As it would appear that α-lactalbumin protein is not present in skin (section 5.4), the lack of data pertaining to skin-specific expression is explained.
The reasons for the lack of guinea pig α-lactalbumin protein in extracts prepared from lactating transgenic mouse skin could be severalfold. In the first instance, the message may not be translated. Such an event is probably not due to rapid degradation of the mRNA: this is not evident in Northern blot analysis. A specific block of translation, such as that described by Casey et.al. (1988), is a possibility, but such a system has never been shown to operate in milk protein gene expression. Another consideration is that α-lactalbumin may be translated, but subsequently degraded. This situation is appealing, as the skin cells in which the α-lactalbumin message is transcribed may lack the sub-cellular secretory organelles necessary for correct processing of α-lactalbumin polypeptides; and furthermore the differentiation of skin epithelium into keratinised epidermal cells and sebum in sebaceous glands involves the extensive degradation of protein (Wheatley, 1986).

Additional interest is generated if it is considered that skin and mammary tissues share a closely related embryology: both tissues are derived from the germinal layer of the foetal ectoderm (Beck et.al.,1985). Furthermore, the skin is a highly diverse organ, consisting of a variety of specialised structures; the mammary gland can be regarded as a particular skin appendage with a very specialised function (Raynaud, 1961). Given the knowledge that milk protein gene expression requires the secretory epithelial cells to achieve certain levels of structural organisation (Bissel and Hall, 1987), it appears desirable to determine the site of α-lactalbumin mRNA synthesis in the skin of transgenic mice. To this end the localisation of guinea pig α-lactalbumin mRNA in various tissues of lactating transgenic mouse was examined by in situ hybridisation. These experiments are described in chapter 6.
Chapter 6
The identification of cell-specific α-Lactalbumin transgene expression by in situ hybridisation

Introduction

6.1 The relationship between mammary gland and skin

6.1.1 Mammary gland and skin epithelia share a common evolutionary background

The early embryo is covered by a single layer of cuboidal cells known as the ectoderm (see Beck et al., 1985). This is subsequently covered by the periderm, a layer of flattened cells. Beneath this layer, the ectoderm develops into the skin, and its associated organs; the specialised developments of the ectoderm - the mammary glands, hair follicles, sweat glands and nails - all develop from thickenings of the germinal layer. The mammary gland has been described as a specialised sweat gland, sebaceous gland, hair follicle or undifferentiated cutaneous gland (see Sakakura, 1987); yet none of these definitions is really satisfactory (Raynaud, 1961), as the organ appears to develop as a specialised structure in its own right, via epithelial-mesenchyme interactions (Sakakura, 1987).

The development of the mouse mammary gland has been covered by Topper and Freeman (1980) and Sakakura (1987), but the development of the human gland is similar and is discussed by most embryology texts (Beck et al., 1985; Baxter, 1953).

6.1.2 The development of the mammary gland

The mammary glands of all species develop from an ectodermal thickening, which is present in the very early embryo, and which corresponds to the mammary band or ridge (also known as the milk lines) which run along the ventral surface of the body from the axillary region to the inguinal area. In the mouse this thickening is known as the mammary crest, and is formed on the 11th day of pregnancy. This crest subsequently concentrates into ten discrete entities, three
on each side of the thorax and two pairs in the inguinal region, corresponding to the eventual mammary glands of the adult female mouse. The segmentation is brought about by the migration of cells, as opposed to extensive localised growth. These rudiments then develop into bulb-shaped structures, the mammary buds, which are connected to the epidermis by an ectodermal collar, and grow rapidly downwards into the mammary fat pad precursor. These larger structures are known as the mammary sprouts, and it is at this stage, between the 13th and 15th days of pregnancy, that the mammary glands of male mice are inhibited. The mesenchyme which surround the mammary sprouts and is believed to 'instruct' their formation (Topper and Freeman, 1980; Sakakura, 1987) grows rapidly around the ectodermal collar of the mammary sprout in the male mouse, in response to fetal androgens. The collar is effectively pinched in two, isolating the sprout bulb from the epidermis; it remains implanted in the subepidermal mesenchyme (the mammary fat pad precursor). Further development in the male is slow, or strain-specific regression may even occur: about 50% of male BALB/c mice completely lack mammary glands. Nipple formation does not occur in the male mouse, and any surviving epithelium will persist as a blind duct (Sakakura, 1987).

In the pregnant female mouse, the sprout proliferates further during the last five or six days of gestation, and gives rise to the mammary cord. Invagination of the epidermis around the mammary cord gives rise to the nipple sheath, and a duct is formed in the mammary cord via fusion of intercellular vacuoles; this opens to the exterior. The cord may branch to a small extent before birth, and will continue to branch after birth during a period of isometric growth (Daniel and Silberstein, 1987). The ends of the branches subsequently develop characteristic end buds, which are responsible for an acceleration in branching and allometric growth. Unlike female mammary embryogenesis, the development of the mammary gland in the newborn female is dependent on the action of hormones (Topper and Freeman, 1980).

6.1.3 The development of other skin appendages

The development of the skin is described in the majority of embryological texts, and has been dealt with in depth in a series of books edited by Jarrett (1986).

Although skin is derived from both ectoderm and mesoderm, the specialised skin organs are exclusively of ectodermal origin. The pilar complex, which includes the hair follicle, the
sebaceous gland, the hair muscle and sometimes an apocrine gland, is derived from differentiating epidermal basal cells in the germinal layer, which come into contact with mesenchymal components. The early hair germ has an appearance close to that of a mammary bud, and grows at a slant through the dermis into the subcutaneous adipose tissue. At the same time, the follicle duct extends upwards through the epidermis to open to the exterior. From the hair germ develop three buds, which give rise to the hair bulb, the sebaceous gland, and the apocrine gland. These buds are enveloped in a fibrous layer of mesenchymal cells, which direct the further differentiation of the epithelial cells into the more specialised hair follicle structures.

The eccrine or sweat glands originate independently of the hair follicle, from a thickening of the germinal layer which is similar in appearance to those which give rise to the hair follicles themselves and the mammary glands. Thereafter the morphology differs, as the eccrine bud grows into the dermis as a thin cord; this subsequently develops a lumen, and the distal end of the cord develops into the secretory element.

It is therefore evident that the mammary gland and the skin possess an extensive developmental relationship; and that perhaps the expression of milk proteins in the skin is not unexpected. However, in the mammary gland specialised lobular structures are formed during pregnancy in order to synthesise milk; the synthesis of milk proteins is confined to these structures (Topper and Freeman, 1980). It is therefore of interest to examine the cell-specific expression of the α-lactalbumin transgene in transgenic mouse mammary gland, and in transgenic mouse skin; the technique of in situ hybridisation between cellular mRNA and recombinant complementary RNA probes permits this examination.
6.2 The localisation of mRNA by *In situ* Hybridisation

*In situ* hybridisation (ISH) involves the detection of mRNA within the cells of a tissue section, typically 5-7μm thick, by hybridisation with complementary RNA or cDNA probes. The techniques central to ISH were developed to allow the mapping of genomic sequences on chromosome preparations by DNA hybridisation (Gall and Pardue, 1971); the availability of cDNA clones has increased the specificity of this technique, allows the detection of specific mRNA sequences within cells (Venezky *et al.*, 1981), and the study of the level of expression and the distribution of differentially expressed transcripts in heterologous cells within a tissue or whole organism (Cox *et al.*, 1984). The use of single-stranded RNA probes greatly increases the sensitivity of these techniques (Cox *et al.*, 1984), and the identification of a single expressing cell within a tissue is possible (for example, see Hayashi *et al.*, 1986).

Various protocols have been developed to allow *in situ* probing of fixed or frozen tissues, and cultured cells (reviewed by Harper and Marselle, 1987). Sections from paraffin-embedded fixed tissues have the advantages of easy handling and superior conservation of cell morphology when compared to frozen tissue sectioning, and this was therefore the method of choice for this study. The tissue sections are treated with xylene and alcohol to dewax them, and then rehydrated; hybridisation solution is applied directly to the section, which is subsequently washed to the required stringency and coated with a layer of photographic emulsion in gel form. Upon developing, silver grains can be seen to have been precipitated in the proximity of signal from the $^{35}$S-labelled probe.
Results

6.3 *In situ* hybridisation of mouse sections

6.3.1 Subcloning of the guinea pig α-lactalbumin cDNA into a GEMINI vector.

The GEMINI series of vectors (Promega Biotec) contain both phage T7 and SP6 RNA polymerase promoters capable of *in vitro* transcription of RNA molecules from cDNA templates, arranged so as to direct transcription in opposing directions, and separated by a multiple cloning sequence (MCS). A cDNA inserted into the MCS can therefore be transcribed in either orientation, yielding mRNA or cRNA as required. Before transcription, GEMINI-based plasmids should be linearised by endonuclease digestion at a unique polylinker restriction site located at the distal end of the cDNA as viewed from the RNA polymerase promoter which is to be used to direct transcription. This causes the polymerase to terminate at the linearisation site, thus producing specific transcripts of a predetermined length.

The guinea pig α-lactalbumin cDNA clone pgpK9 (Hall *et al.*, 1982) was digested with *Bam H1* and *Nae 1* to release a cDNA fragment free of non-coding sequences and poly-A tail (Fig. 6.1), which was purified by gel electrophoresis on a 1.5% agarose gel and band elution with the Geneclean kit. 1 μg of the GEMINI vector pGEM-4z, which incorporates the lac-Z blue/white selection system, was digested in the first instance with *Bam H1*; subsequently, the *Bam H1* digestion buffer was made 100mM with respect to NaCl and the DNA was digested with *Sma 1*. The vector ends were dephosphorylated by addition of 1 μ Calf Intestine Alkali Phosphatase (CIAP) to the digestion reaction 10 minutes before the reaction was terminated, and the DNA was purified by phenol extraction and ethanol/NaAc precipitation. 100ng of vector DNA was ligated with between 10 and 100ng of K9 cDNA insert overnight at 15°C using 1 μ T4 DNA ligase and commercial ligation buffer, in a total reaction volume of 10μl. 50ng of each ligation mix was transformed into *E.coli* JM83 as described (section 2.2.3), and the cells were plated onto ψ-agar plates incorporating 100μg/ml ampicillin and spread with X-gal. After overnight incubation at 37°C, white colonies were picked and grown overnight in 5ml ψ-broth plus 100μg/ml ampicillin. Plasmids were prepared from these cultures (section 2.2.4b), restricted with *Eco R1* and *Hind III* to release the 438bp cloned cDNA insert, and analysed by electrophoresis on a 1.5% agarose gel. One of
the isolates showing successful ligation of the α-lactalbumin cDNA into pGEM-4z was grown up in 1L culture, and plasmid DNA was extracted by the CsCl gradient centrifugation method (section 2.2.4a).

Transcription of cRNA from this plasmid, pgpK98, proceeds from the T7 promoter. The RNA product of cold transcription (section 2.2.15b) was analysed by gel electrophoresis, and appeared to be of the correct size, approximately 450 nucleotides. Radiolabelled 35S cRNA probes were transcribed as described in section 2.2.15a, and the reaction was stopped by the addition of 200μl 2.5M NH₄Ac, 10mM DTT. 2μl of this solution were TCA precipitated and counted, to determine the activity of the probe; the rest of the probe was precipitated and taken up in a volume which resulted in the dilution of the probe to 2x10⁶cpm/μl.

6.3.2 Preparation of tissue sections

A lactating transgenic mouse and a lactating normal mouse, both immediately post-partum, were dissected, and the mammary glands, tail and abdominal skin, liver, kidney, heart, lung, oesophagus, small intestine, salivary gland, spleen and thymus were removed and fixed by immersion in 1% paraformaldehyde in PBS for several days. Mammary gland, being a fatty tissue, requires a long fixing period. When fixed, the tissue pieces were perfused with paraffin wax using a Miles Tissue-Tec VIP processing machine. The process involves post-fixation of the tissue in 2 changes of 10% formalin, followed by thorough dehydration through an alcohol series and clearing with xylene. The dehydrated tissue is then perfused with liquid paraffin wax. These tissue pieces were subsequently set in one-inch wax blocks using a Tissue-Tec embedding centre, at which hot paraffin wax is poured into moulds containing the tissue and a plastic base, and allowed to set. The mammary and skin tissues were mounted in separate blocks, but the organs were mounted together, making a total of six blocks.

Sections 5μm thick were cut on a Leitz base sledge microtome. The ribbon of sections obtained was floated on a 50°C water bath to allow the wax to expand, and any creases in the section to fall out. Individual sections were separated with forceps. These were floated onto poly-L-lysine coated microscope slides, previously prepared as described (section 2.2.16). Excess water was drained from the slides, and they were allowed to dry overnight at 37°C. Slides were
stored in a slide box at room temperature for one week or longer, to allow the sections to become firmly attached to the slide.
A 438bp fragment was excised from pgpK9 by digestion with *Bam H1* and *Nae I*. This fragment was ligated into the polylinker of pGEM4z at the *Bam H1* and *Sma I* sites, thus placing the cDNA insert in such an orientation that complementary RNA would be synthesised from the T7 promoter after template linearisation with *Eco RI*. 
guinea pig α-lactalbumin cDNA — pAT 153

pGEM4z

SP6

MCS

T7

Sma 1 Bam H1

Eco R1

Eco R1

Sma 1

Bam H1

Linearise with Eco R1

transcription from T7

α-lactalbumin cRNA

pgpK98
6.3.3 Hybridisation of tissue sections with α-lactalbumin cRNA

Sections were prepared basically as described by Angerer and Angerer (1981). Slides were dewaxed with xylene as described (section 2.2.17), rehydrated through an alcohol series, and treated with proteinase K to increase accessibility of the probe to the cytoplasmic mRNA species. After prehybridising the sections with 50% formamide, 2x SSC for 15 minutes at 37°C, 20μl of hybridisation solution containing 2×10^6 cpm of ^35S-labelled riboprobe was placed on a prewashed glass coverslip. Each slide was carefully wiped dry of prehybridisation solution, and was lowered section side down onto the drop of hybridisation solution on the cover slip. The hybridisation solution was allowed to spread evenly over the slide, taking care to exclude air bubbles. Incubation was carried out in plastic boxes, containing six slides each, lined on all sides with Whatman 3MM absorbent paper soaked in 2x SSC, 5% formamide. This creates a vapour-saturated environment, and prevents the evaporation of the hybridisation solution during overnight hybridisation at 43°C.

After hybridisation, slides were immersed into 4xSSC to allow the coverslips to float off. The sections were then washed as described in section 2.2.17, and treated with RNAse A to remove traces of unhybridised single stranded riboprobe. This step was followed by a high-stringency wash in 2x SSC at 55°C, with shaking, followed by 0.1x SSC at 55°C without shaking. The tissue sections were then dehydrated through an alcohol series, and allowed to air dry at room temperature for 2 to 3 hours.

6.3.4 Autoradiographic analysis of hybridised tissue sections

5-10ml of Ilford K5 gel-form photographic emulsion was spooned into a dipping chamber and melted at 43°C in a water bath. An equal volume of 2% glycerol, also at 43°C, was added slowly to the emulsion and the two liquids were mixed gently by slow stirring with a glass rod, taking care to avoid the formation of air bubbles. The emulsion was then left to stand in the 43°C water bath to eliminate any bubbles which had been formed. In order to control for the presence of air bubbles, a clean slide was dipped into the emulsion, and held up to the safelight; in this way any bubbles can be seen clearly in the emulsion coating the slide. When this appeared free of air bubbles, the test slides were dipped by sliding them into the emulsion and withdrawing them in one smooth movement. Emulsion was immediately wiped off the back of each slide, and they
were placed on an ice-cold metal plate for 10 minutes to allow the emulsion to set. They were then left to dry in a horizontal position, at room temperature, for 4 hours. Exposure was carried out in a light-proof slide box, kept dry by a layer of desiccated silica gel incorporated into the lid, for one week at 4°C.

The exposed slides were developed in undiluted Kodak D-19 developer for 2 1/2 minutes at room temperature, rinsed in 2% acetic acid for 30 seconds to stop the developing process, and fixed for five minutes in Ilford Amfix photographic fixer. When the emulsion had completely cleared, the slides were washed very gently under a running R.O. water tap for 15 minutes, stained with hematoxylin and eosin, dehydrated through an alcohol series and mounted with histomount.

6.4 Microscopic examination of developed and stained slides

The mounted slides were allowed to dry overnight, and examined by light microscopy. Both colour and black and white photographs were taken, taking advantage of the retractile nature of the silver grains in phase-contrast to show up the signal as white grains against a black background under a dark field view.

6.4.1 Mammary sections

The structure of the lactating mouse mammary gland differs considerably from that of the pre-pubescent organ. The virgin animal possesses a branched ductal structure, which is sheathed in myoepithelial cells; the ducts terminate in end buds (Daniel and Silberstein, 1987). During pregnancy, the end buds develop under the influence of hormones into the lobuloalveolar structures which secrete milk (Topper and Freeman, 1980; see fig. 6.2). Each lobule contains 10 to 100 alveoli, which consist of a single spherical layer of epithelial cells surrounding a central lumen. These cells differentiate into secretory cells, probably under the influence of progesterone and prolactin (Topper and Freeman, 1980) and undergo extensive development of secretory structures such as the golgi and RER. The alveoli or acini are covered
by a net of myoepithelial cells, which contract under oxytocin-mediated stimulation from suckling to express the milk from the lumen (Anderson, 1978; Tindal, 1978).

Mammary sections from a lactating transgenic mouse and a lactating non-transgenic littermate containing a view of a single mammary lobule were photographed under dark field and light field illumination (fig. 6.3; fig. 6.4). The transgenic mouse mammary gland shows strong expression of signal in the secretory epithelium of the acini, which is notably localised to certain cells. The signal is visible under light field illumination (fig. 6.3a), and is emphasised under the refractile conditions of dark-field phase contrast illumination. No signal is apparent in the adipocytes surrounding the lobules or in normal mouse mammary gland (fig. 6.4), even under dark field illumination.

A section showing the juncture between a mammary lobe and a lymph node was photographed in colour under bright field illumination (fig. 6.5). The signal is clearly restricted to the expressing cells of the mammary gland, and absent from mammary the lymph node. Thus in transgenic mouse mammary gland the guinea pig $\alpha$-lactalbumin message is expressed exclusively by the secretory epithelium, and is absent from other mammary cells: the guinea pig probe does not hybridise to mammary sections of non-transgenic mice.
Figure 6.2

The structure of the mouse mammary gland

Schematic diagram to show the spatial arrangement of mammary structures within the mouse mammary gland.

A. The ductal tree. This structure is formed during a period of ductal growth both before and after birth, and is not thought to be dependent upon the action of Hormones (Topper and Freeman, 1980).

B. The lobule is a collection of secretory acini.

C. Each acinus or alveolus is lined with mammary epithelium, and is the site of milk synthesis. Milk is secreted into the central lumen of the acinus, and flows into the ductule. These vessels are collected into subsegmental ducts and finally into segmental ducts, which terminate in a lactiferous sinus and the nipple.

D. The mammary gland is composed of a number of lobes, each consisting of a collection of lobules. The lobes are separated by layers of connective tissue, and each lobe secretes into a single segmental duct. The mouse mammary gland is distinct from those of humans or rabbits in having a single sinus, into which all the segmental ducts flow, rather than an individual sinus at the end of each segmental duct.
**Figure 6.3**

*In Situ* hybridisation of transgenic mouse mammary sections

5μm sections of lactating (CBA/J C57/B6)F₁ transgenic mouse mammary gland were hybridised *in situ* with radiolabelled guinea pig α-lactalbumin cRNA. The photographs show:

A. Mammary lobule viewed under bright field illumination. The black pepper-like grains are the silver grains deposited in the emulsion by localised radioactive emissions from the ³⁵S labelled probe. The surrounding adipose tissue is negative.

B. The same mammary lobule viewed under dark field phase-contrast illumination. The silver grains are highly refractile, and appear an intense white.

*Original magnification 160x*
5μm sections of lactating mammary gland from a CBA/J mouse were hybridised with a radiolabelled guinea pig α-lactalbumin cRNA probe.

A. Bright field illumination view of a mammary lobule.

B. Dark field view of the same section. The absence of signal in both photographs is evident when they are compared to those in fig. 6.3

Original magnification 160x
Figure 6.5

Colour view of mammary sections hybridised with guinea pig α-lactalbumin cRNA

A 5µm section of lactating transgenic mouse mammary gland was hybridised in situ with guinea pig α-lactalbumin cRNA. The section was then stained in haematoxylin and eosin, dried, and photographed using Kodak Ektachrome EP-50 (tungsten) colour transparency film.

The photograph shows the boundary between the lobules of expressing mammary tissue (ML) and a mammary lymph node (LN), which is not expressing the transgene.
6.4.2 Skin sections

Skin was cut from both the tail and the abdomen of lactating transgenic and non-transgenic mice. Examination of hybridised and stained sections by light and dark field microscopy revealed the association of signal with the pilar complex, in particular with the sebaceous gland. Fig. 6.6 shows a longitudinal section through a hair follicle, and in both light and dark field views signal is seen to be concentrated in the basal cells of the sebaceous gland, at the juncture between the sebaceous gland and the hair sheath epithelium. The holocrine glands of the sebaceous glands, whose nuclei are clearly visible in the light field view (fig. 6.6a) are negative. A similar structure viewed in transverse section (fig. 6.7) shows the presence of 2 sebaceous glands associated with the same hair follicle. No grains are visible under bright field illumination (fig. 6.7a), but a strong signal is apparent in the dark field view (fig. 6.7b). Once again, the signal is concentrated around the ducts between the hair sheath and the sebaceous glands. All other parts of the skin are negative, with the exception of a small island of signal visible just beneath the hair sheath. It is possible that this could identify a third, smaller sebaceous gland. Photomicrographs taken under light field illumination with colour film confirm the localisation of signal in the skin, and provide a clearer image of the surrounding morphology (fig. 6.8).

A section through a hair follicle of a non-transgenic mouse, clearly showing a sebaceous gland, confirms the specificity of probe hybridisation to transgene sequences (fig. 6.9).

6.4.3 Sections from other tissues

The sections cut from liver, kidney, heart, lung, oesophagus, small intestine, salivary gland, spleen and thymus of both transgenic and non-transgenic mouse were negative when photographed under light field illumination using colour film (fig. 6.10; most data not shown). This was confirmed under dark field illumination (data not shown), even in salivary gland, lung and heart, which had appeared positive on overexposure of northern blots (fig. 5.3).
In Situ hybridisation of sections from transgenic mouse skin

5μm sections were cut from skin from a lactating (CBA/J C57/B6)F1 transgenic mouse, and hybridised with radiolabelled guinea pig α-lactalbumin cRNA.

A. Light field view of a longitudinal section through a hair follicle in the skin of lactating transgenic mouse, showing the hair root structures, the hair itself emerging through the epidermis (E), which has a very thick keratinised layer, and a sebaceous gland (S) underneath the hair follicle (F). The hair (H) has been broken at the point of exit from the follicle, and lies at 90° to the lower part of the hair shaft at this point. Signal is visible at the periphery of the sebaceous gland, at the juncture of the gland and the hair follicle epithelium. There is no signal in the sebaceous gland cells themselves.

B. Dark field view of the same section. The high collagen content of skin causes a large amount of refraction from dermal tissues, and the surface of the hair appears refractile. Nevertheless, signal is clearly visible at the periphery of the sebaceous gland, in the basal layer. The signal is concentrated around the duct which leads from the sebaceous gland into the hair sheath. Note also confirmation of the relative absence of signal in the body of the sebaceous gland, the basal layer of the epidermis, and the hair root cells.

Original magnification 400x
Figure 6.7

_in situ_ hybridisation of sections from transgenic mouse skin

5 μm sections of abdominal skin from lactating transgenic mouse were hybridised with guinea pig α-lactalbumin cRNA.

A. Light field view of a transverse section through a hair follicle in the abdominal skin of a lactating transgenic mouse. Signal is not visible in this view, but the hair shaft (H) and 2 sebaceous glands (S) can be identified.

B. Dark field view of the same section. Signal is very strong around the sebaceous gland, but absent from the holocrine cells themselves and from the cells lining the hair follicle. A separate patch of signal is visible below the hair follicle; this could possibly be a small sebaceous gland or a sebaceous duct.

Original magnification 400x
Figure 6.8

Colour view of skin section hybridised with guinea pig $\alpha$-lactalbumin cRNA

$5\mu$m sections of skin from lactating transgenic mouse were hybridised \textit{in situ} with the $\alpha$-lactalbumin cRNA probe transcribed from pgpK98. The sections were stained in haematoxylin and eosin, dried and photographed under bright field illumination using Kodak Ektachrome EP-50 tungsten colour transparency film.

The photograph shows a transverse section through a hair follicle. The hair itself (H) has stained pink, and beside it the basal cells at one end of a sebaceous gland (S) are expressing $\alpha$-lactalbumin mRNA.
5 μm sections of abdominal skin from a lactating CBA/J mouse were hybridised with guinea pig α-lactalbumin cRNA.

A. Light field view of a longitudinal section through a hair follicle in normal mouse abdominal skin. This section avoids the hair itself and sections the sebaceous gland (S), in which the nuclei of differentiating cells are clearly visible, and the cells of the hair sheath (HS) and root bulb (R).

B. Dark field view of the same section. A large amount of refractile collagen is evident in this section, yet the absence of signal from the basal cells of the sebaceous gland, and every other part of the pilary structure, is evident.

Original magnification 400x
In situ hybridisation of non-expressing tissues shows a complete absence of signal.

5μm sections of lactating transgenic mouse heart, kidney, liver, salivary gland, thymus, spleen, oesophagus, small intestine and lung were hybridised in situ with guinea pig α-lactalbumin cRNA. Sections were stained in haematoxylin and eosin, dried and photographed using Kodak Ektachrome EP-50 colour transparency film.

A. Section through the kidney of transgenic mouse.

B. Section through lung of transgenic mouse. This tissue is negative, notwithstanding the appearance of faint α-lactalbumin bands in lung RNA in overexposed northern blots (fig. 5.3a).
Discussion

6.5 Correct transgene expression in mammary cells

Although several groups have reported the mammary-specific expression of milk protein transgenes (see section 5.1), only Simons' group (Harris, 1988) have shown expression of a transgene (sheep β-lactoglobulin) to be localised to the secretory epithelium in transgenic mice by immunohistochemical staining of mammary sections. The α-lactalbumin transgene mRNA is similarly localised to the secretory epithelium in mammary tissue. This demonstrates that the transgene expression is localised not only to the correct tissue, but also to the relevant cell types in the expressing tissue.

The transgene message is localised to roughly \( \frac{1}{2} \) to \( \frac{1}{3} \) of the epithelial cells rather than being uniformly expressed (fig. 6.3). It is known that mammary explants contain epithelial cells at different stages of differentiation, which can be synchronised by the addition of different hormones to the growth medium (Topper and Freeman, 1980). The cells vary between A-type, which have undergone little secretion-orientated differentiation, and C-type, which are fully active. In vivo, this development is noticed to proceed with some degree of asynchrony, the various acini developing at different rates (Mills and Topper, 1970). The epithelial cells within each individual acinus appear to develop at the same rate, an observation which is reflected in the message distribution in transgenic mouse mammary gland (fig. 6.3).

6.6 Transgene expression in the skin

6.6.1 The development of sebaceous glands

As is the case with the mammary gland and other skin appendages, the entire hair follicle, with which the sebaceous gland is associated, is derived from the stratum germinativum of the epidermis (Baxter, 1953), which is the fetal ectoderm. The sebaceous gland usually develops as a bud from the epithelial wall of the follicle, although it is sometimes observed as an independent ectodermal development. The human gland comprises one or more acini, attached to a common epithelial secretory duct which is continuous with the hair follicle, or pilo-sebaceous canal. The structure of the mouse sebaceous gland is both smaller and simpler, consisting of a single small
acinus which secretes directly into the hair follicle (Wheatley, 1986). The sebaceous gland is a holocrine gland, which implies that the process of sebum formation involves the continuous differentiation of cells from undifferentiated basal cells into lipid-filled mature cells, which lyse to release sebum. The undifferentiated cells are found as a single basal layer around the acinus, and differentiate inwards towards the centre; in so doing they start to accumulate sebum vacuoles, and the nucleus shrinks. The endoplasmic membranes are also lost, the cell becoming a sac of sebum.

6.6.2 Localisation of transgene expression to the sebaceous basal cells

The basal cells of the sebaceous gland retain their differentiative potential, in order to form the eventual holocrine secretion. They have been differentiated from the germinal layer via interactions with the mesenchyme (section 6.1.3), but have not reached a terminally specialised state. These cells evidently retain the ability to express milk proteins in the mouse. It is particularly interesting that they only do so at the juncture of the sebaceous and hair follicle epithelia, where they are arranged into a duct-shaped structure: this correlates well with the data from tissue culture experiments, which have suggested that primary mammary epithelial cells must be allowed to form ductal structures in collagen or adipocyte substrata before they will express milk proteins (see section 4.1; Emerman and Pitelka, 1977; Wiens et al., 1987).

The basal cells of the epidermis develop gradually into keratinised outer cells, and eventually die; they therefore have a similar differentiative potential to the basal cells of the sebaceous gland. The same is true of the cells which give rise to hair roots, yet neither of these cell types express the ο-lactalbumin transgene. It is reasonable to suggest that this is due to the lack of duct formation in these epithelia. The localisation of transgene message to the duct of the sebaceous gland is a striking example of the necessity for correct stromal interactions for the expression of milk protein genes in mammary epithelia.
Chapter 7
Conclusion

7.1 The study of milk protein gene expression

7.1.1 The commercial importance of milk-specific transgenes

In 1987 Simons et al. published a paper reporting the expression of sheep β-lactoglobulin in transgenic mice. This publication, as well as further articles and presentations (Harris, 1988; Simons, 1988; Whitelaw, 1988; Wilmut et al., 1988; Simons et al., 1988) makes the point that the production of proteins in the milk of animals is a commercially desirable objective; a valuable pharmaceutical, for example human factor IX or factor VIII, can be expressed in the milk of a large animal such as a cow or sheep and easily separated from the other components of milk. Such a protein would be free from the pharmacologically unacceptable contamination which haunts microbial production processes, and would be available in large quantities. This aspect has added another dimension to the study of milk protein transgene expression, and has led to the cloning and sequencing of many milk protein genes from larger farm animals (Jamieson et al., 1987; Gorodetsky et al., 1987). Clearly, the objective is to maximise expression of milk protein transgenes; this requires an understanding of how milk protein genes are regulated in the context of the mammary gland, which, although simple in function (it secretes milk) is exceptionally complicated in its regulation (see Larson and Smith, 1974; Neville and Daniel, 1987).

7.1.2 Approaches to the study of milk protein gene regulation

Accordingly, mammary gland biologists have addressed the question of the regulation of milk protein gene expression in several ways: firstly, by established cell culture techniques. Although these seemed promising for a time, especially with the development of floating collagen gel substrata (Emerman and Pitelka, 1977) and culture on adipocytes (Wiens et al., 1987), they failed to allow anything more than constitutive expression of transfected milk protein genes to be reproducibly observed. Hormonal regulation of endogenous genes,
although possible on collagen and adipocyte substrata, is not instructive in this instance as it is impossible to manipulate said genes.

A second approach has involved the sequencing of a variety of milk protein genes and the comparative study of their structure, concentrating especially on promoters (see chapter 3). Although a number of conserved sequences of potential regulatory activity have been identified and tested in in vitro systems (Laird et al., 1988; Lubon and Hennighausen, 1987; chapter 3), the nature of their involvement in the regulation of milk protein gene expression remains unclear. This is due to the absence of a functional assay for milk protein gene expression, a deficiency which is ascribable to the unavailability of responsive cell lines.

A third approach has been to use transgenic animals, not only mice (see chapter 4) but also sheep (Simons et al., 1987). This has seen the successful expression of foreign milk protein transgenes in the transgenic animals, and of heterologous genes attached to milk protein gene promoters (Andres et al., 1987; Pittius et al., 1988). The low levels of expression achieved with manipulated constructs imply that the sequence requirements for correct expression extend outside the immediately identifiable promoter regions, either 5' to the promoter or 3' into the structural part of the gene.

7.2 Gene expression of the guinea pig α-lactalbumin gene

7.2.1 In Vitro analysis of the guinea pig α-lactalbumin gene promoter

DNA probes from the 5' upstream region of the guinea pig α-lactalbumin gene gave rise to a specifically competable, if not tissue-specific, band when incubated with guinea pig nuclear extracts and subjected to gel retardation analysis. This band was not formed after incubation with nuclear extracts prepared from virgin guinea pig mammary gland, but was detected after incubation with nuclear extracts prepared from virgin guinea pig liver and kidney. As mentioned in section 3.5, it remains a strong possibility that extracts prepared from virgin mammary tissue do not consist of nuclear protein, as virgin animals are almost completely devoid of epithelial tissue from which to prepare nuclei.

These results are inconclusive, as it remains impossible to identify functionally important regions of DNA. The only concrete fact emerging from this analysis is that nuclear proteins bind
this DNA region in a specific manner. Attempts at footprinting this region of the guinea pig α-lactalbumin gene using both the DNase 1 protection (Hennighausen and Lubon, 1987) and DMS interference (protocol supplied by Dr. M. Edbrooke, Northwick Park Hospital, London) techniques were unsuccessful.

Accordingly, studies employing in vitro techniques were not pursued further. The study of the expression of cloned genes in transgenic animals remains a much more promising route to the understanding of milk protein gene regulation.

7.2.2 The expression of guinea pig α-lactalbumin in transgenic mice

A 3.5kb guinea pig α-lactalbumin transgene was correctly and tissue-specifically expressed in the mammary glands of transgenic mice, at levels similar to those observed in the guinea pig. Expression was also detected at extremely low levels in the heart, lung and salivary gland of a transgenic mouse, and at high levels in the skin. At all times, the expression was regulated by hormones and thus restricted to pregnant and lactating female transgenic mice. These results establish that the cloned guinea pig α-lactalbumin gene as isolated by Laird (1985) is capable of being correctly regulated in an experimental system, and that the negative results published by Laird et al. (1988) were the consequence of the unsuitability of the cell culture system employed.

Of specific interest is the expression of the guinea pig α-lactalbumin transgene in the skin of pregnant and lactating female transgenic mice. The endogenous mouse β-casein gene was found to be likewise expressed, suggesting that the expression of the transgene in the skin of lactating transgenic (B10xCBA/J)F₁xCBA hybrid mice is not an abnormality specific to the transgene. Furthermore, β-casein was found to be transcribed in the skin of normal lactating CBA/J female mice, indicating that the process of transgene insertion was not responsible for the observed skin-specific expression. Experiments with RNA extracted from guinea pig skin suggest that lactating female guinea pigs may express their endogenous α-lactalbumin gene in the skin at extremely low levels.

The site of expression in the skin was localised by in situ hybridisation to the basal cells which surround the ducts of the sebaceous glands; these are cells which have not yet fully differentiated, and are organised into a ductal structure as is the case in the lactating mammary
gland. No expression was detected in any other cell type in the skin, suggesting that milk protein expression in the skin is a highly specific process.

7.2.3 The extracellular matrix and milk protein gene expression

The localisation of skin-specific α-lactalbumin expression to the undifferentiated cells surrounding the sebaceous gland is a clear manifestation of the importance of the effects of tissue morphology and the extracellular matrix (ECM) on the expression of milk protein genes. The basal cells of the epidermis and the hair follicle are similar to those of the sebaceous glands in terms of differentiated state and origin; yet these cell types do not express the α-lactalbumin transgene.

It is known that during the development of the mammary gland the mesenchyme cells play an important role in the determination of epithelial differentiation (Topper and Freeman, 1980), and that the ECM is an essential factor in the expression of milk protein genes in tissue culture, and in the regulation of the response of cultured mammary epithelial cells to hormones (Bissell and Hall, 1987). It has been proposed (see Bissell and Barcellos-Hoff, 1987) that the functional unit of the mammary gland is best defined in terms of a cell plus the associated basement membrane, and that the structure of the entire tissue is essential for correct tissue-specific gene expression to occur. These arguments can lead to the view (see Bissell and Hall, 1987; Bissell and Barcellos-Hoff, 1987) that the morphology of a cell is instrumental in the determination of its function, and that the ECM, via cytoskeletal interactions, directly influences milk protein gene expression by regulation of the stability and translational efficiency of RNA transcripts.

The expression of guinea pig α-lactalbumin in the sebaceous glands of transgenic mice supports these theories, by suggesting that the expression of the transgene can only occur in epithelial cells which find themselves in a particular environment, that is in a mammary lobule or surrounding a sebaceous gland. Of particular interest is the observation that the expression of the transgene appears to be limited to the basal cells surrounding the duct of the sebaceous gland which leads directly in to the hair follicle. This suggests that the organisation of the tissue into a ductal form is important for α-lactalbumin expression, and is in agreement with data from cell culture studies which suggest that α-lactalbumin is not secreted by cells grown on flat collagen
gels in the absence of the formation of quasi-ductal structures (see Lee et al., 1984; Bissell and Barcellos-Hoff, 1987). Other work has suggested that \( \alpha \)-lactalbumin is expressed in the absence of such ductal formations (Chen and Bissell, 1989), and that the interactions with the ECM are the central element in the regulation of \( \alpha \)-lactalbumin gene expression; if the observed localisation of signal to the ductal areas of the sebaceous glands is not an artifact, then the data presented here seems to argue against this.

7.3 Further Experiments

7.3.1 Introduction of manipulated genes into transgenic mice

Having demonstrated that the guinea pig \( \alpha \)-lactalbumin gene is correctly expressed in the transgenic mice, it becomes possible to begin to map the gene in terms of the functional elements contained within its sequence. One of the more intriguing questions in this area concerns the precise role of the promoter in the control of the expression of milk protein genes. Not only is this a commercially relevant question (see section 7.1.1) but various experiments have indicated that significant control functions reside outside this section of the gene. Transgene constructs in which all or part of the structural part of the gene is replaced do not seem to be expressed at realistic levels, or in a strictly correct manner (see section 5.1; Andres et al., 1987, Pittius et al., 1988, Lee et al., 1988, 1989), whereas whole-gene transgenes are (Simons et al., 1987). Of particular interest are the experiments of Lee et al., which demonstrate that a \( \beta \)-casein minigene, containing the first exon and intron 1 of the genomic casein gene but with the downstream portion replaced by a \( \beta \)-casein cDNA, is not expressed at a realistic level. This implies that important control regions may be placed at some distance from the 5' end of the gene.

Comparative sequence analysis of the guinea pig \( \alpha \)-lactalbumin gene and other milk protein genes (see chapter 3) shows there exists an extensive degree of sequence conservation in the 3' untranslated regions of these genes. Bearing in mind the importance of the 3' UTR in modulation of mRNA stability observed in iron metabolism genes (Casey et al., 1988) and the reported importance of mRNA stability in the regulation of milk protein genes (Guyette et al., 1979) it was decided to construct two different transgenic mouse lines to determine the importance of the 3' UTR in the regulation of the guinea pig \( \alpha \)-lactalbumin gene.
One transgene includes the guinea pig α-lactalbumin promoter, with a bacterial CAT gene inserted downstream of the TATA box but replacing the α-lactalbumin ATG. The SV40 polyadenylation signals present in the vector pBLCAT-3 from which the CAT gene was derived are retained in the transgene. The second transgene is identical save for the insertion of the guinea pig α-lactalbumin 3' UTR and polyadenylation signals downstream of the CAT termination site, but upstream of the SV40 polyadenylation signals. The construction of these transgenes is depicted in fig. 7.1.
Figure 7.1

Construction of transgene plasmids pαLACAT-2 and 3

The α-lactalbumin promoter was excised from pgpH68 by digestion with Hph I and ligated into the Sma I site of the Multiple Cloning Site (MCS) of pBluscript (Promega Biotech). It was subsequently re-excised at Hind III and Bam HI sites, and ligated into the same sites in the MCS of pBL-CAT 3 (Lucklow and Schütz, 1987) to form pαLACAT-2. This plasmid was digested with Hpa II to allow the insertion of the α-lactalbumin 3' UTR downstream of the CAT stop codon, but upstream of the SV40 polyadenylation signals included in pBL-CAT 3, to form pαLACAT-3.
promote

pgpH68

digest with Hph 1
fill in

Sma 1

pBluscript

MCS

Rsa 1
Pvu II

pBL-CAT 3

CAT

Hind III
Bam H1

Hind III
Bam H1

MCS

p α LACAT-2

CAT

Hpa II

p α LACAT-3

UTR

224
7.3.2 The construction of a new transgenic mouse line

\( \text{p} \alpha \text{LaCAT-3 DNA, consisting of the } \alpha \text{-lactalbumin promoter fused to the CAT gene and an } \alpha \text{-lactalbumin 3'} \text{ UTR and polyadenylation signal as shown in fig. 7.1 was microinjected into 100 CBA/JxCBA/J mouse ova by Dr. Andy Mellor, N.I.M.R., Mill Hill, London. The pregnant foster mothers were transferred to the animal house at University College, London, where 25 pups were born. DNA was prepared from the tails of these mice (section 2.2.7), and restricted with } \text{Bam} H1. \text{ On analysis by Southern blotting (section 2.2.10) and probing with a CAT-specific DNA probe, three mice were observed to possess a transgene fragment of the expected 1.8kb size. Two more showed evidence of rearranged transgenes, possessing CAT-specific bands of 7kb. All the remaining DNA samples, including DNA prepared from guinea pig and from normal mouse, did not hybridise to the CAT-specific probe.} \)

Of the five mice which had the transgene DNA integrated into their genomes, only one was female; unfortunately, this mouse carried a rearranged transgene. In order to test for the expression of the transgene, this mouse was mated at the age of 6 weeks with a CBA/J male, and killed 1 day post-partum. No CAT mRNA could be detected in total RNA prepared from mammary or any other tissue of this animal.

Due to the rearrangement of the transgene in the animal in question, it is not possible to derive any expression data for the transgene from this negative result. Such data must be obtained from the analysis of expression in the transgenic female offspring of the male transgenic founder mice carrying an unrearranged transgene. Should the \( \text{p} \alpha \text{LaCAT-3 transgene prove to be correctly expressed, at a realistic level comparable to that of the whole-gene } \alpha \text{-lactalbumin transgene, the control sequences important for expression of the } \alpha \text{-lactalbumin gene will have been localised to the 5'} \text{ and 3'} \text{ untranslated regions of the gene. A further line of transgenic mice carrying the } \text{p} \alpha \text{LaCAT-2 transgene would confirm the importance of the 3'} \text{ UTR. In the eventuality that the } \text{p} \alpha \text{LaCAT-3 transgene should prove to be less than optimally expressed, then the structural portion of the gene will be implicated in the regulation of } \alpha \text{-lactalbumin gene expression.} \)
7.3.3 Skin-specific $\alpha$-lactalbumin expression

The expression of the guinea pig $\alpha$-lactalbumin gene in the skin of transgenic mice appears to mimic the expression of the gene in the mammary gland. This also appears to be true in the guinea pig, although the expression in skin is at a level far inferior to that in mammary gland. It is possible that the modulation of milk protein gene expression in response to a variety of stimuli, such as the injection of hormones and drugs, or of social conditions such as stress, could be studied in the whole animal by measuring the levels of $\alpha$-lactalbumin mRNA in skin biopsies. The distinct advantage of such a method lies in the fact that the animal need not be sacrificed, or subjected to major surgery, in order to analyse tissue mRNA levels. This would benefit the continuity of experiments, and avoid the problems imposed by differences in the physiologies of different individual animals. A further opportunity arises in the eventuality that milk protein mRNA is detectable in humans during pregnancy. This would provide for a means of non-invasive determination of milk protein levels during the course of human pregnancy, a valuable aid in the study of human lactation.

Furthermore, it is observed that hormonal changes, for example those occurring during pregnancy, have a profound effect on the skin, and indeed on the sebaceous gland in particular as is evidenced by the relation of acne ($acne$ vulgaris) to pregnant or non-pregnant states and to general stress; the skin-specific expression of guinea pig $\alpha$-lactalbumin may provide us with a model system in which to study these phenomena.
References

Adler, S., Waterman, M., He, X. and Rosenfeld, M., 1989, Cell 52, p.685-695


Dynan, W., 1986, TIG 2, pp. 196-197.


Guyette, W., Matusik, R. and Rosen, J., 1979, Cell 17, pp. 1013-1023.


McKnight, S., Bustin, M. and Miller, O., 1978, Cold Spring Harbor Symposium on quantitative biology 42, pp. 741-754


Suard, Y., Kraehenbuhl, J-P. and Aubert, M., 1979, J. Biol. Chem. 254, pp. 10466-10475.


Tsai, S., Tsai, M-J. and O'Malley, B., 1989, Cell 57, pp. 443-448.


White, J. and Bauer, W., 1989, Cell 56, pp. 9-10.


