PHYSICAL MAPPING OF THE MURINE CASEIN LOCUS

SISILAMMA GEORGE

Thesis submitted for the degree of
Doctor of Philosophy

University of Edinburgh

1996

EX BIBL. UNIV. EDINBURGEN.

The essential part in Science is not a complicated mathematical formalism or a ritualised experimentation. Rather the heart of science is a kind of shrewd honesty that springs from wanting to know what the hell is going on !

Saul-Paul Sirag

ABSTRACT

Gene or genome mapping is important for the study of gene structure and function, and the planning of subsequent genetic manipulations. The analysis of complex genomes such as those of the mammalian species has revealed that genes can be grouped into families and in many gene families the members are found in clusters e.g. the globin, casein and fibrinogen genes. It is evident that these clusters arise from gene duplication events. Some of these gene clusters are expressed in a co-ordinated manner. Therefore, the analysis of complex loci composed of related genes can contribute to both our understanding of the long range regulation of gene expression and the mechanisms of gene evolution. In the present study, experiments were carried out 1) to construct a physical map of the murine casein locus 2) to analyse expression levels of various casein genes during pregnancy and lactation and 3) to identify genetic variation at the casein loci amongst different mouse strains.

1) The murine casein locus has been characterised by long range restriction mapping and the analysis of long fragment genomic clones. Cloned sequences from five mouse casein genes (α, β, γ, ε, κ) were used to screen a murine (strain 129) genomic library in a bacterial artificial chromosome vector (BAC). Of the nine clones isolated, two contained 3 casein genes α, β, γ and γ, ε, κ respectively. The following combinations were found in other clones - α + β; β + γ; γ + ε. Thus, I deduced that the gene order in the locus is α– β– γ– ε–κ. I have confirmed this order by restriction analysis of the clones. A contig map of the clones and flanking sequences has been established by characterising seven BAC clones, which together span ~470 kb. Long range restriction analysis of genomic DNA indicates that the murine casein locus is confined to a 250 kb *Xho*I partial fragment. The orientations of the casein genes, except one are also established. The α, γ and κ genes have the same 5'-3' orientation while the β gene is orientated in the reverse pattern. The present data indicate that the murine casein locus is larger than the bovine locus with an additional (fifth) gene, ε.

2) Expression studies of casein genes at various time points during pregnancy and lactation revealed a co-ordinate expression pattern for the three (α, β and γ) calcium sensitive genes from mid pregnancy to parturition. Interestingly, the fourth calcium sensitive casein gene, ε exhibited a totally different expression profile. The ε-casein mRNAs were detectable only from first day of lactation. The κ-casein mRNA levels also increased from mid pregnancy onwards but at a lower rate as compared to α, β and γ casein mRNAs. The reason for the transcriptional suppression of the ε casein gene is not clear and the mechanism can be unravelled only by further studies.

3)The genetic variation in the casein loci of *Mus musculus* (eight different strains) and *Mus spretus* was also examined as restriction fragment length variants (RFLVs) using five restriction enzymes, *Bam*HI, *Sfi*I, *Hin*dIII, *Eco*RI and *Xba*I. No variation was observed between the inbred strains (C57/BL6, C3H, NIH, CBA, DBA, 129, SJL, BALB/c) of *Mus musculus*. Between *Mus musculus* and *Mus spretus* RFLVs were observed at the casein loci - α (*Bam*HI, *Sfi*I, *Xba*I and *Eco*RI), β (*Bam* HI and *Sfi*I) γ (*Bam*HI, *Sfi*I, *Xba*I, *Hin*dIII and *Eco*RI) ε (*Bam*HI, *Sfi*I, *Hin*dIII and *Eco*RI) and κ (*Bam*HI, *Xba*I and *Hin*dIII).

In the longer term, the information obtained from the present mapping study and the clones isolated (129 strain) can be used to manipulate the casein locus in embryonic stem (ES) cells as most gene targeting experiments are carried out in ES cells isolated from 129 mice. Gene targeting is more effective if the targeting DNA is prepared from isogenic DNA. It is also advantageous to use isogenic DNA derived maps in targeting experiments. Thus, the results of this study could not only contribute to basic studies on genome structure and function but also in the longer term underpin applications in biotechnology and agriculture/industry.

DECLARATION

I hereby declare that the work presented in this thesis is the product of my own efforts, and has not been submitted in any previous application for a degree. The work on which it is based is my own except where stated in the text and in acknowledgements section.

Sisilamma George

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr. Alan Archibald and Dr. John Clark for their invaluable guidance and constant encouragement to carry out this research.

I am also grateful to my University supervisor, Dr. Linda Mullins for her advice and support.

My sincere thanks to Dr. Bruce Whitelaw for his constructive criticism and advice.

I am thankful to Miss. Roberta Wallace, Ms. Frances Thomson, Mrs.Julie Thomson, Miss. Heather Warnock and Mrs. Carrie Owen for providing the mice for my studies.

My thanks to Anthea Springbett for assisting me with the statistical analysis.

I am grateful to Dr. Jeffrey Rosen, Miss. Katrina Gordon for providing the murine cDNA clones and Dr. Christine Watson for supplying the lactating mouse mammary cDNA library.

My sincere thanks to Mrs. Sandra Couperwhite, Dr. Judy Brown, Dr. Heather McQueen, Dr. Edward Hitchin and Mr. John Webster for their technical help and advice.

I am thankful to Mr. Norman Russel and Mr. Rodney Field for their constant efforts in processing and printing the photographs throughout my Ph.D programme. My thanks are also to Mr. Elliot Armstrong for helping me with graphics.

My sincere thanks to my friends and colleagues, Kenneth Dobie, Simon Boa, Gaetano Donofrio, Katrina Gordon, Anona Turnbull, Jenny Fenty, Sheelagh Strachan, Graham Harold, Jennifer Cattermole, Claire Neil, Simon Hovart, Suminder Swahney, Parveen Goel, George Russel, Usha, A.P, Sreekumar, K.P, Ray Ansell, Heather Finlayson, Alison Downing, Kirsty Maitland, and Susan Anderson for every little help and for providing a happy and cordial atmosphere all the way through my Ph.D programme.

I am obliged to the Association of Commonwealth Scholarship Commission in U.K for providing financial assistance to carry out this project.

Lastly, my heartfelt gratitude to my husband, Ponnachan and daughter, Sophiya for their patience and constant encouragement in completing this study.

Dedicated to my husband and daughter

CONTENTS

	Page
Abstract	
Declaration	
Acknowledgements	
Dedication	
Contents	i
List of tables	vii
List of figures	viii
Abbreviations	x
***Chapter 1* - GENERAL INTRODUCTION**	
1.1. Milk proteins	2
1.2. Caseins	3
1.2.1. Structure	4
1.2.1.1. Proteins	4
1.2.1.2. mRNA/cDNA	6
1.2.1.3. Genes	6
1.2.2. Comparison	8
1.2.2.1. Proteins	8
1.2.2.2. mRNA/cDNA	9
1.2.2.3. Genes	12
1.2.3. Expression	13
1.2.3.1. Native genes	13
1.2.3.2. Transgenes	15
1.2.3.3. Locus control region (LCR)	16

1.2.4. Regulation 22

1.2.4.1. Role of hormones 22

1.2.4.2. Role of introns 25

1.2.4.3. Role of mammary nuclear factors 26

1.2.5. Polymorphism 29

1.3. Whey proteins 35

1.3.1. α-Lactalbumin gene 35

1.3.2. β-Lactoglobulin gene 36

1.3.3. Whey acidic protein gene 37

1.4. Chromosomal location of milk protein genes 39

1.5. Gene families and evolution 40

1.5.1. Multigene families 40

1.5.2. Casein genes 43

1.6. DNA methylation and CpG islands 48

1.6.1. Methylation and gene expression 48

1.6.2. CpG islands 51

1.7. Research project 57

1.7.1 Justification 57

1.7.2 Objectives 60

1.7.3 Experimental approach 60

***Chapter 2* - MATERIALS AND METHODS** 63

2.1. Plasmid DNA manipulations 63

2.1.1. Preparation of LB (Luria-Bertani) plates 64

2.1.2. Bacterial transformation 64

2.1.2.1. Calcium competent DH5 αcells 64

2.1.2.2. Electroporation 65

2.1.3. Plasmid DNA preparation 66

2.1.3.1. Mini plasmid preparation 66

2.1.3.2. Maxi plasmid preparation 67

2.1.4. Caesium chloride gradient purification of plasmid DNA 68

2.1.5. DNA Sequencing 69

2.1.5.1. Using T7 sequencing kit (Pharmacia) 69

2.1.5.2. Using Taq Dyedeoxy terminator cycle sequencing kit (Applied Biosystems) 73

2.1.6. Purification of DNA from agarose gel 75

2.1.7. Polymerase chain reaction (PCR) 76

2.2. RNA analysis 77

2.2.1. Total RNA isolation 77

2.2.2. Isolation of poly (A)$^+$ RNA 78

2.2.3. Synthesis of cDNA 78

2.2.4. Electrophoresis of RNA 79

2.2.5. Northern blotting 80

2.3. DNA analysis 80

2.3.1 Preparation of high molecular weight (HMW) genomic DNA 80

2.3.1.1. Phenol chloroform extraction 81

2.3.1.2. Embedding DNA in agarose plugs 82

2.3.2. Isolation of large genomic clones from murine genomic DNA libraries 83

2.3.2.1. Screening bacteriophage P1 and BAC libraries 83

2.3.2.2. Colony transfer and hybridisation 83

2.3.2.3. P1 DNA preparation 85

2.3.2.4. BAC DNA preparation 85

2.3.3. Restriction digestion 86

2.3.3.1. Plasmid DNA 86

(preparation of casein probes)

2.3.3.2. Genomic DNA 86

2.3.3.3. P1/BAC clones 88

2.3.4. Electrophoresis 88

2.3.4.1. Conventional agarose gel electrophoresis 88

2.3.4.2. Pulsed-field gel electrophoresis 89

2.3.5. Southern bloting 90

2.4. Radioactive labelling of probes 90

2.4.1. Using [α^{32}P]dCTP 90

2.4.2. Using [γ^{32}P]ATP 91

2.4.3. Measurement of incorporation and calculating the specific activity of labelled DNA 92

2.5. Hybridisation and Autoradiography 92

2.6. BAC end fragment isolation 94

2.6.1. Plasmid rescue 95

2.6.2. Inverse polymerase chain reaction (IPCR) 96

2.7. Long range PCR 97

***Chapter 3* - MURINE CASEIN LOCUS - ISOLATION AND CHARACTERIZATION OF CLONES FROM LARGE FRAGMENT MURINE GENOMIC LIBRARIES.**

3.1. Introduction 99

3.2. Results 100

3.2.1. Validation of murine casein probes 100

3.2.1.1. Northern blot analysis 100

3.2.1.2. Sequence analysis of the casein probes 100

3.2.2. Analysis of the bacteriophage P1 library 102

3.2.2.1. P1 library screening 102

3.2.2.2. Restriction analysis of P1 clones 103

3.2.3. Analysis of murine BAC library 103

3.2.3.1. BAC library screening 104

3.2.3.2. Generation of end probes from BAC clones 105

3.2.3.3. Analysis of clones 490 H23 and 555 N16 107

3.2.3.4. Orientation and contig map of the clones 108

3.2.4. Exploration of the locus for putative additional genes 111

3.3. Discussion 111

Chapter 4 - **MAPPING THE MURINE CASEIN LOCUS - HIGH MOLECULAR WEIGHT (HMW) GENOMIC DNA ANALYSIS**

4.1. Introduction 115

4.2. Results 115

4.2.1. Restriction digestion and pulsed-field analysis 115

4.3. Discussion 117

Chapter 5 - **EXPRESSION ANALYSIS OF VARIOUS CASEIN GENES - DURING PREGNANCY AND LACTATION**

5.1. Introduction 119

5.2. Results 120

5.2.1. Northern blot analysis of the murine caseins 120

5.2.2. Statistical analysis 122

5.3. Discussion 123

Chapter 6 - **IDENTIFICATION OF GENETIC VARIATION AT THE CASEIN LOCI BY RESTRICTION FRAGMENT LENGTH VARIATION**

6.1. Introduction 130

6.2. Results 131

6.3. Discussion 134

Chapter 7 - **GENERAL DISCUSSION**

7.1. Evolutionary analysis of casein genes 139

7.2. Regulation of expression 140

7.3. Application in transgenic studies 141

7.4. Application to enrich mouse genetic map 144

REFERENCES 146

LIST OF TABLES

1.1	Composition of milk proteins from different species	2a
1.2	Molecular weight of mouse caseins	3a
1.3	Percentage similarity between αS1-casein cDNAs from different species	10a
1.4	Percentage similarity between αS2-casein cDNAs from different species	10a
1.5	Percentage similarity between β-casein cDNAs from different species	10a
1.6	Percentage similarity between κ-casein cDNAs from different species	11a
1.7	Percentage similarity between cDNAs of mouse caseins	11a
1.8	Various mammary gland specific transgenes and their levels of expression	16a
2.1	Plasmid clones used to generate mouse specific probes	63a
2.2	PCR primers used to generate probes and for screening libraries	63c
3.1	Mouse casein mRNA sizes	100a
3.2	Positive pools identified from the BAC library	104a
3.3	Clones isolated from the BAC library	105a
3.4	Primer pairs used to amplify the BAC end fragments	106a
5.1	Expression levels (%) of the five casein genes in mouse mammary gland	121c
5.2	Parameter estimates for Gompertz and Logistic curves	122a

LIST OF FIGURES

1.1	Sub-unit of a casein micelle model	4a
1.2	Structural organization of the four bovine casein transcription units	7a
1.3	Organization of the beta casein gene - comparison between species	7b
1.4	Signal peptide amino acid sequence of caseins - species comparison	9a
1.5	Signal peptide encoding nucleotide sequence - species comparison	11b
1.6	Tentative evolutionary relationship between calcium sensitive casein genes and between fibrinogen and κ-casein genes.	46a
2.1	Genomic DNA fragment (6.6 kb) of β-casein showing the positions 5' and 3' probes.	63b
2.2	Organization of the P1 and BAC libraries	83a
2.3	Diagramatic representation of P1 vector (pAd10SacB II)	83b
2.4	Diagramatic representation of BAC vector (pBeloBAC 11)	83c
2.5	Diagramatic representation for left and right end-fragment generation (IPCR and Plasmid rescue respectively) from BAC	94a
3.1	PCR amplification - screening of P1 library (ethidium picture)	102a
3.2	Restriction digested P1 clones (ethidium picture)	103a
3.3.a	Ethidium picture of restriction digested BAC clones (490 H23 and 555 N16)	107a
3.3.b	Southern blot analysis of clone 490 H23 (contains the α, β and γ casein genes)	107b
3.3.c	Southern blot analysis of clone 555 N16 (contains the γ, ε and κ casein genes)	107c
3.4	Restriction maps of clones 490 H23 and 555 N16	107d
3.5	Contig map of 7 BAC clones	108a
3.6.a	Restriction analysis of clone 365 E21 (contains the α and β casein genes)	108b
3.6.b	Restriction analysis of clone 201 E9 (contains the α and β casein genes)	108c
3.7	Long range PCR (to orient the κ-casein gene)	109a
3.8.a	Restriction analysis of clone 490 H23 (contains the α, β and γ casein genes; to orient γ-casein gene)	109b
3.8.b	Restriction analysis of clones containing 3' region of γ-casein gene (clones 185 O18 and 555 N16; to orient γ-casein gene)	109c
3.9	Long range PCR (to orient ε-casein gene)	109d

3.10	Restriction analysis of clone 156 J3 (contains κ-casein gene)	110a
3.11	Southern blot analysis of five BAC clones span the entire mouse casein locus - for additional putative genes	111a
3.12	Contig map of five BAC clones - showing the position of additional putative gene	111b
4.1.a	Ethidium picture of restriction digested genomic DNA (*Xho*I, *Sfi*I and *Xho*I/*Sfi*I)	116a
4.1.b	Autoradiogram of restriction analysed genomic DNA	116b
4.2.a	Ethidium picture of restriction digested genomic DNA (*Sfi*I and *Xho*I/*Sfi*I, resolution of fragments from 5 - 75 kb)	116c
4.2.b	Autoradiogram of restriction analysed genomic DNA	116d
4.3.a	Ethidium picture of restriction digested genomic DNA (*Bam*HI and *Bam*HI/*Sfi*I)	116e
4.3.b	Autoradiogram of restriction analysed genomic DNA	116f
5.1	Lactating mouse mammary and liver RNAs probed with casein probes	120a
5.2.a	Northern blot analysis of mammary RNA at 9 time points (Virgin, gestation and lactation, probed with the α, β and γ probes)	121a
5.2.b	Northern blot analysis of mammary RNA at 9 time points (Virgin, gestation and lactation, probed with the ε and κ probes)	121b
5.3	Percentage expression levels of the casein genes	121d
6.1	RFLP analysis of α-casein locus *M.spretus* and *M.musculus*	132a
6.2	RFLP analysis of β-casein locus *M.spretus* and *M.musculus*	132b
6.3	RFLP analysis of γ-casein locus *M.spretus* and *M.musculus*	133a
6.4	RFLP analysis of ε-casein locus *M.spretus* and *M.musculus*	133b
6.5	RFLP analysis of κ-casein locus *M.spretus* and *M.musculus*	133c
6.6	Linkage map of mouse chromosome 5	138a

ABBREVIATIONS

α-la	Alpha lactalbumin
Amp^r	resistant to Ampicillin
APS	Ammonium persulphate
BAC	Bacterial artificial chromosome
β-lg	Beta lactoglobulin
bp	basepair
BSA	Bovine serum albumin
^oC	Degree Celsius
CAT	Chloramphenicol acetyl transferase
cDNA	Complementary DNA
CFTR	Cystic fibrosis transmembrane conductance regulator
CHEF	Contour clamped homogenous electric field
cm	centimetre
cM	Centimorgan
dATP	deoxy Adenosine triphosphate
DCR	Dominant control region
dCTP	deoxy Cytosine triphosphate
DEPC	Diethyl pyro carbonate
dGTP	deoxy Guanosine triphosphate
dITP	deoxy Ionosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	deoxy Nucleotide triphosphate
ds	double stranded
DTT	Dithiothreitol
dTTP	deoxy Thymidine triphosphate
EDTA	Ethylene diamine tetraacetic acid
ES cells	Embryo derived stem cells
FPLC	Fast protein liquid chromatography
FSH	Follicle stimulating hormone
g	gram
HCl	Hydrochloric acid
HMW	High molecular weight
hrs	hours

HTF	*Hpa*II tiny fragments
IPCR	Inverse polymerase chain reaction
IPTG	Isopropyl-β-D-thiogalactopyranosid
Kan^r	resistant to Kanamycin
kb	kilo basepair
KCl	Potassium chloride
LAR	Locus activation region
LCR	Locus control region
LMP	Low melting point
M	Molar
Mb	Mega basepair
mCi	millicurie
MeCP	Methyl CpG binding protein
mg	milligram
µg	microgram
$MgCl_2$	Magnesium chloride
MGF	Mammary gland factor
$MgSO_4$	Magnesium sulphate
min	minute
µl	microlitre
ml	millilitre
mM	millimole
MOPS	3-[N-Morpholino] propane - sulfonic acid
MPBF	Milk protein binding factor
mRNA	messenger RNA
MUP	Major urinary protein
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NEB	New England Biolab
NF	Nuclear factor
ng	nanogram
nM	nanomole
OD	Optical density
PAC	P1 based artificial chromosome
PACE	Programmable autonomously controlled electrodes
p.c	post conception

PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
PGK	Phospho glycerate kinase
PMF	Pregnancy specific mammary nuclear factor
pmol	picomole
PMSF	Phenyl methyl sulfonyl fluoride
RBP	Retinol binding protein
RFLP	Restriction fragment length polymorphism
RFLV	Restriction fragment length variant
RNA	Ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
s	second
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SHSS	Super hypersensitive site
ss	single stranded
SSC	Sodium chloride/Sodium citrate
Stat	Signal transducers and activators of transcription
SV40	Simian virus 40
TACS	Terminator ammonium cycle sequencing
TAE	Tris/Acetate/EDTA
TBE	Tris/Borate/EDTA
TBP	TATA binding protein
TCA	Trichloracetic acid
TE	Tris/EDTA
TEMED	N, N, N, N Tetramethylethylenediamine
Tet^r	resistant to Tetracycline
Tris	Tris (hydroxymethyl) aminomethane
UV	Ultraviolet
V	Volt
WAP	Whey acidic protein
x g	gravitational force
YAC	Yeast artificial chromosome

Chapter1

INTRODUCTION

Gene or genome mapping contributes to the study of gene structure and function, and the planning of subsequent genetic manipulations. There are two major approaches to genome mapping - i) genetic (linkage) mapping and ii) physical mapping. The assignment of genes to specific linkage groups and the determination of their relative distance to other known genes is called genetic, linkage or recombination mapping. In a physical map, the mapped elements are assigned to physical locations with varying degrees of accuracy, so that each element can be positioned in the genome. Physical mapping techniques include *in situ* hybridisation to chromosomes or extended DNA molecules, long range restriction enzyme mapping and contig building. The ultimate physical map is the DNA sequence itself. This may help to unravel the relationship between the arrangement of genes and their expression.

The analysis of complex genomes such as those of the mammalian species has revealed that many genes can be grouped into families of related sequences with an assumed common evolutionary ancestor (reviewed by Dillon and Grosveld, 1993). Moreover, there are many examples of gene families in which the members are found in clusters e.g. the globin, casein and fibrinogen genes. It is evident that these clusters arise from gene duplication events. Some of these gene clusters are expressed in a co-ordinated manner (Dillon and Grosveld, 1993). Therefore, the analysis of complex loci composed of related genes can contribute to both our understanding of the long range regulation of gene expression and mechanisms of gene evolution. In this study the organisation of the murine casein genes will be explored.

The caseins and their genes have been characterised in a number of species and these studies have revealed considerable species difference in the protein and

DNA sequences. Expression of the casein genes is developmentally and hormonally regulated (reviewed by Banerjee and Antoniou, 1985). Large quantities of milk proteins are synthesised and secreted under the control of peptide and steroid hormones, and require the concomitant activation of both milk protein genes and "house keeping" genes. So, it is of interest to examine whether all of the casein genes have a common regulatory sequence which might give rise to common regulatory mechanisms.

1.1. Milk proteins

The protein composition of milk differs qualitatively and quantitatively among mammalian species (Table 1.1). During lactation the mammary gland secretes large amounts of milk protein which consists predominantly of the acid precipitable caseins and several whey proteins (Hennighausen and Sippel, 1982a). Caseins are related proteins that constitute 60-80% of milk proteins. Caseins, which are the major source of amino acids also transport calcium and phosphate to the suckling young (Waugh, 1971). The major whey proteins are α-lactalbumin (α-la), β-lactoglobulin (β-lg) and whey acidic protein (WAP). Alpha-lactalbumin influences lactose synthesis by modifying the substrate specificity of galactosyl-transferase, and thus is important to milk synthesis since lactose is the major osmole of milk (Kuhn, 1983). The role of β-lg is still to be resolved since milk from many mammalian species appears to lack this protein (Pervais and Brew, 1985). The amino acid sequence similarities with human serum retinol binding protein suggests that β-lg may bind and transport retinol or other small hydrophobic ligands to the infant (Godovac-Zimmermann et al., 1985). Whey acidic protein is reported to be present in rodent milk (Hennighausen and Sippel, 1982b) and absent in ruminant milk. However, Simpson et al. (1996) reported the presence of WAP in the milk of swine. Although the function of WAP is

Table 1.1. Protein composition of milk from different species (modified from Martin and Grosclaude, 1993).

Milk proteins	Species (Quantity, g/l)				
	Cow	Sheep	Goat	Mouse	Human
Caseins					
αS1	10.0	7.0	0.7	28.0	0.4
αS2	3.7	7.0	4.0	ND	-
β	10.0	28.0	10.0	21.0	3.0
κ	3.5	3.5	6.0	2.4	1.0
Whey proteins					
α-lactalbumin	1.2	0.8	1.2	trace	1.6
β-lactoglobulin	3.3	2.8	2.3	none	none
whey acidic protein	none	none	none	2.0	none
Serum albumin	0.4	ND	ND	ND	0.4
Lysozyme	trace	ND	trace	ND	0.4
Lactoferrin	0.1	ND	0.1	ND	1.4
Immunoglobulins	0.7	ND	0.5	ND	1.4

ND - Not determined

unknown, certain features of its primary structure suggest that it is similar to a family of proteins known as the "four disulfide core proteins" and in particular it has structural similarity to neurophysin (Hennighausen and Sippel, 1982b). Neurophysin is a part of the precursor to oxytocin and act as a carrier for oxytocin from the hypothalamus to the anterior lobe of pituitary gland. Expression of WAP and β-casein RNAs were also reported in tongue, thymus, liver and pancreas and the neuro-intermediate lobe of pituitary gland. Since the mammary gland is a target tissue for oxytocin, it has been speculated that WAP might be a receptor or a carrier for oxytocin (Hennighausen et al., 1991).

1.2. Caseins

The protein content of mouse milk is relatively high (Meier et al., 1965; Jennes, 1973; Piletz and Ganschow, 1981) - 80-150 mg/ml, of which 63-81% are acid-precipitable caseins (Jennes, 1973). At least five different caseins have been identified in mouse milk by denaturing polyacrylamide gel electrophoresis (SDS-PAGE) and fast protein liquid chromatography (FPLC) (Table 1.2)

The various caseins identified in the milk of different species include α (S1 and S2), β, γ, ε and κ (Hipp et al., 1952; Green and Pastewka, 1976; Mercier et al., 1976 and Hennighausen and Sippel, 1982a). Among these proteins β and κ-caseins are present in the milk of all the species reported so far. Ruminant milk contains two forms of α-casein, αS1 and αS2, which have been shown to be homologous to the α and γ/ε caseins respectively, of rodent milk (reviewed by Fiat and Jolles, 1989), while human milk contains only traces of αS1-casein (Martin and Grosclaude, 1993).

All caseins are calcium sensitive except κ-casein. The caseins raise the calcium and phosphate concentrations in milk by forming stable micelles with these minerals, and thus provide the suckling young with a source of minerals and

Table 1.2. Molecular weight of murine caseins.

Caseins in mouse milk	Molecular weight (kDa)
α	44.1[a]
β	26.3[a]
γ	23.7[b]
ε	14.5[b]
κ	30.9[a]

[a] Stevenson and Leaver, 1994; [b] Hennighausen and Sippel, 1982a

amino acids. The calcium sensitive caseins are precipitated in the presence of low concentrations of calcium and are stably maintained as a micelle suspension as a result of their interaction with κ-casein (MacKinlay and Wake, 1971). Casein micelles contain ~5% of colloidal calcium phosphate, which is sequestered through interactions with the clustered serine phosphate residues of the calcium sensitive caseins (Sleigh et al., 1979). The casein micelles are stabilized by κ-casein, which when cleaved at a specific phenylalanine-methionine bond by chymosin, causes initiation of micelle aggregation, resulting in curd formation (MacKinlay and Wake, 1971). Beta-casein determines the surface properties of the micelle, and is also essential for curd formation when the milk is clotted by chymosin digestion, while α-caseins appear to be less important for curd formation but probably determine the capacity of casein micelles for colloidal calcium phosphate transport (Pearse et al., 1986). Curd formation is physiologically important since it ensures the retention of milk proteins in the stomach of the infant, thus allowing digestion to occur (Jenness and Larson, 1985). A model of a casein micelle is shown in Fig. 1.1

1.2.1. Structure

1.2.1.1. Proteins

Our knowledge of the sequences of the casein proteins are largely predicted and derived from an understanding of the corresponding cDNA sequences. Different variants are reported for each type (αS1, αS2, β and κ) of casein. The variants of each casein type differ from each other either by amino acid substitution or by insertion/deletion of peptide sequences. Some casein variants differ only in their phosphate content (reviewed by Fiat and Jolles, 1989).

The bovine αS1-casein (B variant) contains 199 amino acid residues with a higher proportion of acidic than basic sequences, a uniform distribution of prolyl

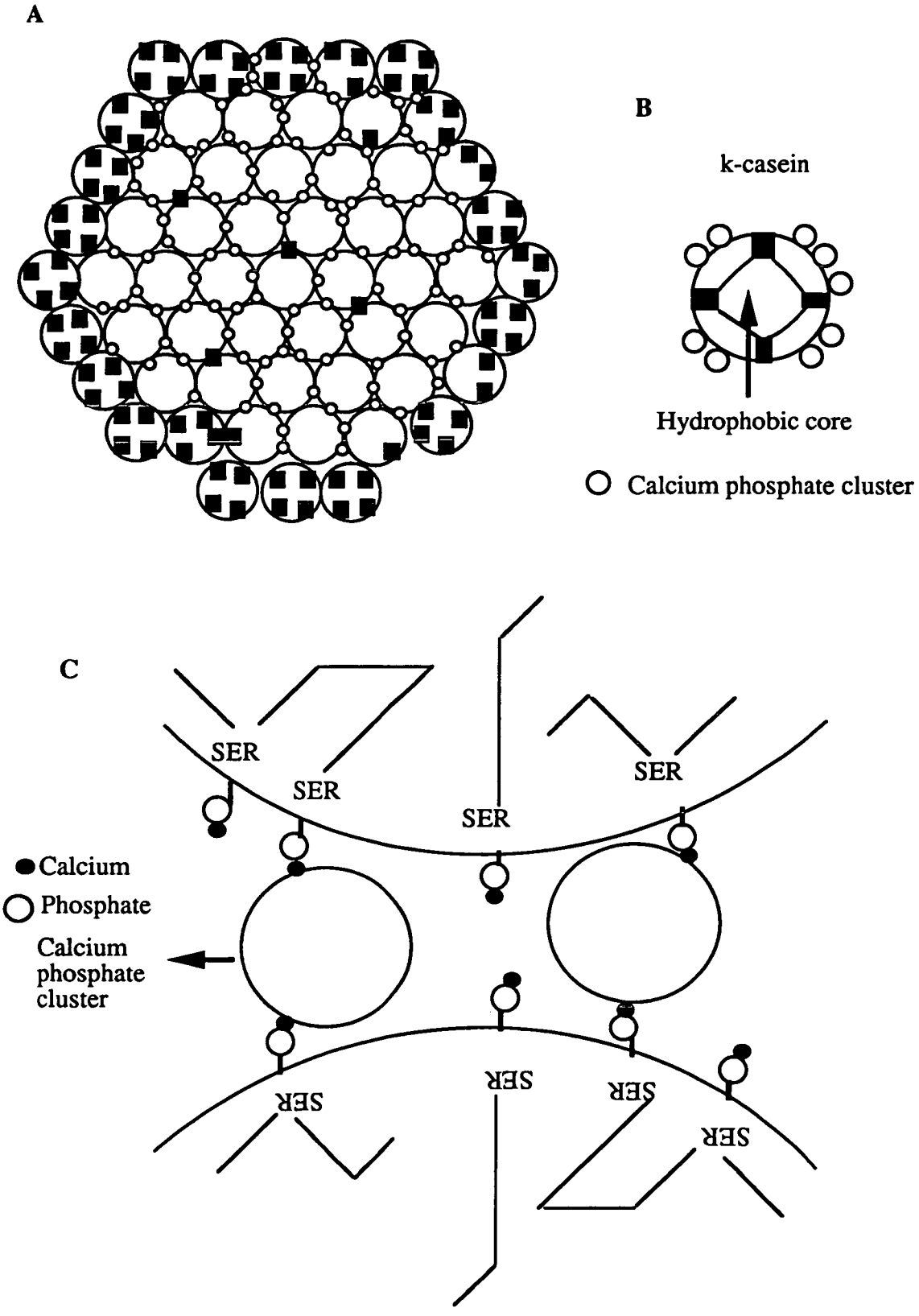


Fig. 1.1. Sub-unit of casein micelle model (Redrawn from Schmidt (1982) and Mepham (1987) .

A. Micelle. **B.** Sub-micelle. **C.** Binding between two sub-micelles

residues (8.5%) and absence of cysteinyl residues (Mercier et al., 1971). The ovine αS1-casein is smaller with 191 amino acid residues as deduced from its mRNA sequence (Mercier et al., 1985). The primary structure of six caprine αS1-casein variants have been established (A, B, C, D, E, F; Brignon et al., 1989, 1990). All of these variants, except D and F have 199 amino acid residues. Rat α-casein is larger and has 269 amino acids as deduced from its mRNA sequence (Hobbs and Rosen, 1982). The deduced length of the two variants of guinea pig casein-B have 165 and 150 residues (Hall et al., 1984a).

The bovine αS2-casein (A variant) has 207 amino acids and carries a large number of positive charges located all along its polypeptide chain especially in the 47 amino acids of the C-terminal fragment. The phosphate groups of αS2-casein are located in three regions - residues 8-16, 56-61 and 129-143. Within the bovine αS2 casein protein there is an internal sequence similarity (38% identical residues) observed between two segments of 76 residues (Brignon et al., 1977). The ovine αS2-casein has 208 residues (Boisnard and Petrissant, 1985), the rat homologue (γ-casein) has 164 residues (Hobbs and Rosen, 1982) and guinea pig casein-A has 223 amino acids (Hall et al., 1984b) as deduced from their respective mRNA sequences. The mouse ε-casein which also exhibits similarity to bovine αS2-casein has 128 amino acids in the mature protein as predicted from its mRNA sequence (Hennighausen et al., 1982a) and contains a potential multiple phosphorylation site.

The bovine β-casein variant A^2 has 209 amino acids including five phosphoseryls. The fragment containing the first 50 residues has a high negative net charge (Ribadeau-Dumas et al., 1972). Human β-casein contains 212 amino acid residues (Greenberg et al., 1984) whilst rat β-casein has 216 amino acid residues, as deduced from its nucleotide sequences (Blackburn et al., 1982).

Bovine κ-casein (B variant) has 169 amino acid residues (Jolles et al., 1972) while human κ-casein has 158 amino acid residues (Brignon et al., 1985). The structures of sheep and goat κ-caseins have also been established (Jolles et al., 1974a, b; Mercier et al., 1976). The mouse κ-casein has 160 amino acids as deduced from its nucleotide sequences (Thompson et al., 1985) while rat κ-casein has 157 amino acid residues (Nakhasi et al., 1984).

An important common or shared feature observed in the signal peptide sequences of various species is the presence of 15 amino acids in case of the calcium sensitive caseins and 21 amino acids in the calcium insensitive κ-caseins. Indeed, the signal peptides of the casein proteins represent the most highly conserved elements within these related proteins.

1.2.1.2. mRNA/cDNA

The cDNA sequences for mouse α-casein (Grusby et al., 1990), ε-casein (Hennighausen et al., 1982a) and γ-casein (Sasaki, 1992) rat and mouse β-caseins (Blackburn et al., 1982; Yoshimura et al., 1986), κ-caseins (Nakhasi et al., 1984; Thompson et al., 1985), four bovine caseins, αS1, αS2, β and κ (Nagao et al., 1984; Stewart et al., 1987; Stewart et al., 1984), porcine β-casein (Alexander and Beattie, 1992), αS2-casein (Alexander et al., 1992) and κ-casein (Levine et al., 1992) have been reported. Sequences have also been determined for the two α-casein cDNAs of guinea pig (Hall et al., 1984a, b) and ovine (Mercier et al., 1985; Boisnard and Petrissant, 1985).

1.2.1.3. Genes

The genes encoding milk proteins have been characterised in several species. The transcriptional units of milk protein genes range in size from 1.7 to 18.5 kb and comprise between 4 and 19 exons. The exons are short and made up

of 21 to 525 bp. The introns are large, consisting of 81 to 5800 bp and often contain repetitive sequences. In contrast to the whey protein genes, in the casein genes no codon is split across an intron and the RNA stretches encoding multiple phosphorylation sites of caseins are generated by splicing (reviewed by Mercier and Vilotte, 1993).

Complete DNA sequences have been determined for the bovine casein genes, αS1 (17.5 kb, 19 exons; Koczan et al., 1991), αS2 (18.5 kb, 18 exons; Groenen et al., 1993) and β (8.5 kb, 9 exons; Bonsing et al., 1988). The bovine κ-casein gene has been partially sequenced (Alexander et al., 1988) and reported to be ~13 kb, comprising 5 exons. All the introns in κ-casein gene are large and both the second, which divides the signal peptide encoding region, and the third contain many *Alu*-like repetitive sequences (Mepham et al., 1992). Furthermore, the 5' flanking region of the κ-casein gene transcription unit lacks the structural motifs shared by the other casein genes. The structural organization of the four bovine casein gene transcription units is shown in Fig. 1.2.

The complete DNA sequence has also been determined for the αS1-casein genes of goat (17.5 kb, 19 exons; Leroux et al., 1992) and rabbit (16 kb, 19 exons; Jolivet et al., 1992). Faruque et al., (1987) partially characterised the mouse α1 casein gene and showed that the majority of the coding sequences of the α1 casein gene are contained in a central 3.5 kb *Hin*dIII/*Bgl*II fragment.

The β-casein gene has been characterised extensively in many different species including cow (Bonsing et al., 1988; Gorodetsky et al., 1988)), sheep (Provot et al., 1995) goat (Roberts et al., 1992), rabbit (Thepot et al., 1991), rat (Jones et al., 1985), mouse (Yoshimura and Oka, 1989) and human (Hansson et al., 1994). The intron/exon organization of β-casein gene is similar in all species (Fig.1.3; reviewed by Mercier and Vilotte, 1993) and has a size of 7-9.2 kb consisting of nine exons except the human β-casein gene which has only eight

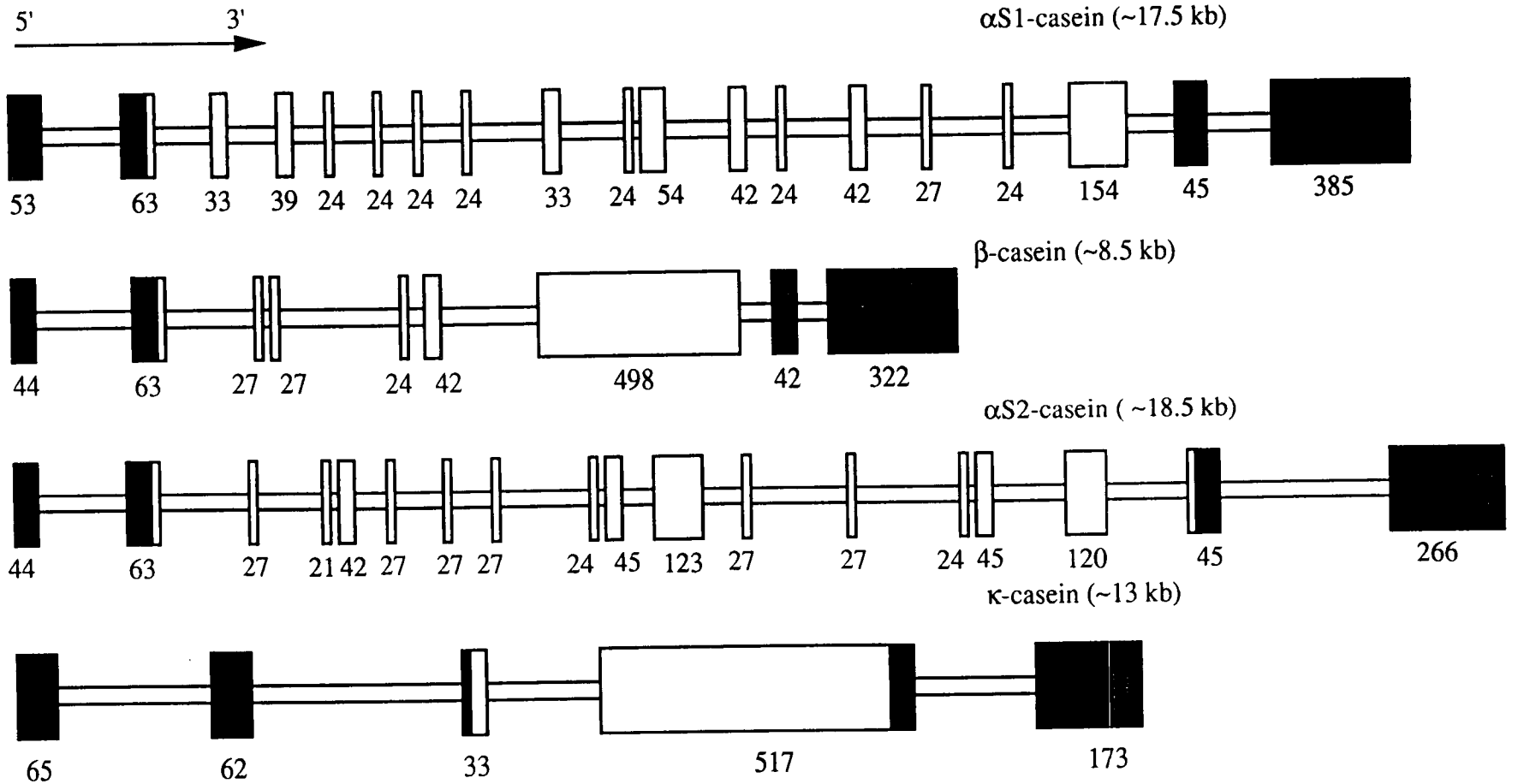


Fig. 1.2. Structural organization of the four bovine casein transcription units.

Horizontal boxes represent introns and vertical boxes represent exons. Black boxes represent 5' and 3' non-coding regions and signal peptide encoding exons. Sizes of exons are indicated in base pairs (modified from Martin and Grosclaude, 1993).

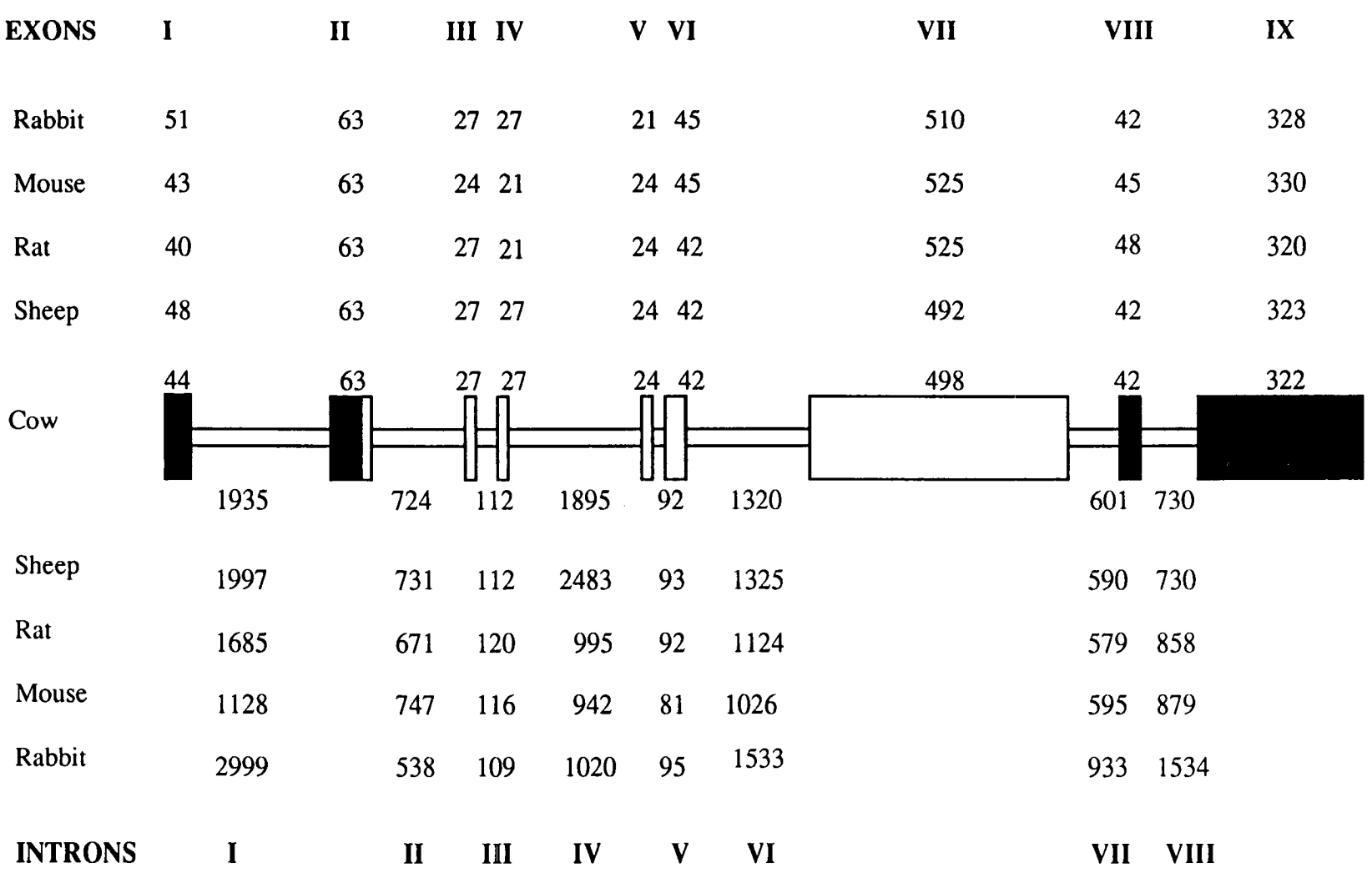


Fig. 1.3. Organization of the beta casein gene- comparison between different species (reference Martin and Grosclaude, 1993; Mercier and Vilotte, 1993). Relative sizes of exons and introns are not comparable.

exons. However, the human beta-casein gene contains sequences similar to exon 3 of β-casein from other species. Therefore, the absence of sequences similar to exon 3 in the mature mRNA of human β-casein suggests the possibility of exon skipping (Menon et al., 1992; Martin and Leroux 1992). Yu-Lee and Rosen (1983) established that rat γ-casein gene comprises nine exons and span a region of 15 kb.

1.2.2. Comparison

1.2.2.1. Proteins

Rosen and Shields (1980) reported that the sequences for homologous caseins in different species are more closely related than sequences for different caseins in the same species. Significant sequence identity is observed between rat α-casein and bovine αS1-casein (Hobbs and Rosen, 1982). The signal peptide sequence and phosphorylation sites of the two variants of guinea pig casein-B are similar to bovine αS1 and rat α-caseins respectively (Hall et al., 1984a) .

Bovine and ovine αS2-caseins are 88% similar at the amino acid level (Boisnard and Petrissant, 1985). Guinea pig casein-A exhibits 30% similarity with bovine αS2-casein (Hall et al., 1984b). The amino acid sequence of mouse ε-casein shows sequence similarity with the N-terminal regions of the bovine αS2-casein. The signal peptide sequences of the mouse ε-casein is similar to the signal peptide sequences of the calcium sensitive caseins from other species (Hennighausen et al., 1982a).

Among the various caseins, β-casein is the most hydrophobic protein and has a high proline content (16.7%). Human β-casein shows 47% similarity with bovine β-casein (Greenberg et al., 1984). Amino acid sequence comparisons of β-caseins between species (rat, ovine, bovine and human) shows that ~38% of the

amino acids are conserved and that the conserved residues occur in clusters throughout the protein.

In comparison with bovine κ-casein, sheep and goat κ-caseins have two insertions and 26 and 28 replacements respectively, whilst the rat and bovine κ-caseins show 49% sequence similarity. The deduced amino acid sequence of mouse κ-casein (Thompson et al., 1985) shows 73% sequence similarity with rat and 43% with ovine κ-caseins. Apart from the signal peptide sequence, the region attacked by the protease, rennin, is also highly conserved.

The amino acid sequences of the signal peptide region of various caseins in different species are shown in Fig. 1.4. It is evident from a comparison of these sequences that the signal peptides are highly conserved. This high degree of similarity among the signal peptides could be attributed to functional constraints on this sequence and/or to a common evolutionary origin of the calcium sensitive caseins.

1.2.2.2. mRNA/cDNA

The α, γ and β murine caseins appear to be homologous to αS1, αS2 and β bovine caseins respectively (Rosen and Shields, 1980). Sequences of αS1-like casein mRNAs from the cow, sheep, rat and guinea-pig have diverged with a large number of point mutations and sequence rearrangements (reviewed by Bonsing and MacKinlay, 1987). In general, for the point mutations found between the three species silent and replacement mutations occur with approximately equal frequency. This may be due to very loose functional constraints which have allowed caseins to diverge at a very rapid rate. Similar comparisons for proteins such as haemoglobins, show silent mutations far out-numbering replacement mutations (reviewed by Bonsing and MacKinlay, 1987). Comparison of the mRNA sequences showed three regions of unusual conservation among the

a. αS1 -casein

	-15					-10									-1
Cow	M	K	L	L	I	L	T	C	L	V	A	V	A	L	A
Sheep	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
goat	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Pig	--	--	--	--	--	F	I	--	--	A	--	--	--	--	--
Rat	--	--	--	--	--	--	--	--	--	--	--	A	--	--	--
Mouse	--	--	--	--	--	--	--	--	--	--	--	A	--	F	--
G-pig (B)	--	--	--	--	--	--	--	--	--	--	--	S	--	V	--
Rabbit	--	--	--	--	--	--	--	--	--	--	--	T	--	--	--

b. αS2-casein

	-15					-10									-1
Cow	M	K	F	F	I	F	T	C	L	L	A	V	A	L	A
Sheep	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
goat	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Pig	--	--	--	--	--	--	--	--	--	--	--	--	--	F	--
Rat	--	--	--	--	--	--	--	--	--	V	--	A	--	--	--
Mouse (ε)	--	--	--	I	--	L	--	--	--	--	--	--	--	--	--
Mouse (γ)	--	--	--	--	--	--	A	--	--	V	V	--	--	--	--
G-pig (A)	--	--	L	--	--	L	--	--	--	--	--	--	--	--	--

c. β-casein

	-15					-10					-5				-1
Cow	M	K	V	L	I	L	A	C	L	V	A	L	A	L	A
Sheep	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
goat	--	--	--	--	--	--	--	--	--	--	--	--	--	I	--
Pig	--	--	L	--	--	--	--	--	F	--	--	--	--	--	--
Rat	--	--	--	F	--	--	--	--	--	--	--	--	--	--	--
Mouse	--	--	--	F	--	--	--	--	--	--	--	--	--	--	--
Rabbit	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Man	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

d. κ-casein

	-21						-15					-10					-5				-1
Cow	M	M	K	S	F	F	L	V	V	T	I	L	A	L	T	L	P	F	L	G	A
Sheep	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
goat	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Pig	--	--	--	--	S	--	--	I	--	P	--	--	--	--	--	--	--	--	--	--	--
Rat	--	--	R	N	--	I	V	--	M	N	--	--	--	--	--	--	--	--	--	A	--
Mouse	--	--	R	N	--	I	V	--	M	N	--	--	--	--	--	--	--	--	--	A	--
G-pig	--	--	--	--	--	L	--	--	--	N	--	V	--	--	--	--	--	--	--	A	--
Rabbit	--	--	--	H	--	L	--	--	--	N	--	--	--	V	--	--	--	--	--	A	--
Man	--	--	--	--	--	L	--	--	--	N	A	--	--	--	--	--	--	--	--	A	--

Fig. 1. 4. Comparison (based on bovine caseins) between the signal peptide amino-acid sequences of caseins in different species (reproduced from Fiat and Jolles, 1989; Mercier and Vilotte, 1993)

members of the casein gene family, the 5' untranslated regions, and the sequences encoding the signal peptide sequences and the major sites of casein phosphorylation and calcium binding (Hobbs and Rosen, 1982; Yu-Lee et al., 1986).

A feature of rat and mouse α-casein sequences is an insertion of 10 repeated elements of 18 nucleotides each, in the coding region, which encodes a relatively hydrophobic peptide. This insertion appears to have occurred 7-12 million years ago, just prior to the divergence of the rat and mouse lineages (Hobbs and Rosen, 1982). The insertion of this sequence before the divergence of the two rodent species would explain the larger size of rodent α-caseins compared to other mammalian α-caseins. Overall sequence similarities between the αS1-like caseins in different species is shown in Table 1.3

The αS2-like casein genes appear to have diverged rapidly as a result of major sequence rearrangements. The duplication observed in the amino acid sequences of αS2-casein is also clearly present in the cDNA sequences of the bovine, ovine, cavine and rat homologues but not in the mouse sequence. Mouse ε-casein has a different, but smaller duplication which gives rise to a repetition of amino acid residues. This comparison suggests that the duplication occurred in a common ancestor, to these five mammalian species. It may be possible that the duplicated sequences present in the other casein sequences have been lost by subsequent deletions in the relatively recently evolved mouse. Examination of the duplication in the cDNAs revealed that one or more small sections of sequence were missing in each species relative to others and may be due to deletion of whole exons in the chromosomal DNA or due to point mutations in the vicinity of the splice junctions which interfere with the correct splicing mechanisms (reviewed by Bonsing and MacKinlay, 1987). Table 1.4 shows the overall identity between the αS2-like caseins in various species.

Table 1.3. % Similarity of αS1-casein cDNAs of different species

	Sheep	Goat	Pig	Rabbit (α)	G.pig (B)	Rat (α)	Mouse (α)
Cow	94.5	94.2	79.6	72.4	69.0	67.1	71.3
Sheep		98.5	79.3	72.8	67.9	67.6	70.1
Goat			79.1	72.6	68.4	66.2	69.5
Pig				74.8	70.2	68.9	68.9
Rabbit					70.9	69.2	72.1
G.pig						67.2	69.1
Rat							88.1

Table 1.4. % Similarity of αS2-casein cDNAs of different species

	Sheep	Goat	Pig	G.pig (A)	Rat (γ)	Mouse (γ)	Mouse (ε)
Cow	94.8	94.6	82.0	68.0	66.1	68.1	74.4
Sheep		98.0	82.8	68.1	65.4	62.2	74.9
Goat			82.4	67.6	66.6	68.8	76.6
Pig				68.6	70.4	65.1	67.7
G.pig					74.9	76.0	72.2
Rat						86.8	62.6
Mouse (γ)							70.8

Table 1.5. % Similarity of β-casein cDNAs of different species

	Sheep	Pig	Rabbit	Rat	Mouse	Human
Cow	96.5	85.5	74.3	69.6	70.3	78.4
Sheep		85.4	74.1	69.9	70.2	78.0
Pig			73.9	71.3	70.4	76.9
Rabbit				69.7	68.7	75.9
Rat					89.6	70.1
Mouse						71.6

Data compiled by a bestfit comparison of the cDNA sequences available in the data base.

The β-casein cDNAs of rat and mouse are 89% identical. Significant similarity is also observed between the rat and bovine β-casein, except in the C terminal encoding region which may be due to a high rate of point mutations and a number of small deletions (Bonsing and MacKinglay, 1987). In contrast to the α-casein, the lack of major rearrangements amongst the β-caseins suggest that the hydrophilic N-terminal and hydrophobic C-terminal domains are important for β-casein function. Beta-casein cDNA sequences are compared between species and the overall similarities are shown in Table 1.5.

The κ-casein cDNA sequences of rat and mouse are closely related (87% similarity) but have diverged considerably from the bovine sequence (Thompson et al., 1985). Since κ-casein stabilizes the casein micelle against precipitation by calcium, the sequence that determines the sensitivity of κ-casein to limited proteolysis by chymosin needs to be maintained. It has also been suggested that conservation of sequence is much more pronounced downstream of the chymosin-sensitive bond than upstream and it appears that hydrophobicity is more important than amino acid sequence in the upstream region of κ-casein (Bonsing and MacKinlay, 1987). Even though point mutations and a number of insertions/deletions have occurred, the overall architecture is preserved presumably as a result of functional constraints on the κ-casein structure (Bonsing and MacKinlay, 1987). cDNA sequence similarities between κ-caseins in different species are summarised in Table 1.6.

The cDNA sequences of various murine caseins have been compared with each other and the results are presented in Table 1.7. Nucleotide sequences from the signal peptide region of various caseins from different species are compared in Fig. 1.5.

Table 1.6. % Similarity between the cDNA sequences of κ-casein in different species.

	Sheep	Goat	Pig	G.pig	Rat	Mouse	Human
Cow	93.9	93.7	82.2	67.8	67.3	65.5	76.7
Sheep		97.7	82.7	68.7	68.0	66.8	77.3
Goat			83.0	68.1	67.4	67.2	77.9
Pig				68.4	67.7	67.4	79.1
G.pig					64.7	62.2	70.0
Rat						87.1	68.3
Mouse							67.0

Data compiled by a bestfit comparison of the cDNA sequences available in the data base.

Table 1.7. % Similarity between cDNAs of mouse caseins.

	β	γ	ε	κ
α	78.3	72.6	66.7	70.5
β		75.3	74.6	76.0
γ			70.8	63.6
ε				70.4

Data compiled by a bestfit comparison of the cDNA sequences available in the data base.

Fig. 1.5. Nucleotide sequences of casein encoding cDNA, in the region specifying the signal peptide. Numbers refer to the codons of the signal peptide (-) and the mature protein (+). Dashes represent nucleotides identical to those of each bovine cDNA taken as reference. Lower case letters refer to substituted nucleotides that did not specify any change of codon meaning (reproduced from review by Mercier and Vilotte, 1993).

a. αS1-casein

Cow	-15 ATG	AAA	CTT	CTT	ATC	-10 CTC	ACC	TGT	CTT	GTG	-5 GCT	GTT	GCT	CTT	-1 GCC	+1 AGG	CCT	AAA	CAT	+5 CCT	ATC	AAG	CAC
Sheep	---	---	---	--c	---	--t	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Goat	---	---	---	--c	---	--t	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--T	---
Pig	---	---	---	--c	---	T-t	-T-	---	---	-Ca	---	---	--c	---	---	---	---	---	-C-	---	C--	-G-	--t
Rat (α)	---	---	---	---	---	---	---	--c	--c	---	---	-C-	---	---	--t	CT-	---	-G-	GC-	-A-	CGt	-GA	A-t
Mouse (α)	---	---	--c	--c	---	---	---	--c	--c	---	---	-C-	---	T--	--t	AT-	--c	-G-	-T-	-A-	-Gt	-GA	A-t
G-pig (B)	---	--g	---	---	---	---	---	--c	--g	---	---	TC-	---	G-g	---	-T-	--g	---	TT-	--c	T--	-G-	---
Rabbit	---	--g	---	--c	---	---	--t	--c	---	---	---	AC-	---	---	---	---	-A-	---	TT-	-A-	T-a	GGA	---

b. αS2-casein

Cow	-15 ATG	AAG	TTC	TTC	ATC	-10 TTT	ACC	TGC	CTT	TTG	-5 GCT	GTT	GCC	CTT	-1 GCA	+1 AAG	AAT	ACG	ATG	+5 GAA	CAT	GTC	TCC
Sheep	---	---	---	---	--t	---	---	---	---	---	--c	---	---	---	---	---	C--	-A-	---	---	---	---	---
Goat	---	---	---	---	--t	---	---	---	---	---	--c	---	---	---	---	---	C--	-A-	---	---	---	---	---
Pig	---	---	---	---	---	---	---	---	---	---	--c	---	--t	T--	---	---	C--	GA-	---	--g	---	---	---
Rat (γ)	---	---	---	---	---	--c	---	---	--g	G--	---	-C-	--t	--g	--t	---	C--	G--	G-A	A-G	G--	AAA	C--
Mouse (ε)	---	---	---	A--	--t	C-G	--t	---	---	---	--c	---	--t	---	---	---	C-G	-G-	---	--g	--A	TA-	AT-
Mouse (γ)	---	---	---	---	---	---	G--	---	--g	G--	-T-	---	--t	--g	---	---	C-C	GAA	--A	A-G	G--	AAG	---
G-pig (A)	---	---	C--	---	---	--G	---	---	--c	---	---	--c	--t	--c	---	--C	C-C	-A-	TCA	--g	--A	CAG	---

c. β-casein

Cow	-15 ATG	AAG	GTC	CTC	ATC	-10 CTT	GCC	TGC	CTG	GTG	-5 GCT	CTG	GCC	CTT	-1 GCA	+1 AGA	GAG	CTG	GAA	+5 GAA	CTC	AAT	GTA
Sheep	---	---	---	---	---	---	---	--t	---	---	---	---	---	---	---	---	---	-A-	---	---	---	---	---
Goat	---	---	---	---	---	---	---	--t	---	---	---	---	---	A--	---	---	---						
Pig	---	---	C--	---	---	---	---	---	T-C	---	---	--t	---	---	---	---	-C-	AA-	---	---	---	---	-C-
Rat	---	---	---	T--	---	---	---	---	--t	---	--a	--t	--t	---	---	--g	---	AA-	--T	-C-	T--	-C-	---g
Mouse	---	---	---	T--	---	--c	---	---	--t	---	--c	--t	--t	---	---	---	---	ACT	AC-	TTT	ACT	GTA	TCC
Rabbit	---	---	---	---	--t	---	---	---	---	---	---	--c	--t	---	---	--g	---	AA-	---	C--	---	-G-	--t
Man	---	---	---	---	---	--c	---	---	---	---	---	--t	--t	---	---	--g	---	ACC	AT-	---	AG-	CTT	TC-

d. κ-casein

Cow	-21 ATG	ATG	AAG	AGT	TTT	TTC	-15 CTA	GTT	GTG	ACT	ATC	-10 CTG	GCA	TTA	ACC	CTG	-5 CCA	TTT	TTG	GGT	-1 GCC	+1 CAG	GAG
Sheep	---	---	---	---	---	---	---	---	---	---	---	--a	---	---	---	---	---	---	---	---	---	---	---
Goat	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Pig	---	---	---	---	-C-	---	A--	---	---	C--	---	---	---	---	--a	---	--t	---	---	---	--a	G--	---
Rat	---	---	-G-	-A-	---	A--	G--	---	A--	-A-	---	--a	---	c--	--t	---	--c	---	---	-C-	--a	G--	-T-
Mouse	---	---	-G-	-A-	---	A--	G--	--g	A--	-A-	--t	---	---	---	--t	---	--c	---	---	-C-	--a	G--	ATa
G-pig	---	---	--a	tc-	---	C-t	---	---	---	-A-	--a	G--	---	---	--t	---	--t	---	---	-C-	--a	G--	-T-
Rabbit	---	---	---	CA-	---	C-t	---	---	---	-Ac	---	---	---	G--	---	t--	--t	---	---	-C-	--a	G-C	ATa
Man	---	---	---	---	---	C-t	---	---	--c	-A-	GC-	---	---	---	---	---	--t	---	---	-C-	-Tg	G--	-TT

Fig. 1.5

1.2.2.3. Genes

Koczan et al. (1991) compared the bovine αS1-casein gene with other casein genes and reported that the first exon is non coding. The leader peptide and the first two amino acids of the mature protein are coded by exon 2. A similar arrangement is found in the mouse and bovine β-casein genes and in all the rat casein genes (Yoshimura and Oka, 1989). A major phosphorylation site within the αS1 sequence is created by splicing of the exons 9 and 10 (Koczan et al., 1991). A similar phenomenon was found in the rat α-casein gene (Yu-Lee et al., 1986) and the β-casein genes of rat, mouse and bovine (Yu-Lee et al., 1986; Yoshimura and Oka, 1989; Stewart et al., 1987). The bovine αS2-casein gene is organized like that encoding αS1-casein, but is more closely related by sequence to the β-casein gene (Groenen et al., 1993)

Characterisation of the 7.2 kb rat β-casein gene revealed three regions of greater than 70% similarity to the rat γ-casein gene within the first 200 base pairs of conserved sequences of the 5' flanking region (Jones et al., 1985). The immediate 5' flanking region (about 200 bp) of goat β-casein gene is 65% identical to mouse and 75% to rat but negligible similarity was observed in the region upstream. Goat and cattle exhibit 95% identity over the entire 1.7 kb 5' flanking region of the β-casein gene (Roberts et al., 1992). Yoshimura and Oka (1989) compared the sequences of the mouse and rat β-casein genes and found that the exon/intron organization of these two genes are highly conserved. Further comparative analysis between mouse β-casein gene and three other rat casein genes (α, β and γ) and bovine αS1-casein gene revealed the presence of four highly conserved sequences, three in the 5' flanking region and one in the intron. The four conserved sequences of these casein genes were not found in the genes of other milk proteins suggesting that these sequences are specific to the calcium sensitive casein gene family. It has been suggested that a consensus sequence

coined the 'milk box' might be shared by proximal 5' flanking regions of α-lactalbumin and the calcium sensitive casein genes (Hall et al., 1987; Laird et al., 1988).

The structure of the 13 kb long bovine κ-casein gene is not related to the other casein genes. Both the length and sequence of the signal peptide encoding region differ from the other casein genes. Also the 3' untranslated region is shorter as compared to the calcium sensitive casein genes. Many potential binding sequences for various factors (reviewed by Maga and Murray, 1995) like, mammary gland specific factor (MGF; Schmitt-Ney et al., 1991) or milk protein binding factor (MPBF; Watson et al., 1991), nuclear factor1 (NF1; Goyal et al., 1990; Li and Rosen, 1995) and for hormones such as glucocorticod and progesterone are present in the 5' untranslated region of the bovine calcium sensitive casein genes, but not in the κ-casein gene.

1.2.3. Expression

1.2.3.1. Native genes

Hobbs et al. (1982) examined the level of expression of different casein and WAP genes in the mammary glands of virgin, pregnant and lactating rats by analysing the corresponding mRNAs. They reported markedly different basal levels of α, β, γ caseins and WAP mRNAs in the virgin mammary gland, with α-casein mRNA present at levels of several hundred molecules per cell and WAP mRNA at only a few molecules per cell. During early gestation, the levels of these mRNAs increase, but not in a co-ordinated manner, and by mid-gestation β-casein mRNA is the predominant rat casein mRNA. WAP mRNA levels increase more slowly in early gestation but by late gestation appear to increase co-ordinately with the three calcium sensitive casein mRNAs. Non-co-ordinate expression of β casein and WAP genes has also been observed in mice, both in terms of basal

level of expression and the point at which mRNA first accumulates during gestation (Harris et al., 1991). Beta casein is expressed at low levels in virgin mammary gland and then rapidly induced from day 10 of gestation, while the onset of expression of WAP is delayed until the 14th day of pregnancy (Pittius et al., 1988; Harris et al., 1991). Differences in the kinetics of α and κ-casein mRNA accumulation have also been shown during mammary development in rodents by *in vitro* studies (Vonderhaar and Nakhasi (1986). The level of κ-casein mRNA is low but detectable until mid-gestation and then increase markedly in concert with α-casein mRNA, maintaining a relative level two to three fold less than α-casein mRNA at later stages of gestation and lactation. There is no evidence of co-ordinate expression of α-lactalbumin and caseins in rat mammary gland during pregnancy and lactation (Nakhasi and Qasba,1979).

Expression analysis of native or modified genes using *in vitro* transcription systems, mammary cell lines and transgenic animals has greatly improved the knowledge of the functioning of milk protein genes. Most of the expression studies have been carried out using β-casein genes from various species. Mammary epithelial cell lines such as murine COMMA 1D (Danielson et al., 1984) and HC11 (Ball et al., 1988) provide the best current *in vitro* model systems for the study of regulated expression of the endogenous β-casein gene, native, shortened/truncated β-casein genes from various species, and β-casein promoter-driven hybrid genes. The synergistic action of lactogenic hormones and extracellular matrix on β-casein gene expression is well established (reviewed by Mercier and Vilotte, 1993). Comparative expression analysis of native and altered rat (Doppler et al.,1989), murine (Yoshimura and Oka, 1990) and bovine (Schmidhauser et al., 1990) β-casein genes transfected into mammary cell lines suggested that most prevalent *cis* regulatory elements mediating the hormonal and extracellular matrix effects might occur within 2.6 kb of the 5' flanking region.

1.2.3.2. Transgenes

Although, *in vitro* systems like cell culture/tissue culture are efficient systems for studying the regulated expression of genes, these systems lack the natural physiological environment. The development of transgenic animals has improved the study of foreign gene expression *in vivo* in the correct physiological environment. Regulated expression of experimentally modified genes can be evaluated in a better way in whole animals (transgenic animals) than in cultured cells. Mammary specific expression of a transgene requires the presence of promoter and regulatory regions from a milk protein gene, which is expressed during lactation.

Transgenic studies have been carried out in various species using native, shortened or hybrid genes. Lee et al. (1988) reported very low expression of a rat β-casein transgene comprising the transcription unit plus 3.5 kb of 5' and 3 kb of 3' flanking region in transgenic mice. Low level expression was also observed in transgenic rabbits carrying a human interleukin-2 gene driven by 2 kb of rabbit β-casein 5' flanking region (Buhler et al., 1990). Mammary gland specific expression of bovine αS1-casein derived transgenes (bovine cDNA plus 1.35 kb 5' and1.5 kb 3' flanking regions of bovine αS1 casein gene and another construct containing 1.35 kb 5' flanking region, the bacterial CAT gene and a polyadenylation signal derived from SV40) was also reported, but at a very low level (Clarke, et al., 1994). In contrast, very high, stage specific expression was observed in transgenic mice carrying a caprine β-casein transgene with 3 kb and 6 kb respectively of 5' and 3' flanking sequences (Persuy et al., 1992). Greenberg et al. (1991) reported the production of biologically active follicle stimulating hormone (FSH) in the milk of transgenic mice carrying a fusion gene comprising a FSH cDNA and a modified rat β-casein gene. DiTullio et al. (1992) observed mammary gland specific expression of a cystic fibrosis transmembrane

conductance regulator (CFTR) cDNA under the control of the goat β-casein gene promoter in transgenic mice.

A fusion gene comprising the human lactoferrin cDNA with 6.2 kb or 14.2 kb 5' and 6.5 kb 3' promoter/regulatory sequence from the bovine αS1-casein gene was expressed at low levels in the mammary gland of transgenic mice (Platenburg, et al., 1994). Fusion genes comprising the cDNA for human lysozyme and 5' promoter elements from either the bovine αS1 (21 kb 5' region and 2 kb 3' region) or β-casein (3.7 kb 5' region and 4.5 kb 3' region) gene have been used to produce transgenic mice. Tissue specific expression of the transgene was observed only in those mice containing the promoter element from the αS1-casein gene. The expression of human lysozyme in bovine milk could be beneficial, altering the overall microbial level and the functional and physical properties of milk, as lysozyme is positively charged and antimicrobial (Maga et al., 1994). Expression analysis of transgenic mice carrying individual bovine αS2 (gene with 8 kb of 5' and 1.5 kb of 3' flanking sequences), κ (gene flanked by 5 kb of 5' and 19 kb of 3' sequences) and β (gene flanked by 16 kb of 5' and 8 kb of 3' sequences) casein genes showed that neither the bovine αS2 nor the κ-casein transgenes were transcribed at detectable levels in the mammary gland. However, the bovine β-casein transgene was expressed in all lines of mice and relatively high level of expression was detected in 50 % of these lines (Rijnkels et al., 1995). Table 1.8 shows the expression levels of milk proteins - caseins and whey proteins - derived transgene constructs in transgenic animals.

1.2.3.3. Locus Control Region (LCR)

Random integration of transgenes into the host genome can display varying levels of expression depending upon the site of integration. The phenomenon, where the presence of other regulatory regions in the host genome at

Table 1. 8. Various mammary gland specific transgenes and their levels of expression (modified from Maga and Murray, 1995).

Source of promoter	Length of promoter used for transgene	Genes used for transgene construct	Animals selected for transgene expression	Levels of protein expression	Reference
WAP-rat	Entire gene	Rat WAP	Mice	27% of endogenous	Bayna and Rosen, 1990
WAP-mouse	2.5 kb 5'	Ha-ras oncogene	Mice	Produced tumors	Andres et al., 1987
WAP-mouse	2.6 kb 5', SV40 polyA	cDNA t-PA	Mice	300 ng/ml	Gorden et al., 1987
WAP-mouse	2.6 kb 5', SV40 polyA	cDNA t-PA	Goat	3 μg/ml	Ebert et al., 1991
WAP-mouse	Entire gene	Mouse WAP	Pigs	1 mg/ml	Wall et al., 1991
WAP-mouse	2.3 kb5', 1.6 kb 3'	cDNA human SOD	Mice	0.7 mg/ml	Hansson et al., 1994
WAP-mouse	2.6 kb 5', 1.6 kb 3'	cDNA human proteinC	Mice	3 μg/ml	Velander et al., 1992
WAP-mouse	2.6 kb 5', 1.6 kb 3'	cDNA human proteinC	Pigs	0.001-1 mg/ml	Velander et al., 1992
WAP-mouse	4.2 kb 5'	gDNA human proteinC	Mice	0.1-0.7 mg/ml	Drohan et al., 1994
WAP-rabbit	17.6 kb 5'	gDNA human α1-AT	Mice	10 mg/ml	Bischoff et al., 1992
WAP-rabbit	6.3 kb 5'	gDNA human GH	Mice	4-22 mg/ml	Devinoy et al., 1994
α-lactalbumin-g-pig	1195 bp 5', 395 bp 3'	G-pig α-la	Mice	NQ	Maschio et al., 1991
α-lactalbumin-bovine	Entire gene	Bovine α-la	Mice	1.5 mg/ml	Bleck and Bremel, 1993
α-lactalbumin-bovine	750 bp 5', 336 bp 3'	cDNA bovine α-la	Mice	0.0025-0.45 mg/ml	Vilotte et al., 1989
α-lactalbumin-bovine	477 bp 5', 336 bp 3'	cDNA bovine α-la	Mice	0.1 mg/ml	Soulier et al., 1992
α-lactalbumin-goat	Entire gene	Goat α-la	Mice	1.2-3.7 mg/ml	Soulier et al., 1992
β-lactoglobulin-sheep	4 kb 5', 1.9 / 7.3 kb 3'	Sheep β-lg	Mice	23 mg/ml	Simons et al., 1987
β-lactoglobulin-sheep	4 kb 5'	Human α1-AT	Mice	0.4-7.3 mg/ml	Archibald et al., 1990
β-lactoglobulin-sheep	4 kb 5'	Human α1-AT	Sheep	35 mg/ml	Wright et al., 1991
β-lactoglobulin-sheep	3/5.5/10.8 kb 5', 8 kb 3'	Sheep β-lg	Mice	1-8.5 mg/ml	Shani et al., 1992
β-lactoglobulin-sheep	3 kb 5'	cDNA human-SA	Mice	No expression	Shani et al., 1992
β-lactoglobulin-sheep	3 kb 5'	gDNA human-SA	Mice	2.5 mg/ml	Shani et al., 1992

(Contd)

β-lacoglobulin-sheep	1.8 kb 5', 4.6 kbWAP 3'	cDNA human SOD	Mice	10 ng/ml	Hansson et al., 1994
β-lactoglobulin-sheep	4.3kb/3.3 kb/2 kb/800bp/ 406 bp 5', 1.9 kb 3'	Sheep β-lg	Mice	5%-300% of endo-genous sheep gene	Whitelaw et al., 1992
β-lactoglobulin-sheep	146 kb 5', 1.9 kb 3'	Sheep β-lg	Mice	very low level	Whitelaw et al., 1992
β-lactoglobulin-sheep	76 bp 5' and 1.9 kb 3'	Sheep β-lg	Mice	No expression	Whitelaw et al., 1992
β-casein-rat	3.5 kb 5', 3 kb 3'	Rat β-casein	Mice	0.01-1%(endogenous)	Lee et al., 1988
β-casein-rat	2.3 kb5', 5.5 kb 5'	Bacterial CAT	Mice	Variable	Lee et al., 1989
β-casein rabbit	2 kb 5'	Human interleukin-2	Rabbits	430 ng/ml	Buhler et al., 1990
β-casein goat	3 kb 5', 6 kb 3'	Goat β-casein	Mice	12-24 mg/ml	Persuy et al., 1992
β-casein-goat	4.2 kb 5', 5.3 kb 3'	Goat β-casein	Mice	1 mg/ml	Roberts et al., 1992
β-casein-goat	4 kb 5', 4.5 kb 3'	cDNA CFTR	Mice	NQ	DiTullio et al., 1992
β-casein-goat	6.2 kb 5', 7.1 kb 3'	cDNA bovine κ-casein	Mice	0.9-1.5 mg/ml	Gutierrez et al., 1995
β-casein-bovine	16 kb 5', 8 kb 3'	Bovine β-casein	Mice	1-20 mg/ml	Rijnkels et al ., 1995
αS1-casein bovine	1.35 kb 5', 1.5 kb 3'	cDNA bovine αS1-casein	Mice	0.1% of endogenous	Clarke et al., 1994
αS1-casein bovine	1.35 kb 5', SV40 3'	Bacterial CAT	Mice	3 ng/ml	Clarke et al., 1994
αS1-casein bovine	2.9 kb 5', 3.5 kb 3'	cDNA human IGF-1	Rabbits	1 mg/ml	Brem et al., 1994
αS1-casein bovine	21 kb 5', 2 kb 3'	gDNA human urokinase	Mice	1-2 mg/ml	Meade et al., 1990
αS1-casein bovine	21 kb 5', 2 kb 3'	cDNA human lysozyme	Mice	0.25-0.71mg/ml	Maga et al., 1994
αS1-casein bovine	6.2/14.2 kb 5', 6.5 kb 3'	cDNAhuman lactoferrin	Mice	0.1-36 μg/ml	Platenburg et al., 1994
αS1-casein bovine	15 kb 5', 6 kb 3'	cDNA human lactoferrin	Cow	NQ	Krimpenfort et al., 1991
αS2-casein-bovine	8 kb 5', 1.5 kb 3'	Bovine αS2-casein	Mice	No expression	Rijnkels et al ., 1995
κ-casein-bovine	5 kb 5', 19 kb 3'	Bovine κ-casein	Mice	No expression	Rijnkels et al ., 1995

WAP- Whey acidic protein, β-lg- beta lactoglobulin, α-lg- alpha-lactalbumin, α1-AT- α1-Antitrypsin, Human SA - Human serum albumin, SOD- Super-oxide dismutase, CFTR- Cystic fibrosis trans-membrane conductance regulator, IGF- Insulin like growth factor, GH- Growth hormone, t-PA- tissue-Plasminogen acivator , CAT- Chloramphenicol acetyl transferase. NQ - Not quantitated

or near the site of integration of the construct influences the level and specificity of expression of the transgene, is called a position effect. However, some transgenes exhibit position-independent, copy number dependent, high levels of tissue specific expression (Grosveld et al., 1987; Forrester et al., 1987; Greaves et al., 1989; Bonifer et al., 1990). Such transgenes appear to be insulated from position effects. These observations suggested the presence of a locus control region (LCR) or dominant control region (DCR) (Grosveld et al., 1987) or locus activation region (LAR) (Forrester et al., 1987) within the transgene sequences near the gene/genes. Originally such a region was identified as a control region upstream from the human β-globin locus (Grosveld et al., 1987).

LCRs are characterised by tissue specific, developmentally stable DNaseI hypersensitive sites. In the human β-globin gene cluster (5'-ε-^{G}γ-^{A}γ-δ-β-3') a different class of super (major) hypersensitive sites for DNase-I was identified far from the individual genes (Tuan et al., 1985; Forrester et al., 1989). These sites have been named super hypersensitive sites (SHSS) because they appear earlier in the course of DNaseI digestion than do the local (minor) hypersensitive sites located either near the individual globin genes or throughout the cluster (reviewed by Gross and Garrard, 1988). Tuan et al. (1985) identified a region of 6-25 kb 5' upstream of the human ε-gene, containing five super hypersensitive sites (-6 kb, -11 kb, -14.5 kb, -17.5 kb and -21 kb) and a single site at about 20 kb 3' of the β-gene. Among these, three 5' upstream sites (-6 kb, -11 kb and -17.5 kb) and the 3' downstream site are erythroid specific sites.

Unlike the stage specific minor hypersensitive sites, these erythroid specific major hypersensitive sites are found in a variety of erythroid cells regardless of whether the principal expressed gene is embryonic, foetal or adult globin. Thus a region of ~90 kb is demarcated within the family of human β-globin gene cluster. The chicken β-globin cluster has four upstream super

hypersensitive sites, one of which is constitutively present while the other sites are erythroid specific (Reitman and Felsenfeld, 1990). Grosveld et al. (1987) constructed a minilocus linking ~21 kb 5' upstream region (containing all super hypersensitive sites), the human β-globin gene with all its known local regulatory elements, and a 12 kb 3' region containing the downstream super hypersensitive site. Transgenic mouse carrying this minilocus expressed the human β-globin gene in a position independent, tissue specific, copy number dependent manner. In contrast, it has also been shown that transgenic mice carrying β-globin gene with the local regulatory sites alone, either was not expressed at detectable levels or was expressed at a very low level irrespective of the copy number (reviewed by Evans et al., 1990).

Not all the hypersensitive sites are required for the action of the LCR. Expression analysis of human β-globin transgenes with deletions of various super hypersensitive sites showed that the presence of the -11 kb site alone is sufficient to provide the LCR function and transgene expression in mice in a copy number dependent manner (reviewed by Evans et al., 1990). Studies carried out in two erythroid cell lines such as mouse MEL cells and human K562 cells showed that fusion genes (homologous and heterologous) containing the human β-globin LCR exhibited a tissue specific, copy number dependent, high level of expression (van Assendelft et al., 1989). This suggests that the β-globin LCR can confer tissue specificity on a variety of genes, provided the genes possess appropriate local regulatory elements. Tuan et al. (1989) carried out transient expression studies in K562 cells using fusion gene constructs composed of shortened human β-globin LCR and a CAT gene. They observed a 300 fold increase in CAT activity when the construct contained 0.8 kb of human β-globin LCR having the -11 kb Super hypersensitive site, suggesting some kind of powerful enhancer activity, although the LCR clearly raises transcription levels in a way that the local β-globin family

enhancers and promoters do not. Jarman and Higgs (1988) have identified nuclear scaffold attachment regions in the human β-globin LCR, but these do not lie near the -11 kb super hypersensitive site.

Festenstein et al. (1996) identified a locus control region within the 2 kb 3' flanking region of human CD2 gene. They identified three clusters of DNase-I super hypersensitive sites in this 2 kb region, of which the upstream cluster (SHSS-1) functions as the major enhancer. Expression studies in transgenic mice showed that deletion or truncation of the downstream cluster (SHSS-3) resulted in mosaic expression due to position effect variegation. In contrast, integration of the transgene with a complete LCR at the centromere did not show variegation of expression. Bonifer et al. (1990) have reported the presence of a regulatory region, possibly a locus control region in the chicken lysozyme minigene. All the seven founder mice carrying the chicken lysozyme transgene with its 11.5 kb 5' and 5.5 kb 3' flanking sequences showed an integration site independent, copy number dependent, tissue specific high level expression of the transgene.

Thus, the LCR has a dominant effect on the chromatin structure, insulating the gene/genes in its domain against the presumed suppressive effects of surrounding heterochromatin.

As yet, a locus control region in the casein locus has not been identified. Unlike the human β-globin locus, where the individual gene expression is temporally and spacially regulated, all the genes in the casein locus are expressed in the pregnant and lactating mammary gland. Moreover, it has been shown that mice homozygous for a deletion/knockout of the β-casein gene, did not show any alteration in the mammary gland phenotype and/or the micelle forming ability of caseins (Kumar et al., 1994). Since deletion/alteration of the globin locus can result in anaemia, it appears that the selection pressure on the globin locus is much more strong as compared to the casein locus.

As illustrated in Table 1.8 very low levels of expression were observed for the bovine αS1 and the rat β casein transgenes and casein promoter driven heterologous genes, while both the bovine κ and αS2-casein transgenes did not show any detectable level of expression in transgenic mice. This may be due to the lack of sufficient 5'/3' sequences containing the regulatory elements and/or due to some kind of position effect exerted on the transgene with respect to its integration site in the genome. It may also be due to the absence of a distant regulatory region in the transgene constructs, as was the case for the human β-globin gene.

Studies carried out on the human β-globin gene showed that most of the transgenic animals (70%) failed to express detectable amounts of transgene when only local regulatory sites accompanied the introduced β-globin gene. Of those in which the transgene was expressed, the level of expression per copy was less than 1% of the endogenous gene (reviewed by Evans et al., 1990). When the globin mini locus (including the LCR) was introduced into transgenic animals, the transgene was expressed in a position independent and copy number dependent manner.

Among the casein transgenes, bovine β-casein with 16 kb of 5' flanking sequence (Rijnkels et al., 1995) and bovine αS1 driven human urokinase transgenes with 21 kb of 5' flanking sequence (Meade et al., 1990) showed integration site dependent, high level of expression in a few lines of transgenic mice, whilst a goat β-casein (Persuy et al., 1992) transgene with 3 kb 5' and 6 kb 3' flanking sequences showed copy number independent expression. In all other experiments, the casein transgenes or casein gene promoter driven transgene constructs contained flanking sequences ranging from 2-8 kb (5') and 1.5-19 kb (3').

So far, no reports are available on transgenic studies using a 'casein mini locus' containing regions from the 5' upstream and 3' downstream sequences of the locus. Further experiments need to be carried out to determine whether there are sequences within the casein locus which would confer integration site independent, copy number dependent, mammary gland specific expression of casein or other transgenes. For example, the 5' upstream and 3' downstream regions of the locus could be examined for super hypersensitive sites, which might help to identify any locus control region for the genes in the loci.

Interestingly, the first LCR described, not only confers position independent expression on a single gene but on a gene cluster - the β-globin locus. The regulatory mechanisms for multigene loci can be studied by generating transgenic mice carrying large DNA fragments. In the case of the human beta globin gene cluster the whole locus is required for full expression with the LCR as an essential element. In humans, not in chickens, the globin genes are arranged in the order in which they are expressed during development, with ε-globin showing an embryonic and early foetal profile and γ-globin being expressed in the late embryonic stage and throughout the foetal period. Expression of β-globin increases during the foetal period and becomes predominant in the adult. The LCR is required for activation of all these genes, since in naturally occurring mutants in which the LCR is deleted, the genes are intact, but silent. It has also been reported that a gene proximal to the LCR could compete with a distal gene for its expression, but one located distally will be unable to compete effectively (reviewed by Dillon and Grosveld, 1993).

1.2.4. Regulation

For the sake of completeness and to avoid unnecessary repetition/duplication, the regulation of casein and whey protein genes are reviewed here.

1.2.4.1. Role of hormones

Topper (1970) first demonstrated that mammary tissue explants from mid-pregnant mice synthesise caseins during short term incubation in a chemically defined medium containing prolactin, cortisol and insulin. This observation opened a new perspective in the understanding of hormone inducible gene expression, revealing that the interaction of multiple hormones may be involved in regulation of specific gene expression.

The best characterised of the elements known to be important for gene regulation is the promoter, the region immediately upstream of each transcription initiation site. The promoter consists of binding sites for upstream factors and the basic transcription machinery, which is a very large complex with the TATA binding factor (TBP) as the central component. DNA binding proteins such as transcription factors, CCAAT binding proteins, enhancer binding proteins, octamer binding proteins and hormones act to promote or inhibit the basic machinery (reviewed by Johnson and McKnight, 1989). It has been shown that these elements provide control at many levels, such as tissue specificity, developmental and temporal specificity and in regulating the level of gene expression. But it is important to note that a promoter alone generally cannot drive efficient expression of a gene in cell transfection experiments and often fails to give any expression in transgenic mice. Another type of element that can potentiate gene transcription is the enhancer, which contains a collection of sites that can bind activating or suppressing protein factors (reviewed by Dillon and Grosveld, 1993).

Initial studies carried out *in vivo* and *in vitro* cultures of mammary gland from rabbit (Houdebine and Gaye, 1975; Devinoy et al., 1978) rat (Matusik and Rosen, 1978) and mouse (Terry et al., 1977; Nagaiah et al., 1981) have clearly demonstrated that prolactin is responsible for the primary accumulation of casein mRNAs in the mammary gland. In order to determine the precise level at which prolactin exerts its effect, a comparison of the casein mRNA levels and the rate of transcription of casein genes has been carried out using different experimental approaches. From pulse labelling experiments, in rat mammary gland in tissue culture, it has been shown that prolactin exerts a pleotropic effect by increasing both casein gene transcription and the half life of newly synthesised mRNA molecules (Guyette et al., 1979). Similar results have been obtained by measuring the rate of β-casein transcription and the β-casein mRNA levels in pseudopregnant rabbits treated with prolactin (Teyssot and Houdebine, 1980). Rosen et al. (1986) reported that the major effect of prolactin on the regulation of casein mRNAs appears to be exerted at the level of nuclear RNA processing. Although the prolactin receptors have been identified and cloned (Davis and Linzer, 1989; Edery et al., 1989) little is known about how prolactin regulates transcription.

Although prolactin is primarily responsible for milk protein gene expression, glucocorticoids and progesterone also play major roles in the modulation of the complex events involved in this process. While glucocorticoids alone are unable to induce casein gene expression, the presence of such steroids is absolutely necessary for the full induction of casein mRNAs by prolactin in mammary tissue cultures of rabbit (Devinoy et al., 1978) rat (Matusik and Rosen, 1978) and mouse (Nagaiah et al., 1981). These results as well as those of Chomczynski et al. (1986), suggest that the regulation of β-casein mRNA stability requires the simultaneous presence of both prolactin and glucocorticoid in these species. Inhibition of glucocorticoid receptor complex binding to nuclear acceptor

sites has also been shown to cause a near complete suppression of casein mRNA in *in vitro* cultures of murine mammary gland (Majumder et al., 1983). Houdebine (1980) suggested that progesterone has a potent inhibitory effect on prolactin stimulation of mammary cells. Progesterone antagonism to casein gene expression appears to result from competition with glucocorticoid for binding sites in the mammary cytosol (Banerjee et al., 1982, 1983).

Danielson et al. (1984) reported the synthesis of caseins in the presence of insulin, hydrocortisone and prolactin in the COMMA1D mammary cell culture line. Using protein synthesis inhibitors, Poyet et al. (1989) have shown that post-transcriptional effects of lactogenic hormones on casein gene expression in COMMA1D cells are mediated by labile protein(s) exerting a selective effect on casein mRNA stability. Sankaran et.al. (1984) reported a decrease in casein gene expression on oestrogen depletion, which was reversed by oestrogen administration.

Hobbs et al. (1982) reported that glucocorticoids, in the absence of prolactin, stimulate considerably the level of WAP mRNA in tissue culture preparations of mammary gland, suggesting that the expression of the WAP gene is regulated independently of the casein genes. Welte et al. (1993) conducted *in vitro* DNase I footprinting experiments with a purified preparation of the glucocorticoid receptor from rat liver and showed the presence of multiple binding sites for the glucocorticoid receptors in the promoter regions of both the rat β-casein gene and the mouse WAP gene. The glucocorticoid receptor binding sites are in close proximity or even overlap with the binding sites for candidate factors involved in mammary cell specific gene expression and suggest direct co-operation between glucocorticoid receptor and mammary cell specific transcription factors in mediating the expression of milk protein genes.

Analysis of the 5' flanking regions of casein genes has revealed several structural features representing potential progesterone and glucocorticoid receptor binding sites. Both the γ and β-casein genes contain a 16 nucleotide sequence homologous to the chicken ovalbumin gene 5' sequence involved in progesterone receptor binding and induction of ovalbumin gene expression (Rosen et al., 1985). Rosen and his colleagues also suggested from their experiments with a γ-casein-CAT (chloramphenicol acetyl transferase) fusion gene that casein gene promoters when linked to a heterologous gene are extremely weak even in the presence of hormones. Although cells transfected with an entire rat β-casein gene (David-Inouye et al., 1986) expressed the gene, it was apparently not hormonally regulated. The loss of hormone responsiveness might be a result of transfection, cloning or due to the absence of some key regulatory (distant) sequences.

1.2.4.2. Role of introns

Lee et al., (1989) investigated the relative contribution of the promoter and intragenic sequences in the hormonal regulation of the β-casein gene. They used explant cultures derived from transgenic mice bearing the entire 7.2 kb rat β-casein gene and a rat β-casein promoter-CAT fusion gene, containing either 2300 or 524 bp of 5' flanking DNA, exon I and 490 bp of intron I. They observed a maximal 4.5 fold increase in CAT activity in the presence of lactogenic hormones in contrast to a 25 fold induction of the entire rat β-casein transgene.

The role of introns in the regulation of the β-lactoglobulin (whey protein) gene has also been reported. Whitelaw et al. (1991) reported that a hybrid gene containing 4.3 kb immediate 5' flanking region of sheep β-lg and a genomic minigene of human α1-antitrypsin expressed much more efficiently than a β-lg-α1antitrypsin-cDNA hybrid gene in transgenic mice. They also observed a similar

pattern with the intact β-lg gene compared to its corresponding intronless β-lg minigene in transgenic mice.

These results show that even though the promoter sequences contain elements important for tissue specific expression and prolactin induced transcription, they are clearly not sufficient to elicit the full expression of casein genes. The presence of putative progesterone binding sequences in the introns of bovine αS2-casein gene also suggests a regulatory role for the introns (Groenen et al., 1993)

1.2.4.3. Role of Mammary nuclear factors

Specific nuclear DNA binding factors, which regulate the transcription of differentially expressed genes, have been discovered in many cell types and tissues (reviewed by Johnson and McKnight, 1989). There are very few reports detailing the number of nuclear factors specific or enriched in mammary epithelial cells. Promoter regions of genes that are exclusively or preferentially expressed in mammary epithelial cells have been used to identify such factors. Several attempts have been made to identify the nuclear transcription factors that act on the milk protein gene promoters. Even though, a few nuclear transcription factors have been reported to have binding sites in the milk protein promoter region, these factors are not confined to the mammary epithelial cells (Schmitt-Ney et al., 1992a).

MGF

Mammary gland factor (MGF) is a member of a novel family of transcription factors termed signal transducers and activators of transcription (Stat). Hence, MGF is now known as Stat5. Two Stat5 homologues have been identified, Stat5a and Stat5b, which are 96% similar at the amino acid level but

differ significantly in their C-termini (Lui et al., 1995). The members of the stat family of transcription factors are present in many tissues and mostly participate in the signalling pathways in response to cytokines (reviewed by Ihle and Kerr, 1995).

Schmitt-Ney et al. (1991, 1992b) showed that certain sequences in the promoter region of rat β-casein gene were recognised by a mammary gland specific nuclear factor (MGF), which is important in the regulation of transcription. Activation of MGF is mediated through the action of kinases and phosphatases. Mutation of the MGF binding site in the β-casein gene promoter completely abolishes responsiveness of the promoter to lactogenic hormones in HC11 cells. Watson et al. (1991) have reported the presence of MGF binding sites, upstream from the transcription initiation site of sheep β-lg. Groenen et al. (1992) analysed the bovine casein genes and identified two nuclear factors which are involved in the expression of the αS2-casein gene and the related calcium sensitive αS1 and β-casein genes. One of these factors is an octamer binding protein that binds around -50 and the other, a mammary gland specific factor that binds around -90 of the promoter region. These binding sites are highly conserved in all calcium sensitive casein genes of mouse, rat, rabbit and cattle. Wakao et al. (1992) reported that MGF is a single polypeptide of 89 kDa which recognises a specific DNA sequence that is conserved with slight variations in the α and β-casein gene promoters of cattle and rodents.

Expression studies with native and modified milk protein genes in mammary cell lines, transgenic animals and DNA fingerprinting have demonstrated the occurrence of *cis*-acting regulatory elements in the proximal 5' flanking region of the bovine casein (except κ-casein) and whey protein genes (reviewed by Maga and Murray, 1995). One of these elements, which is recognised by a specific mammary nuclear factor, is common to αS1, αS2, and β-

caseins, β-lg, α-la and rodent WAP genes (Mercier, 1992). Pierre et al. (1992) located a regulatory region between nucleotide -3768 and -3155 in the 5' upstream region of the αS1-casein gene which enhances the prolactin induced promoter activity.

PMF

Experiments in mouse mammary cell culture showed that the progesterone regulation of β-casein gene expression is mediated by a pregnancy specific mammary nuclear factor (PMF). It is reported to be a transcriptional repressor that binds to two separate sites of the β-casein promoter region and mediates the inhibition of β-casein expression by progesterone. An inverse relationship was observed between the gene expression and PMF binding activity (Lee and Oka,1992).

YY1

The results of experiments conducted in the mammary epithelial cell line (HC11) suggests that the nuclear factor YY1 represses transcription of the β-casein gene while the addition of purified mammary gland specific factor (MGF) decreased YY1 binding to its recognition site (Meier and Groner (1994). This indicates that MGF regulates the DNA binding activity of YY1, thereby preventing transcriptional repression.

Casein gene expression and regulation are complex but a number of features common to all or many casein genes have been identified. Indeed, some of these elements/features are also shared by the whey protein genes.

1.2.5. Polymorphism

Milk protein polymorphism was reported as early as 1955 by Aschaffenburg and Drewry. They discovered two electrophoretically distinct β-lactoglobulins in bovine milk. Grosclaude et al. (1972) detected a polymorphism of bovine αS1-caseins by gel electrophoresis. So far, five bovine αS1-casein variants (A, B, C, D and E), two αS2-casein variants (A, D), seven β-casein variants (A^1, A^2, A^3, B, C, D and E) and two κ-casein variants (A and B) have been identified. These casein variants differ from each other either by amino acid substitution or insertion/deletion of peptide sequences (reviewed by Fiat and Jolles, 1989). Deletion of amino acid residues only occurs in two cases, αS1-casein A (deletion of amino acids from positions 14-26 compared to the B-variant) and αS2-casein D (deletion of amino acids from positions 51-59 compared to αS2-casein A). This latter deletion removes a ubiquitous putative multiple phosphorylation site of αS2-casein. Boulanger et al. (1984) reported the unusual complexity of the αS1-casein electrophoretic polymorphisms in the goat. Six variants (αS1-casein A, B, C, D, E, F) have been established for the caprine αS1-casein. Among these, the D and F variants have an internal deletion of 11 and 37 amino acid residues, respectively. Both deletions start at the same position (residue 59) and lead to the loss of a multiple putative phosphorylation site (Brignon et al., 1989, 1990).

Milk production is a sex and age limited trait, as milk production and milk protein synthesis are confined to lactating (i.e. older) females. Thus the main disadvantages in analysing milk protein polymorphism are 1) lactating (i.e. older) females are required to produce the milk 2) genetic studies require more time, as the phenotype (and genotype) of offspring cannot be determined until lactation 3) the milk protein genotypes of males can only be inferred from their female progeny. Moreover, as is true for all polymorphisms revealed as protein

electrophoretic variants, not all changes at the nucleotide level will cause a change in amino acid/protein, and/or an overall change in the charge that can be resolved by electrophoresis. Genotyping at the DNA level should circumvent these difficulties. The study of genetic polymorphism at the nucleotide level has led to the discovery of new alleles, has provided information about the mechanism responsible for the occurrence of deleted caseins, and has enabled animals to be genotyped at birth, which is a useful tool in the breeding programme.

Rando et al. (1987) reported that bovine κ-casein genotypes could be determined even in the absence of the gene product. They carried out a restriction fragment length polymorphism (RFLP) analysis of the DNA of cows using the endonuclease *Pst*I. Different DNA restriction patterns were observed for κ-casein AA and κ-casein BB cows. Both types have two invariant fragments of about 6.8 kb and 1.1 kb in addition to the polymorphic fragments - κ-casein AA type (4.3 kb and 0.3 kb), BB type (4.6 kb alone) and with κ-casein AB cows (all five fragments). RFLP analysis of the κ-casein locus was also carried out by Damiani et al. (1989). They observed polymorphic sites for the restriction endonucleases *Hin*dIII, *Hin*fI, *Mbo*II and *Taq*I and as well as two polymorphic sites for *Mbo*II and *Taq*I in the non coding region which differentiated the A allele into two new variants, A1 and A2. A *Hin*dIII restriction site polymorphism in the bovine κ-casein gene has also been reported (Pinder et al., 1991). In addition to the *Hin*dIII restriction site in the κ-casein, Cronin and Crocket (1993) reported a *Rsa*I polymorphism in cattle and bison. They mapped the *Hin*dIII and *Rsa*I restriction sites and identified three alleles. Preliminary screening revealed one allele specific to cattle, one to bison and one shared by the two species. No fixed allelic differences were observed among cattle breeds or bison breeds.

Threadgill and Womack (1990) reported RFLPs with four restriction enzymes, *Eco*RI, *Hin*dIII, *Msp*I and *Taq*I in the casein loci, in three breeds of

cattle, Brahman, Hereford and Jersey. All four enzymes revealed RFLPs at the αS1 locus (*Msp*I, two fragments; *Eco*RI, five fragments; *Taq*I, six fragments; *Hin*dIII, two fragments). In the case of the αS2 locus *Eco*RI and *Hin*dIII produced monomorphic restriction patterns while *Msp*I (six fragments) and *Taq*I (three fragments) were polymorphic. *Eco*RI showed a monomorphic pattern for the β and κ-casein loci while *Hin*dIII, *Msp*I and *Taq*I were polymorphic. Brahman appeared to exhibit much more variation at the milk protein loci than did either the Hereford or Jersey. Twenty seven restriction fragments were specific to the Brahman, but only eight fragments were specific to the Hereford. In zebu like breeds, two new polymorphic sites for *Hin*fI were observed in the 5' untranslated region of the β-casein gene (Sulimova et al.,1992).

Di-Gregorio et al. (1991) reported nine RFLP in sheep using seven endonucleases and four cDNA probes specific for the αS1, αS2, β and κ-casein genes. In contrast to the low level of variation observed at the protein level, this DNA analysis revealed considerable polymorphisms and a high level of heterozygosity that together represents a useful tool for genetic analysis.

Genetic analysis demonstrated that the extensive polymorphism in goat αS1-casein is under the control of at least seven autosomal alleles, termed αS1-casein A, B, C, D, E, F and O, which segregate according to Mendelian expectations (Grosclaude et al., 1987; Mahe and Grosclaude, 1989). Leroux et al. (1990) confirmed the protein polymorphism of the goat αS1-casein at the DNA level by RFLP analysis. Of the eleven restriction enzymes used, three enzymes, *Pst*I, *Taq*I and *Rsa*I allowed the identification of the various alleles. They also showed that the electrophoretic patterns of the αS1-casein^{E} allele display one fragment which is 0.5 kb longer than its allelic counter parts and suggests that this allele arose from an insertion occurring in a region involved in the control of

expression of the αS1-casein gene, since this allele is associated with low αS1-casein content.

RFLP analysis by restriction digestion and Southern blotting is both slow and time consuming. Methods based on polymerase chain reaction (PCR) technology (Saiki et al., 1988) offer faster and more sensitive alternatives.

The complete structural organization of the goat αS1-casein transcription unit was established from PCR experiments and it was demonstrated that both deletions occurring within variants F and D arise from an improper processing of the primary transcript (Leroux et al., 1992). Ng-Kwai-Hang et al. (1991) confirmed the κ-casein genotypes of Holstein bulls by PCR amplification of a 99 bp region of the κ-casein gene containing nucleotide substitutions diagnostic for the A and B variants. The two variants were confirmed by sequencing the amplified products. These analyses were performed on DNA isolated from semen and were in complete agreement with the analysis of milk samples for the protein variants in the corresponding daughters.

PCR can be combined with restriction analysis such that RFLPs are converted to PCR-RFLPs, in which the polymorphic fragments are revealed by ethidium bromide staining rather than by blotting and hybridisation or by sequencing the amplified product. For example, the κ-casein variants can also be identified by PCR-RFLP (Zadworny et al.,1990; Zadworny and Kuhnlein, 1990). A PCR amplified product (99 bp region) from the κ-casein gene in Holstein cows was restriction digested with *Mbo*II (A-variant specific) or *Taq*I (B-variant specific) and the results were found to be in agreement with known κ-casein phenotypes. Using PCR and restriction analysis Lien et al. (1992) identified multiple alleles (A1, A2, A3, and B) of β casein in cattle. Genotyping of the κ-casein locus of the crossbred Russian Black Pied and cross bred Zebu cattle by protein polymorphism and DNA polymorphism revealed a new allele of the κ-

casein gene (κ-Casein F; Sulimova et al.,1992). Other methods of detection of allelic variation include the use of allele specific primers and hybridising the Southern blotted PCR products with allele specific oligos.

The effect of the different protein variants on the physico-chemical and technological characteristics of milk has been analysed. For example many authors have reported that the A and B variants of κ-casein have different effects on cheese making. Sherbon et al. (1967) observed that milk from cows homozygous for the κ-casein B allele produces a coagulation of better consistency and a firmer curd than the milk from cows homozygous for the κ-casein A allele. Grosclaude (1988) reported that the κ-casein B variant is superior to the A variant in clotting properties and cheese-yielding capacity of milk. Anema and Creamer (1993) reported that micelle solvation (g H_2O/g dry casein micelle pellet) was often found to be lower for bovine milk containing either the B variant of αS1-casein or the A variant of κ-casein. They also found that these two casein variants were associated with a lower κ-casein content of the milk. In milk, micelle solvation is greatly influenced by the κ-casein content. Higher amounts of κ-casein are associated with smaller micelle size and greater solvation.

In the goat, the A, B and C alleles of αS1-casein are associated with a high amount (3.6 g/l), the D and F alleles are associated with low amount (0.6 g/l) and E allele is associated with an intermediate amount (1.6 g/l) of αS1-casein in milk. The αS1-casein^{O} allele may be a true null allele (Grosclaude et al., 1987; Mahe and Grosclaude, 1989). Milk from goats homozygous for the αS1-casein A allele have smaller micelles and provide significantly higher cheese yields than goats homozygous for the F allele. Also the milk from αS1-casein A homozygotes shows increased curd firmness, which may be due to the differences in protein structure (reviewed by Martin and Grosclaude, 1993).

Bovenhuis (1992, 1993) studied the association between milk protein genotypes and production traits using data on milk protein types. He reported that the κ-casein gene tends to be associated with protein content, β-lg with fat content and the β-casein gene with both protein and fat content of milk. Bovenhuis et al. (1992) reported the effect of milk protein genotypes using single gene analysis and multigene analysis. The results indicated that some effects ascribed to certain milk protein genes in the single gene analysis were not effects of the milk protein gene itself but of linked genes. It was found that κ-casein genotypes influence protein percentage, β-casein genotypes - milk yield, fat percentage and protein yield and β-lg genotypes - fat percentage. It was unclear whether this was due to the direct effects of milk protein genes or effects of linked genes. Martin (1993) reported that goat αS1-casein provides a remarkable model in which quantitative allelic variability occurs in addition to the classic structural polymorphism and is controlled by at least 14 alleles at the αS1 casein locus, distributed in 6 different classes of protein variants, which are divided into four levels of expression.

Lien et al. (1993) analysed three bovine casein loci, αS1, β and κ for genetic linkage. A total of 330 individual sperm cells from a triply heterozygous bull were selected and the DNA amplified in a two step polymerase chain reaction to achieve a high specificity of amplification. No recombinants were found among the 330 single sperm cells analysed, giving a lod score higher than 30 (i.e odds of 10^{30}:1 in support of linkage) and providing evidence for very close linkage between the bovine casein genes. They also observed that the frequencies of some casein variants in Norwegian cattle are very low (αS1-C, 0.05; β-A^3, 0.004; β-B, 0.02 and κ-B, 0.15).

1.3. Whey proteins

In addition to caseins, the major whey proteins present in milk are alpha lactalbumin (α-la), beta lactoglobulin (β-lg) and whey acidic proteins (WAP). Among these α-la is the major whey protein in most mammalian species (Brew and Hill, 1975), while WAP is the major whey protein, in mice and rats (Hennighausen and Sippel,1982b).

1.3.1. α-Lactalbumin gene

Alpha-lactalbumin is involved in lactose synthesis and the corresponding gene has been completely sequenced for several species including bovine (Vilotte et al., 1987), human (Hall et al., 1987) and caprine (Vilotte et al., 1991). The gene has a 2 kb transcription unit with 4 exons. It has been proposed that both α-la and lysozyme (an enzyme that catalyses the hydrolysis of a glycosidic linkage in polysaccharides) have arisen from a common ancestral gene (Brew et al.,1967). This hypothesis has been confirmed by a comparison with the structure of the chicken egg white lysozyme gene (Qasba and Safaya,1984). Both genes contain three introns at similar positions and the first three exons are quite similar in sequence - 43 to 56% identity. The positions of introns are conserved between human and rat α-la genes, but the intron sizes differ markedly between the two species.

In mice mammary explant culture, induction of α-la gene expression requires the synergistic action of insulin and prolactin (or placental lactogen) and is maximal in the presence of 3×10^{-8} M glucocorticoid, while high concentrations of glucocorticoid inhibit α-la gene induction at both the RNA and protein levels (Ono and Oka, 1980). This inhibitory effect can be reversed by thyroid hormone and prostaglandins (Bhattacharjee and Vonderhaar, 1984). Mammary tissue-specific expression of bovine (Vilotte et al., 1989) and caprine

(Soulier et al., 1992) α-la genes in transgenic mice has been reported (Table 1.8). Expression analysis of the bovine transgene shortened at the 5' end (Soulier et al., 1992) or substituted with a trophoblast interferon cDNA in the coding frame (Stinnakre et al., 1991) indicated that 0.4 kb of 5' and 0.34 kb of 3' flanking sequences are sufficient for mammary specific and correct developmental expression but not for high levels of expression.

1.3.2. β-Lactoglobulin gene

The complete sequence has been determined for the ovine β-lg gene (Harris et al.,1988) while the bovine gene has been partially sequenced (Silva et al., 1990). The ovine gene has a transcription unit of 4.7 kb, containing 7 exons. The organization of the gene is similar to that of genes encoding mouse major urinary protein (MUP) and retinol binding protein (RBP) (Ali and Clark, 1988). In sheep the β-lg gene is expressed initially at mid-pregnancy, increases slowly until parturition and more rapidly thereafter (Harris et al., 1991). In cultured ovine mammary explants, expression of β-lg appears to be less dependent on lactogenic hormones than are the casein genes. Moreover, in transgenic pigs it is shown that glucocorticoid and insulin have only a slightly synergestic effect on prolactin induction of β-lg expression (Shamay et al., 1992). In transgenic mice expression of β-lg transgene and the endogenous β-casein gene appears to be similar (Harris et al. 1991). In marsupials induction of α-la and β-lg depends only on prolactin (Collect et al., 1990, 1991).

Expression levels of β-lg transgenes are shown in Table 1.8. Tissue specific expression of the native ovine β-lg gene (Simons et al., 1987) and derivatives thereof have been reported in transgenic mice. Transgenic studies carried out with progressive resection of the 5' flanking sequences of the ovine β-lg gene indicated that 0.4 kb upstream from the transcription unit was sufficient

for high, tissue specific expression. The region -406 to -149 appears to be essential, but not necessarily sufficient for high tissue specific expression (Whitelaw et al., 1992). *In vitro* binding assays (Watson et al., 1991) identified at least five binding sites for nuclear factor 1 (NF1) and three sites for a specific mammary gland factor termed milk protein binding factor (MPBF) in the -406 bp promoter region of the β-lg gene. The recognition site for this milk protein binding factor might be a 13 bp palindromic nucleotide stretch, GATTCCNGGAACC, that is structurally similar to motifs shared by the proximal 5' flanking regions of the genes encoding other major milk proteins (Vilotte and Soulier, 1992).

Ovine β-lg gene sequences have been used in attempts to target the expression of human proteins to the mammary gland of transgenic animals. A hybrid gene consisting of 4.3 kb 5' flanking sequences including the promoter from the ovine β-lg gene fused to a mini gene encoding human α1-antitrypsin have been highly expressed in the mammary gland of one line of transgenic mice (7 mg/ml; Archibald et al., 1990) and in transgenic sheep (35 mg/ml; Wright et al., 1991). In contrast, for a hybrid gene comprising the human antihaemophilic factor-IX cDNA inserted into a ovine β-lg gene, expression was very low (25 ng/ml) in transgenic sheep (Clark et al., 1989).

1.3.3. Whey acidic protein (WAP) gene

The amino acid sequence inferred from cDNA sequences clearly indicates that WAP belongs to the 'four disulfide core' family, which includes wheat germ aggluttinin, several snake venom toxins and the neurophysins (Hennighausen and Sippel, 1982b). All these proteins share a common pattern of cysteine residues. The arrangement of the cysteines is quite similar in the two domains of WAP and not much different from the arrangement observed in bovine neurophysin. No

similarity was reported in other amino acid or nucleotide sequence between the two domains of WAP or between WAP and neurophysin (Hennighausen and Sippel, 1982b). The two cysteine domains of WAP gene are encoded by exons 2 and 3 which might have resulted from an intragenic duplication (Hennighausen et al., 1982b). The WAP gene has been sequenced for the mouse, rat (Campbell et al., 1984) and rabbit (Thepot et al., 1990). The gene extends over 2.8 kb in the rat and 3.3 kb in the mice and is composed of four exons divided by three introns. The difference in sizes of the WAP genes in mice and rats is primarily due to a difference between the third intron, 1.1 kb in mice but only 0.5 kb in rat.

WAP is abundantly expressed in mammary epithelial cells and the mRNA accounts for 10 to 15% of mammary polyadenylated RNA in lactating rat, mouse and rabbit (Hennighausen and Sippel, 1982b; Hennighausen et al., 1982b; Hobbs et al., 1982). Unlike β-casein, expression of the WAP gene is delayed in mouse mammary tissue until the 14th day of pregnancy (Harris et al., 1991., Pittius et al., 1988) and a several thousand fold increase occurs at mid-lactation as a result of proliferation and differentiation of epithelial cells. Induction and maintenance of endogenous WAP gene expression depends upon the synergistic action of lactogenic hormones (prolactin, glucocorticoid and insulin) and cell-cell and cell-extracellular matrix interactions. When mammary epithelial cells lack the correct spatial structure and cellular environment, secretion of at least one inhibitor might be responsible for the specific post transcriptional suppression of WAP expression (Chen and Bissell, 1989).

Analysis of WAP hybrid gene constructs e.g. WAP-*myc* (Schnenberger et al., 1990), WAP-CAT (Doppler et al., 1991) in transfected mammary organ culture and mammary epithelial cell line HC11, respectively revealed the presence of several regulatory motifs scattered along the 2.5 kb 5' region flanking the WAP gene. These features were confirmed by analysis of transgenic animals carrying

native and modified WAP genes. Levels of expression of native and hybrid WAP genes in transgenic animals are shown in Table 1.8. Transgenic mice carrying a rat WAP gene with 949 bp of 5' and 70 bp of 3' flanking DNA was expressed at levels comparable to or higher than that of the endogenous gene (Dale et al., 1992).

Lubon and Hennighausen (1987) reported the presence of four mammary nuclear protein-binding sites in the -175 to -88 region, upstream from the murine WAP transcription unit, which shares common motifs with proximal 5' flanking regions of other milk protein genes.

1.4. Chromosomal location of milk protein genes

Casein loci have been assigned to chromosome 5 in mouse (α, β, γ; Gupta et al., 1982), 12 in rabbit (α, β; Gellin et al., 1985), 4 in human (β; Menon et al., 1992), 4 in sheep (αS1, αS2, β, κ; Hayes et al., 1993a, b, Ghali et al., 1991; Leveziel et al., 1991), 6/4 in cattle (αS1, αS2, β, κ; Womack et al., 1989; Threadgill and Womack, 1990; Hayes et al., 1993a) and 8 in pigs (αS1, αS2, β, κ; Archibald et al., 1995; Rohrer et al., 1994). Chromosomes 4 and 6 are difficult to discriminate in domestic ruminants, hence, the discrepancy between the chromosomal assignment of casein loci in cattle (reviewed by Mercier and Vilotte, 1993). The assignment of the casein loci to a single chromosome supports the possibility that they are members of a multigene family. The organization of the bovine κ casein gene, together with its upstream sequence confirms that it is unrelated to the genes encoding the calcium sensitive caseins to which it is linked (Alexander et al., 1988). Long range restriction analysis revealed that in bovines the casein genes are arranged in the order of αS1-β-αS2-κ within a 200 kb fragment (Ferretti et al., 1990; Threadgill and Womack, 1990). The close proximity of the casein genes shows that the genes are tightly linked and might be

dependent on a locus control region as in the case of human β-globin locus. Leveziel et al. (1991) reported that casein genes are also very closely linked in the sheep genome.

Some of the genes encoding the major whey proteins have also been assigned to chromosomes. The α-la locus is present on chromosome 5, 3 and 12 in bovine, ovine and human respectively and the murine WAP locus occurs on chromosome 11. The β-lg locus was assigned to chromosome 3 in sheep and 11 in goat and cow and one pseudogene has been reported for β-lg in the ovine and caprine species (reviewed by Mercier and Vilotte, 1993).

1.5. Gene families and evolution

1.5.1. Multigene families

Multigene families are groups of genes with similar sequences and generally have related or overlapping functions that are thought to have arisen by gene duplications. Gene duplications due to unequal crossing over are well documented (Meada and Smithies, 1986). It is also reported that gene duplication may occur at substantial rates, 10^{-4} to 10^{-6}, per locus per generation (Shapira and Finnerty, 1986). New genes may evolve when mutations occur in the members of a multigene family for which selection has been relaxed (Ohno, 1970). Other mechanisms involved in the evolution of multigene families include gene conversion, replication slippage and duplicative transposition. As a result of the continued occurrence of such interactions, 'concerted evolution' of gene families often takes place, so that genes belonging to a family within a species, even if their chromosomal locations differ, are more similar than those in different species (Ohta, 1989). The various gene families range from those with extremely diverse functions, such as the protein kinase family (Hunter, 1987), to those with very variable members, e.g. immunoglobulin genes (Hood et al., 1975), and those

consisting of uniform members, like ribosomal RNA genes (Dover, 1982; Arnheim, 1983).

All existing gene families are the products of millions of years of evolution. The organization and diversity of each family reflect the effects of natural selection. Therefore it appears that the variability contained in each gene family is the product of natural selection for optimum diversity. The simplest way to acquire new functions would be through duplication of a whole gene with subsequent differentiation by nucleotide substitution. Many multigene families exemplify this mechanism - immunoglobulins, cytochrome P450, haemoglobins, growth hormone, interferon-α, esterase, lactic dehydrogenase, chorion, vitellogenin, kallikrein and others. In some cases, differentiation of regulatory regions such as enhancers or promoters is more important to the organisms than that of coding regions as seen in lactic dehydrogenase or esterase (Markert, 1987). Differentiation in both regions may be significant in still other cases as in the haemoglobin and chorion genes (Perutz, 1983; Goldsmith and Kafatos, 1984).

Multigene families with uniform members like those of ribosomal RNA and histone are special cases of gene duplication, in which functional uniformity rather than diversity of gene copies is essential. The redundancy is acquired because of the need for a large amount of the corresponding gene products. Thus, the expression capacity has been the main subject for evolutionary adjustment in such gene families. However, in some cases there is spatial or temporal differentiation in expression among subgroups of a gene family such as oocyte or somatic cell types of *Xenopus* 5SrRNA (Federoff, 1979), and therefore, functional differentiation exists even in such gene families.

Whether or not the members of a gene family are uniform, they would have evolved to meet the needs of an organism. The rat γ-crystallin gene family represents an example of a mixed pattern of concerted and divergent evolution

(den Dunnen et al., 1986). There are six tandemly arranged genes in this family, each contains internal repeats, separated by introns. By comparing the six gene copies, it was found that the sequences of the second exons are very similar presumably as a result of concerted evolution, whereas the third exons are very different and thus show evidence of divergent evolution. Thus even in the same gene family, different evolutionary patterns are observed. In general, it is thought that the molecular interaction mechanisms, such as unequal crossing-over and gene conversion, are responsible for concerted evolution. When the rate of occurrence of such interactions is high, a uniform gene family is expected, whereas when it is low, the family becomes variable (Ohta, 1980, 1983a, b). Diversity between members of multigene families reflects the need of the organism, and must have evolved by natural selection. It is also conceivable that when sequences diverge enough, molecular interactions are inhibited and genes become more and more differentiated to form a supergene family (reviewed by Ohta, 1991).

Increases in functional diversity have been attained even without gene duplication. The two main mechanisms are diversification of gene expression and post-translational modification. Some of the major taxon-specific lens crystallins are closely related to enzymes (de Jong et al., 1989). ε-crystallin that occurs in birds is shown to be a lactate dehydrogenase (Wistow et al., 1987), δ2-crystallin in birds appears to be an argininosuccinate lyase (reviewed by Ohta, 1991), and τ-crystallin in fish and birds appears to be an α-enolase (Wistow et al., 1988). It seems that, in such cases, without gene duplication, the same gene may code both for an enzyme and for a structural protein, and is thought to arise through acquisition of new regulatory elements. The egg-laying hormone of *Aplysia* is encoded by a member of a multigene family of three tandem genes (Scheller et al., 1982). The proteins encoded by this family are characterized by extensive post

translational modification, i.e., proteolytic cleavage. Many different small peptides are produced from the three gene products and their functions depend upon the site of cleavage. Other examples in which proteolytic cleavage is important may be found in the precursors of the opiate peptides (Nakanishi et al., 1979; Noda et al., 1982). Corticotropin-β-lipoprotein precursor contains several duplicated coding regions of peptide hormones and is cleaved to yield a number of different hormones, such as corticotropin, endorphin and melanotropin. The gene encoding adrenal enkephalin has quite a similar organization. Thus, proteolytic cleavage is an important mechanism through which a single gene or transcription unit can have many functions or yield multiple products.

Finally, another mechanism through which additional functions can be acquired without recourse to gene duplication is differential splicing. The immunoglobulin gene family is an example for this (reviewed by Ohta,1991).

1.5.2. Casein genes

The casein genes seems to be structurally quite different since the sizes of the transcription unit ranges from 8.5 to 18.5 kb and the number of introns ranges from 4 to 18 (Mercier and Vilotte, 1993). Interspecies comparison of milk protein genes in 12 species confirmed their high rate of evolution, but that the overall gene organisation is conserved for each class of casein e.g. β-casein genes (Mercier, 1992). Based on the rate of divergence of the casein signal peptides, a calcium sensitive casein-like gene appears to have originated 300 million years ago, at the time of the appearance of primitive mammals (Dayhoff, 1976). This casein-like gene may have been the result of exon recruitment bringing together the basic elements of modern casein: a 5' non coding region, a signal peptide exon, an exon with a minor phosphorylation site at its 3' end and a hydrophobic domain (Jones et al., 1985).

Between the appearance of primitive mammals (300 million years ago) and mammalian radiation (75 million years ago) two types of duplication have occurred. The phosphorylation exon was duplicated intragenically to create the series of small exons observed in the rat β casein gene. Intergenic duplication also occurred creating the individual members of the calcium sensitive caseins (Fiat and Jolles, 1989). Caseins are one of the most rapidly evolving protein families, studied so far. Jones et al. (1985) reported that in many species only three regions of the casein mRNAs are conserved: the 5' non coding region, the signal peptide coding region and the region encoding the phosphorylation sites. These regions correspond exactly to the nutritional role of the caseins which requires three functions: 1) to be secreted 2) to form protein aggregates termed micelles and 3) to be phosphorylated to allow Ca^{2+} binding and transport. The conserved 5' non coding region and the signal peptide allow the secretion of the protein. The caseins aggregate to form micelles by the interaction of their carboxy-terminal hydrophobic domains and the conserved phosphorylation sites allow the binding of Ca^{2+}.

Nucleic acid and amino acid sequence comparisons indicate that the three calcium sensitive caseins evolved from a common primordial gene (Stewart et al., 1987). A common ancestor hypothesis for the αS1, αS2 and β casein genes is supported by the finding of common sequence motifs in the proximal 5' flanking regions and the similar organisation pattern of the first four exons. In particular, the second exon comprises the remaining part of the 5' untranslated region and the coding frame for the signal peptide and the first two amino acids of the mature polypeptide chain (reviewed by Mercier and Vilotte, 1993). Comparison of the sequences of the bovine casein genes (αS1,αS2 and β) revealed that αS1 has diverged much further than αS2 and β and the latter genes could have evolved from a common ancestor by means of gene duplication (reviewed by Bonsing and

MacKinlay, 1987) (Fig.1.6). The observations of Groenen et al. (1993) provide evidence for a closer evolutionary relationship between αS2 and β-casein genes on the basis of nucleotide sequence and exon size similarities. Stewart and MacKinlay (1984) compared the cDNA sequences of the four bovine caseins with those of rat and mouse and were of the opinion that the casein gene family has undergone unusually rapid evolution and the bovine κ-casein might have evolved from a different ancestral gene since it does not share any common pattern with other casein genes.

The κ-casein gene, is believed to have evolved from the same primordial gene as the γ-chain gene of fibrinogen (Stewart et al., 1984). Fibrinogen is the serum protein which when cleaved by thrombin, serves an analogous function in the blood clotting cascade as κ-casein does in milk curd formation. Blood and milk clotting processes are two physiologically important coagulation process. Jolles and Henschen (1982) described the following common features in the coagulation mechanism of blood and milk: a) limited proteolysis- thrombin cleaves only two Arg-Gly bonds in the Aα and Bβ chain of fibrinogen, similarly chymosin cleaves a unique Phe-Met bond in κ-casein; b) during the coagulation process short soluble peptides are released- the fibrinopeptides A and B and the κ-caseinoglycopeptide c) the structures of both peptides are highly variable from one species to another d) some amino acids have never been found in any fibrinopeptides as well as in κ caseinoglycopeptides, Cys and Trp e) all fibrinopeptides carry a substantial negative charge and the κ-caseinoglycopeptides are also acidic.

Nucleotide sequence comparison, mouse κ-casein cDNA was shown to share similarity only with the second half of the rat γ-fibrinogen cDNA. While with human γ-fibrinogen cDNA the similarity spanned two regions, one between nucleotides, 1-328 and the second between nucleotides, 591-726 (Thompson et al.,

1985). Alexander et al. (1988) reported that the most striking similarity observed between the cDNA involves a nucleotide stretch corresponding to the 5' end of κ-casein exon IV and the 3' end of γ-fibrinogen exon II. Jolles et al. (1978) hypothesised that κ-casein and γ-fibrinogen may have evolved from a common ancestor. The strong similarity between β and γ chains of fibrinogen indicates that they have a common ancestor that should have existed more than 450 million years ago. As κ-casein seems to have appeared much later during evolution than the fibrinogen genes it is reasonable to speculate that it could have evolved from the γ chain of fibrinogen (Fig.1.6).

Except histone and interferon most eukaryotic nuclear genes (Breathnach and Chambon, 1981) contain intervening sequence or introns (Gilbert, 1978). The introns are transcribed into pre mRNAs but are excised prior to export of the final mRNA from the nucleus. Although some intronic sequences appear to be conserved, their biological and evolutionary significance is unclear (Sharp, 1985). It has been suggested that introns provide a mechanism for generating diversity during evolution (Gilbert, 1985).

One of the best studied examples of a gene family is the globin family where an initial duplication gave rise to the ancestors of the modern α and β-globins which in turn underwent further duplications to generate clusters of linked genes (Efstratiadis et al., 1980). Duplication of a gene creates the opportunity for one of the duplicate copies to diverge from the other and to acquire additional or modified functions. In the case of the β-globins individual members of the cluster are adapted to function under different physiological conditions and are expressed at different stages of development.

The human β-globin gene cluster contains five active globin genes arranged in the order , ε, Gγ, Aγ, δ and β together with two pseudogenes and extensive tracts of DNA between the genes (Efstratiadis et al., 1980). All five

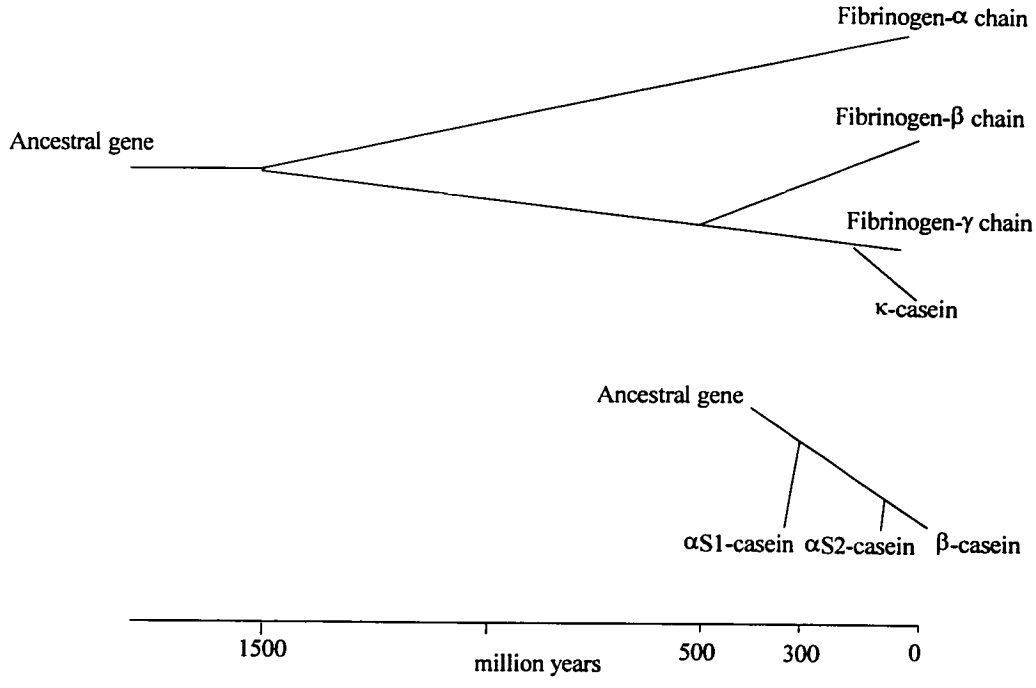


Fig. 1.6. Tentative evolutionary relationship between fibrinogen and κ-casein and the calcium sensitive casein genes by gene duplication (redrawn from Fiat and Jolles, 1989 and Bonsing and MacKinlay, 1987).

genes are interrupted by two introns at identical locations, the first is located between codons 30 and 31 and the second is between codons 104 and 105. Since the human α-globin genes, as well as mouse α and β-globin genes, the chicken β-globin genes and the four rabbit β-globin genes also contain two introns at homologous locations, it is reasonable to assume that these interruptions pre-date the emergence of separate α and β-globin genes about 500 million years ago. In human the β-globin gene cluster is located in chromosome 11 and the α-globin gene cluster is located in chromosome 16. All the genes in each cluster are transcribed from the same DNA strand and are arrayed on the chromosome in the order of their expression during development: 5'-ε-$^{G}\gamma$-$^{A}\gamma$-δ-β-3' (β-globin gene cluster) and 5'-ζ_2-ζ_1-α_2-α_1-3'(α-globin gene cluster) (reviewed by Maniatis et al., 1980). However, the four chicken β-globin genes (5'-ρ-β^{H}-β-ε-3') are not arranged in the order of their expression during development, as the two embryonic genes are found on either side of the adult genes (reviewed by Evans et al., 1990).

In the human β-globin gene family, the two closely related foetal genes (Gγ and Aγ) have clearly arisen by a recent dupliction within the cluster (Maniatis et al., 1980). Comparison of the duplicated DNA segments can reveal something about the modes by which genic and intergenic DNA diverge in evolution. Phylogenetic comparisons of primate β-globin gene clusters show that the duplication probably occurred 20 to 40 million years ago in the lineage leading to old world monkeys, apes and man (Barrie et.al., 1981). Since then, the sequences of most of the two block repeats have diverged about 14% from each other and have also accumulated numerous very small deletions and insertions (reviewed by Jeffreys and Harris, 1982). However, one substantial region (1500 bp long) including most of the Gγ-globin gene is more than 99% identical to the corresponding region around Aγ-globin gene and it has been suggested that the region has been homogenised by gene conversion between misaligned γ-globin

repeats very recently in evolution. Gene conversion between related DNA sequences seems common in evolution and appears to operate not only within protein coding gene families but also in middle and highly repetitive DNA sequences. It may serve to maintain sequence homogeneity within a family (reviewed by Jeffreys and Harris, 1982).

Gene clusters are inherently unstable entities that can expand, contract and disperse during the course of evolution. Natural selection will eliminate the unfavourable, leaving neutral and advantageous changes to be fixed in evolution. Although it is difficult to distinguish between neutral and selective changes, some degree of adaptive change has accompanied the Gγ - Aγ globin gene duplication in man and the $^{A}\beta$ - $^{C}\beta$ duplication in the goat, in which the duplicated genes tend to be expressed at different stages of development.

1.6. DNA methylation and CpG islands

1.6.1. Methylation and gene expression

DNA methylation provides one mechanism for stably altering the local structure of a gene and in this manner may play a role in the regulation of gene activity in cells. In order to understand how methylation influences gene expression it is instructive to follow the path of a typical tissue specific gene through development. Animal somatic cell DNA is characterised by a bimodal pattern of methylation, the tissue specific genes are almost fully methylated in sperm and probably in the female germ line and remain methylated throughout early development and in all non-expressing adult somatic tissues. Demethylation is usually observed in the specific tissue of expression. While constitutively expressed housekeeping genes generally contain CpG islands (see 1.6.2) that are completely unmethylated in the germ line and in all somatic cells. Thus the DNA

methylation pattern is a relatively fixed element of the genome that distinguishes at the molecular level the inactive and active genes of each cell type (Bird, 1986).

Methylation inhibits gene expression by affecting the protein-DNA (gene) interactions required for transcription. Unmethylated DNA adopts a DNAase-I sensitive structure typical of active genes, while fully methylated sequences are resistant to DNAase-I and are structurally similar to inactive genes (Keshet et al., 1986). Thus the presence of methyl moities appears to be capable of altering interactions between the DNA and the factors involved in the generation of the DNAase-I sensitive or insensitive configuration. DNA methylation not only alters the overall structure of a gene but may also interfere with the binding of protein factors at discrete *cis*-acting sites (Bird, 1986).

Studies on the γ-globin gene have shown that methylation plays a role in the repression of this gene. In fibroblasts the endogenous gene is highly methylated and unexpressed but unmethylated copies of the gene introduced in transient transfection experiments are transcriptionally active when introduced into the same cell types. It has also been shown that the exogenous gene is inactivated by *in vitro* methylation (Busslinger et al., 1983). Yisraeli et al. (1986) have reported that methylation represses the expression of the muscle specific α-actin gene when introduced into fibroblasts but not in myoblast cells. The myoblasts obviously contain the information required to recognise and activate the α-actin gene despite the presence of methyl moities. In myoblasts the gene undergoes site specific demethylation as does the endogenous gene.

Treating cell types (e.g. embryonic mouse fibroblasts, 10T1/2 cells) with 5-azacytidine - a potent demethylating agent - results in the activation of certain genes presumably by reversing the methyl mediated state of repression (Gartler and Riggs, 1983). It causes a marked transient genome wide demethylation in treated cells. It is not clear whether it is this demethylation that is responsible for

the activation process, as this drug has other biological activities and has been shown to induce genes in organisms that lack 5-methyl cytosine (Tamame et al., 1983). In most systems where 5-azacytidine has been used stable undermethylation is restricted to activated genes, and raises the possibility that demethylation may actually be secondary to specific gene recognition events, mimicking what may occur during normal development (Konieczny and Emerson, 1984).

Methylated CpGs are protected against nonspecific nucleases (Solage and Cedar, 1978) and against most restriction enzymes that recognise the dinucleotide CpG (Antequera et al., 1989). Meehan et al. (1989) identified a protein, MeCP1 (Methyl CpG binding protein) in extracts from mouse cells which binds to DNA that is symmetrically methylated at multiple CpGs. Such an indirect model for transcriptional repression by methylated CpGs is also reported by Boyes and Bird (1991). They observed a transcriptional inhibition at low template concentration and could overcome the inhibition either by increasing the concentration of methylated template or by addition of methylated competitor DNA. They pointed out that the transcriptional repression of methylated DNA acts via a mediator protein which is indistinguishable from MeCP1.

Inactivation of the X-chromosome in mammalian females may also be due to DNA methylation (Liskay and Evans, 1980). It has been reported that the X-inactivation is accompanied by heterochromatinization of much of the X-chromosome and methylation of its CpG islands (Wolf et al., 1984; Yen et al., 1984; Keith et al., 1986; Lock et al., 1987). Phosphoglycerate kinase (PGK) and α-globin genes have CpG islands and are normally nonmethylated in animal cells, but when methylated adopt an altered chromatin structure (Boyes and Bird, 1991). However, the CpG island associated with PGK is methylated on the inactive X chromosome (Hansen et al., 1988) and that of α-globin gene is found methylated

in permanent cell lines (Antequera et al., 1990). In their methylated form these genes were shown to become resistant to the methyl insensitive restriction endonuclease *Msp*I, implying that methylated sites in the nuclei are made inaccessible by association with nuclear proteins. It is possible that the protein MeCP1 or a similar protein responsible for both the inaccessibility and inactivity of the CpG methylated gene (Boyes and Bird, 1991).

DNA methylation may also be involved in parental imprinting (Li et al., 1993). In several instances genes inserted into transgenic mice show differential activity depending on whether they are subsequently inherited from the father or the mother. Since this pattern is faithfully inherited in every cell of the offspring, it is essentially the methylation state imprinted in the parent that transmits the information determining the activity state of the gene in somatic cells of the offspring (Cedar, 1988).

1.6.2. CpG islands

In vertebrate genomes the dinucleotide CpG is present at only 25% of the frequency expected on the basis of the G+C content and 80% of the genomic CpG sites are methylated (Bird, 1986). Moreover, the position of the unmethylated CpG residues is not random, but rather they are found in clusters known as CpG islands or HTF (*Hpa*II tiny fragments) islands. In an HTF island the G+C content is over 50% and the CpG and GpC dinucleotides occur with almost equal frequency. CpG islands are short (1-2 kb), rich in unmethylated CpG residues, and are found at the 5' end of genes often including the first few exons as well as the upstream sequences (Bird, 1986). The restriction endonucleases *Hpa*II and *Msp*I, both recognise the sequence CCGG but differ in their sensitivities to methylation. *Msp*I can cleave the sequence irrespective of CpG methylation but *Hpa*II is sensitive to methylation. The average spacing of this site in CpG island is about

EX BIBL UNIV. EDINBURGEN.

100 bp compared with about 2000 bp in bulk DNA (Cooper et al., 1983). *Msp*I cuts naked DNA about four times more frequently than *Hpa*II and behaves as a methyl CpG insensitive enzyme when naked DNA is the substrate, but when nuclear chromatin is the substrate it becomes indistinguishable from a methyl CpG sensitive enzyme. In the nuclei methylated CpGs are rendered resistant to cleavage by binding to macromolecules that block access to nucleases (Antequera et al., 1989).

Regarding the evolutionary aspect of CpG islands, one of the hypotheses is that genes of invertebrates were entirely in a nonmethylated state. During evolution, with the appearance of vertebrates DNA methylation spread through the genome, excluding the promoters, which would be suppressed by dense methylation. Cytosine residue of CpG dinucleotide is methylated at the 5th carbon position and then deamidated to form TpG. If the cell's DNA repair systems do not recognise this change immediately then the thymidine (T) survives in place of cytosine (C) and is entitled to be in DNA. Thus, loss of CpG dinucleotides by mutation and retention of CpG dinucleotides in the nonmethylated regions gave rise to CpG islands. With time, many genes succumbed to methylation and loss of CpG islands without losing their capacity for expression. These are exclusively tissue specific genes since all known house keeping genes are associated with CpG island (Antequera and Bird, 1993).

The CpG dinucleotide is the target for mammalian methyl transferase that methylates the cytosine residue at the carbon 5 position. Several models are proposed for the protection of CpG islands from methylation. The simplest of which is that CpG islands are a poor substrate for the methyltransferase. This idea is supported by the finding of Carotti et al. (1989). They showed that CpG rich sequences are methylated relatively inefficiently *in vitro*, suggesting that the high density of CpG dinucleotides in islands or their G+C richness inhibits the

methyltransferase. But this model does not explain why CpG islands are methylated on the inactive X, at imprinted genes and also at nonessential genes in cultured cell lines (Antequera et al., 1990; Jones et al., 1990). A second model proposes that CpG islands are subject to *de novo* methylation but the modification is removed by an island specific demethylating activity (Frank et al., 1991). This model is based on the finding that partially methylated CpG islands loose methylation when introduced into fertilized mouse eggs or embryonic cells and demethylation may only occur when relatively few CpG sites within the transgene are methylated but not when all CpGs are methylated (Szyf et al., 1990; Frank et al., 1991; Choi and Chae, 1993).

A third model proposes that steric hindrance by protein factors is responsible for excluding the methyltransferase from CpG islands (Bird, 1986). Because CpG islands colocalize with gene promoters, transcription factors may be required for island protection. This idea is supported by *in vivo* footprinting studies of protein DNA interactions within the CpG islands of the human PGK-1 gene (Pfeifer et al., 1990; Pfeifer and Riggs, 1991) and the HTF9 locus (Stapleton et al., 1993). Multiple factors are bound over both of these CpG islands. In contrast, the methylated PGK-1 island of the inactive X chromosome has nucleosome like structures replacing transcription factors (Pfeifer and Riggs, 1991). This model suggests that protein factors are bound across the island and could block the methyltransferase by steric hindrance. Indirect support for this model is also inferred from transgenic experiments showing that the CpG islands of the transgene can become methylated at high copy number, suggesting that there is a limited supply of factors that are capable of protecting the island (Mehtali et al., 1990; Gundersen et al., 1991). A limitation of this model is that it does not explain how inactive genes can nevertheless retain CpG islands that are methylation free as in the case of the human α-globin and retinol binding protein

genes. Both these genes are expressed in only one tissue but have nonmethylated islands in all cell types that have been examined (Bird et al., 1987; Antequera et al., 1990).

Antequera and Bird (1993) suggested that the approximate number of CpG islands per haploid genome is 45,000 in human and 37,000 in mouse. They estimated a total of 80,000 genes in both the organisms and showed that 55.9% and 46.9% of all the genes are associated with CpG islands in human and mouse respectively. The proportion of CpG island associated genes is lower in the mouse genome than in the human genome. This may be due to, *de novo* methylation of the CpG islands in the mouse germ line leading to progressive CpG loss, which occurred some time after the divergence of the mouse and human lineages in evolution. Based on extensive analysis of DNA sequences, Larsen et al. (1992) reported that all the human house keeping genes and 40% of tissue specific genes are associated with CpG islands. Antequera and Bird (1993) also reported an estimate of 22,000 house keeping genes and 58,000 tissue specific genes per human haploid genome. This implies that only half of all CpG islands in the genome are associated with house keeping genes and the other half with tissue specific genes.

Tissue specific genes are usually part of the silent majority of the genome and methylation acts as a mechanism for controlling these genes. The methyl groups may generate a local chromatin configuration that renders the gene inaccessible, and thus transcriptionally inactive. This would provide a general mechanism for transcriptional repression which may operate independent of the requirement for interactions between *cis*-acting regulatory elements and tissue specific factors (reviewed by Bird, 1986). Activation of tissue specific genes from their generalised state of repression may first involve recognition of the genes while they are still methylated and this event initiates the process of transcription

and concomitant demethylation (Benvenisty et al., 1985). In their demethylated state these genes would be stably maintained in an active structure that is generally accessible to the transcriptional machinery of the cell.

Yisraeli and Szyf (1984) reported that methylation restriction mapping of many tissue specific genes revealed that the undermethylation observed in active genes almost always involves all parts of the gene including both the promoter and structural gene regions and suggest that methylation may play a regulatory role over the entire gene domain. Studies on the human β-globin gene showed that CpG sequences are under represented (Konkel et al., 1979) and are relatively distributed in the flanking sequences. Despite the relatively low number of CpG residues in the gene domain, Yisraeli et al. (1988) demonstrated that total *in vitro* methylation of these sites was sufficient to inhibit transcription by a factor of about 20. This repression was observed when methylation occurred in regulatory sequences in the 5' flanking region, the sequences covering the gene body and 3' flanking region and also during localised methylation of any of these regions. But the insertion of a fully methylated 50 bp repeat of CpG had no effect on the activity of the gene suggesting that methylation on the natural CpG residues within the gene sequences cannot be mimicked by the addition of irrelevant methylated moities.

Deletion and site specific mutation analysis of the β-globin gene from several mammalian sources showed that only 90 nucleotides in the 5' region are necessary and sufficient for full transcriptional activity (Myers et al., 1986), while studies of Yisraeli et al. (1988) suggested that in addition to this positive regulatory region there are sequence domains in the gene which have a negative regulatory role and the transcriptional repression is mediated through the interaction of proteins with various *cis*-acting well defined negative regulatory elements that are influenced by DNA methylation.

All house keeping genes reported so far are associated with CpG islands, but not all islands are associated with house keeping genes (Antequera and Bird, 1993). Keshet et al. (1985) reported that the APRT (adenosine phosphoribosyl transferase) gene is repressed by methylation at the 5' region, but not at the 3' region. A similar effect was reported in methylation studies of the human γ-globin gene (Busslinger et al., 1983) and the adeno virus Ela gene (Kruczek and Doerfler, 1983), while the viral tk (thymidine kinase) gene activity was strongly repressed by DNA methylation in both 5' promoter region and the 3' structural region of the gene (Keshet et al., 1985). There is a very high concentration of CpG residues over the entire gene both in the promoter and structural region. In contrast, in the APRT and human γ-globin genes there is a relatively high density of CpG moities at the 5' end but very low density at the 3' end. It may be possible that the degree of transcriptional inhibition is dependant upon the amount of methylation and density of CpG residues in the region (Keshet et.al., 1985). Yisraeli and Szyf (1984) pointed out that the pattern of 5' CpG under methylation and 3' methylation is typical in all house keeping genes.

Embryonic stem (ES) cells in culture are capable of *de novo* methylating and demethylating exogenously introduced DNA sequences. It is also reported that the ES cells can specifically recognise and demethylate island DNA that has been methylated *in vitro* before transfection (Frank et al., 1991). Brandeis et al. (1994) suggested that the bimodal somatic methylation pattern is probably established through the combined action of general *de novo* methylation working together with stage specific factors which either protect island DNA or actively demethylate these sequences. *In vitro* methylation studies on hamster APRT gene suggested that island demethylation is directed by discrete *cis*-acting elements and not merely by the high density of CpG residues in these regions (Brandeis et al., 1994).

Sp1 elements are frequently found up stream of house keeping genes (Gardiner-Garden and Frommer, 1987) and could represent a potential candidate for a demethylation signal (Holler et al., 1988). Mutation and deletion studies on the Sp1 consensus sequence of APRT gene in ES cells and in transgenic embryos also suggested that this element may provide a signal needed to direct demethylation of the methylated DNA or protect CpG islands from *de novo* methylation (Brandeis et al., 1994). Macleod et al. (1994) reported that three GC boxes, probably the attachment sites for Sp1 elements are located at the 5' edge of the CpG island of the mouse APRT gene. They showed in transgenic mouse assay that mutation or deletion of the GC boxes cause *de novo* methylation of the CpG island region. Mummaneni et al. (1993) also proposed that the boundries of CpG islands have sequences that block the spread of methylation from nearby methylation centres.

The relationship between transcription and lack of CpG island methylation, however, remains to be completely resolved. Many genes with CpG islands, such as the human α-globin gene, are highly specific in expression yet are nonmethylated in expressing and nonexpressing tissues (Bird et al., 1987). A possible explanation is that such genes are poised for transcription but do not, for some reason, make stable mRNA (Macleod et al., 1994).

1.7. Research project

1.7.1. Justification

The aim of the present project is to physically map the murine casein locus. The map generated will contribute to the understanding of 1) the relationship between genome organization and gene expression and 2) the evolution of multigene families. The map is also a pre-requisite for direct genetic manipulation of the locus. Thus, the results of this study could not only contribute

to basic studies on genome structure and function but also in the longer term underpin applications in biotechnology and agriculture/industry.

The use of the mammary gland as a bioreactor could revolutionise biotechnology. Pharmaceutically and biologically important proteins could be produced by mammary directed expression of homologous or heterologous genes in transgenic animals. Mammary specific expression of a heterologous gene can be achieved by developing fusion genes in which regulatory sequences from milk protein genes are fused to sequences encoding the protein of interest. A number of transgenic studies have been carried out in our laboratory using the native sheep β-lactoglobulin gene (Simons et al., 1987; Whitelaw et al., 1991; Whitelaw et al., 1992; Wilde et al., 1992) and hybrid genes derived by fusing the promoter region of the sheep β-lactoglobulin to genes, encoding potentially therapeutic proteins such as human clotting factor IX and human α1-anti trypsin (Clark et al., 1989; Archibald et al., 1990).

Currently, many human proteins are produced by recombinant micro-organisms (Primrose, 1987; Sarmeintos et al.1989). However, many human proteins cannot be produced correctly in bacteria or yeast as the microorganisms are incapable of completing the necessary post translational modifications. Other means of producing such proteins include tissue culture of insect cells transfected by baculovirus (Miller, 1988) or mammalian cells (Cartwright, 1987) and transgenic farm animals. The production of therapeutic proteins like human clotting factors, in the milk of farm animals has several advantages over preparations from human blood, including the absence of human pathogens. However, the presence of other infectious agents may not be ruled out. Ultimately, transgenic animals may be a cheaper source of recombinant proteins than mammalian cells maintained in tissue culture. Milk has the advantage over blood that it is very easy to collect and is available in large volumes. Moreover, milk is

relatively isolated from the animal and the presence of large quantities of biologically active molecules is less likely to have any side effects on the producer animal.

Transgenic technology could also be used to alter the composition of milk as a food product. For example, the protein content of milk could be changed by introducing extra copies of existing genes, by using regulatory sequences from a milk protein gene that is expressed at a higher level, by transferring milk protein genes from another species or by transferring modified genes.

Other changes in the nature of milk could have beneficial effects on downstream processing. For example, the stability of casein micelles is important in the processing of milk. Micelles are formed by interactions between calcium phosphate and casein molecules stabilized by a coating of κ-casein. Increasing the amount of κ-casein in milk would be expected to reduce micelle size, enhance the thermal stability of milk and so reduce the danger of coagulation in the sterilized product. More β-casein in milk reduces the time for rennet coagulation, increases the rate of whey expulsion and increases the firmness of the resulting curd (reviewed by Jimenez-Flores and Richardson, 1988). Changes to caseins will also affect the properties of the cheese formed. Deletion of phosphate groups will yield a softer and moist cheese but extra phosphate groups will yield a firmer cheese. Increasing the number of proteolytic sites results in rapid maturation of cheese, and may have a considerable economic advantage (reviewed by Jimenez-Florez and Richardson, 1988).

The production of human milk proteins, especially lactoferrin, in the milk of farm animals might make cows milk a better substitute for human milk (Mercier, 1986). The synthesis of antibacterial proteins, such as lysozyme (Maga et al., 1994; Maga and Murray, 1995) in the milk may reduce the incidence of mastitis in farm animals, which causes great financial loss to the farmers and

distress to the animals. Thus, this project seeks to underpin future genetic manipulation of milk with potential benefits for the agricultural and biotechnology industries and eventually their customers.

1.7.2. Objectives

Objectives of the present study are to-

1) Construct a physical map (restriction enzyme map and contig map) covering the region of the murine chromosome 5 containing the α, β and γ casein loci of mouse casein locus.

2) Map the κ-casein locus in mice

3) Confirm the existence of murine ε-casein locus and map it in mice

4) By comparative mapping analysis of the murine casein loci, examine the evolution of the various casein genes (both the calcium sensitive casein gene family and the κ-casein gene)

5) Examine the expression levels of the casein genes during pregnancy and lactation

6) Identify genetic variation at the casein loci amongst different mouse strains.

1.7.3. Experimental approach

Physical mapping of the casein locus could be carried out by a) restriction enzyme analysis and b) building up a contig map which covers the entire murine casein locus.

The primary factors needed for constructing a long range restriction map are high molecular weight (HMW) DNA, rare cutting enzymes (CpG island cutting, inter island cutting and enzymes with large recognition sequences, which can generate large DNA fragments) and a pulsed-field gel apparatus, enabling the separation of restriction digested DNA.

In order to construct a contig map, suitable large fragment genomic clones are required. The candidate vector systems for such clones are yeast artificial chromosomes (YACs), bacteriophage P1 and bacterial artificial chromosomes (BACs).

Even though YAC vectors can incorporate fragments up to 10 Mb, several major problems are associated with YAC clones (reviewed by Monaco and Larin, 1994). One is the high percentage (40-60%) of chimaeric clones, i.e. they contain two separate segments of non-contiguous DNA in the same YAC clone either due to co-ligation of DNA inserts *in vitro* prior to yeast transformation or due to recombination *in vivo* between two DNA molecules that were introduced into the same yeast cell. A second problem is that some clones are unstable and tend to delete internal regions from their inserts. A third major problem is the difficulty in separating the YAC clones from the yeast host chromosome background by simple methods. All these problems restrict the effective utilisation of YAC clones in genome mapping experiments. In particular, chimaerism and deletions raise problems in chromosome walking and gene-isolation experiments.

The P1 vector can accommodate fragments up to 100 kb. Since one of my supervisors had a P1 library I decided to screen the library, even though it is a relatively small library (~40000 clones, 1 x genome coverage). When the genome coverage is low, the probability of finding the gene/sequence of interest is also reduced. Unfortunately, PCR screening of the P1 library did not identify any positive pool (the organization of P1 library is shown in chapter 2) for the various casein genes except for α-casein gene. Further screening of the positive pool did not reveal the presence of any clone positive for the α-casein gene.

So I switched over to a commercially available BAC library (Research Genetics, USA) containing 221184 clones (7.37 x genome coverage). The BACs can accommodate fragments up to 300 kb with an average size of 100 kb. The

BAC system for mapping and analysis of complex genomes has been developed by Shizuya et al. (1992). There are fewer reports on the BAC system, but results so far, suggest that this system offers high cloning efficiency, easy manipulation and stable maintenance of the cloned DNA (Shizuya et al., 1992; Woo et al., 1994; Li Cai et al., 1995).

The cloning capacity of the BAC vector lies between that of P1s and YACs. As the bovine casein locus is reported to be contained in a 200 kb DNA fragment and if the mouse casein locus is as compact as the bovine locus, it might be possible to isolate the entire locus in a single BAC.

As a long term aim, the information obtained from the present mapping study and the clones isolated can be used to manipulate the casein locus in embryonic stem (ES) cells. Gene targetting is more effective if the targetting DNA is prepared from isogenic DNA. It is also advantageous to use isogenic DNA derived maps in targetting experiments. As most gene targeting experiments are carried out in ES cells isolated from 129 mice, it was quite appropriate to use the BAC library prepared from ES cells of 129 mice for the present study.

To carry out the gene expression study, mammary RNA was prepared at different time points, virgin, pregnancy and lactation. The levels of expression of the casein genes were determined by Northern blot analysis.

Data on the genetic variation of the caseins at the DNA level of various strains and species of mouse are limited. Therefore, restriction fragment length variant (RFLV) studies were carried out on the DNA prepared from eight inbred strains of *Mus musculus* and from *Mus spretus*. The RFLV analysis between the inbred strains will elucidate constant and possible polymorphic restriction enzyme cleavage sites, which will be useful in casein locus manipulation in ES cells.

Chapter 2
MATERIALS AND METHODS

2.1. Plasmid DNA manipulations

cDNA and genomic DNA clones of various mouse casein genes were kindly provided by Dr. J.M Rosen (cDNA clones of α, γ; Gupta et al., 1982), Dr. Satish Kumar (genomic DNA clone of β; obtained from Goodman and Rosen, 1990) and Miss. Katrina Gordon (cDNA clone of κ). Probes were generated from these clones using the restriction enzymes as shown in Table 2.1. Two probes were generated for β-casein from a 6.6 kb genomic DNA clone, a *Hin*dIII/*Sca*I fragment (888 bp; from the 5' end) and a *Sal*I/*Sac*I fragment (643 bp; from the 3' end) (Fig. 2.1). Also a probe from the 5' end of the β-casein gene was prepared by PCR amplification using primers designed from the extreme 5' region of genomic DNA sequences available in the database (see Table 2.2). A cDNA probe for ε-casein was prepared by PCR amplification of a lactating mouse mammary cDNA library (kindly provided by Dr. C.J. Watson). The primers were designed from the ε cDNA sequences available in the data base (see Table 2.2).

3'specific probes for all casein genes were prepared by PCR amplification of the genomic DNA using primers designed from the 3' untranslated region of the cDNAs/mRNAs except for β. The primers for β–casein were designed from the sequence of intron 8. All primer sequences were designed using the programme, Primer version 0.5 (Whitehead Institute for Biomedical Research). The primer sequences designed for the different caseins and the expected PCR product sizes are shown in Table. 2.2

Table 2.1. Plasmid clones used to generate murine specific probes.

casein genes	Cloning vector	Cloning site	Size of Cloned fragment	Resistant antibiotic	Enzymes used to excise cloned fragment[1]
α	1. pUC 19	*Pst*I	1100 bp[2] (cDNA)	Ampicillin	*Pst*I
	2. pBR322(?)	*Pst*I	800 bp (cDNA)	Tetracycline[3]	*Pst*I
β	pGEM.BS	*Eco*RI	6.6 kb (genomic fragment)	Ampicillin	*Hin*dIII/*Sca*I (5' specific)
					*Sal*I/*Sac*I (3' specific)
γ	pBR322	*Pst*I	900 bp (cDNA)	Tetracycline	*Pst*I
κ	pBS.KS	*Eco*RI	900 bp (cDNA)	Ampicillin	*Kpn*I/*Bam*HI

[1] The excised fragments were used as probes

[2] The 1100 bp cDNA clone was proved to be rat α-casein cDNA

[3] The second α-cDNA clone obtained from Jeffrey Rosen was murine specific, but the cloning vector was uncertain because the clone was found susceptable to tetracycline and resistant to ampicillin.

The clones were α- pCMα11 and γ- pCMγ19 (Gupta et al., 1982), β- pB7 (Goodman and Rosen, 1990) and κ- pκ-Cas (Gordon, pers. commun.).

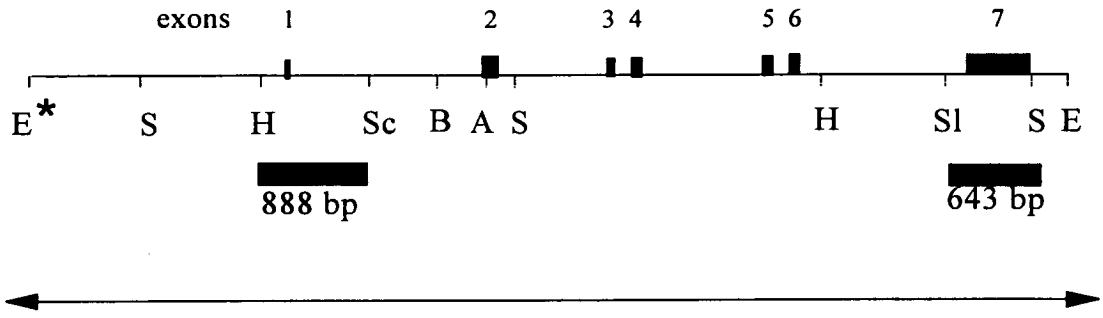


Fig. 2.1. pB7 containing 6.6 kb β-casein gene cloned in pGEM-Blue-Script. Thick lines shown below represent the positions of the DNA probes used in Southern and Northern blot analyses. Thick lines above represent the exons, 1 to 7. E* - *Eco*RI site from the vector. The restriction sites in the cloned DNA are S- *Sac*I, H- *Hin*dIII, Sc- *Sca*I, B- *Bst*EII, A- *Asp*7001, Sl- *Sal*I, E- *Eco*RI. The sequence of β-casein gene including the region corresponding to this fragment is available in the EMBL database (X15991).

Table 2.2. PCR primers used to generate probes and for screening libraries.

Casein cDNA/gene (EMBL: accession number**)	Primer	Primer sequences	PCR product size (bp)
α. (M36780)	3'UT forward	TGACTGGACCCTCCATTCTC	228
	3'UT reverse	GCTCAGATGATGCAACTGGA	
β. (X15991)	intron 8 forward	TTCTCGACCTATGGGTCAGG	205
	intron 8 reverse	AATCAGTGATTCCCAACATTCC	
	3'cDNA probe (f)	CATCCAGCCTATTGCTCAACCC*	415
	3' cDNA probe (r)	ATGTTGTGGAGTGGCAGGGAAT*	
	5' end forward	TGCTTTTCTCCTGAGAATCTCC	325
	5' end reverse	TGTACAATTGACGGAGAAGGG	
γ. (D10215)	3'UT forward	TTGGTTTCTTCGTGGTTTCC	244
	3'UT reverse	GTCTTGGCGAGAGATATTTTGG	
ε. (V00740)	3'UT forward	TGGAAAATTCACATCTCCATTT *	252
	3'UT reverse	TGAAAAGAATGTGGTTCAGAGA*	
	5'cDNA probe (f)	GAGGATGGAGCAATACATCTCC*	373
	5' cDNA probe (r)	GCTGGGGGTAGATGGGTAAT *	
κ. (M10114)	3' UT forward	TCTGGAACCAAATTTCAACTTG	102
	3'UT reverse	AAGGGAAGACGAGAAAGATCTG	
	5'cDNA probe (f)	CTACAGGCCATCTTTGCCTGCT *	353
	5'cDNA probe (r)	ATGCTGCAGTTGAGGACACTGG*	

* Primers used to amplify cDNA sequences from a lactating mouse cDNA library.

** the EMBL accession number for mouse DNA sequence entry from which the primers were designed with primer version 0.5, Whitehead Institute for Biomedical Research.

2.1.1. Preparation of LB (Luria-Bertani) plates

Sterile LB bottom agar (10 g Tryptone; 5 g Yeast extract; 5 g NaCl; 15 g Bacto agar per litre) was mixed (~ 48°C) with required antibiotic (ampicillin - 25-50 μg/ml; tetracycline - 15 μg/ml; chloramphenicol - 12.5 μg/ml; kanamycin - 25-50 μg/ml), poured into sterile petri-dishes of appropriate sizes and left at room temperature for a few minutes to set.

2.1.2. Bacterial transformation

2.1.2.1. Calcium competent DH5α cells

Bacterial transformation of plasmids was carried out using MAX efficiency DH5α competent cells (Gibco-BRL). The cells were thawed on wet ice, 100 μl aliquots transferred into the required number of microfuge tubes kept on ice. Plasmid containing the cloned DNA fragment (1 μl, 1-10 ng DNA) was added to each tube and mixed by moving the pipette through the cells while dispensing. A positive control was prepared by adding 5 μl (0.05 ng) of pUC19 (Amp^r) DNA to one of the tubes containing 100 μl of cells in order to determine transformation efficiency. The cells were held on ice for 30 min and heat shocked for 45 s in a 42°C water bath without shaking. The tubes were placed on ice again for 2 min. An aliquot (0.9 ml) of SOC medium (2% Bacto tryptone; 0.5% Yeast extract; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM glucose, pH 7.0) at room temperature was added and the cells were incubated in a shaking incubator (Gallenkamp, 225 rpm) at 37°C for 1 hour. The transformation control containing plasmid DNA was diluted 1:100 with sterile SOC medium and plated out (100 μl) on LB plates with 50 μg/ml ampicillin (LB/amp). Aliquots (100 μl) of two dilutions (1:100 and 1:10) of the test samples were plated on LB plates containing the appropriate antibiotic and incubated at 37°C overnight.

2.1.2.2. Electroporation

Electroporation was used to transform bacteria with fragments from BAC clones. The fragments were subcloned in the pBeloBAC11 vector and electroporated into DH10B electro-competent cells as follows:

Electroporation competent cells were prepared by inoculating 1 litre of LB medium (10 g Tryptone; 5 g Yeast extract; 5 g NaCl; 2 g $MgCl_2$, pH 7.0 per litre) with 10 ml of freshly prepared DH10B starter culture (Gibco-BRL). The cells were grown at 37°C in a shaking incubator (225 rpm) to an OD_{600} of 0.5-0.8. The cells were chilled on ice for 20 min and centrifuged in a cold rotor (Sorvall F16/250) at 4000 rpm for 15 min. The cells were then resuspended in a litre of pre-cooled 10% glycerol and centrifuged as above. The pelleted cells were washed twice in 500 ml and 20 ml respectively, with cold 10% glycerol and resuspended in 2.5 ml of the same solution. Aliquots of 60 µl were stored at - 70°C.

An aliquot of 60 µl containing electro-competent cells was thawed on ice and an aliquot of 1 µl (5-10 ng) of the ligation mix was added. The cells were mixed by gentle tapping and kept on ice for 1 min. The Cellject basic electroporation system (Flowgen) was set to 2500 V. The mixture was transferred to the bottom of a cold 2 mm electrode gap electroporation cuvette and a pulse of 12500 V/cm applied for 5 milliseconds. The cells were immediately resuspended in 1 ml of SOC medium (room temperature). The cell suspension was transferred to a polypropylene tube and incubated at 37°C in a shaker (200 rpm) for 1 hr. Appropriate aliquots were plated out on LB plates containing chloramphenicol (12.5 µg/ml).

2.1.3. Plasmid DNA preparation

2.1.3.1. Mini plasmid preparation

Small scale preparation of plasmid DNA was performed using the Wizard mini-preps DNA purification system (Promega). Single colonies were cultured overnight in 5 ml of LB medium. The culture (2-3 ml) was pelleted at 5000 rpm (Sorvall HB-4) for 10 min and the cells were resuspended in 200 µl of resuspension solution (50 mM Tris-HCl, pH 7.5; 10 mM EDTA; 100 µg/ml RNAase A). The cells were lysed by the addition of 200 µl of cell lysis solution (0.2 M NaOH; 1% SDS) and mixed by inversion. Neutralisation solution (200 µl; 1.32 M Potassium acetate, pH 4.8) was added to precipitate the proteins and chromosomal DNA. The tubes were mixed well by inversion and centrifuged at 12000 rpm for 5 min in a table top centrifuge (Sorvall MC12V). The supernatant was transferred to a clean microfuge tube, mixed with 1 ml of Wizard mini preps DNA purification resin and poured into a 2 ml syringe fitted on a Wizard mini column. The solution was gently pushed through the column and the column was washed with 2 ml of column wash buffer (200 mM NaCl; 20 mM Tris-HCl, pH 7.5; 5 mM EDTA; 55% ethanol) to remove all contaminants. The syringe was removed and fitted to the column on a 1.5 ml microfuge tube and centrifuged at 12,000 rpm (Sorvall MC12V) for 20 s to dry the resin. The column was transferred to a fresh microfuge tube, added 50 µl TE buffer (10 mM Tris-HCl; 1 mM EDTA, pH 7.4) and kept for one min. The plasmid DNA was eluted by centrifuging at 12,000 rpm (Sorvall MC12V) for 20 s and stored at -20°C. Cloned DNA of more than 10 kb was eluted using TE buffer warmed to 50°C prior to elution.

2.1.3.2. Maxi plasmid preparation

Large scale preparation of plasmid DNA was carried out using a Qiagen plasmid maxi kit. Qiagen plasmid purification protocols are based on a modified alkaline lysis procedure, followed by binding of plasmid DNA to Qiagen anion-exchange resin under appropriate low salt and pH conditions. RNA, proteins, dyes, and low molecular weight impurities are removed by a medium salt wash. Plasmid DNA is eluted in a high salt buffer, concentrated and desalted by isopropanol precipitation.

A glycerol stock (20%) of transformed bacteria for different clones were prepared and stored at -70°C. An appropriate dilution (1×10^5 or 1×10^6) of the glycerol stock was plated out on LB plates (+ appropriate antibiotic). A single colony was picked for preparing a starter cuture of 10 ml with the necessary antibiotic and incubated overnight at 37°C in a shaking incubator (225 rpm). LB medium/antibiotic (500 ml) was inoculated with 0.5 ml of starter culture and incubated overnight in a 37°C shaking incubator. The culture was centrifuged at 5000 rpm (Sorvall F16/250) for 15 min and the bacterial pellet resuspended in 10 ml of the 'resuspension' buffer, P1 (50 mM Tris-HCl, pH 8.0; 10 mM EDTA + 100 μg/ml RNAase) with out any clumps. Lysis buffer, P2 (10 ml; 200 mM NaOH; 1% SDS) was added, mixed by inversion and incubated at room temperature for 5 min. Added 10 ml of neutralisation buffer, P3 (3 M Potassium acetate, pH 5.5), mixed by inversion, incubated on ice for 20 min and centrifuged at 15000 rpm (Sorvall SS-34) for 30 min. The supernatant was transferred to another tube and centrifuged again. The clear supernatant was then poured into a Qiagen tip-500, which had already been equilibrated with 10 ml of 'equilibration' buffer (750 mM NaCl; 50 mM MOPS, pH 7.0; 15% ethanol; 0.15% Triton X-100). The column was washed twice with 30 ml of QC buffer (1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% ethanol). The plasmid DNA was eluted with 15 ml of

elution buffer, QF (1.25 M NaCl; 50 mM Tris-HCl, pH 8.5; 15% ethanol). To the eluted DNA, 0.7 volumes of room temperature isopropanol was added and centrifuged for 30 min at 15000 rpm (Sorvall SS-34). The supernatant was poured off carefully. The clear glassy pellet was washed with 70% ethanol, air dried and dissolved in 2 ml of sterile TE buffer (10 mM Tris-HCl; EDTA, 1 mM, pH 8.0).

2.1.4. Caesium chloride gradient purification of plasmid DNA

Caesium chloride (CsCl; 4.75 g) and ethidium bromide solution (10 mg/ml; 500 µl) were added to the plasmid DNA (2 ml) prepared using the Qiagen columns. The final volume was adjusted to 5.6 ml with TE buffer. The solution was transferred to a Beckman Quick-seal tube and centrifuged in an L8-M ultracentrifuge at 50,000 rpm (NVT 65.2 vertical rotor; Beckman) at 20°C for 16-18 hrs. RNA and proteins were collected at the bottom of the tube. The closed circular DNA band seen just below the nicked DNA (open circular plasmid DNA) was recovered from the tube with a needle and syringe. The ethidium bromide was removed from the DNA solution by repeated extraction with TE saturated butan-2-ol. The DNA was precipitated by adding 3 volumes of water and 2.5 volumes of cold ethanol. The solution was centrifuged at 12000 rpm (Hermle Z231M) at 10°C and the pellet dissolved in 500 µl of TE. The DNA was again precipitated with $1/10^{th}$ volume of 3 M sodium acetate (pH 5.0) and 2 volumes of ethanol and centrifuged. The supernatant was discarded and the pellet was washed with 70% ethanol and air dried. The air dried pellet was dissolved in 250 µl of TE buffer and stored at -20°C. Optical density (O.D) of a suitable dilution was measured at 260 nm in a spectrophotometer to estimate the DNA concentration (An optical density 1 at 260 nm = 50 µg of DNA/ml. Thus, the concentration of DNA in µg/ml = O.D x 50 x dilution factor).

2.1.5. DNA Sequencing

2.1.5.1. Using T7 sequencing kit (Pharmacia)

The T7 sequencing kit is based on Sanger's dideoxy sequencing technique of base-specific termination of enzyme catalyzed primer extension reactions. Four separate reactions are performed, all containing primer, template and the four deoxynucleotides, but each including a different chain-terminating dideoxynucleotide. In each reaction, a mixture of fragments are generated, each terminated with the particular dideoxynucleotide present in that reaction. Thus, the chain-terminated fragments in each reaction represent the occurrence of the corresponding deoxynucleotide in the sequence. When the products of the four reactions are electrophoresed side-by-side, the sequence in which nucleotides were added to the primer could be deduced from the sequence in which successively larger fragments occur in the four lanes. The positions of the separated fragments are detected by virtue of a label (radioactive or fluorescent) introduced either before or during the primer extension reactions. T7 DNA polymerase was the enzyme used for the primer extension reactions. The sequencing reactions can be performed using either the "Read Short" or "Read Long" conditions. Read Short conditions allow sequences to be read up to ~500 nucleotides with either ^{32}P or ^{35}S. The Read Long conditions generate fragments covering the sequence up to 1000 nucleotides or more.

DNA sequencing was carried out in five different steps as detailed below.

i. Denaturing the template DNA and annealing of primer.

Double stranded DNA template was used for sequencing. An aliquot of (32 µl, 1.5-2 µg) the DNA was denatured by the addition of 8 µl of 2 M NaOH solution. The mixture was vortexed, centrifuged briefly and incubated at room temperature for 10 min. The DNA was precipitated by the addition of 7 µl of sodium acetate (3 M, pH 4.8), 4 µl of distilled water and 120 µl of 100% ethanol

and kept at -20°C for 1 hr. The denatured DNA was pelleted by centrifuging at 12000 rpm (Hermle Z231M) at 10°C for 15 min. The pellet was washed with 70% ethanol, air dried and resuspended in 10 µl of distilled water. To the dissolved DNA (10 µl), aliquots of primer (2 µl; 1-2 pmol) and annealing buffer (2 µl; 1 M Tris-HCl, pH 7.6; 100 mM $MgCl_2$; 160 mM Dithiothritol, DTT)) to give a total volume of 14 µl. The mixture was vortexed gently, centrifuged briefly and incubated at 65°C for 5 min. The tube containing the mixture was incubated at 37°C for 10 min and at room temperature for 5 min. The contents of the tube was collected to the bottom by brief centrifugation.

ii. Labelling reaction

To the tube containing the annealed template and primer, added 3 µl of labelling mix, (Mix-dATP; 1.375 µM each dCTP, dGTP, dTTP and 333.5 mM NaCl), 1 µl of [α-^{35}S]dATP and 2 µl of diluted T7 DNA polymerase (0.5 µl of enzyme + 2 µl of enzyme dilution buffer containing 20 mM Tris-HCl, pH 7.5; 5 mM DTT; 100 µg BSA/ ml and 5% glycerol). The contents were mixed gently, centrifuged briefly and incubated at room temperature for 5 min.

iii. Termination reactions.

Aliquots of short mix A, G, T and C (2.5 µl each) were incubated in four separate tubes at 37°C for at least 1 min, while the labelling reaction was in progress. The labelled mix from step ii (4.5 µl) was added into each of the four pre-warmed tubes containing the short mix. The contents were mixed by gentle pipetting and incubated at 37°C for 5 min. Stop solution (5 µl, 0.3% of bromophenol blue and xylene cyanol FF; 10 mM EDTA, pH 7.5 and 97.5% deionized formamide) was added, mixed gently and centrifuged. The mixture was denatured by heating at 80°C for 2 min and 2 µl from each (A, T, G, C) tube loaded into the wells of a sequencing gel.

iv. Electrophoresis.

Two glass plates were thoroughly cleaned with soap and water, rinsed with distilled water and dried. The plates were ethanol cleaned, dried and the inner side of the small plate was siliconised by smearing with a few ml of Sigmacote (chlorinated organo polysiloxane in heptane, Sigma). The plates were assembled by inserting 0.2 mm wedge spacers between the long edges of the plates and held together with clamps. The two sides and bottom of the plates were taped together. 100 ml of acrylamide gel (4.6%) was prepared by dissolving 42 g Urea, 15 ml 30% acrylamide solution, 5 ml 2% bis-acrylamide solution, 10 ml 10 x TBE (1 x TBE contains 90 mM Tris; 90 mM Borate; 2 mM EDTA, pH 8.3); and 21 ml double distilled water. The solution was filtered through a 2 μM cellulose acetate filter unit under vacuum and degased for 5 min. To initiate gel formation, 600 μl of 10% ammonium per sulphate (APS) and 50 μl of tetramethyl ethylene diamine (TEMED, International Biotechnologies, INC) were added and mixed by swirling. The solution was poured immediately into a 60 ml syringe while blocking its spout. The plunger was inserted, any air trapped inside ejected and the solution injected carefully and slowly between the plates avoiding air bubbles. After pouring the gel, the plates were laid in a horizontal position. The surface-former was inserted and kept in that position for 45-60 min to allow the gel to polymerise.

The tape from the bottom of the plates was removed and the plates transferred to the vertical electrophoresis apparatus (Gibco-BRL) keeping the small plate facing inward. The plates were fixed by tightening the clamps on both sides and the buffer draining tap on the side of the apparatus was closed. The buffer (1 x TBE) was poured into the top and bottom buffer tanks. The surface-former was removed carefully and the sample wells rinsed to remove any unpolymerised acrylamide. The shark's-tooth comb was inserted, so that the points

of the comb just touched the surface of the gel. The apparatus was connected to the power supply and pre-run at 40 W constant power for 45 min. After the pre-electrophoresis the sample wells were again rinsed with buffer. The samples (sequencing reactions, G, A, T, C) were denatured at 80°C for 2 min and immediately transferred to wet ice. The power supply was switched off and 2 µl from each tube was loaded into the wells. The power supply was restored and electrophoresis effected at 40 W constant power (~2 hrs) until the bromophenol blue reached the bottom of the gel. The power supply was again cut off and a second loading of the samples was done in adjacent empty wells. The electrophoresis was continued until the bromophenol blue in the latter samples reached the bottom of the gel.

v. Autoradiography

The power supply was cut off and the buffer drained from the top tank. The gel-plate assembly was carefully removed and placed horizontally with the small plate uppermost. The small plate was separated, leaving the gel on the other plate. The DNA on the gel was fixed by placing the gel/plate in a tray containing ~3 litres of a mixture of 10% acetic acid/10% methanol in distilled water for ~30 min. The gel/plate was removed from the solution and a Whatman Number-1 filter paper cut to the size of the gel was placed on the gel. The small plate was placed on top of the filterpaper and the plates with the gel turned upside down, so that the large plate was on top. The large plate was then separated carefully, so that the gel was transferred to the filter paper. The gel was covered with a plastic wrap and dried using a vacuum gel dryer for ~45 min. After drying the gel, the plastic wrap was separated and the gel with its filter paper support was exposed to X-ray film overnight at room temperature. The film was developed in an automatic film developer (X-ograph Compact x 2).

2.1.5.2. Using Taq DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems)

A *Taq* DyeDeoxy Terminator cycle sequencing kit was used for the preparation of samples for sequence analysis on the Applied Biosystems Model 373A automated DNA sequencing system. The major benefits of this cycle sequencing system are 1) no alkaline denaturation is required for double stranded DNA 2) the same protocol is used for single stranded (ss) DNA, double stranded (ds) DNA and PCR products 3) less template is required 4) less enzyme is used and 5) longer lengths of DNA can be read. Moreover, all four termination reactions are performed in a single tube, thereby reducing the tube handling and pipetting steps by a factor of four. The sequence data are captured semi-automatically in an electronic format removing the tedious process of manual reading of autorads and manual data entry.

All the important reagents were supplied for performing fluorescence-based dideoxy sequencing reactions on ss or ds DNA templates or on PCR fragments. The performance of the kit relies on four ABI dye-labelled dideoxynucleotides, G, A, T, C DyeDeoxy Terminators. When these replace standard dideoxynucleotides in enzymatic sequencing, in the presence of thermally stable Ampli*Taq* DNA Polymerase, a dye label is incorporated into the DNA along with the terminating base.

In the present study, this system was used to sequence PCR products, mainly the PCR products generated from the BAC (Bacterial artificial chromosome) end fragments.

The template DNA was prepared by PCR amplification of the desired fragment as detailed in section 2.6 except that the reactions were carried out in a total volume of 100 µl. The amplified products were electrophoresed on a 1% metaphor agarose gel (Flowgen) in 0.6 x TBE buffer. The amplified fragments

were cut out from the gel with sterile scalpel blades and melted at 85°C for 10 min. 10 µl (10-20 ng) of the melted agarose-DNA mix was used as a template for cycle sequencing.

i. Cycle sequencing

The cycle sequencing reactions were carried out in a Perkin-Elmer Gene Amp-9600 thermal cycler. A single reaction mix contained 4 µl of 5 x TACS buffer (terminator ammonium cycle sequencing; 5 x TACS buffer contains 400 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM $(NH_4)_2SO_4$, pH 9.0), 1 µl of dNTP mix (7.5 µM dITP, 1.5 µM of each, dATP, dCTP, dTTP), 1 µl of each, A, G, T, C DyeDeoxy Terminator, 0.5 µl (4 units) Ampli *Taq* DNA polymerase, 0.5 µl (10 pmol) of primer and 10 µl (10-20 ng) template DNA in a total volume of 20 µl. The reactions were initiated by melting the agarose in the reaction mix at 80°C for 5 min followed by rapid thermal ramp to 96°C for 15 s, followed by 50°C, 1 s and 60°C, 4 min for a total of 25 cycles. At the finish of the reactions the temperature was brought to 4°C.

ii. Purification of extension products

The cycle sequenced products were purified by phenol-chloroform extraction as follows: 80 µl of deionized water and 100 µl of phenol : water : chloroform (68 : 18 : 14) reagent were added to the reaction mixture. The mixture was vortexed, centrifuged and the lower organic phase discarded. The aqueous layer was re-extracted with a second aliquot (100 µl) of phenol : water : chloroform. The mixture was vortexed, centrifuged for 1 min and the upper aqueous layer transferred to a clean tube. The extension products were precipitated by adding 15 µl sodium acetate (2 M, pH 4.5), and 300 µl 100% ethanol. The pellet obtained after centrifugation (15 min at room temperature) was washed with 70% ethanol and air dried.

iii. Preparation of samples for loading

The loading samples were prepared as follows: 5 µl deionized formamide and 1 µl EDTA (50 mM, pH 8.0) were mixed and 4 µl from this mixture was added to the ethanol precipitated pellet. The pellet was dissolved by vigorously mixing and briefly centrifuged to collect all the liquid at the bottom of the tube. The mixture was denatured for 2 min at 90°C and loaded into the well of an Applied Biosystems 373A DNA sequencing gel.

iv. Acrylamide gel preparation

An acrylamide gel (6%) was prepared by dissolving 30 g urea, 9 ml 40% acrylamide stock solution, 22 ml distilled water and 1 g mixed bed resin. The solution was filtered through a 0.2 µM cellulose acetate filter unit under vacuum and degased for 5 min. Added 6 ml of 10 x TBE buffer and the final volume made up to 60 ml. Polymerisation of the gel was initiated by the addition of 300 µl of freshly prepared 10% APS and 27 µl TEMED. The solution was gently swirled to mix and immediately poured between the plates without trapping air bubbles. The plates were laid horizontally and a surface-former was inserted and kept in that position for about 2-3 hrs to allow the gel to polymerise. The clamps and tapes from the edges of the plates were removed. The outside of the plates were washed and dried with a clean towel. The gel/plate assembly was placed in the electrophoresis chamber of the 373A DNA sequencer, the surface former removed and replaced with a shark tooth comb just touching the surface of the gel. The gel was electrophoresed in 1 x TBE buffer for 12 hrs as detailed in 373A DNA sequencing system (Applied Biosystems).

2.1.6. Purification of DNA from agarose gel

The fragment of choice from a restriction digested clone or a PCR product was purified from agarose by agarase digestion as follows:

The digested DNAs/PCR amplified products were electrophoresed in 1% low melting point agarose gels (Gibco-BRL), 1-2 volts/cm for 2-12 hrs. The DNA bands were carefully cut out from the agarose gel using sterile scalpel blades and digested with β-agarase (New England Biolabs), 1 unit/200 μl of agarose. The DNA was concentrated by adding $1/10^{th}$ volume of sodium acetate (3 M, pH 4.5), 2 volumes of 100% ethanol and kept at -20°C for 1 hr. The resulting pellet after centrifugation (15 min, 10°C, 12,000 rpm (Hermle Z231M) was washed with 70% ethanol and resuspended in appropriate volume of TE buffer pH 7.4.

2.1.7. Polymerase chain rection (PCR)

Polymerase chain reaction was used to generate the murine casein probes, probes to identify the BAC (Bacterial artificial chromosome) end fragments and to screen mouse genomic DNA libraries in bacteriophage P1/BAC. Target DNA sequences were amplified with the help of murine primers designed from the DNA sequences of α, β, γ, ε and κ casein cDNA/genomic clones available in the data base (refer to Table 2.2). The enzyme used was *Taq* DNA polymerase (Gibco-BRL). A single reaction mixture contained 50-100 ng template DNA, 100 μM of each dNTP, 20 pmoles of each primer, 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 0.75 mM $MgCl_2$, 0.025% detergent and 1unit of *Taq* DNA polymerase in a total volume of 20 μl. Amplification reactions were carried out in Perkin-Elmer Gene Amp-9600 thermal cycler. The reactions were initiated by denaturing at 95°C for 5 min, followed by 30 cycles of reactions; 94°C, 30 s for denaturation; 55°C, 30 s for annealing; and 72°C, 30 s for primer extension. In the last cycle, the primer extension step was performed for an additional 10 min. At the completion of the reactions, 5 μl of gel loading buffer (30% ficoll; 0.01% orange G and 10 mM EDTA) was added to each reaction mix and electrophoresed on

2.5% metaphor agarose gel (Flowgen) in 0.5 x TBE. The PCR conditions for the BAC end fragment amplification are detailed in section 2.6.

2.2. RNA Analysis

2.2.1. Total RNA isolation

Total RNA from mouse mammary and liver tissues were prepared using RNAzol (Biogenesis, Ltd). RNA extraction is based on the disruptive and protective properties of guanidinium thiocyanate. Guanidinium thiocyanate ensures the rapid inactivation of endogenous RNase activity and the complete dissociation of cellular components from the RNA. In the RNA isolation process, sterile DEPC (Diethyl pyro carbonate; 0.1%) treated tips were used.

To analyse casein gene expression, mouse mammary tissues were dissected out at various time points, virgin, pregnancy (5, 7, 9, 11, 13 and 16th day) and lactation (1st and 14th day p.p) periods. The tissues were homogenised in 2 ml (per 100 mg tissue) of RNAzol solution. Added 100 µl chloroform to each ml of the homogenised tissue and mixed by inversion. The tubes were placed on wet ice (15 min) and centrifuged at 12000 rpm (Hermle Z231M) for 15 min at 10°C. The supernatant was carefully transferred to microfuge tubes avoiding contamination with the interphase. An equal volume (~600 µl) of isopropanol was added, mixed and kept at 4°C overnight. The RNA was pelletted by centrifuging at 12000 rpm (Hermle Z231M) for 15 min at 10°C and the pellet was washed with 70% ethanol (prepared in 0.1% DEPC treated water). The pellet was air dried and resuspended in 250 µl of 1 mM EDTA (prepared in 0.1% DEPC treated water). The optical density of a suitable dilution was measured at 260 nm and the concentration was calculated using the formula, optical density 1 at 260 nm = 40 µg RNA. Thus, the concentration of RNA = O.D x 40 x dilution factor.

2.2.2. Isolation of Poly (A)$^+$ RNA

A Quickprep mRNA purification kit (Pharmacia) was used to isolate mRNA from total RNA using oligo (dT) cellulose columns. Guanidinium thiocyanate present in the RNA extraction buffer completely inhibits RNase activity and allows efficient hydrogen bonding between poly (A) tracts on mRNA molecules and oligo (dT) which is attached to cellulose columns. The mRNAs bound to the columns were eluted using the elution buffer.

The matrix of the oligo (dT)-cellulose spun column was resuspended and centrifuged at 350 x g for 2 min after removing the top and bottom closures. The bottom closure was replaced and 4ml of the total RNA mix (total RNA resuspended in 4 ml of elution buffer, 10 mM Tris-HCl, pH 7.5; 1 mM EDTA) added to the surface of the resin of the oligo (dT) column. The column was closed with the top lid, the resin resuspended by gentle mixing (10 min) and centrifuged (350 x g, 2 min). The supernatant was discarded and the matrix was resuspended in 3 ml of high salt buffer (10 mM Tris-HCl, pH 7.5; 1 mM EDTA; ; 0.5 M NaCl) and centrifuged (350 x g, 2 min). The supernatant was discarded and washing with the high salt buffer repeated two more times. The matrix was then mixed with 3 ml of low salt buffer (10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M NaCl), centrifuged and the supernatant was discarded. A final wash was given by adding 3 ml of low salt buffer on top of the resin. The bottom closure was removed and the column centrifuged. The poly (A) RNAs attached to the cellulose matrix were eluted by washing with the 0.75 ml of elution buffer (three times with 0.25 ml each).

2.2.3. Synthesis of cDNA

The polyA RNA (mRNA), isolated as described above was used for the cDNA synthesis. First strand cDNA was synthesised from 1-2 μg mRNA template

using the Superscript preamplification system (Gibco BRL). Added 1 µl (0.5 µg) oligo (dT) primer to the mRNA in a microfuge tube and the volume made up to 12 µl with DEPC treated water. The contents were mixed and incubated for 10 min at 70°C and 1 min on ice. To the RNA/primer mixture added 7 µl of a reaction mixture (2 µl of 10 x PCR buffer (10 x = 200 mM Tris-HCl, pH 8.4; 500 mM KCl), 2 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP mix and 2 µl of 0.1 M DTT), mixed gently and incubated at 42°C for 5 min. Superscript II reverse transcriptase (1 µl, 200 units) was added to the tube, mixed and incubated for 50 min at 42°C. The reaction was terminated by keeping at 70°C for 15 min, chilled on ice and the reaction mix collected at the bottom of the tube by brief centrifugation. RNase H (1 µl) was added to the mix and further incubated at 37°C for 20 min to digest the RNA. The first strand cDNA synthesised by this method was used for radioactive labelling (random priming).

2.2.4. Electrophoresis of RNA

RNA was electrophoresed in a denaturing formaldehyde-agarose system. 10 µg (3-10 µl) RNA was mixed with 25 µl formamide, 5 µl 10 x MOPS (10 x MOPS = 0.2 M 3-[N-Morpholino]propane-sulphonic acid; 0.5 M Na acetate pH 7.0; 0.01 M Na_2EDTA) and 8 µl formaldehyde (37% v/v) in a total volume of 50 µl and incubated at 65°C for 10 min. The mixture was chilled on ice for 20 s and 5 µl of loading dye (0.01% bromophenol blue in 50% (v/v) glycerol) added. The RNA was electrophoresed on a 1% agarose gel containing 1 x MOPS and 6.29% formaldehyde (formaldehyde solution; 37% v/v was added to 50°C cooled agarose-MOPS mixture, mixed and poured immediately). Electrophoresis was carried out for 16 hrs at 1-2 V/cm. After electrophoresis the gel was stained with ethidium bromide (0.5 µg/ml; 30 min), destained in distilled water (30 min) and photographed.

2.2.5. Northern blotting

Northern blotting refers to the transfer of RNA from agarose gels to nylon or nitrocellulose membrane. A capillary blot was set up by filling a tray with blotting buffer, 20 x SSC (3 M NaCl; 0.3 M Sodium citrate) and preparing a platform by placing a glass plate over the tray. The plate was covered with a wick made from two sheets of Whatman 3 mm filter paper, saturated with the blotting buffer. The gel was placed upside down on the wick. The air bubbles trapped beneath the gel were rolled out with a pipette. The gel was surrounded with parafilm to prevent the blotting buffer being absorbed directly in to the paper towels above. A positively charged nylon membrane (Boehringer-Mannheim) cut to the size of the gel, wetted with blotting buffer was placed carefully on top of the gel. The bubbles trapped below the membrane were rolled out and three sheets of Whatman 3 mm paper (cut to the size of the gel) placed on the membrane, after wetting in the blotting buffer. A stack of absorbent paper-towels (approximately 5 cm high) and a 0.75-1 kg weight were placed on top. The RNA was transferred for ~16 hrs. When the blotting process was complete, the apparatus was carefully dismantled and the positions of the wells marked on the membrane for identification of tracks. The RNA was fixed onto the membrane by autocrosslinking in a Stratalinker-1800 (Stratagene), keeping RNA side towards the UV source. The membrane was wrapped in saran wrap and stored at room temperature for use later.

2.3. DNA Analysis

2.3.1. Preparation of high molecular weight (HMW) genomic DNA.

DNA from the 129 strain of mice (*Mus musculus*) was used for long range mapping of the casein locus. For restriction fragment length variants (RFLVs) studies DNA was prepared from *Mus spretus* and seven more strains of *Mus*

musculus, C57/BL6, NIH, C3H, SJL, BALB/c, CBA and DBA. The high molecular weight genomic DNA was prepared from liver tissue (phenol-chloroform extraction) and spleen cells (DNA embedded in agarose plugs).

2.3.1.1. Phenol-chloroform extraction

Freshly dissected liver was washed with SSCT pH 7.0 (0.15 M NaCl; 0.015 M Na_3 citrate; 10 mM Tris-HCl, pH 7.4), minced and transferred to a pre-cooled glass dounce. The tissue was homogenised in 6 ml of pre-cooled RSBE/NP40 (10 mM Tris; 10 mM NaCl; 2 mM EDTA, pH 7.4; 0.5% Nonidet P40; pH 7.0) containing 1 mM PMSF (phenyl methyl sulfonyl fluoride). The homogenised tissue was transferred into a 30 ml corex tube and centrifuged at 4500 rpm (Sorvall SS-34) for 5 min at 4°C. The cell pellet was resuspended in 6 ml of RSBE/NP40 and washed twice. To the washed pellet 2 ml of RSBE was added and the pellet resuspended by vortexing, adding 2 ml of SNET (0.6 M NaCl; 10 mM EDTA; 20 mM Tris, pH 7.4; 1% SDS) while vortexing. The enzyme, proteinase-K was added to a final concentration of 200 µg/ml and digested at 37°C for 2 hrs (or overnight at room temperature). The digested tissue was mixed with an equal volume of phenol and centrifuged at 4000 rpm (Sorvall SS-34) at 4°C. The aqueous phase (top phase) was transferred to a fresh tube, mixed with equal volume of 1:1 phenol : chloroform/isoamyl alcohol (24 : 1 chloroform : isoamyl alcohol) and centrifuged. The aqueous phase was transferred again to a fresh tube and mixed with an equal volume of isopropanol. The precipitated DNA was hooked out with a sterile loop and washed with 70% alcohol. Air dried DNA was dissolved in 2 ml of TE buffer and stored at 4°C. The DNA concentration was estimated by measuring the absorbance of a suitable dilution at 260 nm in a spectrophotometer.

2.3.1.2. Embedding DNA in agarose plugs

A CHEF mammalian genomic DNA plug kit (Bio-Rad.) was used to prepare intact DNA in agarose plugs. Fresh spleen tissue was homogenised in normal saline and the cells were pelletted by centrifuging at 5000 rpm (Sorvall SS-34) for 10 min at 4°C. The cells were resuspended in 5 ml of normal saline and the total number of cells were calculated using a haemocytometer. The cells were washed twice in normal saline, resuspended in 3.78 ml of cell resuspension solution (10 mM Tris, pH 7.2; 20 mM NaCl; 50 mM EDTA) and equilibrated to 50°C for 15 min. To the cells, 2.22 ml Clean cut agarose (2%, 50°C) (Bio-Rad) or Incert agarose (Flowgen), was added to give a final cell count of 1-2.5 x 10^7 cells/ ml. This resulted in a final concentration of 0.75% agarose. The cell suspension was mixed well, avoiding any clumps. The cell-agarose mixture was kept at 50°C and aliquots (~100 µl each) were pipetted into sterile agarose plug moulds. Solidification of the agarose was effected by placing the moulds at 4°C for 15 min. The plugs were pushed into proteinase-K reaction buffer (100 mM EDTA, pH 8.0; 0.2% sodium deoxycholate; 1% sodium lauryl sarcosine; 1 mg/ ml proteinase-K; 5 ml for each ml of agarose plugs) and incubated for two days at 50°C without agitation. The plugs were then washed with 50 ml of 1 x plug wash buffer (20 mM Tris, pH 8.0; 50 mM EDTA) for 1 hr at room temperature with gentle agitation. The washing was repeated four times. During the third wash, 1 mM PMSF was added to inactivate residual proteinase-K. The plugs were stored in 0.5 M EDTA at 4°C.

2.3.2. Isolation of large genomic clones from murine genomic DNA libraries

2.3.2.1. Screening bacteriophage P1 and BAC libraries

A mouse P1 library containing ~40,000 clones (1 x genome coverage) was kindly provided by Dr. Linda Mullins, University of Edinburgh. The library was arranged in 9 secondary pools and each secondary pool consisted of 10 primary pools. About 400 clones were present in each primary pool.

The mouse BAC library was supplied by Research Genetics, Inc, USA, consisting of 221184 clones, organized in 72 superpools, having a genome coverage of 7.37. Each super pool represents arrayed colonies pooled from eight 384 well plates. Each plate pool was made by pooling the 384 colonies from a single plate. The superpools and plate pools were supplied as purified DNA preparations. Specific plates were requested from Research Genetics after preliminary screening of the superpools and platepools. A PCR based technique (detailed in 2.1.7) was followed to screen the DNA pools.

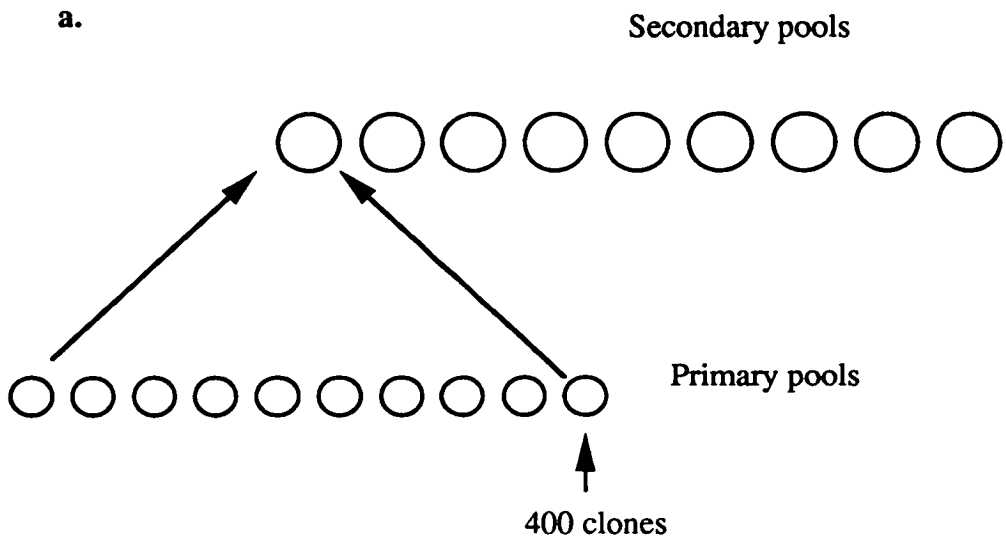
The organizations of the P1 and BAC libraries are shown in Fig. 2.2. The P1 and BAC vectors are shown in Fig. 2.3 and 2.4 respectively.

2.3.2.2. Colony transfer and hybridisation

i. Bacteriophage P1 library

Once a positive pool was isolated from the P1 library, the pool was plated out on a LB, kanamycin (25 μg/ml), Sucrose (5%) plate (~1500 to 2000 colonies). A colony lift and hybridisation technique was used to isolate the positive clones as detailed below:

1. Gridded hybond-N nylon membranes (Amersham) were placed on LB/kanamycin/sucrose plates, smeared with 50 μl of 0.1M IPTG and kept for 20 to 30 min.



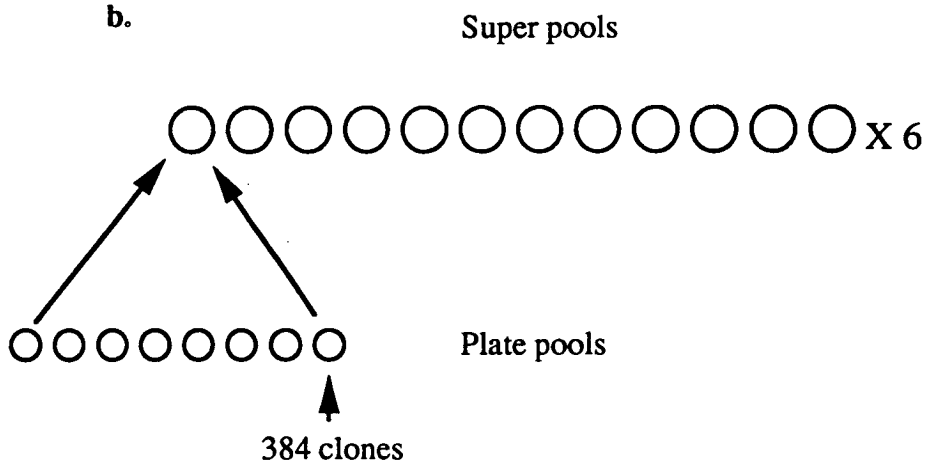


Fig. 2.2. Organizations of a. P1 library b. BAC library

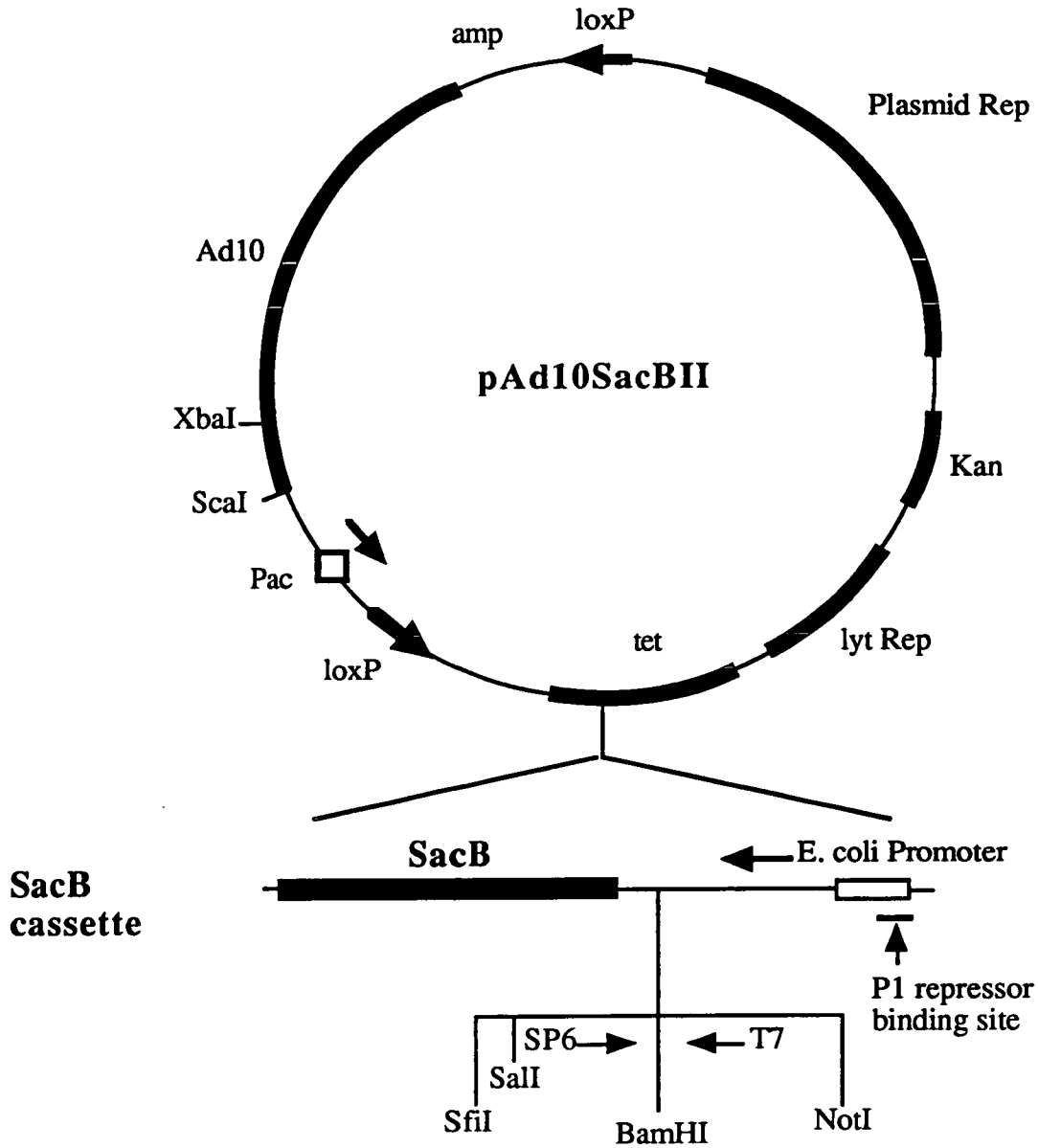


Fig. 2.3. Diagramatic representation of bacteriophage P1 vector, pAd10SacBII

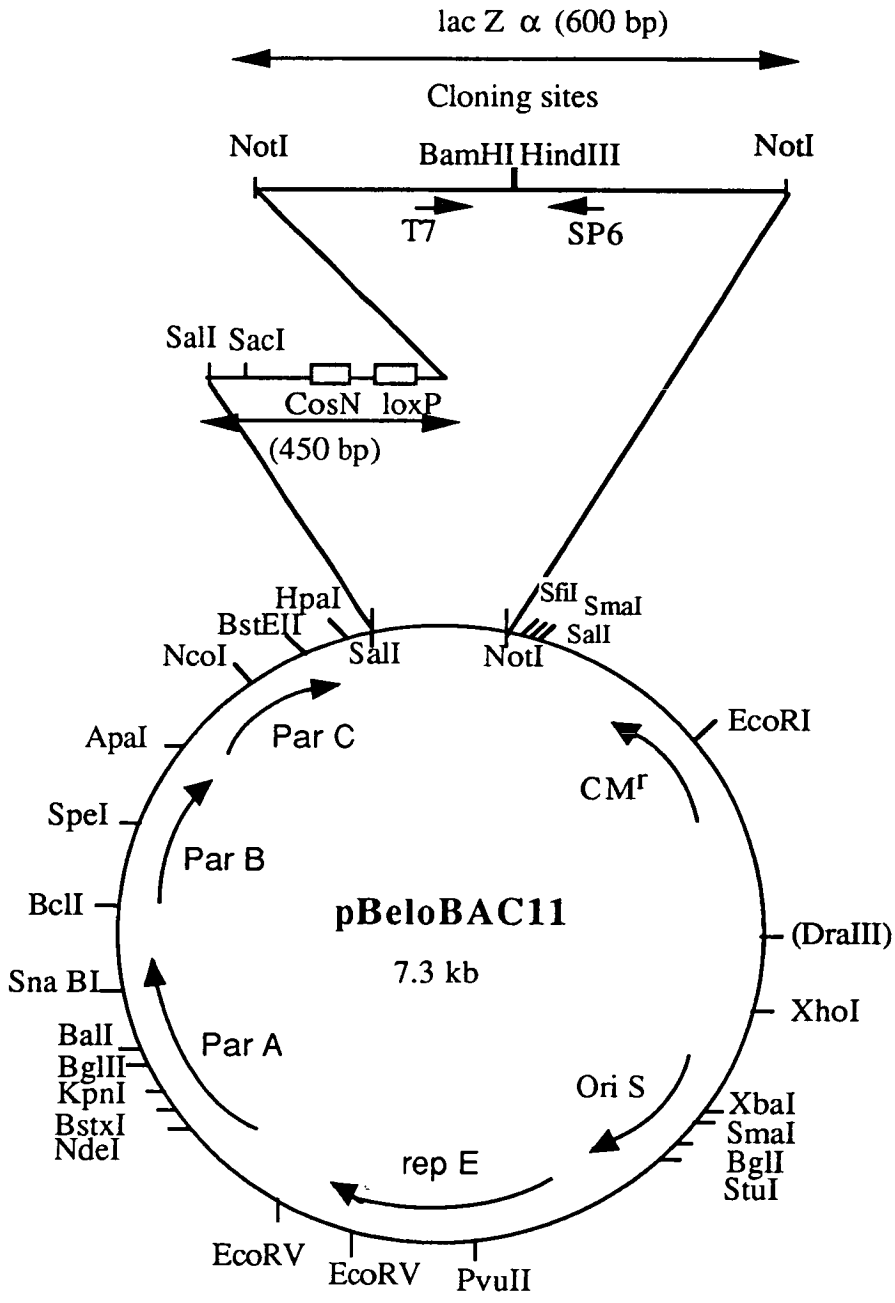


Fig. 2.4. Schematic representation of BAC vector, pBeloBAC11

2. The colonies were carefully lifted in duplicate, from the master plate, using the IPTG treated filter membranes. The filter membranes were placed back on the same agar plates (mentioned in 1), colony side up and incubated for 6 to 7 hours at 37°C.

3. When the colonies were of a pin head size, each filter was processed by placing the filter on Whatman 3 mm filter paper saturated with the following solutions and for the specified time, keeping the colony side up.

a) 0.5 N NaOH + 1.5 M NaCl, 4min;

b) 1 M Tris-HCl, pH 7.4 + 1.5M NaCl, 4min (twice)

The filters were rinsed in 2 x SSC and U.V cross linked in UV Stratalinker-1800 (Stratagene). The membranes were finally washed with a solution containing 0.1% SDS and 2 x SSC. All cell debris were removed manually with a sponge.

ii. BAC library

Before starting any manipulations in the mouse BAC library plates, a replica of each plate was prepared. The colonies were transferred, to a fresh plate containing 60 µl of medium (LB + chloramphenicol, 12.5 µg/ml + 10% glycerol), using a 384 well replica plating device (Genetix). The plate was incubated at 37°C (stationary phase) for 48 hrs, sealed with plastic seal and stored at -70°C.

Each plate was screened by a colony transfer and hybridisation technique. The colonies were transferred, in duplicate, on to positively charged nylon membranes (Hybond-N^+, Amersham) already placed on LB/chloramphenicol plates using the 384 well replica plating device. The plates were incubated at 37°C overnight. The nylon membranes were then treated as follows:

1) The membranes were placed, colony side up, on 3 mm Whatman filter paper saturated in 0.5 N NaOH + 1.5 M NaCl for 4 min at room temperature followed by 4 min at 85°C on a steaming water bath

2) The membranes were transferred to a 3mm Whatman filter paper saturated in 1 M Tris-HCl, pH 7.4 + 1.5 M NaCl for 4 min at room temperature.

3) The membranes were washed with 2 x SSC, pH 7.0 and all cell debris removed mannually with a sponge

e) The membranes were dried on 3 mm Whatman filter paper and autocrosslinked (UV Stratalinker-1800, Stratagene).

2.3.2.3. P1 DNA preparation

P1 DNA was prepared using a Qiagen maxi plasmid purification system as described in 2.1.3 with the following modifications. LB/Kanamycin medium (500 ml) was inoculated with the pre-culture (2.5 ml) and incubated for 1 hr at 37°C in a shaking incubator. After one hour, IPTG (500 µl of 0.1M solution) was added to the culture which was incubated for a further period of ~5 hrs. Pre-warmed (50°C) elution buffer was used to elute the P1 DNA, from the Qiagen column in the final stages of the purification procedure.

2.3.2.4. BAC DNA preparation

A pre-culture was prepared by inoculating 5 ml of LB/chloramphenicol (12.5 µg/ml) medium with a single colony (incubated for 8 hrs). LB/chloramphenicol medium (500 ml) was inoculated with 2 ml of pre-culture and incubated over night. The cells were pelleted and the Qiagen protocol for the plasmid DNA isolation was followed (section 2.1.3) with slight modifications. The bacterial pellet obtained from 250 ml of culture was resuspended in 10 ml buffer P1 and the volumes of buffers P2 and P3 were also increased correspondingly. Moreover, the elution buffer was warmed to 50°C prior to elution of the plasmid DNA. The pelleted DNA was dissolved in 400 µl of TE

buffer and further purified by phenol-chloroform extraction and ethanol precipitation. The purified pellet was dissolved in 250 µl of TE buffer.

2.3.3. Restriction digestion

The restriction enzymes used in the present study were purchased either from New England Biolabs (NEB) or Boehringer Mannheim unless mentioned otherwise. The restriction digestions were carried out according to the manufacturer's instructions except that a 3-5 fold enzyme excess was added for each digestion. Whenever double digestions were performed, the DNA was first digested with one of the enzymes and the second digestion was carried out after ethanol precipitation of the DNA. In cases where all the digestion conditions (e.g. buffer, temperature) for both the enzymes were the same, the double digestions were carried out simultaneously. After digestion, the reactions were stopped by placing the tubes at 65°C for 10 min and then 5 µl of gel loading buffer (30% Ficoll 400; 0.1% SDS; 40 mM Na_2 EDTA, pH 8.0; 0.12% bromophenol blue) was added for every 25 µl of reaction mix.

2.3.3.1. Plasmid DNA (for preparation of casein probes)

For the preparation of murine casein probes, the plasmid DNAs were digested with the restriction enzyme of choice (α-cDNA and γ-cDNA - *Pst*I, β-5' probe - *Hin*dIII/*Sca*I, β-3' probe - *Sal*I/*Sac*I, κ-cDNA - *Kpn*I/*Bam*HI). After digestion, the cloned fragments of interest were verified by electrophoresing 1 µl (100 - 500 ng DNA) of the digested DNA on a agarose (0.8-1%; Gibco BRL) mini gel.

2.3.3.2. Genomic DNA

Restriction enzymes with recognition sequences present in CpG island and inter-island region and enzymes with large recognition sequences were selected to

generate large fragments. The fragments thus generated were analysed by pulsed-field gel electrophoresis. The enzymes *Asc*I and *Not*I are eight base cutters with 100% G+C recognition sequences and two CpGs, while *Bss*HII, *Eag*I and *Sac*II are six base cutter enzymes with two CpGs and 100% G+C recognition sequences. Moreover, the sites for *Bss*HII and *Eag*I are internal to the sites of *Asc*I and *Not*I respectively. *Nar*I and *Sma*I have 100% G+C recognition sequence and one CpG in their site. Enzymes with 66.7% G+C in their recognition sequence and two CpGs (*Mlu*I, *Pvu*I, *Nru*I) and one CpG (*Xho*I, *Sal*I) were also selected. Other enzymes like *Cla*I (33.3% G+C, one CpG), *Bam*HI (66.7% G+C, no CpG), *Sfi*I (large recognition sequence), *Eco*RV, *Apa*I, *Dra*I, *Bgl*I were also used.

Restriction fragment length variant (RFLV) studies were carried out on genomic DNA prepared from the two different species and eight different strains of mice. The enzymes used were *Sfi*I, *Bam*HI, *Eco*RI, *Hin*dIII and *Xba*I.

For each restriction digestion 10 μg of genomic DNA was used. The reaction mixture was incubated at appropriate temperature for about 5 to 6 hours. The reaction was stopped by placing the tubes at 65°C for 10 min. Gel loading buffer, 5 μl was added for every 25 μl of reaction mixture.

Very high molecular weight DNA, prepared by embedding the DNA in agarose plugs, was digested as follows: The DNA in agarose plugs was washed thrice, an hour each time, in TE buffer (pH 7.4) with a mild rocking movement and equilibrated in the appropriate 1 x restriction buffer for at least an hour. The plugs were cut with sterile scalpel blade and ~5 μg DNA was used in each digestion. Added 5-8 units of enzyme/μg of DNA and 5μg bovine serum albumin (BSA) to 1 x restriction buffer in a total volume of 200 μl. The digestion was carried out for ~16 hrs (overnight). The reaction was stopped by washing the plugs with 1 x plug wash buffer for 30 min and equilibrated in 0.5 x TBE. The

plugs were then loaded into the wells of 1% Pulsed-field certified agarose (Bio-Rad) prepared in 0.5 x TBE and the wells were sealed with the same agarose.

Double digestions were also carried out in DNA plugs. After the first digestion, the plugs were washed extensively in TE buffer and equilibrated in the second restriction buffer for 1 hr prior to the addition of the second enzyme cocktail.

2.3.3.3. P1/ BAC clones

Bacteriophage P1 clones and BAC clones were digested with a few selected enzymes, *Not*I, *Sal*I, *Sfi*I, *Eco*RI, *Hin*dIII, *Pst*I, *Bgl*II, *Eco*RV, *Bam*HI, *Sac*I, *Xba*I, *Apa*I and *Not*I, *Sal*I, *Sma*I, *Sfi*I, *Xho*I, *Pvu*I and *Bam*HI respectively.

The cloned DNA (3-5 µl, ~200 ng) was digested with 10-20 units enzyme in presence of 2 mM spermidine, in a total volume of 50 µl of appropriate 1 x reaction buffer. The reaction mix was incubated at the relevant temperature for 5-6 hrs. The reactions were stopped as detailed earlier.

2.3.4. Electrophoresis

2.3.4.1. conventional agarose gel electrophoresis.

Conventional horizontal gel electrophoresis was carried out to separate fragments up to 20 kb. Agarose gel (0.8 - 1%, Gibco-BRL) was prepared in 1 x TAE buffer (40 mM Tris; 40 mM acetate; 1.0 mM EDTA, pH 8.0). Ethidium bromide (0.5 µg/ml) was added to the melted agarose and mixed well. The agarose was cooled to 50°C, poured into the gel casting mould fitted with a suitable comb and left at room temperature to set. The combs were carefully removed and the gel/casting mould was placed in the electrophoresis tank containing 1 x TAE buffer. The samples were loaded into the wells along with

suitable size markers and electrophoresed for 8-16 hrs at low voltage (1-2 V/cm) according to the size of fragments to be separated.

2.3.4.2. Pulsed-field gel electrophoresis

Pulsed-field electrophoresis is a technique for resolving large DNA fragments up to 10 Mb. By alternating the electric field between spatially distinct pairs of electrodes, megabase sized DNAs are able to reorient and move at different speeds according to size through the pores in an agarose gel.

Pulsed-field gel electrophoresis was carried out in a CHEF (Clamped homogenous electric fields) DR-III system (Bio-Rad). The CHEF-DR III system is based on two technologies, the clamped homogenous electric fields (CHEF) and programmable autonomously controlled electrodes (PACE). The CHEF provides highly uniform or homogeneous electric fields within the gel through an array of 24 electrodes which are clamped or held to intermediate potentials to eliminate lane distortion. The electrodes sense changes in local buffer conductivity due to buffer breakdown, change in buffer type, gel thickness or temperature. Potentials are readjusted immediately to maintain uniform fields, thus providing high resolution during run and between runs. The voltage potential of each of the 24 electrodes is regulated independently by PACE. Due to the hexagonal geometry of the electrode array, the system can generate reorientation field angles from 90 - 120° (detailed in the Instruction manual and applications guide of CHEF DR-III pulsed-field electrophoresis systems, Bio-Rad).

The samples were electrophoresed on 1% pulsed-field certified agarose gel (Bio-Rad) in 0.5 x TBE for 18 hours, 6V/cm, 120° angle, at 14°C with initial and final switch intervals of 5 s and 40 s respectively. These conditions effectively separated fragments up to 550 kb. After electrophoresis, the gel was stained with ethidium bromide (0.5 μg/ml, 30 min) nicked by UV irradiation (total = 60

mJoules of energy), destained for half an hour and photographed using very short exposure (<1 s). Small fragments (5-160 kb; P1/ BAC clones) were separated by changing the duration of electrophoresis (12.5-13 hrs) and switch intervals (0.2 s and 8.0 s respectively) while keeping all other parameters the same as above.

2.3.5. Southern blotting

Capillary blots were set up in the same way as detailed in Northern blotting except for the transfer buffer. An alkali buffer (1.5 M NaCl, 0.4 M NaOH) was used as the transfer buffer and also for denaturing the DNA in the gel (15 min) prior to blotting. The gel was placed upside down to avoid trapping of air bubbles and for efficient transfer of the DNA. Capillary blotting was carried out for ~36 hours using positively charged nylon membranes (Boehringer Mannheim). After blotting, the membranes were neutralised in 0.5 M Tris-HCl, pH 7.0 for 5 min, rinsed in 2 x SSC and baked at 80°C for an hour.

2.4. Radioactive labelling of probes

2.4.1. Using [α^{32}P]dCTP

Labelling with [α^{32}P]dCTP was based on the use of random sequence hexanucleotides to prime DNA synthesis on denatured template DNA at numerous sites along its length (Feinberg and Vogelstein, 1983, 1984). The 'Klenow' fragment of DNA polymerase I was used for this reaction. The absence of the 5'-3' exonuclease activity associated with DNA polymerase I ensures that labelled nucleotides incorporated by the polymerase are not subsequently removed as monophosphates. The amount of newly synthesised DNA often exceeds the amount of input DNA.

DNA labelling systems from Amersham (Multiprime DNA labelling system) and Boehringer Mannheim (High prime DNA labelling kit) were used to

label the probes. In each labelling reaction ~25 ng DNA probe was labelled using [α^{32}P]dCTP.

In the Multiprime DNA labelling system, the labelling reaction was performed in a total volume of 50 µl. DNA (25 ng in 33 µl of distilled water) was denatured for 3 min in boiling water bath and immediately chilled on ice for 2 min. To the denatured DNA, added 5 µl multiprime buffer solution (dATP, dTTP, dGTP in a concentrated buffer solution of Tris-HCl, pH 7.8, Magnesium chloride and 2-Mercaptoethanol), 5 µl primer solution (random hexanucleotides in an aqueous solution containing nuclease-free BSA), 2 µl enzyme solution (1 unit/µl DNA polymerase I 'Klenow' fragment cloned in 50 mM Potassium phosphate, pH 6.5, 10 mM 2-Mercaptoethanol and 50% Glycerol) and 5 µl (50 µCi) [α^{32}P]dCTP. The contents were mixed gently, centrifuged briefly and incubated at 37°C for 30 min. The reaction was stopped and the DNA was denatured by adding 5 µl (1/10th volume) 2 M NaOH.

The assay conditions were slightly different in the High prime DNA labelling system. The reaction mix was prepared by adding 8 µl DNA (25 ng), 4 µl high prime reaction mixture (containing random primer mixture, 1 unit/µl Klenow polymerase, stabilized buffer in glycerol, 50% v/v), 3 µl of dNTP mix (500 µM of each dNTPs, dATP, dGTP, dTTP) and 5 µl (50 µCi) [α^{32}P]dCTP in a total volume of 20 µl. The reaction mix was incubated for only 10-15 min.

2.4.2. Using [γ^{32}P]ATP

Oligonucleotides were end labelled using [γ^{32}P]ATP. The enzyme, polynucleotide kinase catalyses the transfer of the terminal phosphate group of ATP to the 5' hydroxylated terminus of DNA or RNA. The assay mixture was in a total volume 20 µl, containing 2 µl reaction buffer (10 x buffer contains 500 mM Tris-HCl, 100 mM $MgCl_2$, 1 mM EDTA, 50 mM dithiothreitol, 1 mM

spermidine, pH 8.2) , 1 µl (100 ng) oligonucleotide, 6 µl (60 µCi) [$\gamma 3^2$P]ATP and 1 µl (10 units) polynucleotide kinase (Boehringer Mannheim). The reaction mix was incubated at 37°C for 30 min and the reaction was stopped by chilling on ice.

2.4.3. Measurement of incorporation and calculating the specific activity of labelled DNA

From the reaction mixture, 1 µl was spotted on the centre of a 25 mm GF/C Whatman glass microfibre filter paper and kept in a plastic scintillation vial for the determination of total input radioactivity. Another 1 µl of the reaction mix was placed in a microfuge tube, to which added 200 µl of 0.2 M EDTA, 50 µl of carrier DNA solution and 2 ml of ice-cold 10% trichloracetic acid (TCA). Complete precipitation of the labelled DNA is ensured by the carrier DNA. The contents were mixed well and allowed to stand on ice for 10 min. The precipitated DNA was collected on a Whatman GF/C-filter paper by vacuum filtration. The filter disc was washed with excess of 10% TCA. The filter paper containing the precipitated labelled DNA was placed in a plastic scintillation vial. Radioactivity of input and labelled nucleotides were measured in a β-counter. Probes in which incorporation of 60% or more of the radioactivity was achieved were used for hybridisation.

Specific activity of the labelled DNA was calculated using the formula

[% of incorporation/total amount of DNA (ng)] x 10^3 dpm per µg

2.5. Hybridisation and Autoradiography

Membranes (for either Southern or Northern blots) were pre-hybridised in 20 ml of pre-hybridising solution containing 6 x SSPE (1.08 M NaCl; 60 mM sodium phosphate; 6 mM Na_2EDTA), 0.5%SDS, 5 x Denhardt's solution (bovine

serum albumin, ficoll and polyvinylpyrrolidone, each one at 0.1% (w/v)) and 100 μg/ml of denatured salmon sperm DNA for at least one hour.

The labelled probe was treated with mouse COT-I DNA (Gibco-BRL) to suppress cross-hybridisation to mouse repetitive DNA during filter hybridisation (only for Southern blots). Mouse COT-1 DNA was prepared from CD1 mice by extracting, shearing, denaturing and annealing DNA under conditions that enrich for repetitive DNA sequences such as the B1, B2 and L1 families.

The labelled probe (50 μl or 20 μl) was made up to 100 μl and passed through a Sephadex G-50 column (Sambrook et. al., 1989) to eliminate the unincorporated radioactive nucleotide. To the filtrate, added 5 μl (100 μg) mouse COT-1 DNA, 50 μl 20 x SSC, 25 μl H_2O and 20 μl 1% SDS. The contents were mixed, briefly spun and placed in a boiling water bath for 5 min to denature the DNA and immediately transferred to 65°C and incubated for 20 min.

After pre-hybridisation, the solution was poured off and the membranes were hybridised in 20 ml of the hybridising solution (6 x SSPE, 0.5% SDS, and 100 μg/ml of denatured salmon sperm DNA) containing the labelled probe/COT-1 DNA mix. The membranes were hybridised for 36-48 hrs at 65°C. After hybridisation, the membranes were washed twice in a solution containing 2 x SSC and 0.1% SDS at 65°C for 20 min each and a final wash was given in 0.1 x SSC and 0.1% SDS (65°C for 15 min). The membranes were dried (just to remove the dripping solution) on 3mm filter paper and exposed against AGFA CURIX X-ray film with two Dupont Cronex Lightning Plus intensifying screens at -70°C for one to two weeks and autoradiographed in an X-Ograph Compact x 2.

A hybridisation temperature of 55°C was used for end labelled probes (i.e. oligonucleotide probes), keeping all other parameters the same as above.

Southern blots of cloned DNA fragments were hybridised for 6 to 16 hrs and were exposed to X-ray film for very short periods (15 min to 3 hrs).

The membranes were stripped by placing them in boiling 0.1% SDS and allowed to cool to room temperature. The membranes were reprobed after stripping and exposing to film.

2.6. BAC end fragment isolation

Right and left ends of insert DNA were isolated following the method described by Woo et.al. (1994) to facilitate chromosome walking. A schematic representation of BAC end isolation is shown in Fig. 2.5.

In the BAC library supplied by Research Genetics, the *Hin*dIII restriction site was used as the cloning site. The generation of right and left end fragments of the BAC clones was based on the presence of unique/single restriction enzyme sites on either side of the cloning vector. Both right and left end fragments can be isolated by a technique called plasmid rescue. For the right end fragment isolation, the unique restriction site of the selected enzyme should lie near the left end of the cloning site (e.g. *Bam*HI, *Sph*I and *Sac*I; see Fig. 2.4 and 2.5) and can have one or more sites in the cloned fragment. Digesting with such an enzyme will cut and remove the insert DNA including the left end fragment, leaving the right end fragment intact with the vector DNA. Since the vector is intact except for a small portion near the left side of the cloning site, the resulting vector + end fragment can be self ligated. The subclones generated in this way can be electroporated into DH10B cells. Selection for the uptake of DNA molecules consisting of vector sequences plus the right end fragment is effected by growing the bacteria on LB/chloramphenicol plates. The size of the fragment can be assessed by a double digestion with the selected enzyme and *Hin*dIII.

The left end fragment can also be isolated in a similar way. In this case, the unique enzyme site should be present near the right end of the cloning site. In pBeloBAC11, *Sfi*I is the only enzyme having a unique restriction site near the

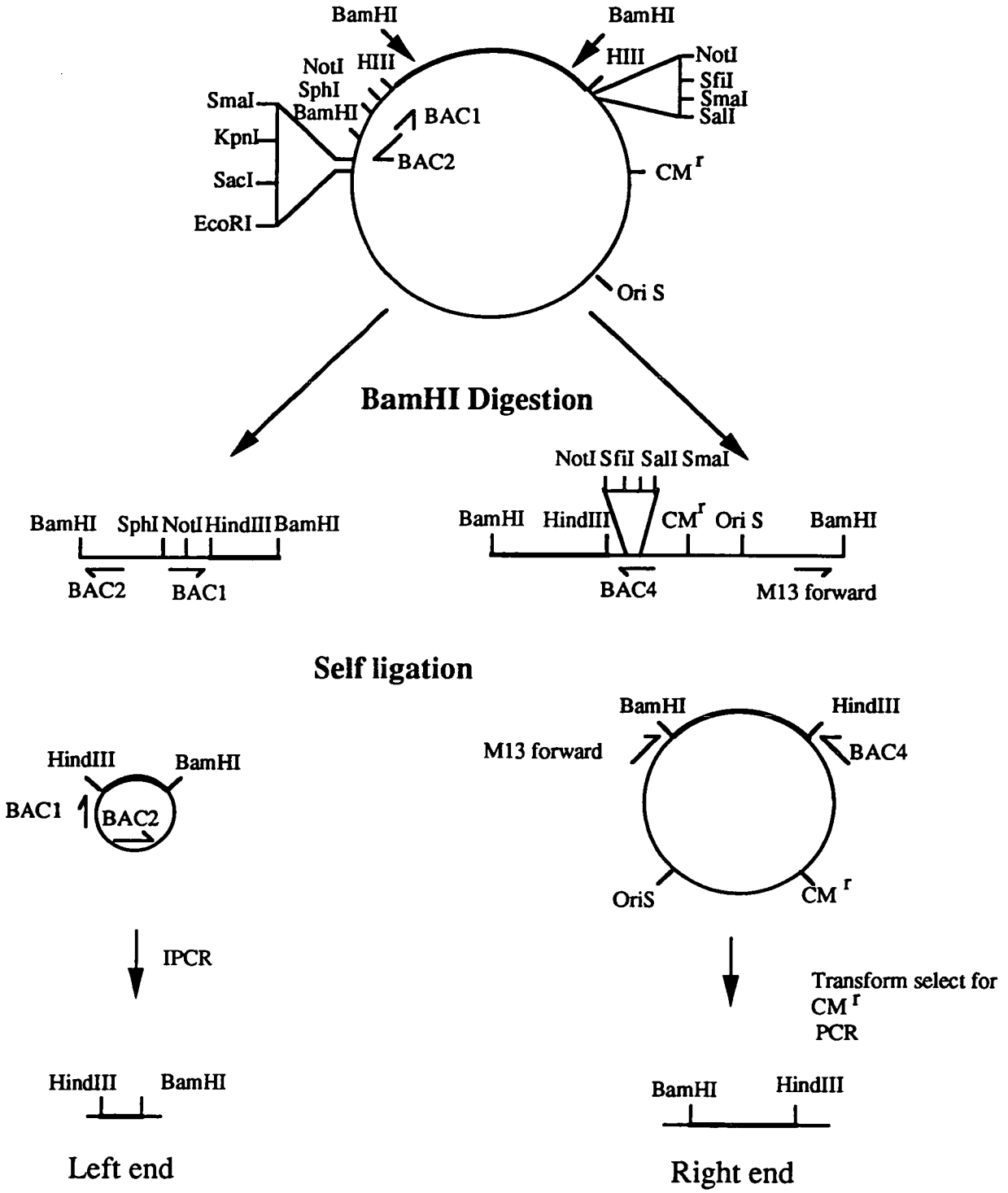


Fig.2.5 Schematic diagram of the generation of left and right end specific probes from BACs with IPCR (left end) and plasmid rescue (right end) and the restriction enzyme BamHI. BAC1, 2, 4 and M13 forward are oligonucleotide primers specific to pBeloBAC11. (modified from Woo et al., 1994)

right end of the cloning site. The difficulty with this enzyme is that there is some redundancy in the *Sfi*I recognition site (GGCCN/NNNNGGCC). Thus it is necessary to fill in the recessed 3' end, to enable a blunt ligation.

The subcloned end fragments (size < 5 kb) can be PCR amplified with suitably designed primers from the vector sequences, depending on the restriction enzyme selected. A large sized subcloned fragment can be sequenced using the promoter sequences (left end - T7; right end - SP6). Primers can be designed from such sequenced fragments and subsequent PCR amplification can generate probes for the respective ends.

Inverse polymerase chain reaction (IPCR) can also be used to isolate the end fragments. Unlike the plasmid rescue, to carry out IPCR the selected enzyme must have a site near to the cloning site of the end fragment of choice. The size of the end fragment + the vector sequences must be less than 5 kb for effective IPCR. The resulting fragments, following restriction digestion (for left end fragment, e.g. *Bam*HI, *Sac*I, *Sma*I, *Sal*I, *Sph*I; for right end fragment, e.g. *Sal*I, *Sma*I) contain the vector + the end fragment and can be self ligated. These self ligated circular fragments can be amplified with the oppositely oriented neighbouring primers (designed from the vector sequences alone) by IPCR. The IPCR is easier, less time consuming and does not require any subcloning procedure.

2.6.1. Plasmid rescue

Right end fragments of the cloned DNA were generated by digesting the BAC clones with enzymes such as, *Sac*I, *Bam*HI, *Sph*I and *Not*I for ~6 hours at 37°C (since the sites of *Not*I and *Sal*I flank the cloning site digesting clones with these enzymes release the vector from the insert, so also does *Hin*dIII). The digested DNAs were loaded side by side on a 1% agarose gel in 0.5 x TBE and electrophoresed (2 V/cm) for 12 to 15 hrs. The gel was Southern blotted onto

positively charged nylon membrane (Boehringer Mannhiem) and hybridised with ^{32}P labelled pBeloBAC11 DNA for 12 to 15 hrs as detailed in section 2.5. The hybridisation patterns were compared relative to the *Not*I control and the digestion that produced vector + end fragment slightly larger (1-5kb) than the *Not*I control was then self ligated as follows.

The digested DNA was ethanol precipitated (~50ng) and ligated using 40 units of T4 DNA ligase (New England Biolabs) in a total volume of 20 µl at 16°C for 12-15 hrs. The ligated DNA (1µl) was electroporated into DH10B electro-competent cells and transformants were selected on LB/chloramphenicol plates. Mini preparations of plasmid DNA were made from ~10 colonies and the right end fragment was verified by digesting with the selected enzyme and *Hin*dIII (cloning site). Probes were generated from plasmids containing the right ends by PCR amplification with M13 forward sequencing primer (5'-GTAAAACGACGGCCAGT-3') and the reverse primer BAC4 (5'-CACTATAGAATACTCAAGC-3'). The reactions were initiated by denaturing at 94°C for 4 min followed by 94°C, 1 min 30 s - denaturation; 55°C, 2 min - annealing and 72°C, 3 min - primer extension, for 30 cycles. In the last cycle, primer extension step was done for an additional 7 min. PCR amplified products were purified from the gel (1% LMP agarose Gibco-BRL) by digesting with β-agarase.

2.6.2. Inverse polymerase chain reaction (IPCR)

To isolate the left ends of insert DNAs, BAC clones were digested with *Ava*I, *Bam*HI, *Eco*RI, *Kpn*I, *Sac*I and *Sma*I for 6 hrs at 37°C. DNA was ethanol precipitated and self ligated in a total volume of 20 µl as described in section 2.6.1. The circularised DNA fragments were used as templates for IPCR. The ligated mixture (1µl) was used for an IPCR reaction in a total volume of 25 µl

containing 20 mM Tris-HCl (pH 8.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 μM primers, 60 μM dNTPs, 0.1mg/ml detergent and 1.25 units of *Taq* DNA polymerase. The primers were BAC1: 5' CTGCAGGCATGCAAGC; BAC2: 5' GTCGACTCTAGAGGATC 3'). The reactions were carried out in the same way as detailed in section 2.6.1 except for an annealing temperature of 60°C. Left ends of the insert DNAs were verified by digesting the IPCR products with *Hin*dIII and the enzyme used for creating the restriction fragments and loaded on a gel side by side with uncut IPCR products. The IPCR product was gel purified and used as probes.

Left end fragment isolation was also carried out by plasmid rescue. The clones were digested with *Sfi*I and purified by ethanol precipitation. The cut ends were blunted by T4 DNA polymerase (0.5 units of enzyme, 4 μl of 5 x reaction buffer, 1 μl of 2 mM dNTPs in a total volume of 20 μl and incubated at 37°C for 5 min) and the DNA was again purified by ethanol precipitation. The cut ends were blunt ligated in a reaction mix (20 μl) containing 100 units of T4 DNA polymerase (16 hrs, 16°C). The ligated DNA was electroporated as detailed in section 2.1.2.2. Plasmid DNAs were prepared and the subcloned fragments were sequenced (detailed in section 2.1.5.2) using T7/M13 promoter sequences. PCR primers were designed from the sequenced fragments and amplified products were used as probe.

2.7. Long range PCR

Expand™ Long template PCR system (Boehringer Mannheim) was used to amplify fragments of larger size (~15 kb). The system is composed of a unique enzyme mix containing thermostable Taq and Pwo DNA polymerases, of which the latter enzyme has a proof reading capacity (3' exonuclease activity). Thus Pwo

polymerase can reduce the error frequency by a factor of 10 as compared to *Taq* DNA polymerase.

The reaction mix was prepared by mixing two separate master mixes of 25 µl each. This circumvents the need of a hot start and ensures successful amplification. Master mix 1 composed of dNTPs, primers and distilled water and master mix 2 contained buffer, enzyme mix and distilled water. The reaction was carried out in a total volume of 50 µl, consisting of 300 nM of each primer, 350 µM of each dNTP, 5 µl of 10 x buffer 1 (17.5 mM $MgCl_2$; 500 mM Tris-HCl, pH 9.2, 160 mM $(NH_4)_2SO_4$), 0.75 µl of enzyme mix and 100 ng (2 µl) template DNA. PCR amplifications were carried out in a Perkin-Elmer Gene Amp-9600 thermal cycler. The reaction was initiated by denaturing the template DNA for 2 min at 92°C followed by 10 cycles of 92°C, 10 s denaturation, 55°C, 30 s annealing and 68°C, 8 min elongation. This was followed by 15 cycles of 92°C, 10 s denaturation, 55°C, 30 s annealing and 68°C, 8 min elongation plus cycle elongation of 20 s for each cycle (i.e. cycle 11 has an elongation time of 8 min 20 s, 12 has 8 min 40 s and so on) and a final elongation time of 7 min at 68°C. When the reaction was completed an aliquot (3-5 µl) was mixed with 5 µl of gel loading buffer and electrophoresed on 1% agarose gel in 1 x TAE buffer.

Chapter 3

Murine casein locus - Isolation and characterization of clones from large fragment murine genomic libraries.

3.1 Introduction

Mouse milk contains five types of caseins, α, β, γ, ε and κ for all of which cDNA clones have been isolated and sequenced. In other species such as cattle, sheep, goat and pigs only four types of caseins have been identified so far, αS1, αS2, β and κ (reviewed by Mercier and Vilotte, 1993). The sequences of the various casein cDNAs exhibit varying degrees of similarity between each other in the same species and between the putative homologous caseins of different species (shown in section 1- Tables 1.3 to 1.7). In bovines, it has been reported that the casein genes are arranged in a linear array on a 200 kb genomic DNA fragment (Ferretti et al., 1990; Threadgill and Womack, 1990). In mice, the casein loci (α, β and γ) have been assigned to chromosome 5 (Gupta et al., 1982). Even though two more casein mRNAs (ε and κ) have been identified from murine mammary gland (Hennighausen and Sippel, 1982a), so far there are no reports on the identification of the corresponding genes or their chromosomal locations. Moreover, the existence of the ε-casein gene remains controversial or at least unconfirmed.

Therefore, the present study is focused on constructing a map of the murine casein locus by restriction analysis and by contig mapping. The map should allow 1) the identification of the genes for all known murine caseins 2) confirmation of the existence of ε-casein gene 3) determination of the order and arrangement of the casein genes in the locus and 4) assignment of the ε and κ-casein genes to specific mouse chromosomes. Large fragment murine genomic libraries are an essential resource for these studies and in particular for the construction of a contig map(s) around the casein genes. Since, one of my

supervisors provided a bacteriophage P1 library, I decided to screen the library, even though the genome coverage was low (1.0). Unfortunately, the casein genes were not represented in this P1 library. Therefore, I switched to a commercially available bacterial artificial chromosome (BAC) library which has a genome coverage of 7.37. This chapter describes the isolation and characterization of large fragment murine genomic clones containing the casein genes. As an initial step, I validated the various casein probes obtained from different sources.

3.2. Results

3.2.1. Validation of murine casein probes.

3.2.1.1. Northern blot analysis.

Total RNA isolated from mid-lactating mouse mammary gland and liver (negative control) was electrophoresed on a formaldehyde gel with suitable RNA size markers (ranging from 0.2 kb to 7.5 kb). The RNA was Northern blotted to nylon membranes. Radioactive hybridisation of the Northern blot with the murine α, β, γ, ε and κ casein probes (listed in Table 2.1 in section 2, the ε-casein probe was prepared by PCR amplification, see Table 2.2) showed that each casein probe hybridised to mRNAs of different size. No cross hybridization was observed. None of the probes hybridised to the liver RNA. The sizes of the mRNAs hybridised to the various casein probes are shown in Table 3.1.

3.2.1.2. Sequence analysis of the casein probes

Sequence analyses were carried out on the murine probes prepared from cDNA and genomic DNA clones. Partial DNA sequences were determined for each casein probe by Sanger's dideoxy sequencing technique in order to confirm the identity of the genes/probes. About 250 base pairs of each probe were sequenced. As the probe sequences for α, β and κ (1100bp, 6.6 kb and 900bp

Table 3.1. Murine casein mRNA sizes.

Casein probes used	mRNA size observed (kb)	Reported mRNA size (kb)
α	1.45	1.45[a], 1.6[b]
β	1.20	1.15[a], 1.45[b]
γ	1.00	0.95[a], 0.88[b]
ε	0.89	0.86[b]
κ	0.95	0.84[c]

Identified by Northern blot analysis.

[a]Gupta et al., (1982). [b]Hennighausen et al.,(1982).

[c]Thompson et al., (1985).

respectively) were cloned in the multiple cloning sites of pUC19, pGEM Blue Script and pBS-KS vectors, I used the universal forward primer for the sequencing of these probes/clones. As the γ-casein cDNA (900 bp) is cloned in pBR322 sequencing was carried out with a primer designed from the cDNA sequences available in the data base (EMBL: D10215, 5'-GTGAGCATTTGGGAAGGTGT-3'). The sequences obtained were compared with the mRNA/cDNA sequences available in the data base. Sequence comparisons revealed a complete identity with the published murine mRNA/cDNA sequences except for the α-casein probe. The sequences of the α-casein probe showed 100% identity to the rat α-casein cDNA sequences and 88% identity to the mouse α-casein cDNA. Southern blot analysis with this α-casein probe also gave some nonspecific hybridisation. Hence, the mouse α-casein cDNA clone (800 bp) was obtained afresh from Dr. Jeffrey Rosen. Even though the cloning site was *Pst*I, there was some uncertainty about the identity of the vector in which the α-cDNA sequences were cloned. Therefore, α-specific primer sequences were also designed from the published cDNA sequences (Forward- 5'-TGACTGGACCCTCCATTCTC-3') for the confirmatory sequence analysis. Following sequence analysis 100% identity was observed with the published mouse α-casein cDNA sequences.

Epsilon-casein probes (specific for 5' and 3' regions) were generated by PCR amplification of a 7 day lactating mouse cDNA library using primers designed from the published cDNA sequences (see Table 2.2 in section 2). 3' specific probes were also generated for all other casein genes by PCR amplification of genomic DNA (see Table 2.2 in section 2). The probes generated from the validated clones and the PCR amplified (5' and 3' specific) fragments were used in all subsequent hybridisation experiments.

3.2.2. Analysis of the bacteriophage P1 library

The bacteriophage P1 library was kindly provided by Dr. Linda Mullins, Centre for Genome Research, University of Edinburgh. The library contained ~40,000 clones and was organized as secondary and primary pools (refer to Fig. 2.2 in section 2).

3.2.2.1. P1 library screening

Mini prep DNA from the secondary and primary pools were used for PCR based screening. Primers designed from the 3' untranslated region of α, γ, ε and κ casein cDNA sequences and intron 8 of β-casein gene (see Table 2.2) were used to search the library by PCR for the corresponding gene sequences. The secondary pools were named alphabetically from 'a to j' and the primary pools were numbered from 1 to 10. Secondary pool 'b' and primary pool '9' appeared to be positive for the presence of α-casein sequences since both the pools yielded a PCR product of the correct size with the α-casein primers (Fig. 3.1).

After identifying the positive pool, the selected pool was plated out on a LB/kanamycin/sucrose plate (1500-2000 colonies). The positive clones were identified by a colony lift and hybridisation technique using the α-casein cDNA probe. Two positive clones were identified (named as P1CL-1 and P1CL-2). DNA was prepared from both the isolated clones. The presence of α-casein sequences was further verified by PCR amplification of the purified clone DNA with the same primer pairs used for the library screening (Fig. 3.1). A double digestion carried out on P1CL-2 with restriction enzymes *Not*I and *Sal*I showed that the cloned insert had a size of ~80 kb.

3.2.2.2. Restriction analysis of P1 clones.

The P1 DNA prepared from P1CL-1 and P1CL-2 were restriction digested with the enzymes *Eco*RI, *Hin*dIII, *Pst*I, *Bgl*II, *Eco*RV, *Bam*HI, *Not*I, *Sac*I, *Pst*I, *Xba*I and *Apa*I (Fig. 3.2). The digested DNA was Southern blotted and probed with the α-casein cDNA probe. The hybridisation signal was very faint and did not produce any detectable fragment on the film even after overnight exposure. Since the DNA prepared from the clones contains large number of copies of the same region, if the region contains the gene corresponding to the probe, it is expected to give a strong hybridisation signal which could be visualised after short exposures (few minutes or hours). However, faint hybridisation bands were only observed after prolonged exposure (for 7 days). This might be due to some limited sequence similarities between the probe and the DNA sequences of the clones.

Thus, even though initially two P1 clones appeared to be positive for the α-casein gene, further analysis suggested that the product amplified in PCR was an artefact. Alternatively, clone instability and loss of the α-casein gene specific sequences may account for this inconsistency.

3.2.3. Analysis of murine BAC library

The BAC library was supplied by Research Genetics, Inc, USA, consisting of 221,184 clones, organized in 72 superpools, having a genome coverage of 7.37. Each super pool represents arrayed colonies pooled from eight 384 well plates. Each plate pool was made by pooling the 384 colonies from a single plate. The superpools and plate pools were supplied as purified DNA preparations. Once positive plate pool/pools were identified, the specific plates containing 384 clones were obtained from Research Genetics, USA, for further isolation of the clone/clones.

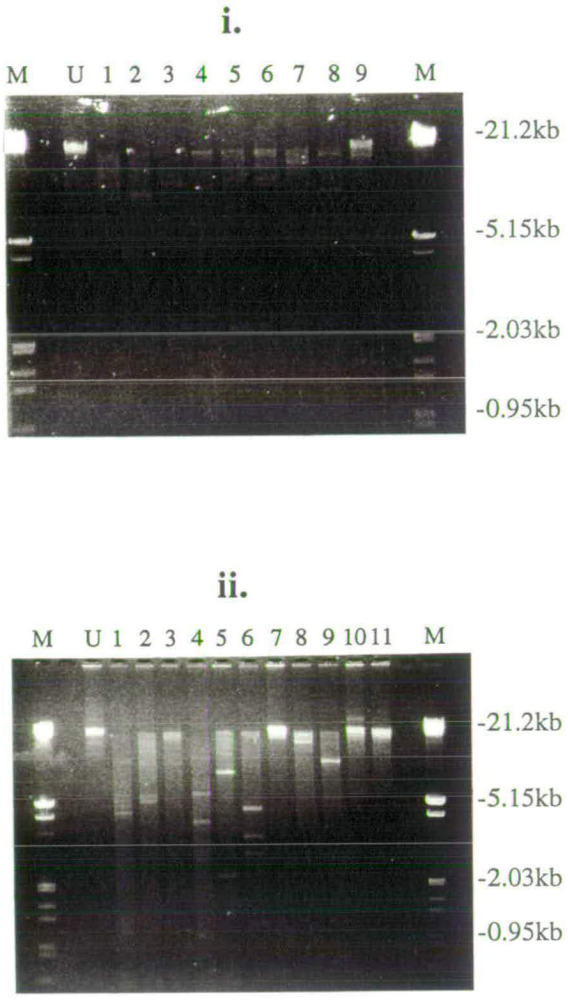


Fig. 3.2.

Ethidium picture of restriction digested P1 clones,

i. P1CL-1, digested with 1-*Eco*RI, 2-*Hin*dIII, 3-*Pst*I, 4-*Bgl*II, 5-*Eco*RV, 6-*Bam*HI, 7-*Not*I, 8-*Sac*I, 9-*Pst*I,

ii. P1CL-2, digested with 1-EcoRI, 2-HindIII, 3-PstI, 4-BglII, 5-EcoRV, 6-*Bam*HI, 7-*Not*I, 8-*Sac*I, 9-*Pst*I, 10-*Xba*I and 11-*Apa*I.

M - is the λ *Hin*dIII/*Eco*RI marker and U- is the uncut DNA

3.2.3.1. BAC library screening

The DNA from super pools and plate pools of mouse BAC library were screened by PCR amplification. The primers used were the same as described in section 3.2.2.1. The identified positive pools are shown in Table 3.2. A few plates which appeared to cover the entire locus with overlapping clones (shown in Table 3.2) were selected for clone isolation by colony transfer and hybridisation technique. The frozen individual plates containing arrayed single colonies were obtained from Research Genetics, USA. The colonies were transferred from the plate into a second 384 well plate in which each well containing 60 µl of LB/chloramphenicol (to make a back up copy) and onto the nylon membrane placed on LB/chloramphenicol plates using a sterile 384 well replica plating device in duplicate. The colonies in plate and on membranes were grown overnight at 37°C. None of the replicas yielded all 384 colonies. Varying number of colonies were found missing in the replica made from each plate (missing clones range from 10-129 colonies). Repeated examination confirmed that it was not due to any defect in the transfer process, but due to the non viability of the bacteria in the microplate wells.

Even though 129 clones were missing from plate 156, which appeared positive for κ-casein gene, I could identify a positive clone for κ-casein gene (156 J3). While from the plate 346, which appeared positive for γ and ε-casein genes, ~20 clones were missing and none of the clones was found positive for either γ or ε gene. The difference in observation regarding plate 346 was discussed further with Research Genetics. After examining the master plate, I was informed that out of the 20 clones only 4 were found viable in the master plate too. The 4 viable clones were provided as bacterial stabs for further examination. But even after screening those 4 clones, no positive clone (for either γ or ε) could be identified from plate 346. Therefore, the PCR products amplified from the pooled DNA

Table 3.2. Positive pools identified from the BAC library.

Super pool No.	Casein genes present α	β	γ	ε	κ	Plate pool No.	Casein genes present α	β	γ	ε	κ
2					*	9 - 16					no positive
6	*	*				41 - 48	no positive				
8			*	*		57 - 64			no positive		
10			*	*		73 - 80			no positive		
12			*	*		89 - 96			no positive		
15					*	113 - 120					no positive
20					*	153 - 160					156[S]
24			*	*		185 - 192			185[S]	185	
26	*	*				201 - 208	201[S]	201			
44			*	*		345 - 352			346[S]	346	
46		*				361 - 368	365[S]	365			
50	*	*	*	*		393 - 400	400	400	394[S]	394	
51					*	401 - 408					402
52	*	*				409 - 416	411	411			
53					*	417 - 424					418[S]
55	*	*				433 - 440	436	436			
56			*	*	*	441 - 448			447[S]	no positive	
57			*	*		449 - 456			no positive		
58	*	*	*			457 - 464	461	461,459[S]	459		
60			*			473 - 480			477		
62	*	*	*			489 - 496	490[S]	490	490		
68	*					537 - 544	540				
70			*	*	*	553 - 560			555[S]	555	555

Pools were identified by PCR screening. Each super pool corresponds to 8 plate pools. * Represent the presence of right sized PCR products corresponding to each gene. [S]- represent the plates selected for clone isolation by colony transfer and hybridisation. Additionally plate 280 was also screened to identify clones containing fragments extending from the 5' region of α-casein gene.

from plate 346 might be either due to the presence of γ and ε in the missing clones (i.e. if all clones were viable when the pool DNA was prepared) or due to the false priming from a region other than the intended priming sites. Examples of the lack of any positive plate pool from a positive super pool or lack of any positive clone from a positive plate was observed throughout the screening process (Table 3.2 and 3.3).

In total, nine clones were isolated from the BAC library after screening the individual plates containing arrayed clones (Table 3.3). More than one casein gene was found in several individual clones. Clones containing α + β; β + γ and γ + ε, genes revealed the organization of the four genes as α– β– γ– ε. Two other clones consisting of three genes, α, β, γ and γ, ε, κ confirmed this organization and also allowed the κ gene to be mapped in relation to the other casein genes. Thus, from the isolated clones I could deduce the organization of the genes in the locus as α–β–γ–ε–κ.

3.2.3.2. Generation of end probes from BAC clones

Clone 156 J3 (~98 kb) contained only the κ-casein gene and probing with an end fragment (see 3.2.3.2 below) revealed that the gene is situated at one end of the cloned insert. The clone was examined for the presence of other known casein genes and was found to be negative. Comparing this clone with clone 555 N16 (~105 kb) containing the γ, ε and κ gene also showed that the κ- gene is situated at one end of the locus with clone 156 J3 containing ~85 kb of flanking region (detailed in section 3.2.3.4). To identify more clones in the region beyond the α-casein gene, primers were designed from the right end fragment (5' upstream region of α-casein gene) of clone 365 E21 (forward- 5' ATCCAGCTCAAGGGGAAGTTCC 3', reverse- 5' GAGAGGCATGAGAGGGGTTGAA 3'). Another round of screening the library

Table 3.3. Clones isolated from the BAC library.

Plates Screened	clones identified	genes present				
		α	β	γ	ε	κ
365	365 E21	*	*			
201	201 E9	*	*			
459	459 N19		*	*		
185	185 O18			*	*	
394	394 C5			*	*	
490	490 H23	*	*	*		
555	555 N16			*	*	*
156	156 J3					*
280	280 M14	beyond 5' of α				
346	no positive					
418	no positive					
447	no positive					

The clones were isolated by PCR screening, colony transfer and hybridisation. * - represents the presence of corresponding gene in the respective clone.

allowed the isolation of a positive clone, 280 M14, which was also found negative on probing with the casein probes.

The left and right end fragments of BAC clones were generated by IPCR and plasmid rescue (for details refer to section 2.6). The end probes of each BAC clone facilitated the identification of the respective end fragments after restriction digestion. Thus, (detailed) maps of each analysed clone have been constructed. BAC1 and BAC2 primers were used to generate left end probe by amplifying fragments of restriction digested (490 H23- *Sph*I, 555 N16- *Sac*I, 365 E21- *Sac*I, 201 E9- *Bam*HI, 185 O18- *Sma*I, 156 J3- *Sac*I) self-ligated left end fragments. The restriction digested (490 H23- *Sph*I, 555 N16- *Sph*I, 365 E21-*Bam*HI, 201 E9-*Bam*HI) self-ligated right end fragments were sub-cloned. Plasmid DNA prepared from the subclones were PCR amplified using M13 and BAC4 primers. Thus, both right and left end probes were generated by PCR amplification of the respective end fragments.

To carry out plasmid rescue and further PCR amplification, the end fragment size should be less than 5 kb. Wherever that was not possible, e.g. clones 185 O18, *Sac*I digested self-ligated (~7 kb insert) and 156 J3, *Sph*I digested self-ligated (~10 kb insert) ~350 base pairs of the subclones were sequenced using SP6 promoter sequences. Primers (Table 3.4) were designed from the sequences and subsequent PCR amplified products were used as right end probes. T7 and SP6 oligos (Table 3.4) (promoter sequences flanking the left and right ends of the cloning site respectively) were also used to probe restriction digests of clone DNA in certain cases to confirm the end fragments.

The probes generated were sequenced and compared with the sequences available in the database. The right and left end probes of clone 201 E9 showed similarity to many murine repeat family sequences and resulted in nonspecific

Table 3.4. Primer pairs used to amplify BAC end fragments (right and left, wherever the end fragment size was more than 5 kb).

Name of BAC sub-clone	Primers	Product size (bp)
185 O18 *Sac*I (right end)	F- GCTCTGTCCAACCTTTGGCTGT R- CCAGCCAAGTGAGTGTGAGGAA	143
156 J3 *Sph*I (right end)	F- GCATCTGCAGAAACCTATTACATG R- TTCATTCGGATCATTCAGTGACC	183
490 H23 *Sfi*I (left end)	F- AAGAAATTCTTCCTGCTCATTCA R- GTTGCCTGAATTAATCTCAAGTTG	105

Promoter sequences used for sequencing BAC ends

Name of BAC end	Name of promoter	Sequence
Right end	SP6	ATTTAGGTGACACTATAG
Left end	T7	TAATACGACTCACTATAGGG

hybridisation. The problem was solved by treating the labelled probes with mouse COT1-DNA (detailed in section 2.5)

The left end fragment of clone 490 H23 (*Sfi*I digested and self ligated, for details refer to section 2.6) was also generated by plasmid rescue and sequenced using T7 promoter sequences. Primers designed from the sequences (Table 3.4) were used for long range PCR, in attempts to orientate the ε-casein gene.

3.2.3.3. Analysis of clones 490 H23 and 555 N16

The clones, 490 H23 and 555 N16 (containing α, β, γ and γ, ε, κ genes respectively) were mapped with a few selected restriction enzymes (*Not*I, *Sal*I, *Sma*I, *Sfi*I, *Xho*I, *Pvu*I and *Bam*HI). Since *Not*I does not have a recognition site within the cloned mouse sequences and has sites in the vector flanking the cloning site, it cuts the insert out of the vector as a single fragment which helped to assess the size of the cloned fragment. The inserts of clones 490 H23 and 555 N16 are of ~155 kb and ~105 kb respectively. Restriction digestion and Southern blot analysis of the clones are shown in Fig. 3.3.a, b, c. Clone 490 H23 contains complete α, β and γ genes while clone 555 N16 contained only the 3' region of γ-casein gene. This was evident from the *Bam*HI fragment present in the clones and genomic digest, two fragments (40 and 25 kb) in α, one (40 kb) in β and two (25 and 10 kb) in γ casein genes.

The end fragments of each clone were identified by probing the digested clones with probes generated from the left and right ends of the respective clone. Thus a detailed map of the clones (490 H23 and 555 N16) (Fig. 3.4) was generated and positions assigned for the five different casein genes. The α and β genes are located close to each other with an inter-genic region of ~8 kb and ε gene is located ~20 kb away from γ gene. The distance between β and γ is ~55 kb and between ε and κ is ~40 kb. The map shows that the murine casein locus is

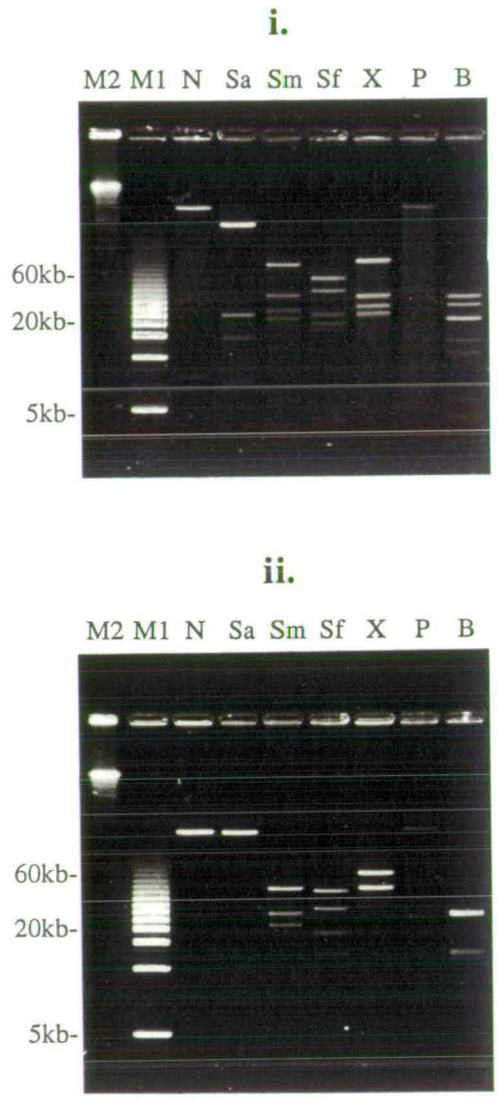


Fig. 3.3.a.
Photograph of ethidium stained gel containing restriction digested clone **i.** 490 H23. **ii.** 555 N16.
Electrophoresis was performed in a CHEF DR-III system (Bio-Rad) for 12.5 hrs, 14°C in 0.5 x TBE at 6 V/cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively. The enzymes used were the same for both the clones and are indicated above each lane as follows: N-*Not*I, Sa- *Sal*I, Sm-*Sma*I, Sf-*Sfi*I, X-*Xho*I, P- *Pvu*I, B- *Bam*HI. M1 and M2 are molecular size markers, 5 kb and 39 kb ladders respectively.

Fig. 3.3.b. Southern blot analysis of clone 490 H23 which contains the α, β and γ genes. The probes used were cDNA (α and γ) and the - intron 8 PCR (β). Fragments which hybridised with more than one probe are marked with common superscripts.

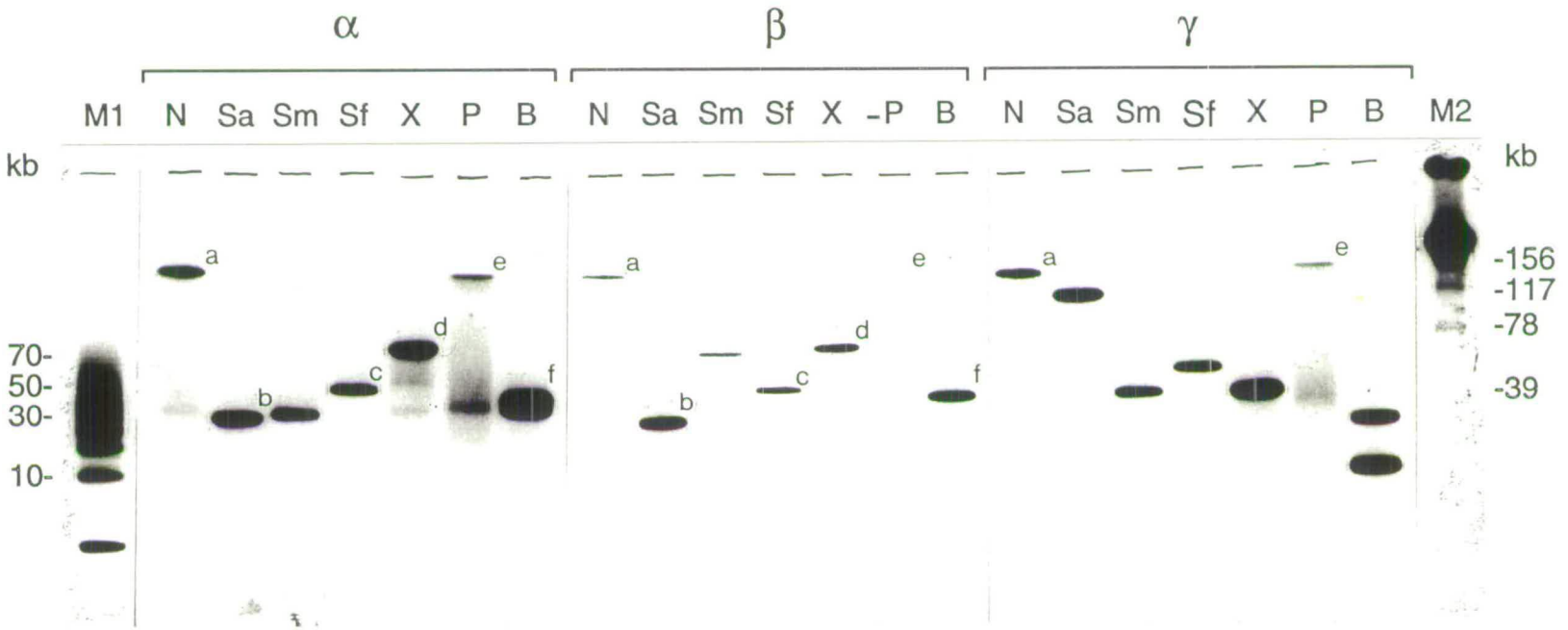
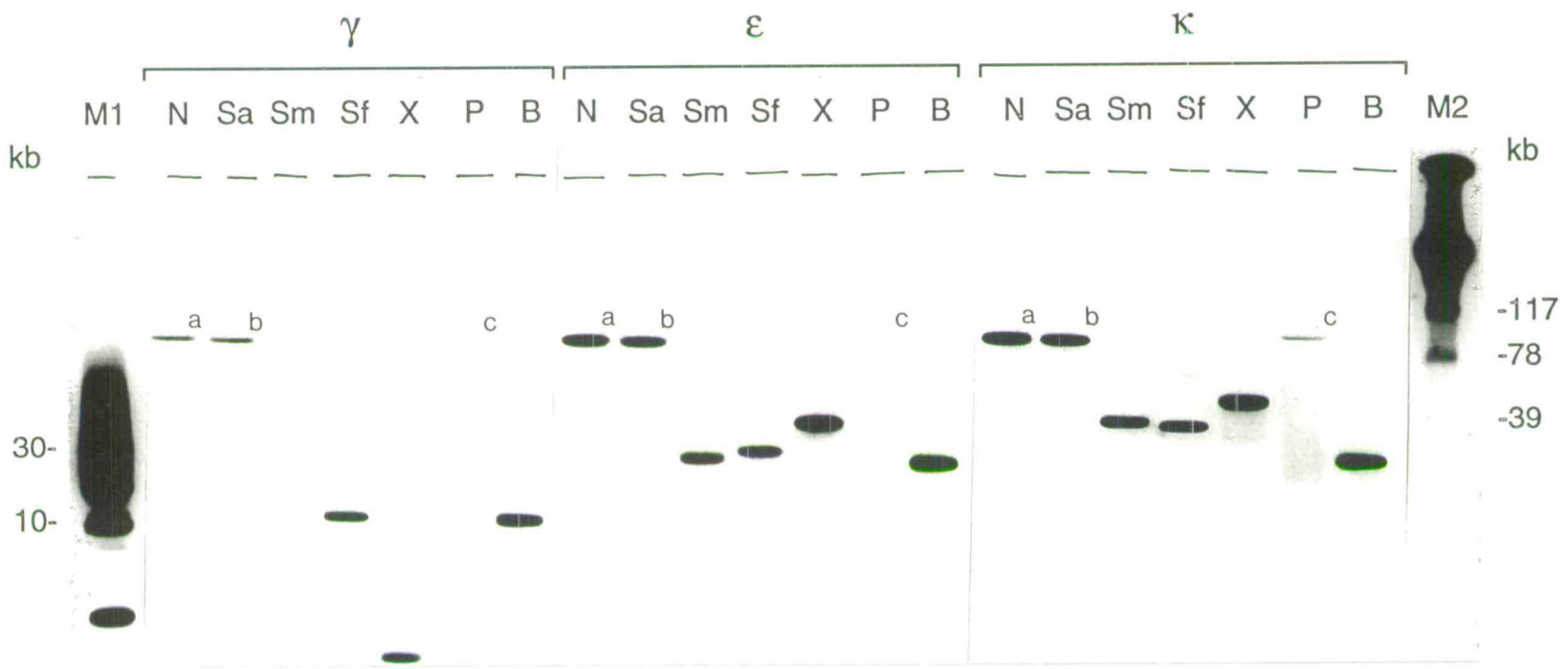

α
β
γ
M1
N
Sa
Sm
Sf
X
P
B
N
Sa
Sm
Sf
X
-P
B
N
Sa
Sm
Sf
X
P
B
M2
kb
kb
70-
50-
30-
10-
a
b
c
d
e
f
-156
-117
-78
-39

Fig 3.3b

Fig. 3.3.c. Southern blot analysis of clone 555 N16 which contains the γ, ε and κ genes. cDNA - γ and κ and 5' PCR (cDNA) - ε probe were used. Fragments hybridised with more than one probe are marked with common superscripts. The absence of any fragment in lane 3 (*Sma*I digestion) when probed with the γ-cDNA was due to the small size (<2 kb) of the hybridising fragment which ran out of the gel under the electrophoretic conditions described (The fragment was identified under suitable electrophoretic conditions).


γ
ε
κ
M1
N Sa Sm Sf X P B
N Sa Sm Sf X P B
N Sa Sm Sf X P B
M2
kb
kb
a b c
a b c
a b c
-117
-78
-39
30-
10-

107c

Fig 3.3c

Fig. 3.4. Restriction map of the clones 490 H23 and 555 N16 showing the organization of the genes in the murine casein locus. The dotted line (probes) below the maps indicate the limits of hybridisation of the cDNA or other probes. The solid bars (genes) below the maps indicate the maximum size of each gene i.e the restriction fragments to which the corresponding probes hybridise. V- vector site

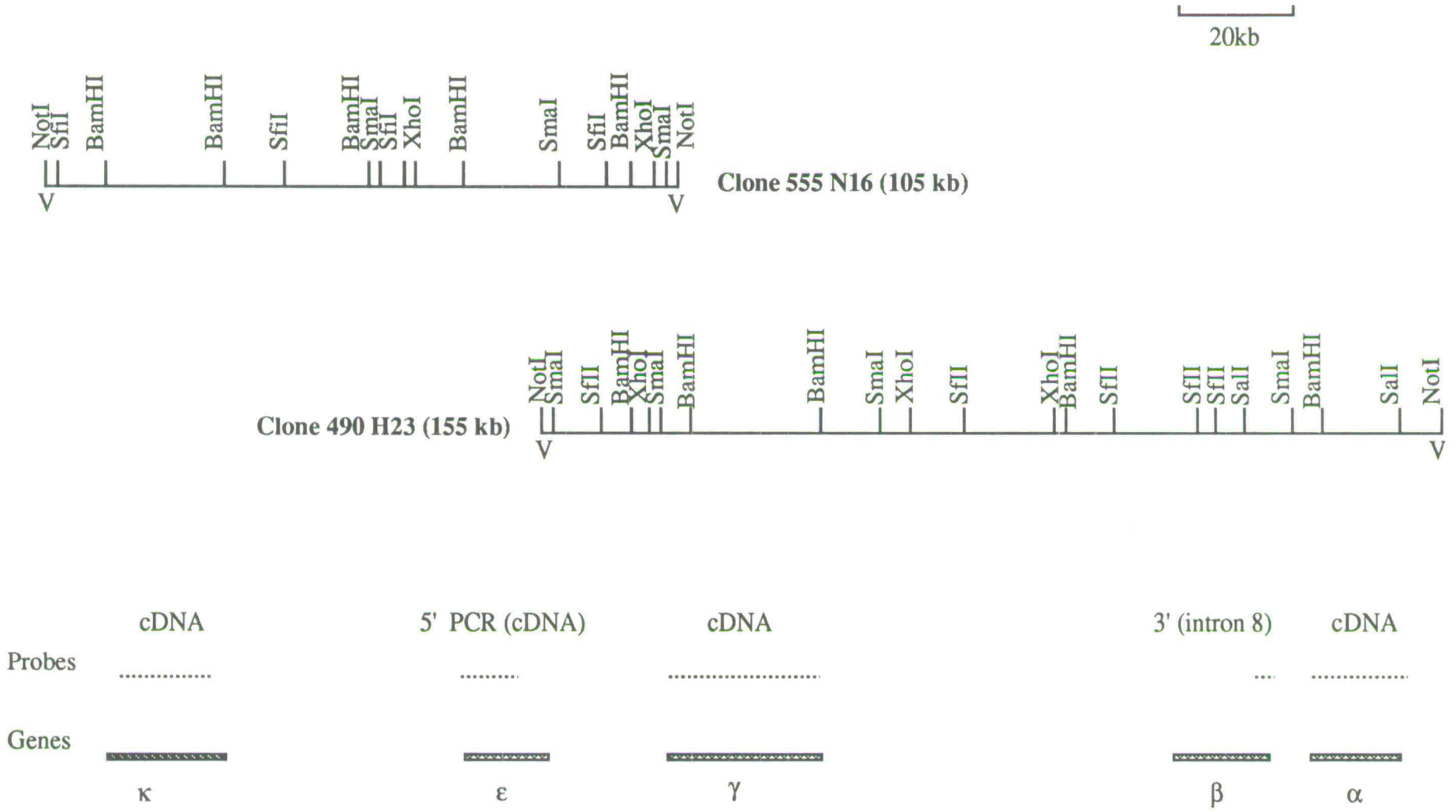

20kb
NotI
SfiI
BamHI
BamHI
SfiI
BamHI
SmaI
SfiI
XhoI
BamHI
SmaI
SfiI
BamHI
XhoI
SmaI
NotI
V
V
Clone 555 N16 (105 kb)
Clone 490 H23 (155 kb)
NotI
SmaI
SfiI
BamHI
XhoI
SmaI
BamHI
BamHI
SmaI
XhoI
SfiI
XhoI
BamHI
SfiI
SfiI
SfiI
SalI
SmaI
BamHI
SalI
NotI
V
V
cDNA
5' PCR (cDNA)
cDNA
3' (intron 8)
cDNA
Probes
Genes
κ
ε
γ
β
α

Fig. 3.4

contained in a 250 kb fragment and the genes are arranged in the order α–β–γ–ε–κ, fixing α and κ at the ends of the locus.

3.2.3.4. Orientation and Contig map of the clones

As well as clones 490 H23 and 555 N16, clones, 365 E21, 201 E9 (both ~115 kb and contain α and β genes), 185 O18 (~80 kb, γ and ε genes) and 156 J3 (~98 kb, κ gene), were mapped using the enzymes described earlier. A contig map of the locus is shown in Fig. 3.5. This analysis also served to orientate the various genes within the casein locus.

1. α-β orientation

Two *Bam*HI fragments of ~40kb and 25 kb were identified by the α cDNA probe while the 5' and 3' specific β probes and the 3' specific α probe hybridised to an identical 40 kb fragment (Fig. 3.6.a, b). In addition, an identical *Sal*I fragment of ~25 kb was identified by the α cDNA, α-3' and β-3' (intron 8) probes. The β-5' probe identified different sized fragments in various clones. This was due to the absence of other *Sal*I sites in the clones containing α and β genes (Fig. 3.5) (but the cloning vector, pBeloBAC11 has two *Sal*I sites on either side of the cloning site, refer to Fig. 2.4 in section 2). Thus, the 40 kb *Bam*HI fragment containing the β-casein gene and the 3' region of α-casein gene and the 25 kb *Sal*I fragment containing the α-casein gene and the 3' region of β-casein gene indicated a tail to tail orientation of the two genes.

Long range PCR experiments were performed in the clones containing the α and β genes to determine the relative orientation of these genes. Various primer combinations, α-3' forward + β- intron 8 reverse; α-3' reverse + β- intron 8 forward, α-3' forward + β- intron 8 forward and α-3' reverse + β- intron 8 reverse (refer to Table 2.2 for primer sequences) were used in PCR experiments. The

Fig.3.5. Contig map of 7 BAC clones covering the murine casein locus containing all five known casein genes. The *Xho*I (X) and *Sal*I (S) sites are shown. The results of the partial *Xho*I digests of genomic 129 DNA were consistent with the *Xho*I sites mapped in these seven BAC clones.

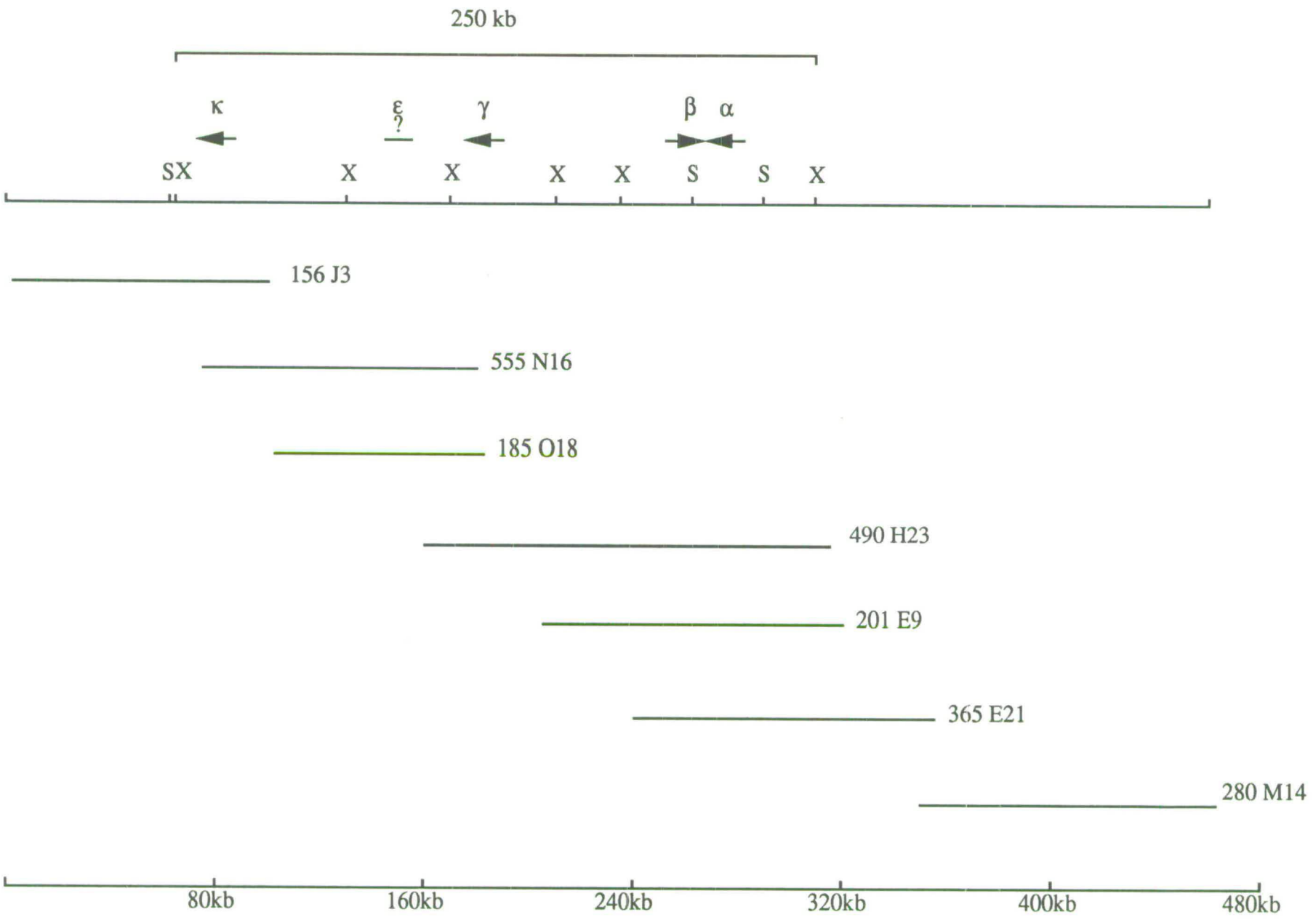

250 kb
κ
ε
?
γ
β
α
SX
X
X
X
X
S
S
X
156 J3
555 N16
185 O18
490 H23
201 E9
365 E21
280 M14
80kb
160kb
240kb
320kb
400kb
480kb

Fig. 3.5

Fig. 3.6.a. Restriction analysis of clone 365 E21, containing α and β-casein genes.

i. Photograph of ethidium stained gel. Electrophoresis was performed in a CHEF DR-III system (Bio-Rad) for 12.5 hrs, 14°C in 0.5 x TBE at 6 V/cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively. The enzymes used are indicated above each lane as follows: N- *Not*I, Sa- *Sal*I, Sm- *Sma*I, Sf- *Sfi*I, X- *Xho*I, P- *Pvu*I, B- *Bam*HI. M1 and M2 are molecular size markers, 39 kb and 5 kb ladders respectively. Smearing of the lanes was due to bacterial DNA contamination, which did not interfere with hybridisations.

ii. Southern blot was hybridised with the α (cDNA) and β (5' end, PCR) probes.

iii. Same blot was reprobed with the α (3' UT) and β (intron 8) PCR probes.

*- represent the *Sal*I and *Bam*HI fragments compared in orientating the α, β-casein genes.

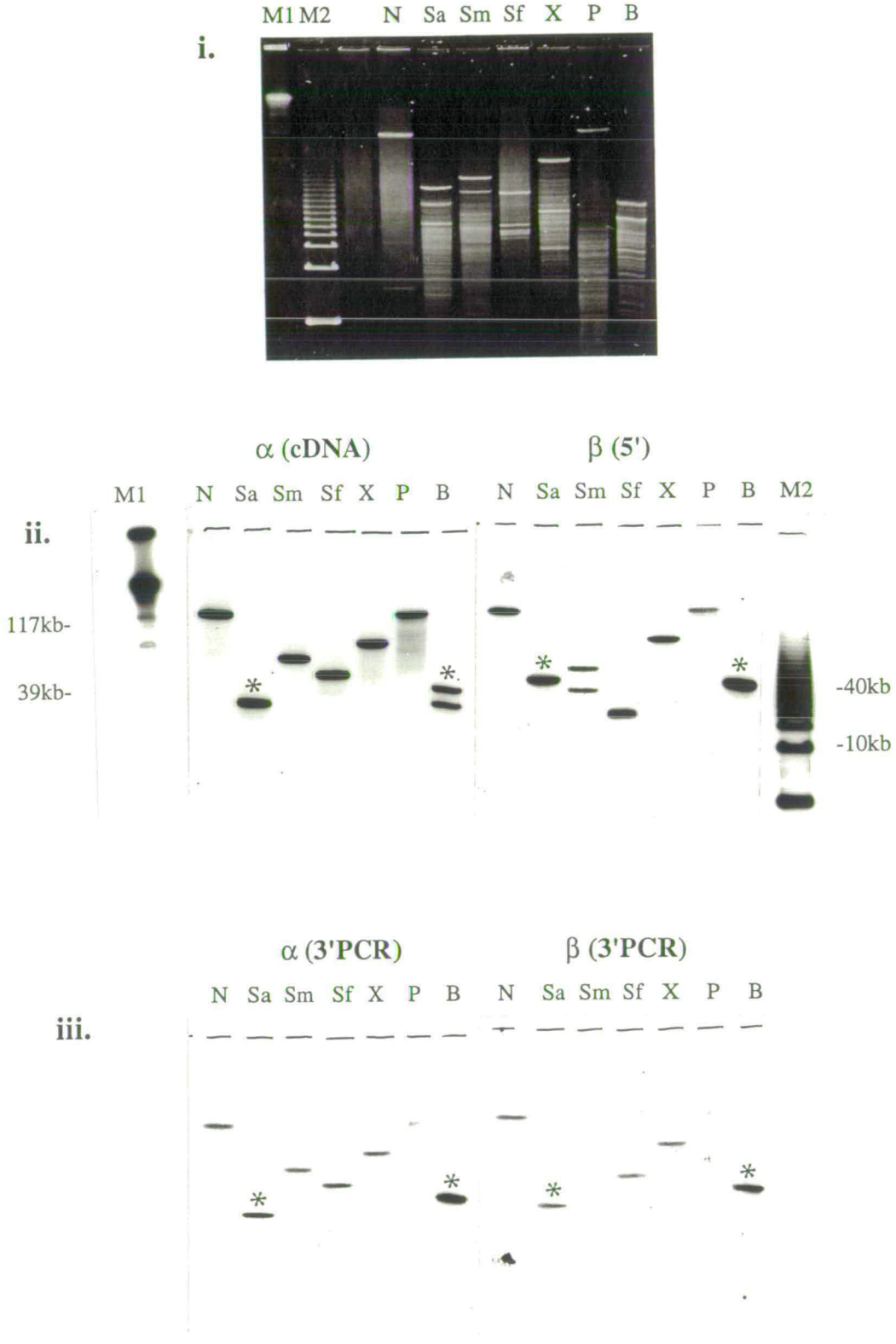

i.
M1 M2 N Sa Sm Sf X P B
ii.
α (cDNA)
β (5')
M1 N Sa Sm Sf X P B N Sa Sm Sf X P B M2
117kb-
39kb-
-40kb
-10kb
iii.
α (3'PCR)
β (3'PCR)
N Sa Sm Sf X P B N Sa Sm Sf X P B

Fig. 3.6.a

Fig. 3.6.b. Restriction analysis of clone 201 E9, containing the α and β-casein genes.

i. Photograph of ethidium stained gel. Electrophoresis was performed in a CHEF DR-III system (Bio-Rad) for 12.5 hrs, 14°C in 0.5 x TBE at 6 V/cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively. The enzymes used are indicated above each lane as follows: N- *Not*I, Sa- *Sal*I, Sm- *Sma*I, Sf- *Sfi*I, X- *Xho*I, P- *Pvu*I, B- *Bam*HI. M1 and M2 are molecular size markers, 39 kb and 5 kb ladders respectively. Smearing of the lanes was due to bacterial DNA contamination, which did not interfere with hybridisations.

ii. Southern blot was hybridised with α (cDNA) and β (5' end, PCR) probes.

iii. Same blot was reprobed with the α (3' UT) and β (intron 8) PCR probes.

*- represent the *Sal*I and *Bam*HI fragments compared in orientating the α, β-casein genes.

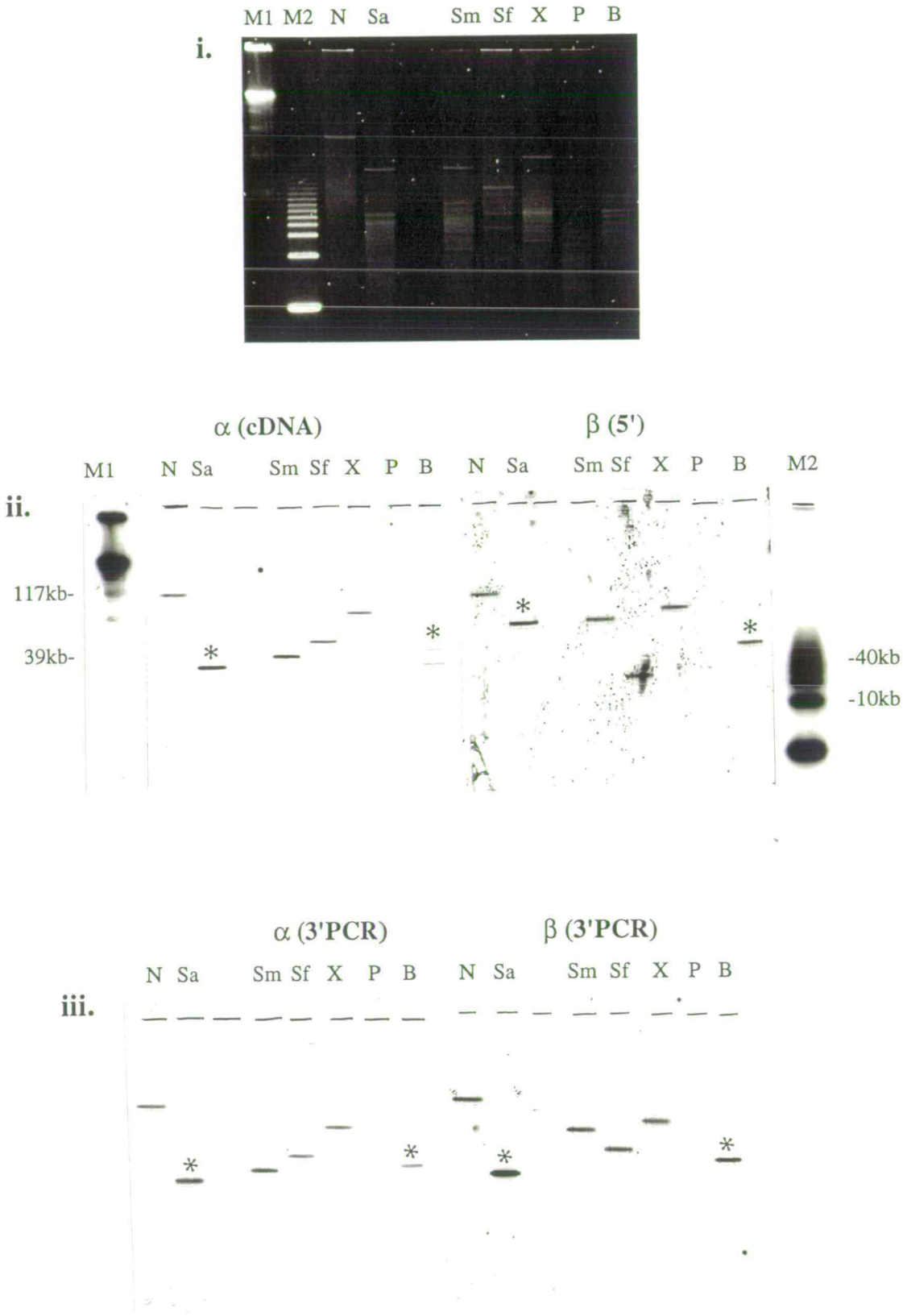

i.
M1 M2 N Sa Sm Sf X P B
ii.
α (cDNA)
β (5')
M1 N Sa Sm Sf X P B N Sa Sm Sf X P B M2
117kb-
39kb-
-40kb
-10kb
iii.
α (3'PCR)
β (3'PCR)
N Sa Sm Sf X P B N Sa Sm Sf X P B

Fig. 3.6.b

positions of the α and β genes and the forward primers are schematically presented in Fig. 3.7. The forward primers from the 3' regions of the two genes (α-5'-TGACTGGACCCTCCATTCTC-3'; β-5'-TTCTCGACCTATGGGTCAGG-3'), yield a fragment of ~11 kb (Fig. 3.7). These observations demonstrated the tail to tail orientation of the α and β genes and showed that the two genes are situated very close to each other with an intergenic region of ~8 kb.

2. γ-orientation

Two *Bam*HI fragments of ~25 and 10 kb were revealed with the γ cDNA probe (clone 490 H23). The smaller fragment also hybridised to the 3' specific γ probe in this and two other clones analysed in the contig (clones 185 018 and 555 N16) (Fig. 3.8. a, b). This result shows the α and γ genes have the same orientation.

3. ε-orientation

The orientation of the ε gene has yet to be determined. None of the enzymes so far tested enable the determination of this gene's orientation with respect to others. As a final attempt a long range PCR analysis was also carried out to orientate the ε-casein gene. Primers designed from the left end fragment of clone 490 H23 (Table 3.4; Fig. 3.9) (region very close to the ε-casein gene) and ε-casein cDNA derived primers (refer to Table 2.2 in section 2) were used in this experiment. Different primer combinations were used to amplify fragments using the various cloned DNA as templates.

Even though small quantities of various sized fragments were amplified in different experiments, substantial amplification was observed when two sets of primer pairs were used, 490 H23 reverse + ε-5' forward and also 490 H23 forward + ε-5' reverse (template DNA-clones 185 O18 and 555 N16) (Fig. 3.9). The first

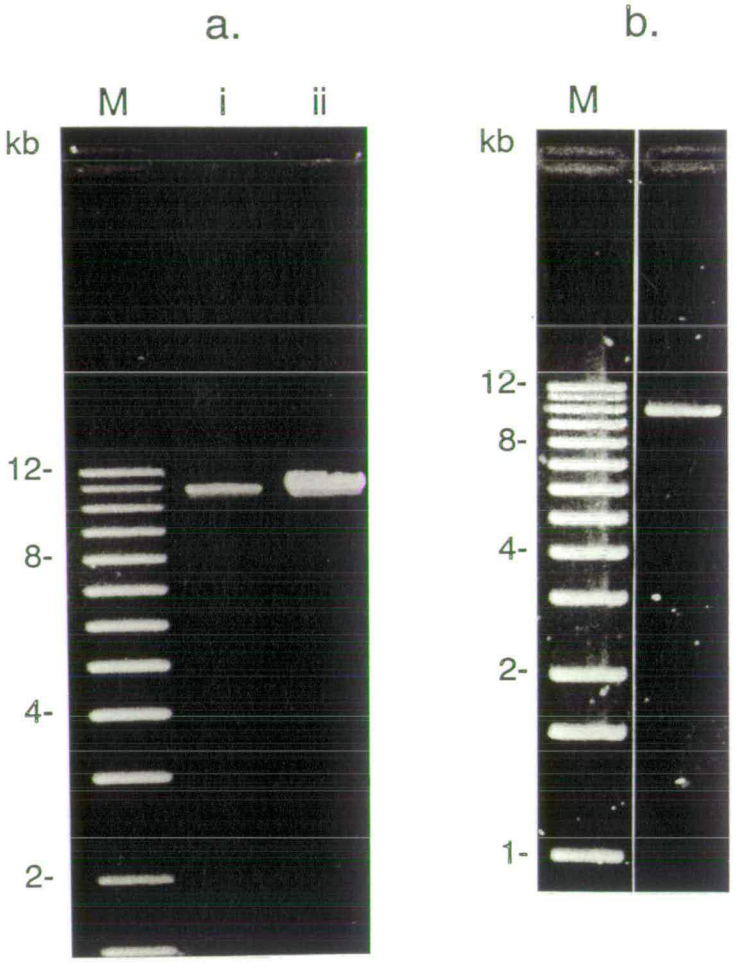
Fig. 3.7. Long range PCR.

a. Using the forward primers from the 3' regions of the α (3' UT) and β (intron 8) genes in clones,

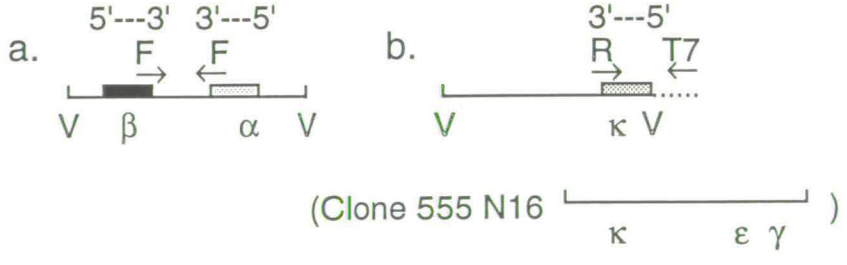
i. 365 E21 ii. 201 E9

b. Using T7 promoter sequences and reverse primer from the 3' untranslated region of κ cDNA in clone 156 J3

M - Molecular size marker (1 kb ladder)



Schematic representation of the primers used for the long range PCR



V- Vector site F- Forward primer R- Reverse primer

Fig 3.7

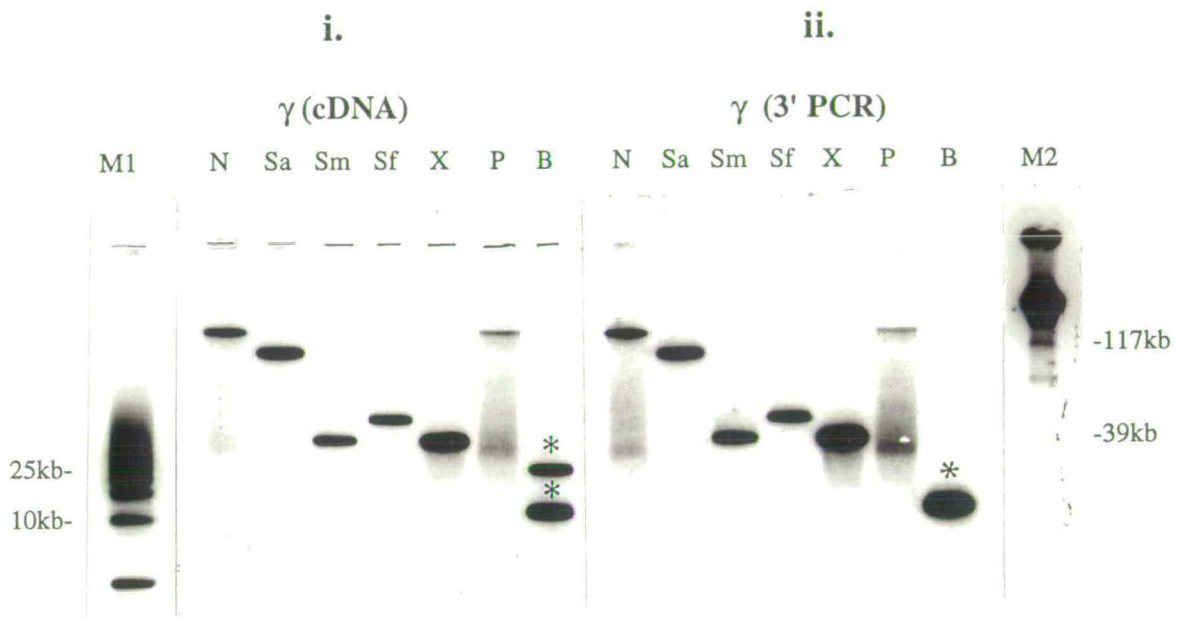


Fig. 3.8.a.
Restriction digested clone 490 H23 containing the α, β and γ-casein genes. The southern blot shown in Fig. 3.5.a was reprobed with

i. the γ - cDNA probe

ii. the γ - 3' UT PCR probe.

The enzymes used are indicated above each lane as follows: N-*Not*I, Sa-*Sal*I, Sm-*Sma*I, Sf-*Sfi*I, X-*Xho*I, P-*Pvu*I, B-*Bam*HI. M1 and M2 are molecular size markers, 5 kb and 39 kb respectively. *- represent the *Bam*HI fragments compared in orientating the γ-casein gene.

Fig. 3.8.b. Restriction analysis of the clones containing the 3' region of the γ-casein gene.

i. Photograph of ethidium stained gel containing restriction digested clone 185 O18 (contains γ, ε genes). Electrophoresis was performed in a CHEF DR-III system (Bio-Rad) for 12.5 hrs, 14°C in 0.5 x TBE at 6 V/cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively. The enzymes used are indicated above each lane as follows: N- *Not*I, Sa- *Sal*I, Sm- *Sma*I, Sf- *Sfi*I, X- *Xho*I, P- *Pvu*I, B- *Bam*HI. M1 and M2 are molecular size markers, 5 kb and 39 kb ladders respectively. Smearing of the lanes was due to bacterial DNA contamination, which did not interfere with hybridisations.

Southern blots were hybridised with γ-cDNA and 3' PCR probes.

ii. 185 O18

iii. 555 N16 (blot shown in Fig 3.5.a was reprobed).

After the first probing the blots were stripped and probed with the second probes.

*- represent the fragments compared in orientating the γ–casein gene.

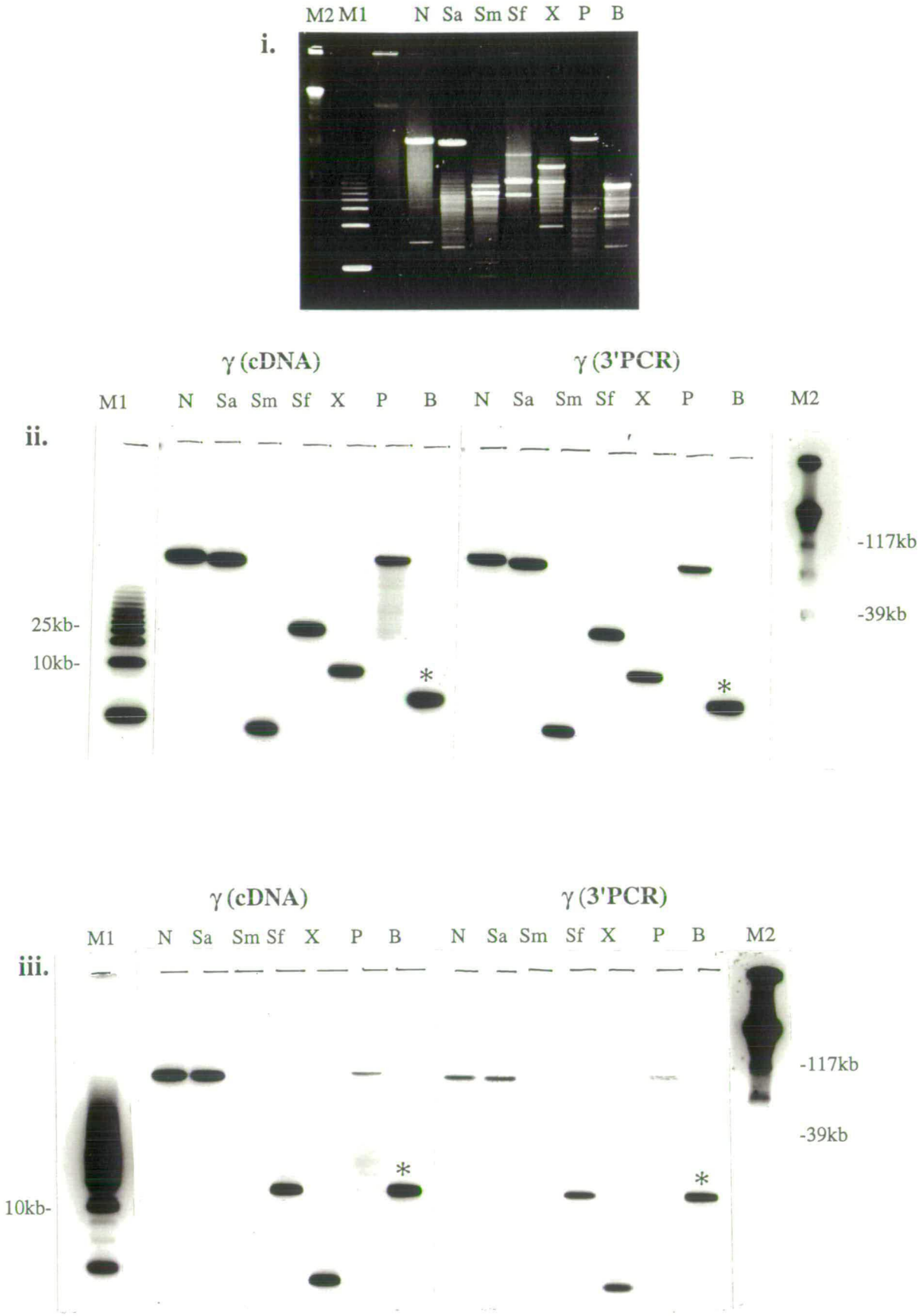

i.
M2 M1 N Sa Sm Sf X P B
ii.
γ (cDNA)
γ (3'PCR)
M1 N Sa Sm Sf X P B N Sa Sm Sf X P B M2
-117kb
-39kb
25kb-
10kb-
*
*
iii.
γ (cDNA)
γ (3'PCR)
M1 N Sa Sm Sf X P B N Sa Sm Sf X P B M2
-117kb
-39kb
10kb-
*
*

Fig. 3.8.b

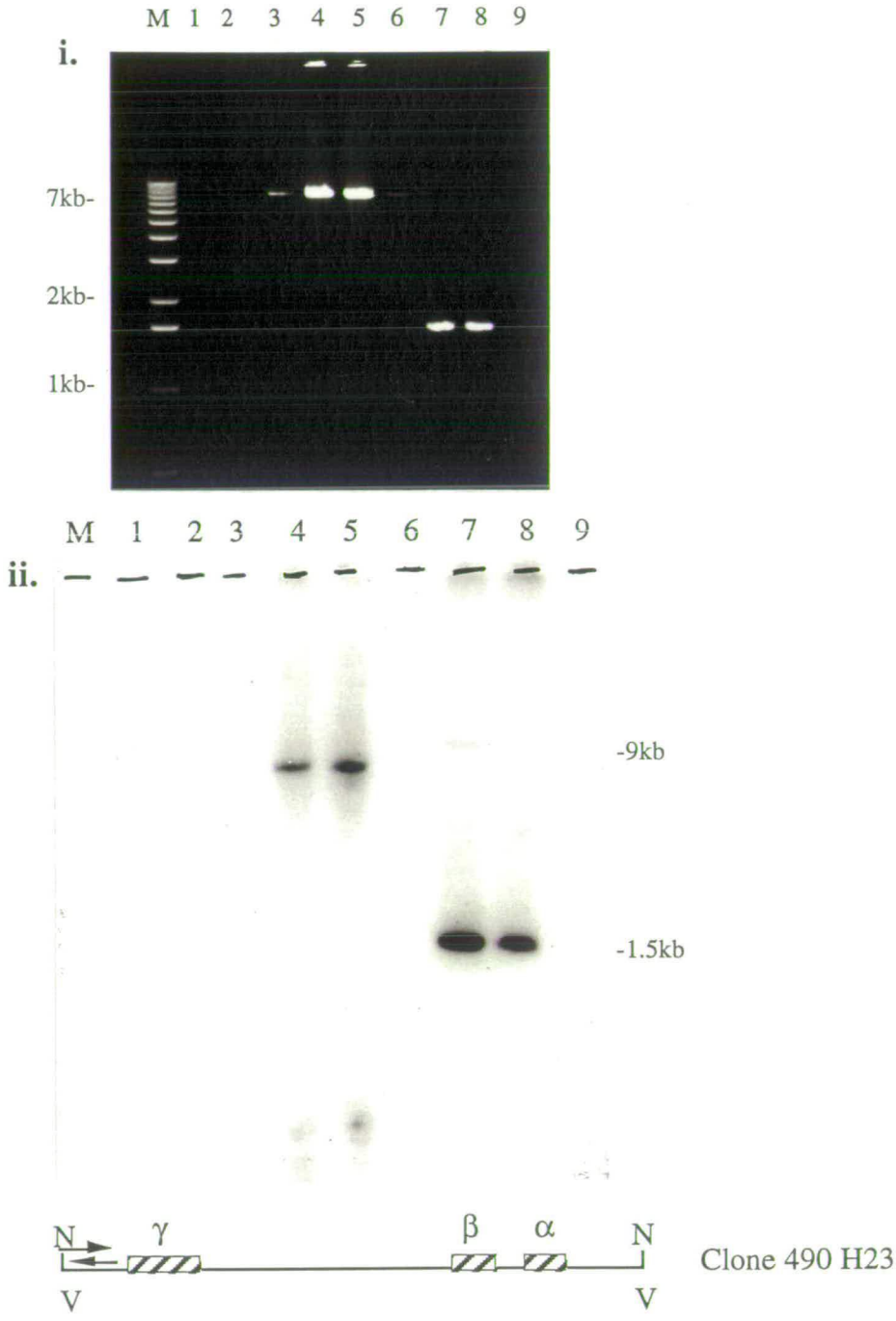
Fig. 3.9. Long range PCR carried out using various primer combinations and different template DNAs, inorder to orientate the ε-casein gene. The primers were designed from the left end fragment of clone 490 H23 and from 5' region of ε-casein cDNA. The 5'-3' direction of the 490 clone primers are known and as shown in the diagram (F- forward, R- reverse). regarding ε-casein, primer 1 is named as F and 2 as R)

i. Ethidium stained gel photograph. 5µl of gthe PCR amplified reaction mix was loaded in each well. M- 1 kb ladder.

Lane No	Primer pairs used	Clones used as template DNA	Amplified fragment size (kb)
1.	490R + ε-R	185 O18	nil
2.	" "	555 N16	15
3.	" "	490 H23	9
4.	490R + ε-F	185 O18	**9**
5.	" "	555 N16	**9**
6.	" "	490 H23	9, 0.2
7.	490F + ε-R	185 O18	12, 4, **1.5**, 0.8
8.	" "	555 N16	12, 4, **1.5**, 0.8
9	" "	490 H23	0.8

ii. Southern blot analysis using ε-casein cDNA probe (The hybridised blot was exposed to X-ray film for 30 min).

Fragments in bold letters were hybridised to ε-casein probe.



Diagramatic representation of the primers designed for the long range PCR to orient epsilon-casein gene. The arrows represent the forward and reverse primers from the left end of clone 490 H23. The primers for the epsilon-casein gene were designed from the cDNA sequence. N- *Not*I site, V- Vector sequence, α, β, γ, ε and κ- locations of the casein genes.

Fig. 3.9

primer pair amplified a fragment of ~9 kb and the second pair amplified a fragment of ~1.5 kb. Since the orientations of the 490 H23 reverse (5'-3' towards the ε gene) and forward (5'-3' towards the γ gene) primers are known, logically it is not possible to amplify a fragment using the latter primer pair irrespective of the orientation of the ε-casein gene. The amplified products were further analysed by Southern blotting and hybridisation using the ε-casein probe. Hybridisation of the probe to both the fragments (9 kb and 1.5 kb) made it difficult to orientate the gene. Since the small fragment (1.5 kb) also hybridised to the ε-casein probe, this might be an amplified fragment of the ε-casein gene due to some kind of false priming. Even then, considering the size of the maximal restriction fragment (Fig. 3.4, 15 kb *Sma*I/*Bam*HI) containing the ε-casein gene and the primer pair used, the 9 kb fragment might be the true amplified product. If that is true, then the ε-casein gene also has the same orientation as the β-casein gene.

4. κ-orientation

The orientation of the κ-casein gene was identified by long range PCR performed in the clone 156 J3 (~ 98 kb). Probing the clone with a probe generated from the left end fragment placed the κ gene near the T7 promoter sequence, which flanks the left end of the cloning site (Fig. 3.10). A primer pair consisting of T7 promoter sequence and the reverse primer from the 3' untranslated region of the κ-cDNA amplified a fragment of ~10 kb (Fig. 3.7). Also the clone was found negative for any of the other known casein genes. So, positioning this clone in the contig map in relation to clone 555 N16 (which also contains the ε and γ genes) showed that the κ gene has the same orientation as the α and γ genes.

Fig. 3.10. Restriction analysis of clone 156 J3.

i. Photograph of ethidium stained gel containing restriction digested clone 156 J3 (contains κ-casein gene). Electrophoresis was performed in a CHEF DR-III system (Bio-Rad) for 12.5 hrs, 14°C in 0.5 x TBE at 6 V/cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively. The enzymes used are indicated above each lane as follows: N- *Not*I, Sa- *Sal*I, Sm- *Sma*I, Sf- *Sfi*I, X- *Xho*I, P- *Pvu*I, B- *Bam*HI. M1 and M2 are molecular size markers, 39 kb and 5 kb ladders respectively. Smearing of the lanes was due to bacterial DNA contamination, which did not interfere with hybridisations.

ii. Southern blot hybridised with κ-cDNA and left end probes.

Similar sized fragments were hybridised in both the hybridisations.

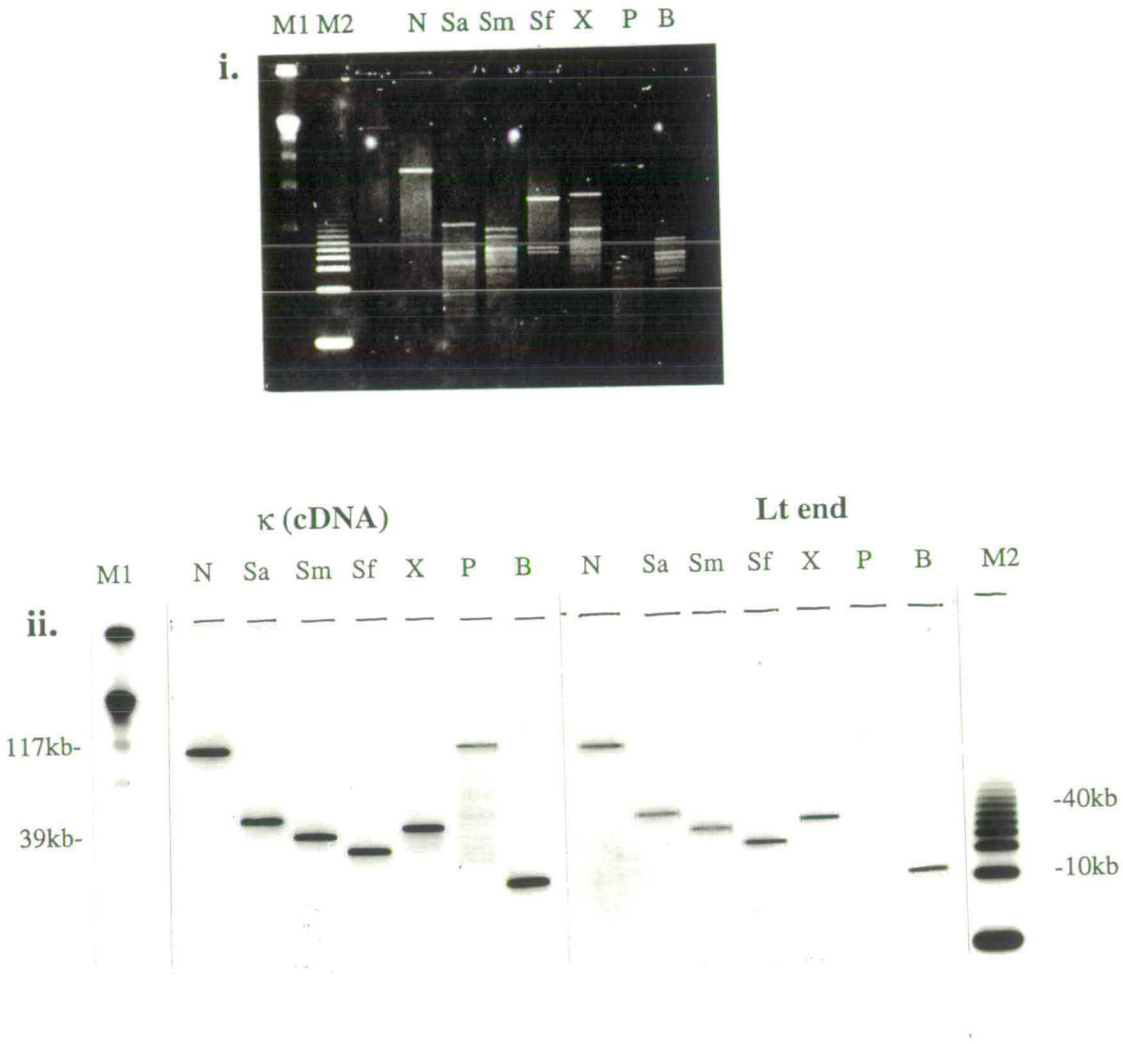

M1 M2 N Sa Sm Sf X P B
i.
κ (cDNA)
Lt end
M1 N Sa Sm Sf X P B N Sa Sm Sf X P B M2
ii.
117kb-
39kb-
-40kb
-10kb

Fig. 3.10

3.2.4. Exploration of the locus for putative additional genes.

The locus was examined for the presence of additional gene sequences expressed in the mammary gland. This was carried out by selecting five BAC clones, which span the entire murine casein locus (~470 kb) including ~165 kb 5' (5' to the α-casein gene) and ~85 kb 3' (3' to the κ-casein gene) flanking regions. The clones, 156 J3, 555 N16, 490 H23, 365 E21 and 280 M14, were digested with *Bam*HI and electrophoresed under pulsed-field conditions. The resulting Southern blot was probed with total cDNA synthesised from poly (A) RNA isolated from 14 day lactating mouse mammary gland and later reprobed with a mixture of all known casein probes (cDNA- α, γ, κ; ε - 5' cDNA, and β - 3' cDNA probes) (Fig. 3.11). As there are no internal *Bam*HI sites in the β and ε-casein genes (Fig.3.4) the fragments identified by the 3' and 5' probes respectively contain the entire genes. When the autorad figures were compared, the total cDNA probing exposed the presence of an extra hybridising fragment of ~15 kb in clone 490 H23. Positioning of this fragment in the clone map revealed that the fragment lies between the β and γ genes in the 55 kb intergenic region, near to the γ gene (Fig. 3.12).

3.3. Discussion

The absence of any casein gene in the P1 library is probably due to its limited genome coverage (1.0). Compared to the P1 library, the BAC library has ~7 times more genome coverage (7.37). So, there is an increased probability of incorporating the fragments containing the same genes, thus enabling the identification of more than one clone containing the same genes from the BAC library. Even though the cloning efficiency is dependent on many factors, from the clones isolated, it was observed that fragments from certain regions of the genome are more efficiently cloned than other parts. For example, 8 plate pools were

Fig. 3.11. Southern blot analysis of five BAC clones, span the entire murine casein locus including ~165 kb 5' (5' to the α-casein gene) and ~85 kb 3' (3' to the κ-casein gene) flanking regions. The clones were digested with *Bam*HI.

i. Photograph of ethidium stained gel. The clones are 1. 156J3 2. 555N16 3. 490 H23 4. 365 E21 5. 280 M14.

M1 and M2 are molecular size markers, 39 and 5 kb ladders respectively.

ii. a. Southern blot hybridised with 14 day lactating mammary cDNA (synthesised from poly (A) RNA) probes. The hybridised blot was exposed to film for 5 days.

b. The blot was reprobed with a mixture of all known casein probes, cDNA - α, γ, κ; ε - 5' PCR (cDNA) and β - 3' PCR (cDNA) . The blot was exposd to film for 2 hours. The faint bands were varified by longer exposure.

*- represents the additional fragment (~15 kb) observed on comparison.

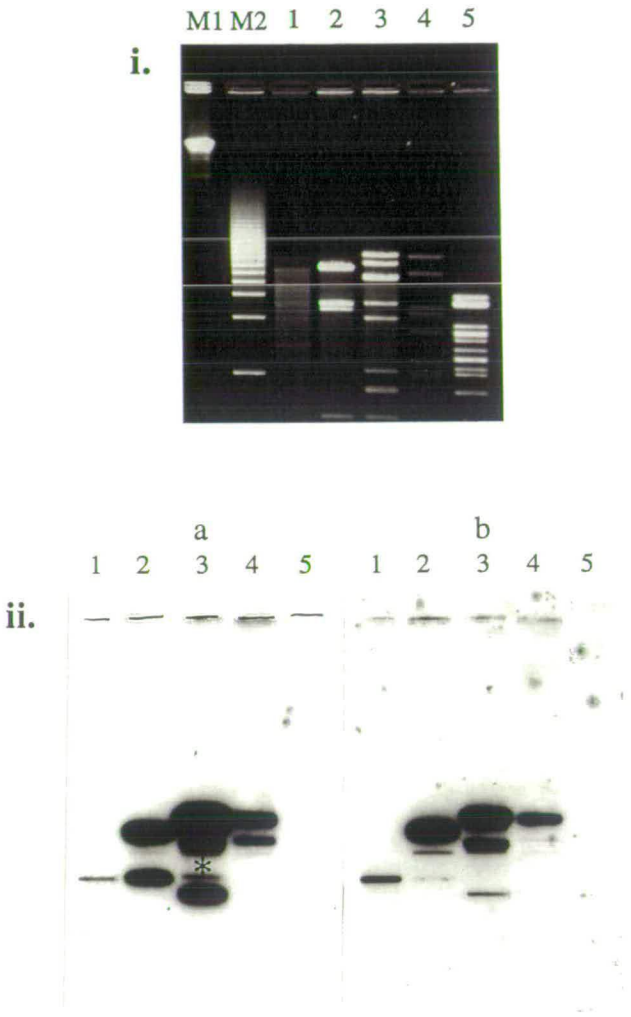


Fig. 3.11

Fig. 3.12. Contig map of the five BAC clones span the entire casein locus including ~165 kb 5' and ~85 kb 3' flanking regions examined for the presence of any additional putative genes. Location of the new 15 kb fragment is shown. B- *Bam*HI sites.

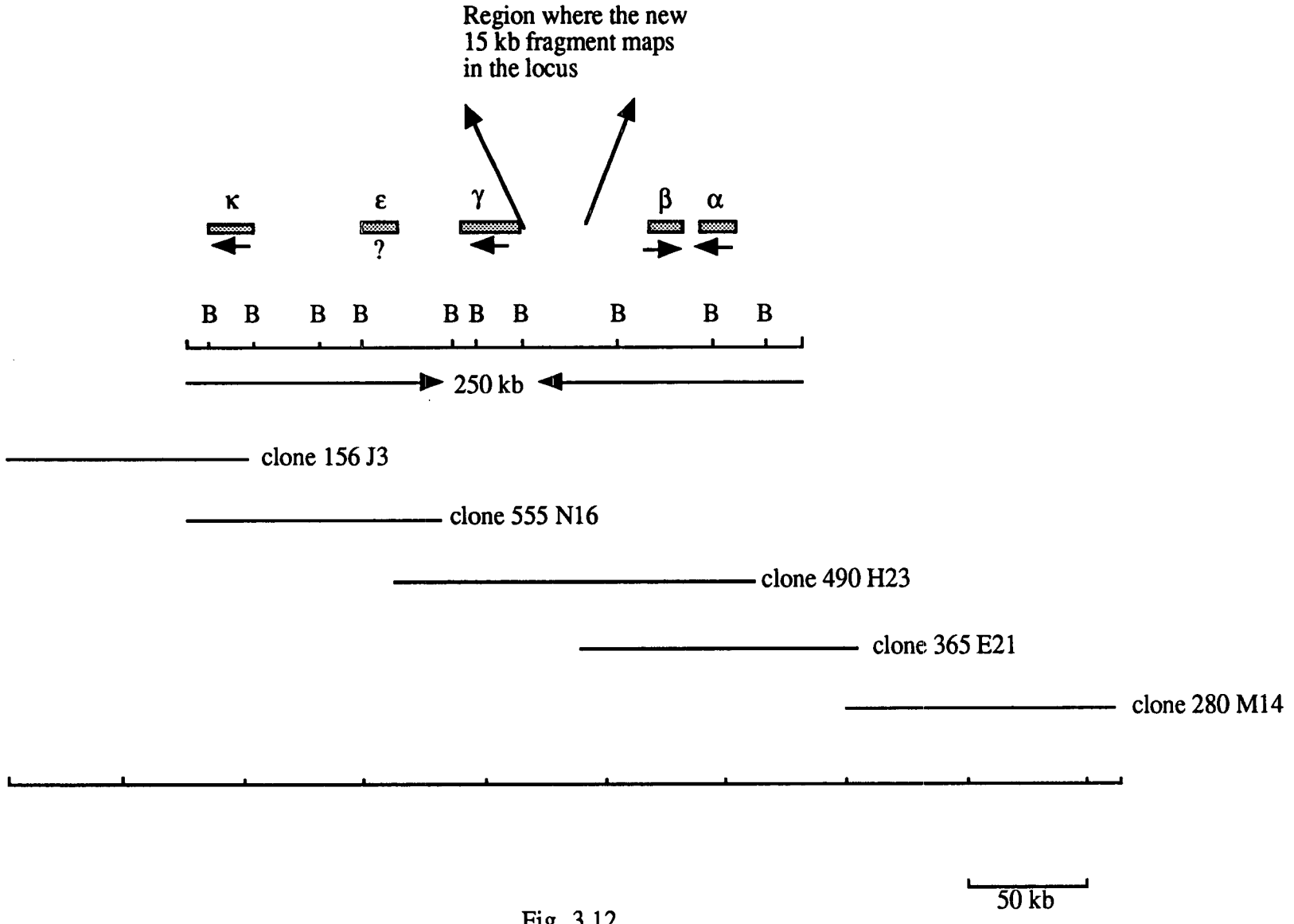

Region where the new
15 kb fragment maps
in the locus
κ
ε
?
γ
β
α
B B B B B B B B B B
250 kb
clone 156 J3
clone 555 N16
clone 490 H23
clone 365 E21
clone 280 M14
50 kb

Fig. 3.12

found positive for the α, β and γ casein genes while from the same library only 4 plate pools were found positive for the ε and κ genes. Therefore, on this basis, it may be possible to isolate at least one clone containing the α, β and γ casein genes from a library having 1 x genome coverage and for ε and κ genes a minimum of 2 x genome coverage is required for identifying one positive clone.

A few super pools and plate pools of BAC library and the pools from P1 library gave positives in which apparently correct sized PCR fragments could be amplified, but failed to yield clones harbouring the relevant genes in subsequent rounds of screening. This failure could be attributed to either clone instability, the loss of clones from the relevant archives or to false priming events.

The bacterial artificial chromosome (BAC) vector system proved to be particularly useful to confirm the detail structure of the murine casein locus. Although the insert sizes of the clones isolated ranged from 80 kb to 155 kb, BACs are capable of accommodating fragments upto 300 kb (Shizuya et.al., 1992; Woo et.al., 1994., Li Cai et. al., 1995).

The ~470 kb casein locus (including 5' and 3' flanking region) has been investigated for the presence of other genes expressed in the lactating mammary gland. The presence of an additional ~15 kb fragment (in clone 490 H23) on probing with total mammary cDNA (corresponding to poly (A) RNA) suggests the presence of one more region in the locus, which is transcribed during lactation. Detailed investigation is needed to clarify whether this encodes another milk protein or casein or is due to the presence of a gene whose expression is ubiquitous. I have only examined the BAC contig for the presence of genes expressed in the lactating mammary gland. Further examinations with cDNA probes from other tissues could confirm the ubiquitous or confined expression of this gene.

By the analysis of large fragment genomic clones I have demonstrated that all five mouse casein genes map to a common 250 kb region of chromosomal DNA. I have further shown that the order of the genes within the casein locus is α-β-γ-ε-κ. Earlier Gupta et.al. (1982) reported that the casein genes, α, β, γ are located in murine chromosome 5. Thus, I can now also assign the ε and κ genes to chromosome 5.

In mice the α and β genes and the γ and ε genes are located close to each other. The α, β genes together constitute 34 -35 kb with an intergenic region of 7 - 8 kb and the genes are in a tail to tail orientation. The distance between the γ and ε genes is less than 20 kb. From the map generated, it is found that the overall sizes of the α and ε genes are not more than 15 kb, κ not more than 20 kb and γ not more than 25 kb, which is consistent with the size of casein genes from other species (Alexander et.al., 1988; Bonsing et.al., 1988; Koczan et.al., 1991; Groenen et.al., 1993; Provot et.al., 1995). Fine mapping with other enzymes should make it possible to define more precisely the limits of these genes. Murine β-casein gene is already sequenced completely and has a size of about 13 kb including its 5' and 3' flanking sequences (Yoshimura and Oka, 1989) which is in agreement with our finding (~14 kb).

The bovine casein locus has been shown to be contained in a genomic DNA fragment of about 200 kb with the genes in the order αS1-β-αS2-κ (Threadgill and Womack, 1990; Ferretti et al., 1990). The data obtained in this study indicate that the murine casein locus is larger than bovine locus with an additional (fifth) gene, ε. The cDNAs of murine α, β and κ caseins exhibit sequence similarity to bovine αS1, β and κ respectively, while γ and ε both share sequence similarity with bovine αS2 (Mercier and Vilotte, 1993).

The caprine casein locus has also been reported to be contained in a 250 kb region of genomic DNA and the four casein genes are tightly linked and

arranged in the order, αS1-β-αS2-κ (Leroux and Martin, 1996). Interestingly, they have also shown that the α and β genes are orientated in a tail to tail fashion ~12 kb apart. The overall size of the caprine casein locus and the tail to tail orientation of the α and β genes are in agreement with the present findings.

Comparison of the bovine/caprine and murine casein loci suggests that a further gene duplication has occurred in the murine lineage with murine γ and ε sharing a common ancestor with bovine/caprine αS2. There is no record of a fifth casein gene in cattle or goats. Thus, the presumed homology between the genes encoding murine α, β and κ and bovine/caprine αS1, β and κ caseins is supported by sequence similarity and conserved gene order.

Chapter 4

Mapping the murine casein locus - High molecular weight (HMW) genomic DNA analysis.

4.1. Introduction

The physical map constructed by BAC clone analyses was confirmed by restriction analysis of genomic DNA prepared from strain 129 mice. The various restriction fragments identified in the BAC clones, on hybridisation with different casein probes were further ascertained in genomic DNA. This analysis was carried out by long range restriction mapping of very high molecular weight genomic DNA under pulsed-field conditions.

4.2. Results

4.2.1. Restriction digestion and pulsed-field analysis

Genomic DNA was digested with restriction enzymes, having recognition sequences in CpG islands, inter CpG island regions or having large recognition sequences, to generate large genomic DNA fragments. The enzymes used were, *Not*I, *Asc*I, *Sgr*AII, *Bss*HII, *Eag*I, *Sac*II, *Nar*I, *Sma*I, *Mlu*I, *Pvu*I, *Nru*I, *Xho*I, *Sal*I, *Cla*I , *Sfi*I and *Bam*HI. Of these enzymes, all except *Sfi*I and *Bam*HI have either one or two CpG motifs in their recognition sites and are sensitive to cytosine methylation.

Restriction digested genomic DNA was electrophoresed in a CHEF DR-III system (Bio-Rad) under pulsed-field conditions (for 18 hrs at 14°C in 0.5 x TBE at 6 V/ cm, 120° angle with initial and final switch intervals of 5 s and 40 s respectively) which separated fragments up to 550 kb. Analysis of Southern blots showed that of these enzymes, only *Xho*I, *Sfi*I and *Bam*HI would be useful in characterising the murine casein locus. *Xho*I gave a partial digestion of the DNA

(Fig. 4.1a, b) with fragments of various sizes. The apparently common hybridising fragments detected were ~340 and ~250 kb with all five casein probes, 190 kb with the α, β, γ and ε probes, 150 kb with the α, β and γ probes, 45 kb with the ε probe and 70 kb with the κ probe. These data are consistent with the BAC clone map.

Restriction digestions with *Sfi*I and *Sfi*I/*Xho*I gave similar sized fragments of ~60 kb with the α, β and γ probes (Fig. 4.1b) while the clone map showed two fragments, ~50 kb hybridised to the α, β probes and ~60 kb to the γ probe. This suggested that the fragment hybridised by the α, β and γ probes represents more than one fragment probably due to the reduced resolution efficiency in the range of 50 to 100 kb. Moreover, in the clone map two *Xho*I sites were detected in the *Sfi*I fragment containing the γ gene. Further analyses were carried out on genomic DNA by resolving fragments in the range of 5-75 kb (electrophoresis was performed in a CHEF DR-III system (Bio-Rad) for 12.5 hrs, 14°C in 0.5 x TBE at 6 V/cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively, Fig. 4.2a, b). These conditions separated two *Sfi*I fragments of ~50 and ~60 kb - the γ-casein probe hybridised to the larger fragment whereas α and β probes hybridised to the smaller fragment. Double digestion with *Sfi*I/*Xho*I reduced the size of *Sfi*I fragment containing the γ gene, consistent with the BAC clone map. These findings clarified the descrepancy in observations between the clones and the genomic DNA.

Southern blot analyses of *Bam*HI digested genomic DNA are shown in Fig. 4.3a, b. The α-casein cDNA probe hybridised to two fragments of 40 and 25 kb while the 3' probe of α and the 5' and 3' probes from the β-casein gene identified a similar sized (40 kb) fragment. Moreover, in the *Sfi*I/*BamH*I double digestion the α-cDNA probe hybridised to two fragments of 25 and 20 kb and a similar sized 20 kb fragment was hybridised to the 3' probe of α and 5' and 3'

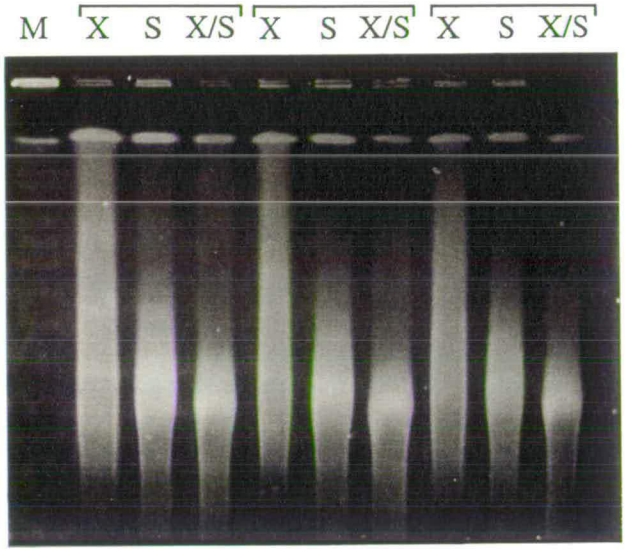


Fig. 4.1. a.

Ethidium stained pulsed-field gel photograph of restriction digested genomic DNA. Electrophoresis was performed for 18 hrs, 14°C in 0.5 x TBE at 6 V/ cm, 120° angle with initial and final switch intervals of 5 s and 40 s respectively. The enzymes were X- *Xho*I, S- *Sfi*I, X/S- *Xho*I/*Sfi*I. M is the molecular size marker (49 kb ladder). The blot was cut into 3 portions, each containing *Xho*I, *Sfi*I and *Xho*I/*Sfi*I digested DNA and were probed with the casein probes.

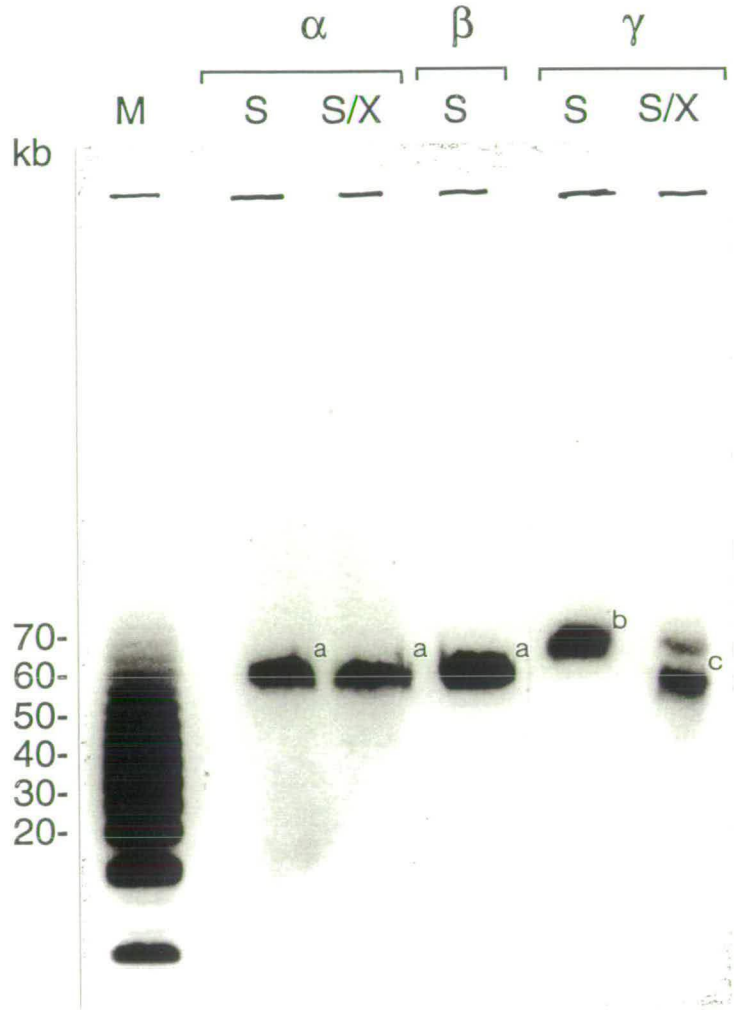


Fig 4.2b

Fig.4.2. b. Pulsed-field gel of restriction digested genomic DNA showing a better resolution in the lower range (5-75 kb), hybridised with α, β and γ probes. The probes used were α, γ - cDNA β - 5' (*Hin*dIII/*Sca*I) genomic fragment. Fragments hybridised with more than one probe are marked with common superscripts. The enzymes were S- *Sfi*I, X/S- *Xho*I/*Sfi*I. M is the molecular size marker (5 kb ladder). The ladder starts from 10 kb onwards. Here, α and β probes hybridised to the same fragment and a slightly larger fragment hybridised to γ probe in *Sfi*I digestion. The γ probe also hybridised to a smaller fragment in *Xho*I/*Sfi*I double digestion which was consistent with the observations in the BAC clone analysis.

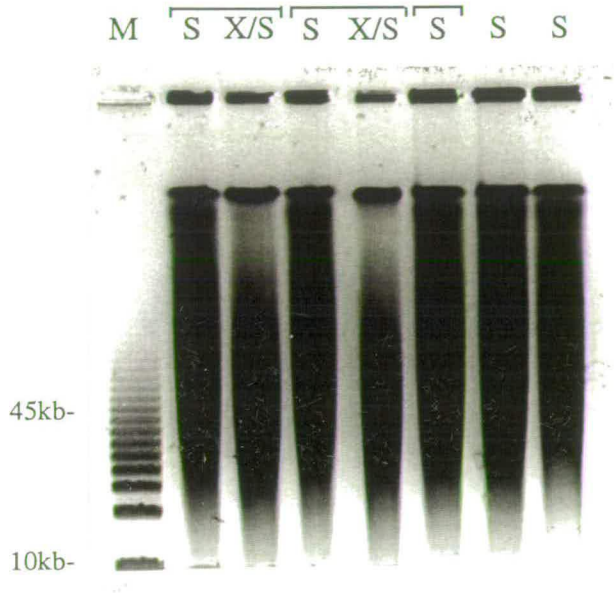


Fig.4.2. a.
Ethidium stained pulsed-field gel photograph of restriction digested genomic DNA. Fragments between 5-75 kb were separated by electrophoresing for 15 hrs, 14°C in 0.5 x TBE at 6 V/ cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively. The enzymes were S- *Sfi*I, X/S- *Xho*I/*Sfi*I. M is the molecular size marker - 5 kb ladder. The blot was cut into 3 portions, 1 and 2 containing *Sfi*I and *Xho*I/*Sfi*I digested DNA (probed with α and γ probes) and 3 containing only *Sfi*I digested DNA (probed with β probe).

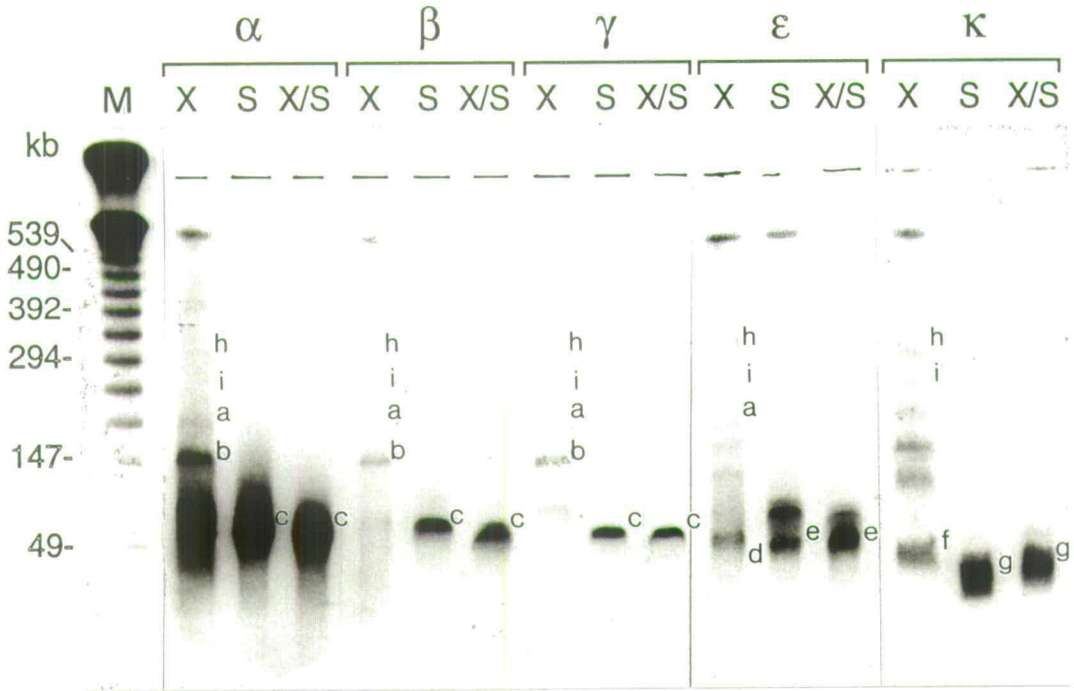


Fig 4.1b

Fig.4.1. b. Pulsed-field gel of restriction digested genomic DNA, hybridised with (a) α, (b) β, (c) γ, (d) ε, (e) κ probes. The probes used were α, γ and κ - cDNA, ε - 5' PCR (cDNA) and β - 5' (*Hin*dIII/*Sca*I) genomic fragment. The enzymes were X-*Xho*I, S- *Sfi*I, X/S- *Xho*I/*Sfi*I. M is the molecular size marker (49 kb ladder). Same superscript is given to similar sized fragments hybridised to various probes.

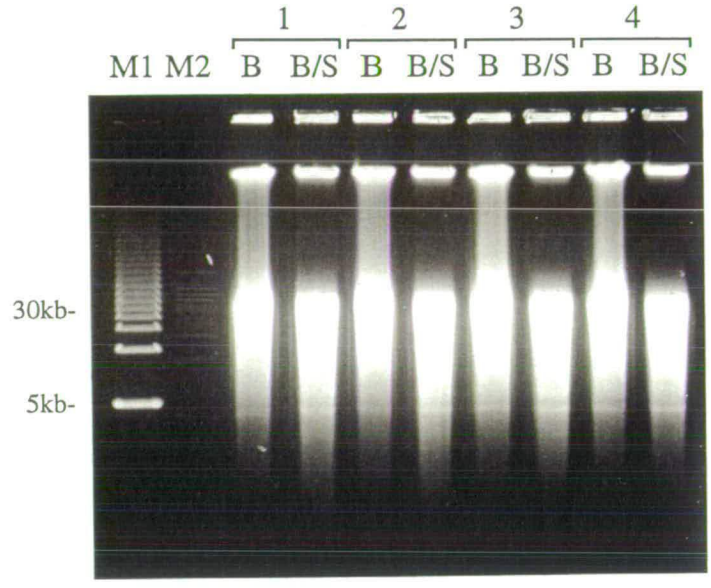


Fig. 4.3. a.

Ethidium stained pulsed-field gel photograph of restriction digested genomic DNA. Electrophoresis was performed for 13 hrs, 14°C in 0.5 x TBE at 6 V/ cm, 120° angle with initial and final switch intervals of 0.2 s and 8.0 s respectively. The enzymes used were B- *Bam*HI, B/S- *Bam*HI/*SfiI*. M1 and M2 are molecular size markers, 5 kb and 8 - 48 kb ladders. The blot was cut into 4 portions, each containing *Bam*HI and *Bam*HI/*Sfi*I digested DNA and were probed with the casein probes.

Fig.4.3. b. Pulsed-field gel of restriction digested genomic DNA, hybridised with the casein probes. i. cDNA/5' probes ii. 3' probes. The probes used were

i. α, γ, and κ - cDNA ; ε - 5' PCR (cDNA); β - 5' (*Hin*dIII/*Sca*I) genomic fragment

ii. α, γ, ε and κ - 3' PCR (3'UT) ; β - PCR (intron 8)

The enzymes were B- *Bam*HI, B/S- *Bam*HI/*Sfi*I. M is the molecular size marker, 5 kb ladder. Fragments hybridised with more than one probe are marked with common superscripts.

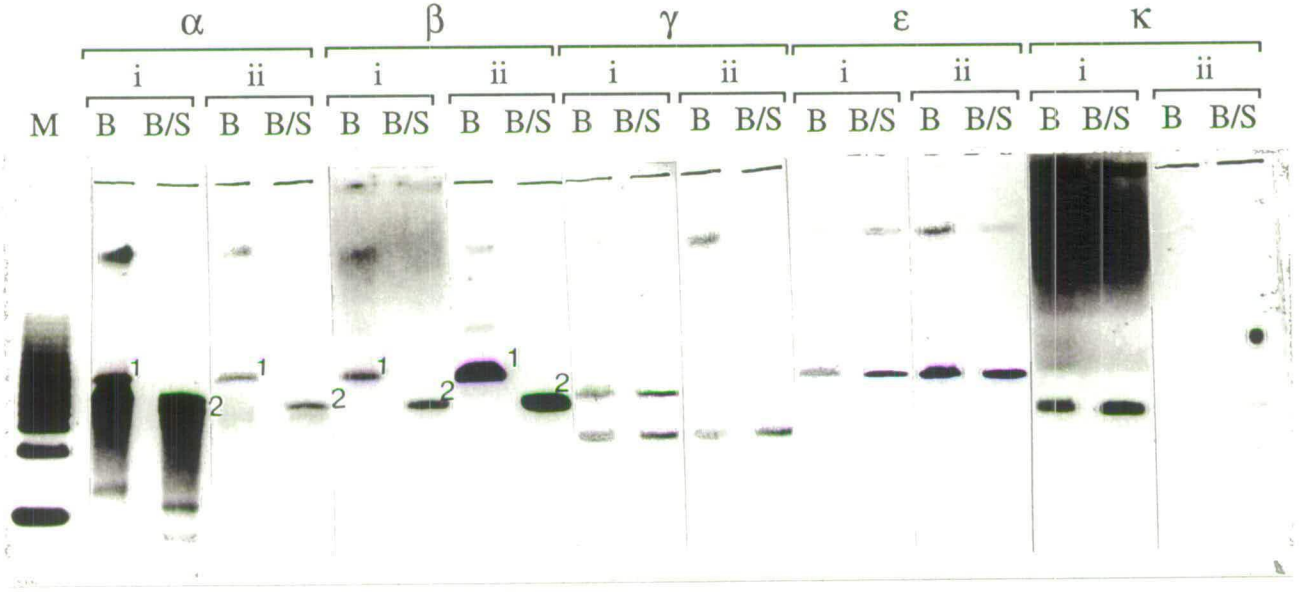


Fig 4.3.b

probes of β-caseins. This suggested that the 40 kb *Bam*HI fragment contains a restriction site for *Sfi*I, which shortened the fragment to 20 kb. Two fragments of 25 and 10 kb were hybridised to the γ cDNA probe and the smaller fragment was identified by the 3' probe. All these findings were consistent with the observations in the BAC clone analysis.

Other enzymes such as *Apa*I, *Bgl*I, *Eco*RV and *Hae*II were also used to analyse the genomic DNA. These enzymes were found not to be effective in characterising the murine casein genes, due to the inability to generate fragments which help either to orientate a gene or to order the various genes in the locus.

4.3. Discussion

The fragments identified in the genomic DNA analysis were compared to confirm the physical map constructed by the BAC clone analysis. The partial fragments obtained in *Xho*I digestion of genomic DNA are consistent with the clone map. It is clear from the map that the murine casein locus is contained within a 250 kb partial *Xho*I fragment. Partial digestion might be due to methylated cytosines in the CpG dinucleotide of the *Xho*I recognition sites, since the genomic DNA used for the restriction analysis was prepared from spleen cells in which casein genes are not expressed. Restriction digestion of clones with *Sfi*I gave fragments of ~50 kb containing α and β, ~60 kb containing γ, ~35 kb containing ε and ~35 kb containing κ casein genes. These data are also consistent with the sizes obtained in *Sfi*I digestion of genomic DNA. *Bam*HI digestion of genomic DNA and clones produced similar sized fragments and clearly indicates the tail to tail orientation of the α and β genes. The presence of a *Sfi*I restriction site in the 40 kb BamHI fragment of α and β-casein genes is in agreement with the findings of Yoshimura and Oka (1989).

In vertebrates the dinucleotide CpG is present at only 25% of the expected frequency. In addition 80% of the genomic CpG sites are methylated (Bird, 1986). Moreover the distribution of unmethylated CpG residues is not random. They are found in clusters in regions known as CpG islands. In mammals CpG islands are rich in G+C residues and are the principal sites of cleavage for methylation sensitive restriction enzymes with G+C rich recognition sequences containing one or more CpG residues (Brown and Bird, 1986). The presence of CpG islands on genomic maps has been instrumental in the localisation and subsequent isolation of genes (Gessler et al., 1990). The occurrence of CpG islands in the vicinity of a probe is often indicated by the detection of hybridising fragments of apparently the same mobility after digestion with a variety of enzymes which have cutting sites within the CpG islands.

Following digestion of genomic DNA with enzymes whose sites include CpG motifs, I was unable to identify any fragments with the casein probes. Most of these enzymes are sensitive to methylation of the CpG sequence. Thus, the failure to find fragments, resolvable even by PFGE, that hybridise to casein probes suggests that either there is an absence of appropriate sites (CpG islands) or that the sites are methylated and resistant to cleavage. Due to the inability to detect in genomic digests the *Sma*I fragments predicted from the clone maps supports the latter interpretation. Hence, it is concluded that there are no CpG islands (i.e. unmethylated clusters of CpG) associated with the murine casein locus. This is not surprising as only about 40% of tissue specific genes (in humans) are associated with CpG islands (Larsen et al., 1992) and the proportion of CpG island associated genes is lower in mouse genome (Antequera and Bird, 1993).

Chapter5

Expression analysis of various casein genes - during pregnancy and lactation

5.1. Introduction

Mammary gland development in the mouse depends upon the interactions of multiple hormones and growth factors. Development is maximal during pregnancy and occurs in several stages. In the early stages of pregnancy, cellular proliferation of the distal portion of the ductal tree leads to the formation of alveoli in response to elevated levels of oestrogen and progesterone. As pregnancy progresses, there is regression of the mammary fat pad, increased vascularisation of the gland and increases in both the size and number of the alveolar lobules. In the later stages of pregnancy, columnar epithelial cells with large vacuoles line the alveoli and the cells become secretory under the influence of lactogenic and adrenal steroid hormones. The acinar luminae become filled with colostrum. At the begining of lactation, there is a sudden decrease in the circulating high levels of oestrogen and progesterone and an increase in prolactin levels. The histological appearance of the mammary gland undergoes very little change during lactation (reviewed by Vonderhaar and Ziska, 1989).

During lactation the various milk proteins, comprising caseins and whey proteins, are produced in response to the interaction of peptide and steroid hormones and growth factors with mammary epithelial cells. Varying levels of mRNA transcripts of different milk protein genes are reported in rat, mouse and rabbit mammary gland during pregnancy and lactation, both *in vivo* and in mammary explant cultures (reviewed by Banerjee and Antoniou, 1985).

Physical mapping of the casein locus (chapters 3 and 4) has shown that the various casein genes are ordered in a tandem array, close to each other, in the

mouse genome. Since the genes are tightly linked and the gene products are present in the milk, it is of interest to investigate whether or not these genes are expressed co-ordinately in the mammary gland. Hence, the expression levels of the five casein genes, α, β, γ, ε and κ were analysed during pregnancy and lactation.

5.2. Results

5.2.1. Northern blot analysis of the murine caseins.

Total RNA was prepared from the mammary glands of virgin, pregnant and lactating F1 females (C57/BL6 x CBA). RNA was isolated at 5, 7, 9, 11, 13 and 16 days of pregnancy and 1 and 14 days of lactation. Two samples were analysed at each time point. Total mammary RNA was fractionated in MOPS/formaldehyde agarose gels along with liver RNA as negative control. The RNA was Northern blotted onto positively charged nylon membranes and analysed with various murine casein probes.

Initially the murine casein probes, α, γ, ε and κ - cDNA probes; β - 3' genomic DNA fragment (*Sal*I/*Sac*I fragment, refer to Fig. 2.1 in section 2) were examined for their hybridisation specificity. The labelled probes were hybridised to Northern blots prepared from lactating mouse mammary gland and liver. Each probe identified a different sized RNA transcript and thus, could be considered specific. (Fig. 5.1).

The accumulation of the five casein RNAs during development of the normal mammary gland from virgin through pregnancy to late lactation was determined. Radioactive hybridisation of the Northern blots with the murine casein probes are shown in Fig. 5.2. a, b. The hybridised blots were exposed for ~16 hrs and subsequently scanned in a phosphorimager. The hybridisation intensities were determined by phosphorimager scanning and normalised to the

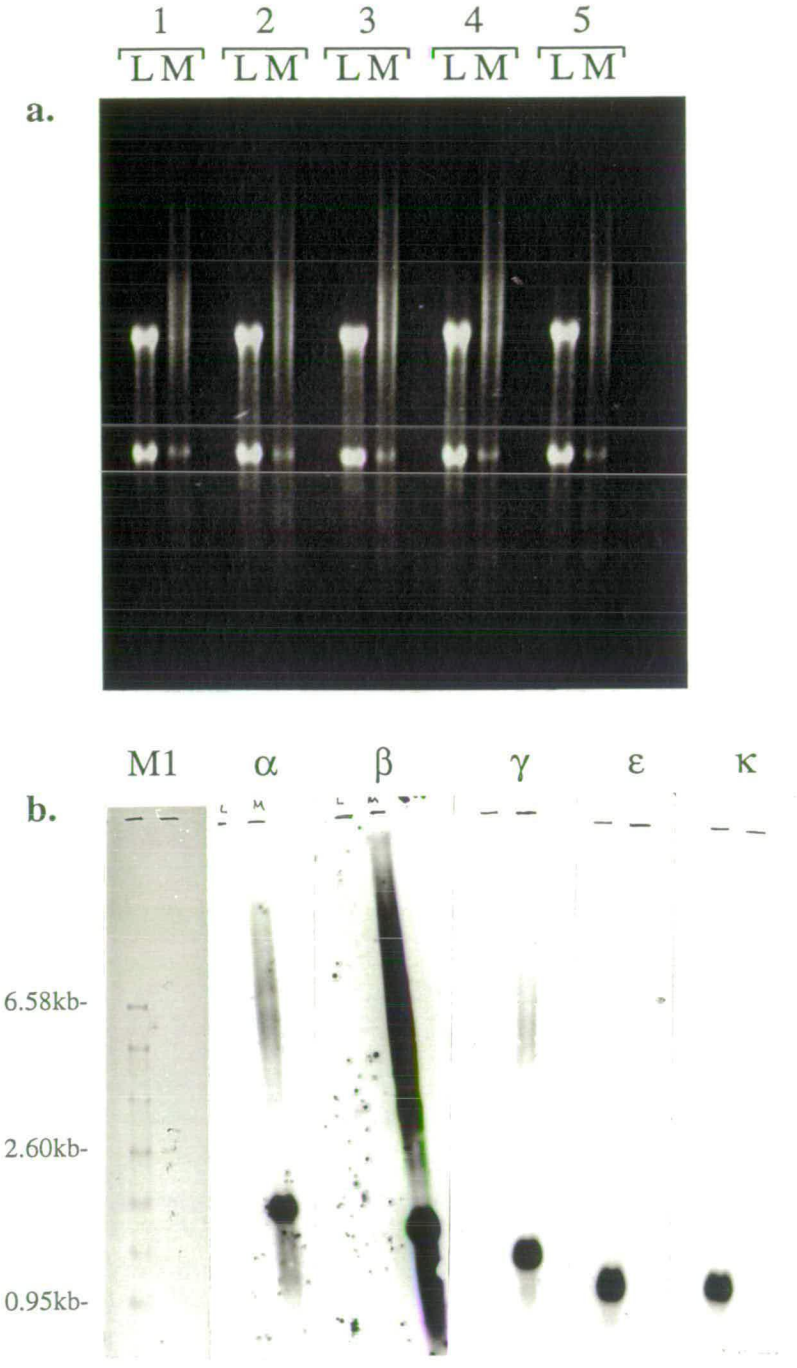


Fig. 5.1.

a. Ethidium stained gel photograph of lactating mammary (M) and liver (L) RNAs. The electrophoresis was performed on a 6.29% formaldehyde gel for 16 hrs at 1-2 V/cm and the gel was Northern blotted. The blot was cut in to 5 portions, each containing mammary and liver RNAs.

b. Northern blots were hybridised with the murine casein probes: (1) α (cDNA) (2) β (5'*Hin*dIII/*Sca*I genomic fragment) (3) γ (cDNA) (4) ε (5' PCR) (5) κ(cDNA) M1- is the RNA size marker.

hybridisation signal of ribosomal RNA probes (Fig. 5.2. a, b). Variation in the levels of the different casein RNAs (α, β, γ and κ) were observed between siblings from mid-gestation onwards. This might be due to variations in placental lactogen level between the two individuals at each time point; each timepoint analysed represents a time-window of 15 hours due to the in-house system for identifying plugs.

The level of expression for each casein gene was determined relative to the level at day 14 of lactation. Thus, the relative expression levels (%) were calculated taking the 14 day level as 100%. The results for the five murine casein RNAs are presented in Table 5.1 and Fig. 5.3. As determined by hybridisation of Northern blots there are no detectable levels of the casein RNAs in virgin and 5 day p.c (post conception) mammary gland. Low levels of the α and β- casein RNAs were detected in the mammary gland at day 7 of pregnancy. The γ-casein RNA was detected initially on the 9th day of pregnancy. The levels of these (α, β and γ) casein RNAs increased gradually and reached ~60, 90 and 100% by the first day of lactation. The ε-casein RNA were detected only on the first day of lactation. Detectable amounts of κ-casein RNA were observed from the 11th day of pregnancy, increased gradually and reached ~60% by the first day of lactation. Thus, all casein RNAs except ε showed a rapid induction from mid gestation while the expression of ε-casein was delayed until the onset of lactation.

Although the α, β and γ caseins appeared to be expressed co-ordinately from mid-pregnancy, the levels of all three casein RNAs varied from 60 to 100 % by the onset of lactation. The κ-casein level increased at a lower rate from mid pregnancy, but not in a co-ordinate fashion with the other three caseins. The ε-casein showed a totally different expression profile, which appeared to be reminiscent of that of α-lactalbumin (Vilotte and Soulier, 1992).

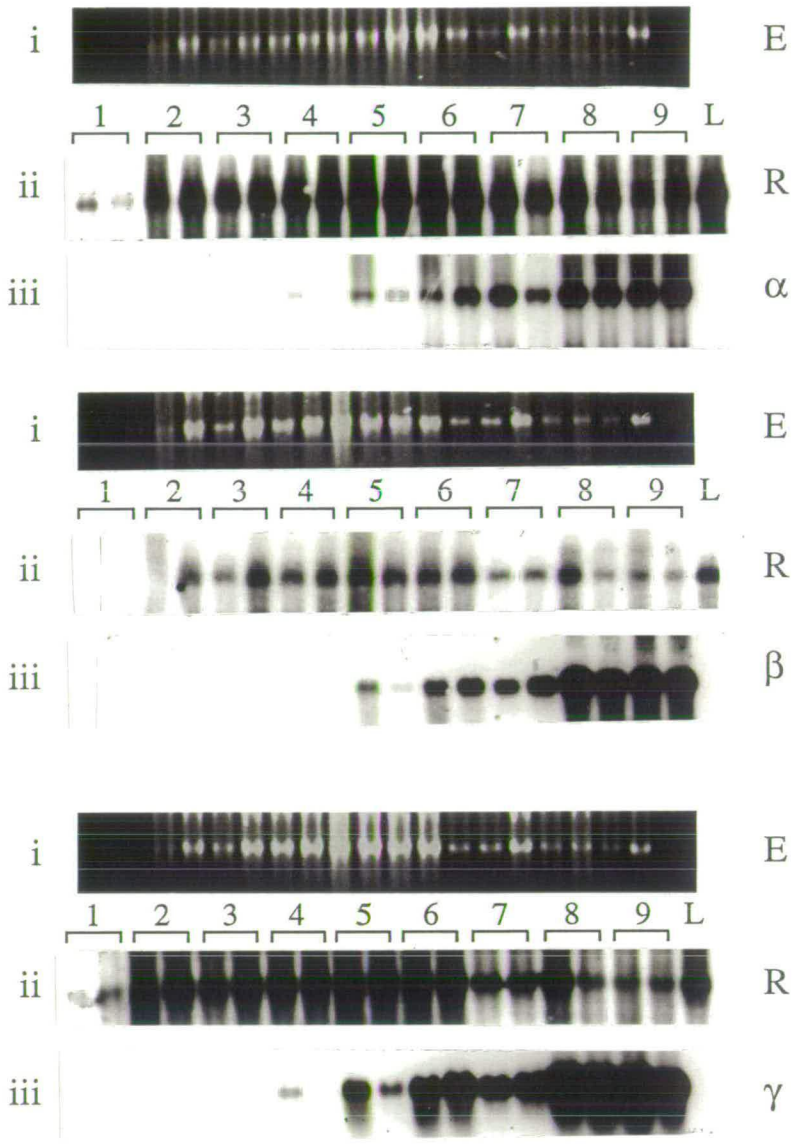


Fig. 5.2. a.

Northern blot analysis of mammary RNAs at 9 time points.

i. Ethidium stained gel (E) photograph of the ribosomal RNA

ii. probed with ribosomal RNA probe (R)

iii. probed with α, β and γ-casein probes (probes as shown in Fig. 5.1)

The time points are 1) virgin 2) 5 dp 3) 7 dp 4) 9 dp 5) 11 dp 6) 13 dp 7) 16 dp 8) 1dl 9) 14dl. (dp - days of pregnancy; dl - days of lactation). L - liver RNA used as a negative control.

The sizes of ethidium and autorad pictures are not comparable.

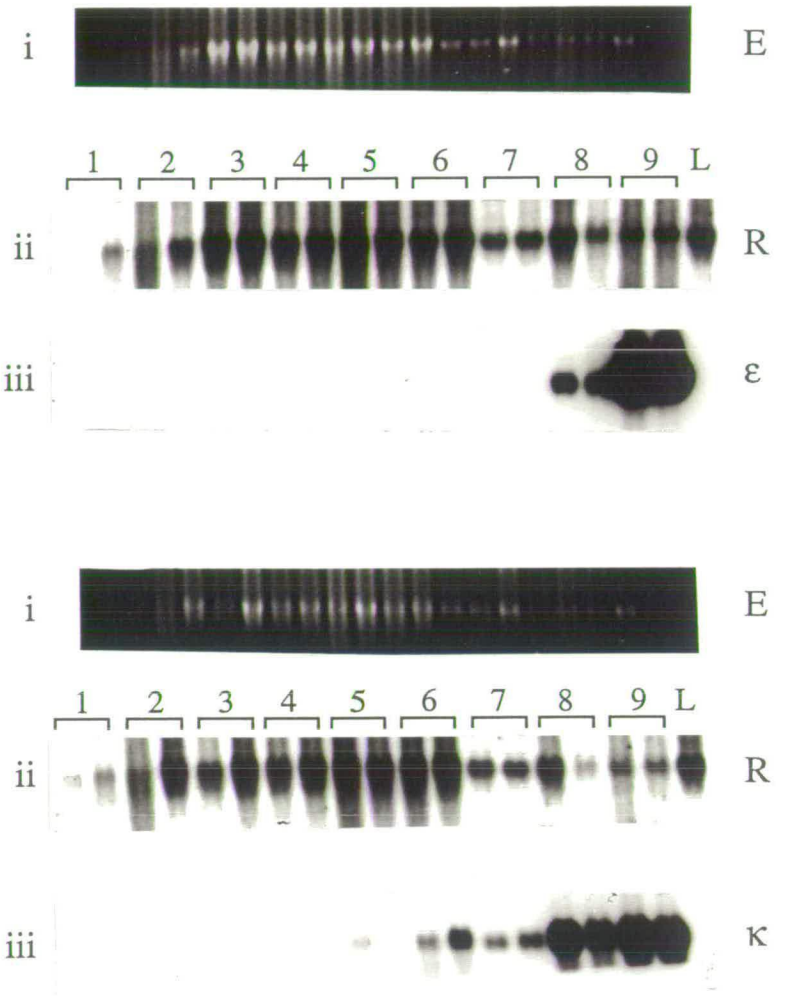


Fig. 5.2. b.

Northern blot analysis of mammary RNAs at 9 time points.

i. Ethidium stained gel (E) photograph of the ribosomal RNA

ii. probed with ribosomal RNA probe (R)

iii. probed with ε and κ-casein probes (probes as shown in Fig. 5.1)

The time points are 1) virgin 2) 5 dp 3) 7 dp 4) 9 dp 5) 11 dp 6) 13 dp 7) 16 dp 8) 1dl 9) 14dl. (dp - days of pregnancy; dl - days of lactation).

L - liver RNA used as negative control.

The sizes of ethidium and autorad pictures are not comparable.

Table 5.1. Expression levels (%) of the five casein genes in virgin, pregnant and lactating mouse mammary gland (v- virgin, dp- days of pregnancy, dl- days of lactation)

Days	α	β	γ	ε	κ
v	<1	<1	<1	<1	<1
5dp	<1	<1	<1	<1	<1
7dp	1	1	<1	<1	<1
9dp	2	2	1	<1	<1
11dp	5	5	4	<1	1
13dp	16	15	16	<1	4
16dp	32	34	35	<1	10
1dl	60	92	100	5	58
14dl	100	100	100	100	100

Relative levels of RNA accumulation was calculated taking the 14 day level as 100 %.

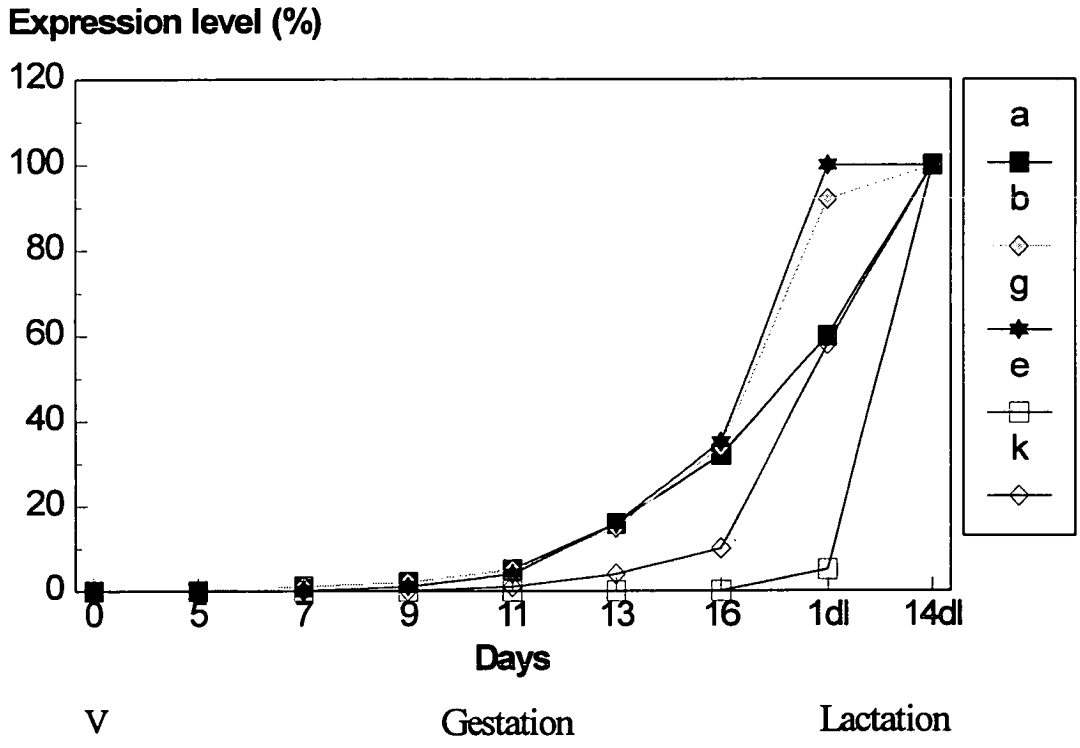


Fig. 5.3. Expression levels of the casein genes in virgin (v), pregnant and lactating (l) mouse mammary gland. Level of RNA in 14 day lactating mammary gland was taken as 100%. a- α, b- β, g- γ, e- ε and k- κ caseins.

5.2.2 Statistical analysis

With the assistance of Ms. Anthea Springbett statistical analyses were carried out on the levels (in virgin, pregnant and lactating mammary gland) of five different casein RNAs measured on 9 different time points (2 mice/day). The analysis indicated a curvilinear relationship between time and expression level for each casein gene. Analyses were also performed for the time at which the expression levels begin to rise and their relative rates of increase. The parameter estimates were fitted to two types of curves, Gompertz curve and logistic curve (Wilson, 1977). Both curves and parameter estimates are similar but not identical. These curves are examples of sigmoid curves which are commonly used to model growth rates and other variables such as hormone levels in animals. Their parameters can be interpreted as initial baseline levels, rate and time of increase and final asymptotic levels. These curves can also be used for interpolation of time points which have not been observed directly.

The Gompertz curve provided a good fit for the α, β and γ caseins, an acceptable fit for the ε-casein and a poor fit for κ-casein. Therefore, κ-casein was analysed by logistic curve, which was found suitable. The parameters can be interpreted as lower and upper asymptotes (corresponding to initial baseline and final asymptotic levels of expression), the rate of rise between the asymptotes (B) and the time at which levels begin to rise (M) (Table 5.2).

The levels of α, β and γ-casein mRNAs start to rise at around 16 to 17 days of pregnancy (Table 5.2). The level of ε-casein mRNA starts to rise much later. However, the estimate of M for ε-casein has a very large standard error and is not significantly larger than the estimates of M for the other caseins. The standard errors of some estimates are large because of the limited number of data and most of the differences are not statistically significant. The rate of increase of expression levels (B) is greatest in γ and second highest in β-casein. The logistic

Table 5.2. Parameter estimates for Gompertz and Logistic curves

Casein mRNAs	Parameter estimates (Gompertz curve)		Parameter estimates (Logistic curve)	
	B	M	B	M
α	0.15 +/- 0.03	17.4 +/- 0.90	0.26 +/- 0.05	19.3 +/- 0.80
β	0.35 +/- 0.05	16.0 +/- 0.30	0.53 +/- 0.05	17.1 +/- 0.20
γ	0.78 +/- 0.61	16.2 +/- 0.30	0.63 +/- 0.13	16.9 +/- 0.40
ε	0.11 +/- 0.27	32.4 +/-34.20	0.49*	27.1*
κ			0.50 +/- 0.07	20.4 +/- 0.30

*- Standard errors for ε-casein were unobtainable for the logistic fit due to convergence problems. B - rate of rise; M - time in days when levels start to rise.

curve of the κ-casein is interpreted in a similar way as Gompertz curve but is not exactly equivalent. Analysis shows that the κ-casein lies between ε-casein and the other caseins in its rate of increase and the time at which transcripts begin to accumulate.

5.3. Discussion

Northern blot analysis of the five casein RNAs in virgin, pregnant and lactating mammary glands did not reveal detectable amounts of any of the casein RNAs either in virgin mice or at 5 days p.c. Induction of all casein RNAs except the ε-casein was observed from mid-pregnancy onwards. Even though very low levels of the β-casein RNAs have been detected in virgin and early pregnant mouse mammary gland (Harris et al., 1991; Baruch et al., 1995), its induction was observed from mid-gestation which is in agreement with the present findings. A similar observation for the α, β, and γ casein RNAs has also been reported for rat mammary gland (Hobbs, et al., 1982) while the κ-casein RNAs were not detected in mammary glands of virgin rats (Nakhasi et al., 1984). It has been estimated that virgin and early pregnant mammary gland contain a low percentage (~25 %) of epithelial cells (Nicoll and Tucker, 1965) of which only a small proportion will be differentiated enough to express milk protein genes. This immature state of the gland presumably, accounts for the low levels of casein RNAs during these periods. As pregnancy advances, considerable morphological changes occur to the mammary epithelia. During the latter half of pregnancy, in mouse mammary gland, the epithelial cell volume increases ~5 fold, as the alveolar structures are formed, and continues into lactation (Foster, 1977). It has also been estimated that mammary epithelial cells divide with a doubling time of about 6 days (Knight and Peaker, 1982). If a cell divides, the two daughter cells have 1/2 as much RNA each as the original cell. The RNA content of each cell increases substantially

with the level of mammary RNA increasing ~6 fold as estimated from the recovery of total RNA/g of mammary tissue (Harris et al., 1991). Thus, in the latter half of pregnancy, an active transcription of the casein genes in a large number of differentiated mammary epithelial cells might account for the increased levels of various casein RNAs.

Analysis of the expression profiles showed an apparent co-ordinated accumulation of α, β and γ casein RNAs from mid pregnancy to the onset of lactation. Co-ordinated expression is also observed in the two globin gene families, which are located on separate chromosomes. For the proper assembly of globin tetramers into haemoglobins, a balanced synthesis of α and β-globins are required (reviewd by Orkin, 1990). Therefore, common regulatory mechanisms may exist to ensure the co-ordinated expression of these unlinked gene clusters. In the present study, the physical map generated has shown that the casein genes are ordered in a tandem array in the mouse casein locus with the genes arranged α - β - γ - ε - κ. The α, β and γ caseins are calcium sensitive caseins and transport quantities of calcium and phosphate to the suckling young. Considering the location of the genes and the similar function of the gene products, it might be possible that these genes are under a common regulatory mechanism which accounts for the co-ordinated expression of these genes in mouse mammary gland. The tail to tail orientation of the α and β genes does not appear to impede co-ordinated expression of these two genes. However, in rats the α, β and γ casein genes are not expressed in a co-ordinated manner (Hobbs et al., 1982). Thus, the mechanism underlying the regulation of casein gene expression appears to be different in these two rodents.

The κ-casein RNAs, increased in a non-coordinate pattern with the other caseins (α, β and γ) from mid-pregnancy onwards, but at a lower rate and reached a concentration of ~30% (of that seen at 14th day of lactation) by 16th day of

pregnancy. This shows that a 3.3 fold increase in the κ-casein mRNA level occurs from parturition through to mid lactation. A similar expression pattern has also been detected in rat mammary gland (Vonderhaar and Nakhasi, 1986) with a 4.5 fold increase in level from mid-gestation to 14 day lactation (Nakhasi et al., 1984).

Unlike the α, β, γ and κ casein genes the ε-casein gene was expressed only in the lactating mammary gland. Being a calcium sensitive casein and having an internal position in the casein locus, close to the γ-casein gene, the reason for the expression of ε-casein gene being limited to the lactating gland (a pattern more or less similar to α-lactalbumin - induction from 18th day of pregnancy; Vilotte and Soulier, 1992) is not clear. From the evolutionary point of view, when a single gene for a particular protein is present, evolutionary pressures tend to retain this gene in a fully functional form and prevent mutational decay and divergence. But if more than one copy is present, then mutational degradation of the excess copies might add these to junk DNA. Selection pressure does preserve intact the physiologically important and useful genes, and occasional retention and preservation of duplicate copies of genes (Markert, 1987). The great "challenge" for any duplicated gene is to quickly acquire the status of an independently regulated gene encoding a protein with somewhat distinctive biochemical properties that will provide an adaptive advantage under evolutionary selection pressure. Otherwise, such duplicated gene might undergo mutative degradation and become a pseudogene.

Evolutionary analysis shows that the casein genes have evolved from two small gene families, the casein gene family (α, β, γ and ε; reviewed by Mercier and Vilotte, 1993) and the fibrinogen gene family (κ; Thompson et al., 1985). By comparison the calcium sensitive casein genes, α, β, γ and ε showed homology to bovine αS1, β and αS2 (γ and ε) respectively. These genes have evolved by gene duplication events and the duplicated copies diverged and acquired new functions.

Beta casein determines the surface properties of milk and is essential for curd formation while α-caseins determine the capacity of casein micelles for colloidal calcium transport. Unlike other species, mouse milk contains three α-type caseins, α, γ and ε. The absence of ε-casein during pregnancy may be because the evolutionary pressure on this gene is substantially different from other casein genes or it may have some other specific function in milk as seen in the case of α-lactalbumin (Brodbeck and Ebner, 1966).

The difference in the expression profile of ε-casein may also be due to the presence of repressor element(s) which suppress the expression of this gene during pregnancy as observed for WAP gene expression. It has been shown that in transgenic mice, due to the absence of repressor element(s) the WAP transgene (Burdon et al., 1991) and WAP promoter driven human tissue plasminogen activator hybrid gene (Pittius et al., 1988) are expressed from 13 and 10 days of pregnancy respectively, in contrast to the endogenous WAP gene which is induced late in pregnancy (from 15 days). Furthermore, transgenic mice which expressed WAP during gestation, prior to normal induction, the mammary gland was underdeveloped, contained few alveoli and resembled glands of non-transgenic mid-pregnant mice. Such animals failed to lactate and nurture their offspring and the phenotype was named 'milchlos' (Burdon et al., 1991). It may be possible that the early expression of ε-casein might have some deleterious effect on the mammary gland development and mammary epithelial cell differentiation as seen in the early expression of WAP gene. However, this seems to be unlikely since the other highly homologous casein proteins are expressed during pregnancy. A clear understanding can only be obtained by detailed studies.

At each time point from mid-pregnancy onwards, variations in the levels of casein RNAs were observed between siblings. In studies like this, such variations are almost inevitable since the expression levels of each gene will vary

between individuals. In total, 18 animals (2 separate animals at each time point) were sacrificed to carry out this study. However, inbred mice strains in which the individual mice are genetically identical make such studies more robust than comparable work with outbred species. The noise in the experimental data can be attributed to environmental effects or limitations of the techniques followed. Alterations in placental lactogen level and circulating corticosterone can cause such variations since placental hormones can replace, at least in part, pituitary homones during mid to late gestation. It has been shown that prolactin can be substituted by placental lactogen (Markoff and Talamantes, 1980). The serum concentration of prolactin is low in mouse during the latter half of pregnancy, whereas the placental lactogen concentration starts to rise on day 9 and reaches a high level on day 14 of pregnancy (Thordarson and Talamantes, 1987). Although the serum concentration of growth hormone increases during the latter half of pregnancy its elevation is not as pronounced as that of placental lactogen. It is also known that the serum concentration of placental lactogen is positively correlated with litter size (or number of placenta). Thus, one explanation for some of the variations in casein gene expression between individuals could be the number of embryos/foetuses each female was carrying or the number of pups they are suckling. Moreover, circulating corticosterone also increases slightly, before day 10 of pregnancy but thereafter a sharp increase in plasma concentration occurs reaching the highest values at day 16 of gestation (Barlow et al., 1974). Mammary gland growth and differentiation will not proceed in the absence of mammotropic hormones, prolactin, growth hormone or placental lactogen. It is, therefore, likely that placental lactogen is the most important lactogenic hormone during the latter half of pregnancy in the mouse, since prolactin concentration is low at this time and growth hormone does not appear to be as potent a lactogen as either placental lactogen or prolactin (Thordarson and Talamantes, 1987).

The levels of the casein RNAs (except ε) increased 2 to 3 fold from the later stages of pregnancy through to mid lactation. In mammary epithelial cells, prolactin acts in concert with other lactogenic hormones such as glucocorticoid and insulin to induce various milk protein RNAs. Studies on the promoter regions of β-casein, WAP and β-lactoglobulin identified various transcription factors such as nuclear factor1 (Li and Rosen, 1995), mammary gland factor (MGF) or milk protein binding factor (MPBF), now known to be Stat5 (Schmitt-Ney et al., 1991; Watson et al., 1991) and DNA sequence elements like composite response element (CoRE; Raught et al., 1995) which mediate the response to lactogenic hormones. The elevated levels of lactogenic hormones and their co-operative action with various transcription factors might account for the additional 2-3 fold increase in the level of the casein RNAs from late pregnancy to lactation.

Regulation of the casein genes in rodents appears to be under both positive (Schmitt-Ney et al., 1992a, b) and negative (Altiok and Groner, 1994; Meier and Groner, 1994) influences, as increases in casein mRNA abundance at parturition (Hobbs et al., 1982) has been explained by a lowering of serum progesterone that reduces a negative influence (Lee and Oka, 1992) and allows the mammary cell to respond to the positive influence of increased serum prolactin. Although the casein genes are arranged sequentially on the same chromosome in mouse and in other species like bovines and caprines (Ferretti et al., 1990; Leroux and Martin, 1996) the expression profiles suggest that the casein gene locus is not overall, under complete co-ordinate regulation.

The presence of a locus control region (refer to introduction) in the casein locus is still under investigation. Unlike the human β-globin locus, where expression of each gene is temporally and spacially regulated, all the genes in the casein locus are expressed in the pregnant and lactating mammary gland. Moreover, it has been shown that β-casein deficient mice do not show any

alteration in phenotype and/or the micelle forming ability of caseins (Kumar et al., 1994). Since deletion/alteration of the globin locus results in anaemia, it appears that selection pressure on the globin locus is much stronger as compared to the casein locus. The present study showed that the α, β and γ casein genes exhibited a co-ordinate pattern of expression from mid-gestation, even though the levels of expression varied by the onset of lactation. Perhaps, since the α, β and γ casein genes show the highest and earliest expression, the candidate region to look for a LCR (if such a region is present in the casein locus) is the region encompassing the α, β and γ genes.

Chapter 6

Identification of genetic variation at the casein loci by restriction fragment length variation

6.1. Introduction

As discussed earlier (see section 1.2.5) there have been extensive studies of genetic variation at the casein loci. Most of this research has focussed on the dairy species of domestic animals, namely cattle, sheep and goats. The possibility of relationships between milk protein polymorphisms and dairy performance has driven these research programmes. The value of milk protein polymorphism as a tool in selective breeding has been discussed in detail by Ng-Kwai-Hang and Grosclaude (1992).

Initially genetic variation at the casein loci was studied at the protein level. Casein variants were identified by gel electrophoresis for bovine αS1, αS2, β and κ-caseins (reviewed by Fiat and Jolles, 1989) and caprine αS1-casein (Boulanger et al., 1984). The limiting factors for such studies are 1) only some of the changes in the nucleotide level cause a change in amino acid/protein and an overall change in the charge that cause an alteration in electrophoretic mobility 2) lactating (i.e. older) females are needed for the production of milk.

More recently, variation in milk proteins has been studied at the DNA level, using RFLPs (restriction fragment length polymorphisms), PCR (polymerase chain reaction), PCR-ACRS (polymerase chain reaction-amplification created restriction sites) and PCR-RFLPs.

Studies on genetic variation at the murine casein loci are very limited. Analysis at the protein level did not reveal any casein polymorphism in various inbred strains of mice. However, by urea-gel (pH 3.7) electrophoresis Piletz and Ganschow (1981) observed a protein variant, probably a casein (type not specified) in *Mus musculus castaneus*. Linkage data at the murine casein loci is

also very limited. Geissler et al. (1988) mapped the W^{19H} deletion (W- white spotting locus) proximal to the casein cluster and showed that the α and β casein genes are linked in mice. They observed *Taq*I RFLPs in α and β-casein loci but were unable to detect a RFLP for γ-casein locus. In the linkage map of mouse chromosome 5, the α and β casein gene loci are located at E39 and E40, flanked by W-locus and Hmg17-rs6 locus respectively.

The experiments performed with a few selected enzymes to search for polymorphic sites at the casein loci in eight inbred strains of *Mus musculus* and between *Mus musculus* and *Mus spretus* are described in this chapter.

6.2. Results

High molecular weight genomic DNA was prepared from eight different strains of *Mus musculus*, 129, C57/BL6, C3H, DBA, NIH, BALB/c, SJL and CBA and from *Mus spretus*. A panel of five restriction enzymes were selected, *Bam*HI, *Sfi*I, *Eco*RI, *Xba*I and *Hin*dIII. On the basis of preceding results (chapter 3 and 4) it was expected that in *Bam*HI and *Sfi*I digests the casein gene probes would reveal large DNA fragments. The digested DNAs were electrophoresed on conventional horizontal gels except *Bam*HI and *Sfi*I digestions, which were resolved by pulsed-field conditions. The gels were Southern blotted onto positively charged nylon membranes (Boehringer Mannheim). The blots were analysed with various casein probes - α, γ, ε and κ - cDNA and β - 3' genomic fragment (*Sal*I/*Sac*I fragment, refer to Fig. 2.1 in section 2).

No variation in restriction fragments as generated by the five enzymes was detected with the casein probes amongst the inbred strains of *Mus musculus* (Fig. 6.1 to Fig. 6.5). Between the strains, weakly hybridised fragments were also seen in the various digestions on probing with different casein probes, which appeared

to be due to some incomplete digestion of the DNA (ladder like hybridisation bands).

Between the two species *(Mus musulus* and *Mus spretus)* the five enzymes generated restriction fragment length variants (RFLVs) at the casein loci with the following combinations of restriction enzymes and probes: α-casein - *Bam*HI, *Sfi*I, *Xba*I, *Eco*RI; β-casein - *Bam*HI, *Sfi*I; γ-casein - *Bam*HI, *Sfi*I, *Xba*I, *Hin*dIII and *Eco*RI; ε-casein - *Bam*HI, *Sfi*I, *Hin*dIII, *Eco*RI; κ-casein - *Bam*HI, *Xba*I, *Hin*dIII

α-casein locus:

Of the five restriction enzymes used, all the enzymes except *Hin*dIII, generated polymorphic restriction fragments in the α-casein locus (Fig. 6.1). *Hin*dIII produced a monomorphic fragment of ~4.3 kb in both the species. The fragments generated by the four enzymes showed a two-allele polymorphism with approximate fragment sizes as shown below:

	Fragment size (kb)			
Restriction enzyme	*Bam*HI	*Sfi*I	*Xba*I	*Eco*RI
Mus spretus	25	40 + 70*	7 .0	6.5
Mus musculus	25 + 40	50	9.5	3.8 + 4.1

β-*casein locus*:

Unlike the α-casein locus, the β-casein locus exhibited two allele polymorphisms only with the enzymes, *Bam*HI and *Sfi*I, which generated large fragments (Fig. 6.2). The fragments produced were,

	Fragment size (kb)	
Restriction enzyme	*Bam*HI	*Sfi*I
Mus spretus	20	30 + 70*
Mus musculus	40	50

Fig. 6.1. Restriction fragment length variant (RFLV) analysis of the α-casein locus of *Mus spretus* and eight inbred strains of *Mus musculus*.

They are 1. *Mus spretus* 2. 129 3. C57/BL6 4. C3H 5. DBA 6. NIH 7. BALB/c 8. SJL 9. CBA. The enzymes used are shown on top of each autoradiogram. The sizes of the fragments are shown in kb.

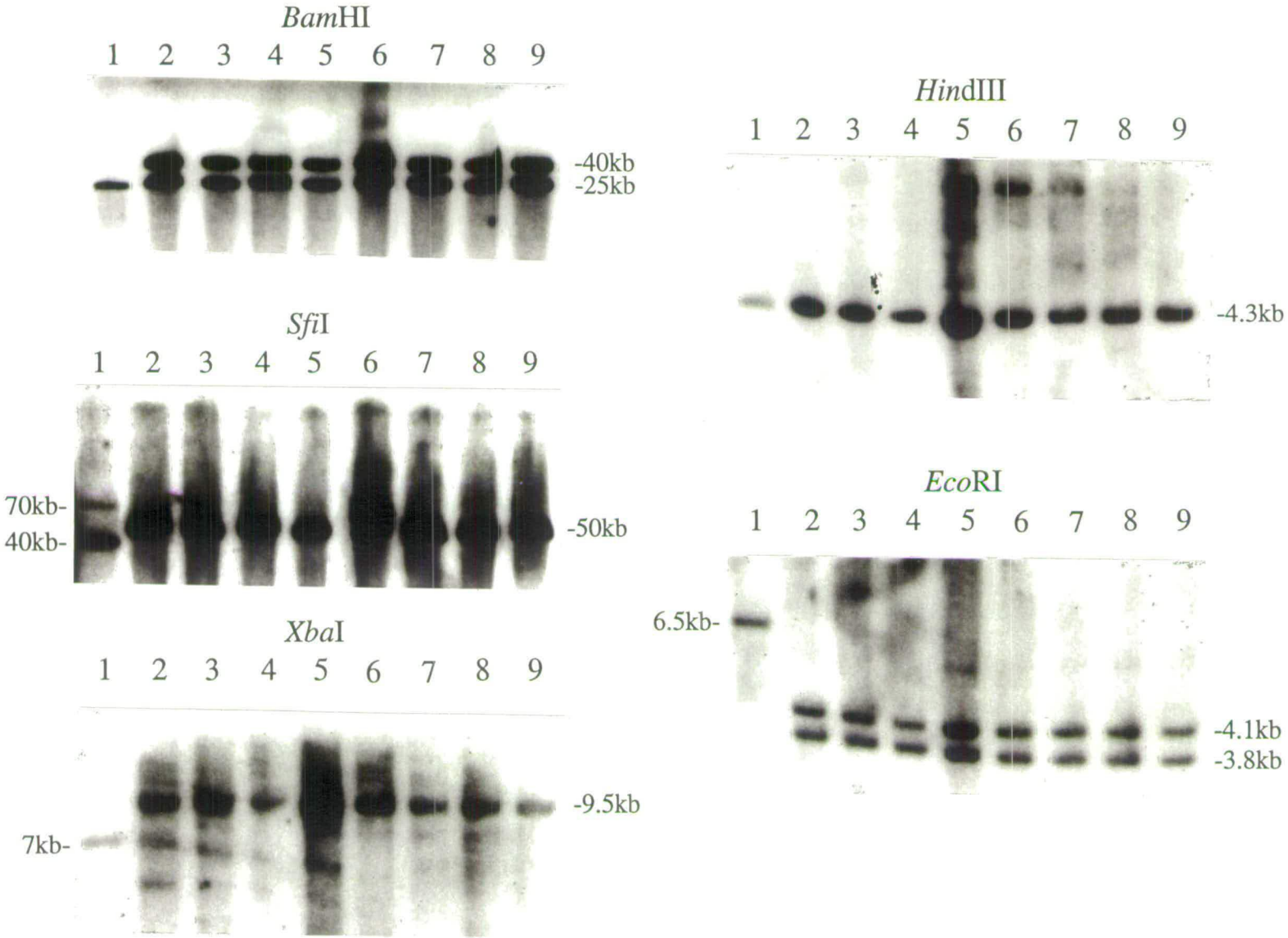

BamHI
1 2 3 4 5 6 7 8 9
-40kb
-25kb
SfiI
1 2 3 4 5 6 7 8 9
70kb-
40kb-
-50kb
XbaI
1 2 3 4 5 6 7 8 9
7kb-
-9.5kb
HindIII
1 2 3 4 5 6 7 8 9
-4.3kb
EcoRI
1 2 3 4 5 6 7 8 9
6.5kb-
-4.1kb
-3.8kb

Fig 6.1

Fig. 6.2. Restriction fragment length variant (RFLV) analysis of the β-casein locus of *Mus spretus* and eight inbred strains of *Mus musculus*.

They are 1. *Mus spretus* 2. 129 3. C57/BL6 4. C3H 5. DBA 6. NIH 7. BALB/c 8. SJL 9. CBA. The enzymes used are shown on top of each autoradiogram. The sizes of the fragments are shown in kb.

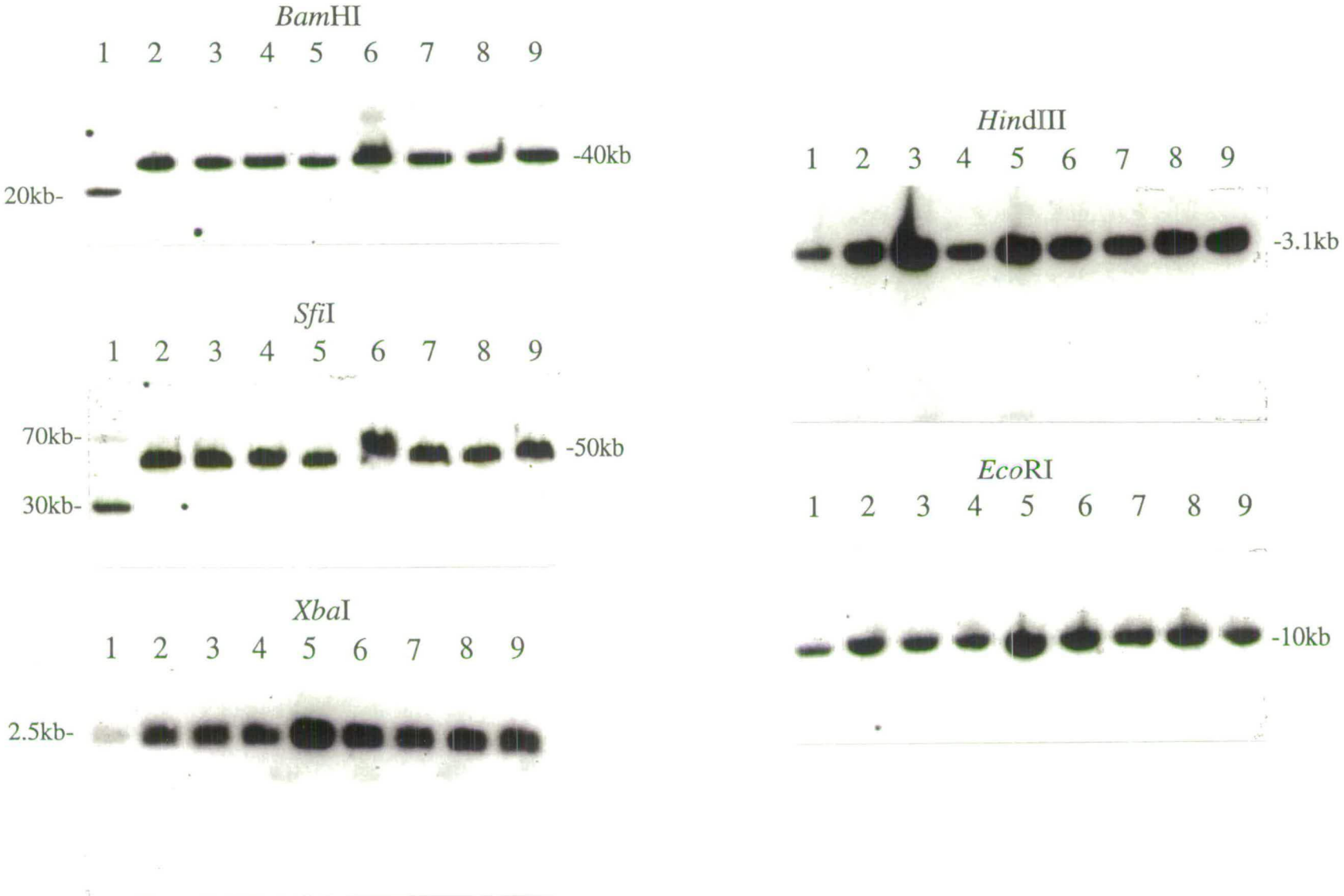

*Bam*HI
1 2 3 4 5 6 7 8 9
20kb-
-40kb
*Sfi*I
1 2 3 4 5 6 7 8 9
70kb-
30kb-
-50kb
*Xba*I
1 2 3 4 5 6 7 8 9
2.5kb-
*Hin*dIII
1 2 3 4 5 6 7 8 9
-3.1kb
*Eco*RI
1 2 3 4 5 6 7 8 9
-10kb

Fig 6.2

The frequent cutting enzymes, *Xba*I, *Hin*dIII and *Eco*RI generated similar sized fragments of 2.5 kb, 3.1 kb and 10 kb respectively in both the species. The 70 kb fragment observed in α and β casein probing appeared to be a partial digestion product due to the low intensity of hybridisation as compared to the 40 and 30 kb fragments.

γ-casein locus:

All five enzymes generated polymorphic fragments in the γ-casein locus and showed two allele polymorphisms (Fig. 6.3). The following polymorphic fragments were observed in the two species.

	Fragment size (kb)				
Restriction enzyme	*Bam*HI	*Sfi*I	*Xba*I	*Hin*dIII	*Eco*RI
Mus spretus	15 + 23	20 + 40	3.5 + 5.5	3.0 + 4.0 + 4.5	10 + 5.5
Mus musculus	10 + 25	60	2.8 + 3.1 + 5.4	2.9+ 3.4 + 5.2	8.5 + 5.5

ε-*casein locus*:

At the ε-casein locus *Xba*I generated monomorphic fragments (13 kb) in both the species. The polymorphic fragments produced (Fig. 6.4) by the other enzymes were as follows:

	Fragment size (kb)			
Restriction enzyme	*Bam*HI	*Sfi*I	*Hin*dIII	*Eco*RI
Mus spretus	25	80	4.5	6.5
Mus musculus	30	35	4.5 + 6.5	9.0

Fig. 6.3. Restriction fragment length variant (RFLV) analysis of the γ-casein locus of *Mus spretus* and eight inbred strains of *Mus musculus*.

They are 1. *Mus spretus* 2. 129 3. C57/BL6 4. C3H 5. DBA 6. NIH 7. BALB/c 8. SJL 9. CBA. The enzymes used are shown on top of each autoradiogram. The sizes of the fragments are shown in kb.

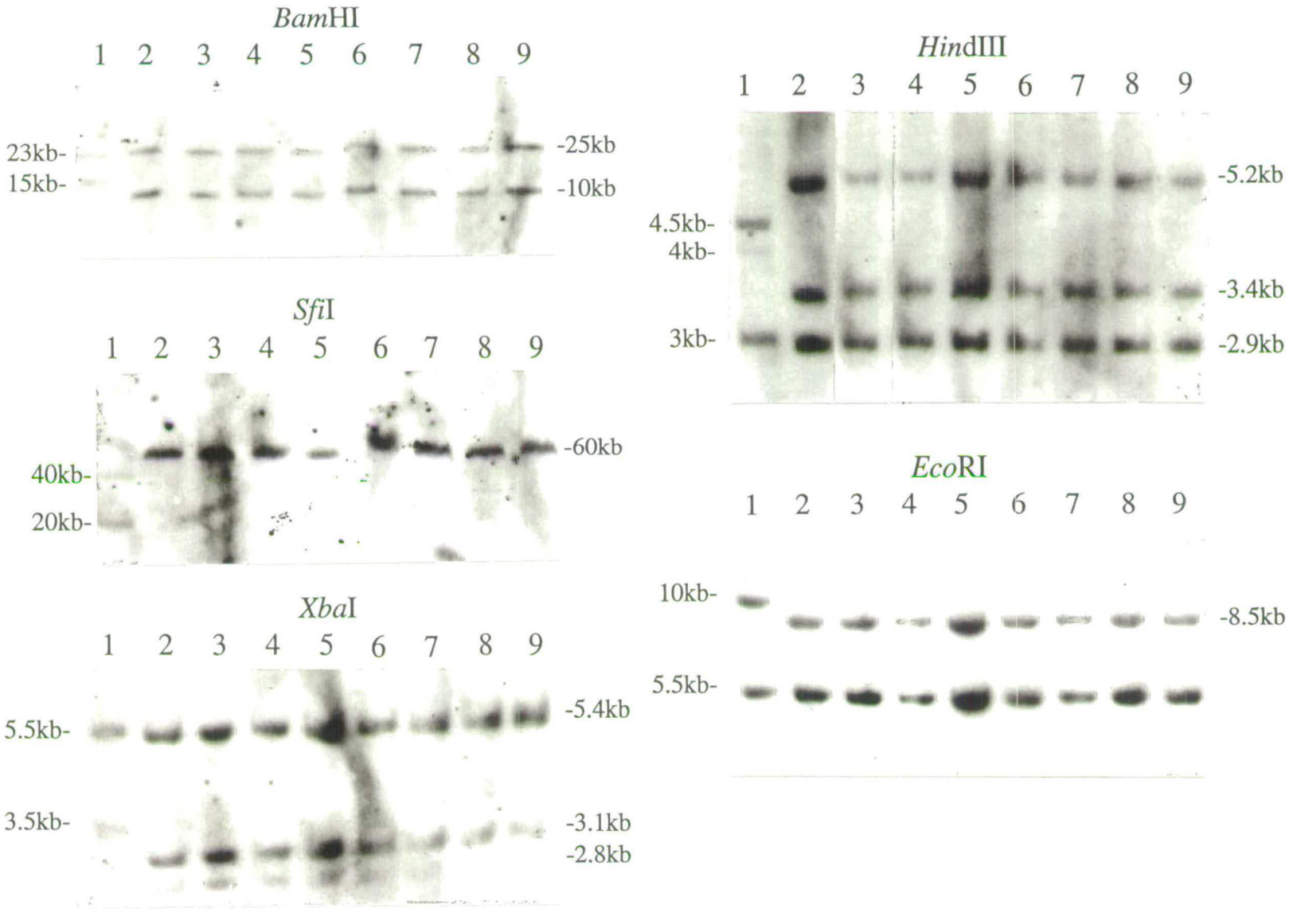

BamHI
1 2 3 4 5 6 7 8 9
23kb-
15kb-
-25kb
-10kb
SfiI
1 2 3 4 5 6 7 8 9
40kb-
20kb-
-60kb
XbaI
1 2 3 4 5 6 7 8 9
5.5kb-
3.5kb-
-5.4kb
-3.1kb
-2.8kb
HindIII
1 2 3 4 5 6 7 8 9
4.5kb-
4kb-
3kb-
-5.2kb
-3.4kb
-2.9kb
EcoRI
1 2 3 4 5 6 7 8 9
10kb-
5.5kb-
-8.5kb

Fig 6.3

Fig. 6.4. Restriction fragment length variant (RFLV) analysis of the ε-casein locus of *Mus spretus* and eight inbred strains of *Mus musculus*.

They are 1. *Mus spretus* 2. 129 3. C57/BL6 4. C3H 5. DBA 6. NIH 7. BALB/c 8. SJL 9. CBA. The enzymes used are shown on top of each autoradiogram. The sizes of the fragments are shown in kb.

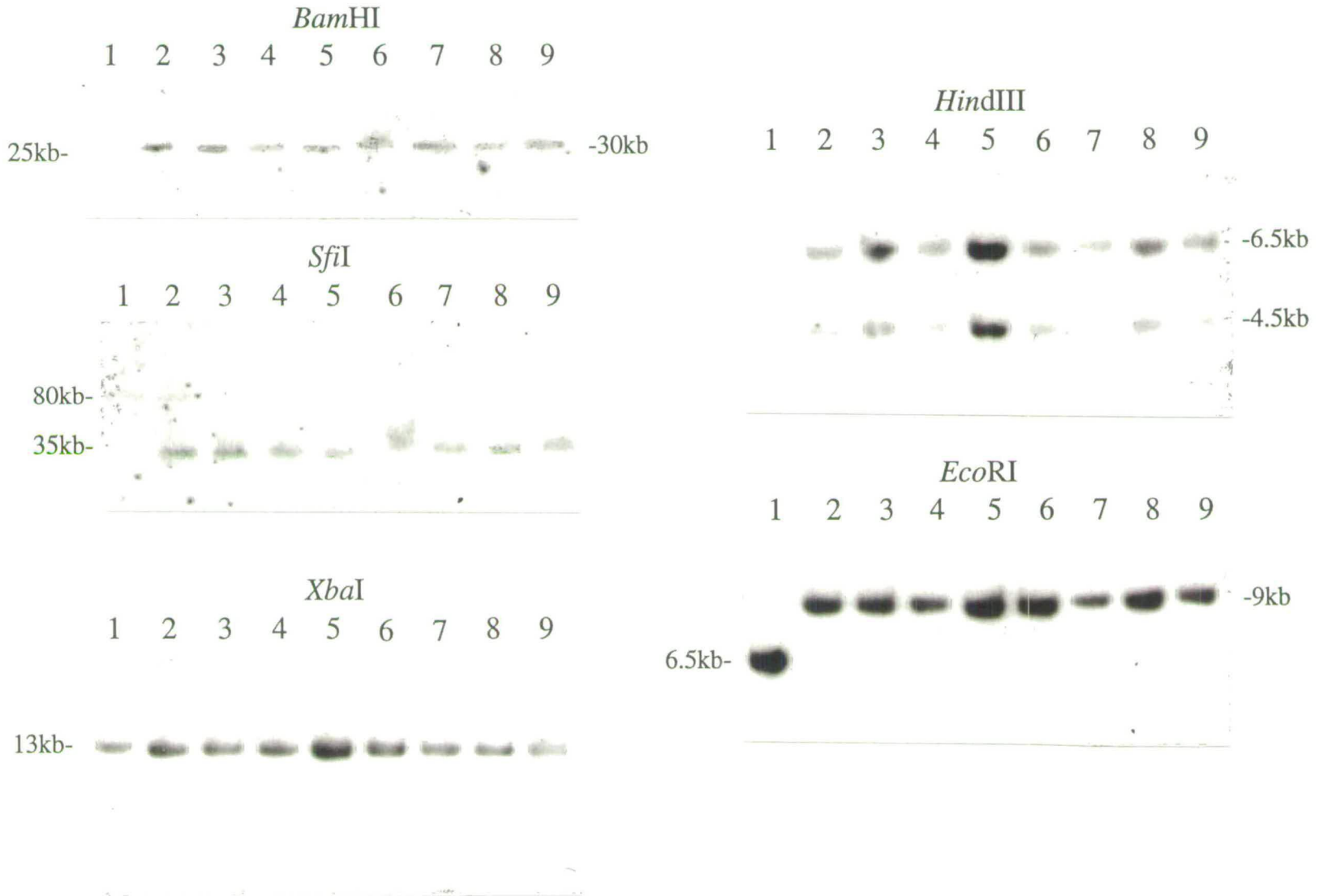

BamHI
1 2 3 4 5 6 7 8 9
25kb-
-30kb
SfiI
1 2 3 4 5 6 7 8 9
80kb-
35kb-
XbaI
1 2 3 4 5 6 7 8 9
13kb-
HindIII
1 2 3 4 5 6 7 8 9
-6.5kb
-4.5kb
EcoRI
1 2 3 4 5 6 7 8 9
-9kb
6.5kb-

Fig 6.4

κ-casein locus:

Three enzymes yielded polymorphic fragments in the κ-casein locus (Fig. 6.5). The enzymes and fragment sizes are detailed below. *Sfi*I and *Eco*RI generated similar sized fragments, 35 and 7 kb respectively in both the species.

	Fragment size (kb)		
Restriction enzyme	*Bam*HI	*Xba*I	*Hin*dIII
Mus spretus	25	2.8 + 3.9	3.0
Mus musculus	20	2.8 + 5.0	2.7 + 3.0 + 3.8

Thus, restriction fragment length variation analyses with five different enzymes indicated biallelic polymorphic differences in the five casein loci between *Mus musculus* and *Mus spretus* while no polymorphism was observed in the inbred strains of *Mus musculus*.

6.3. Discussion

RFLP analysis of the inbred strains of *Mus musculus* showed monomorphic restriction fragments at the casein loci with all selected enzymes. This suggests the occurrence of similar alleles in all the five casein loci, α, β, γ, ε and κ of the analysed strains. Since the various strains under study were generated by inbreeding which allows homogenisation of the DNA, the presence of the same allele in all these strains was expected. Piletz and Ganschow (1981) analysed milk proteins from 58 inbred strains of mice and showed a protein variant, probably a casein which was electrophoretically separated on pH 3.7 urea gels but was not detectable on pH 9.5 urea gels, only in *Mus musculus castaneus*. The absence of casein variants with the other 57 strains supports the present finding.

Polymorphic differences between *Mus musculus* and *Mus spretus* were identified for all five casein loci, with the two species apparently fixed for different alleles at each locus. This finding is in accordance with studies of many

Fig. 6.5. Restriction fragment length variant (RFLV) analysis of the κ-casein locus of *Mus spretus* and eight inbred strains of *Mus musculus*.

They are 1. *Mus spretus* 2. 129 3. C57/BL6 4. C3H 5. DBA 6. NIH 7. BALB/c 8. SJL 9. CBA. The enzymes used are shown on top of each autoradiogram. The sizes of the fragments are shown in kb.

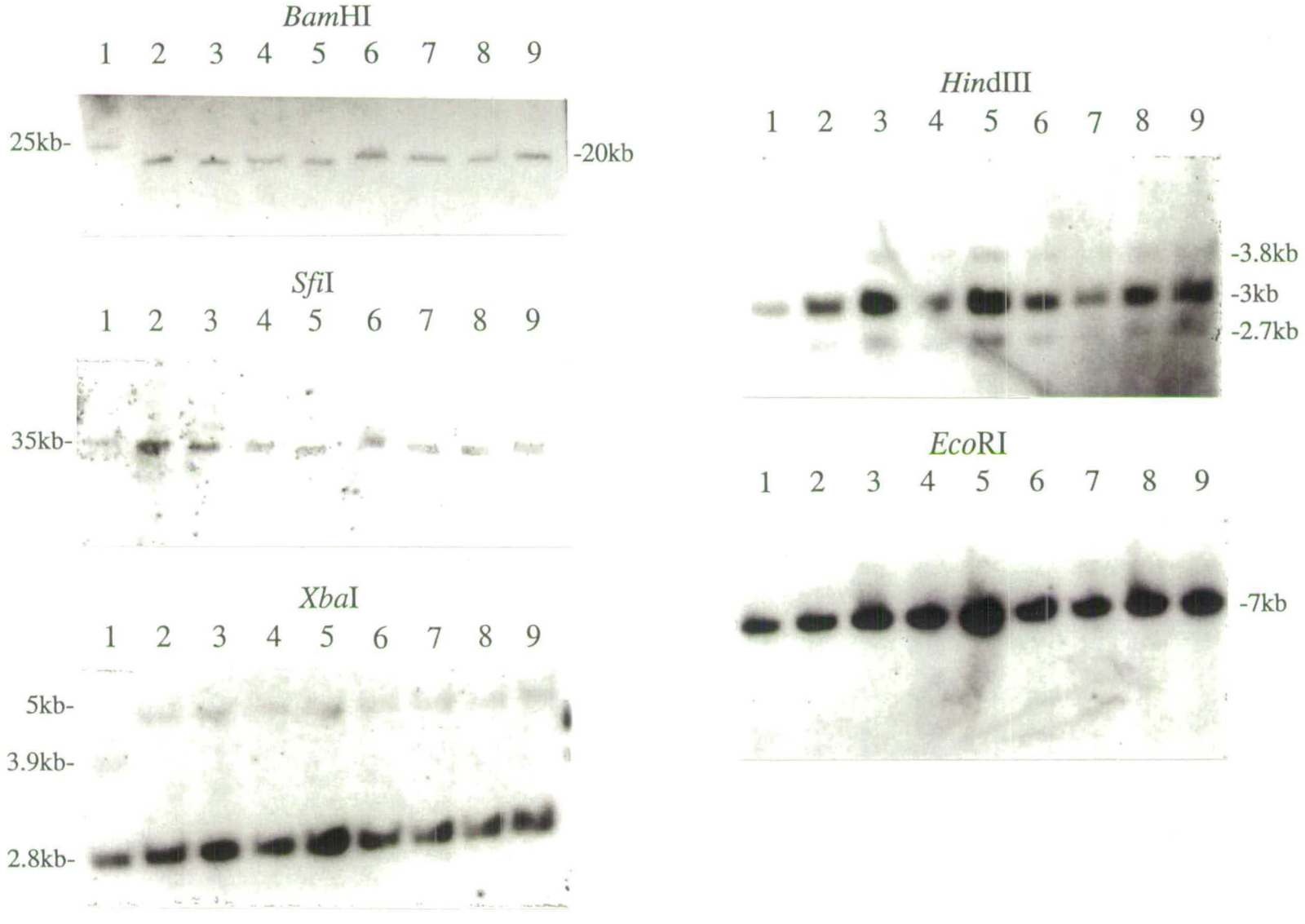

*Bam*HI
1 2 3 4 5 6 7 8 9
25kb-
-20kb
*Sfi*I
1 2 3 4 5 6 7 8 9
35kb-
*Xba*I
1 2 3 4 5 6 7 8 9
5kb-
3.9kb-
2.8kb-
*Hin*dIII
1 2 3 4 5 6 7 8 9
-3.8kb
-3kb
-2.7kb
*Eco*RI
1 2 3 4 5 6 7 8 9
-7kb

Fig 6.5

other loci in these two species of mice. The level of polymorphism was high at the casein loci except at β-locus, in which only *Bam*HI and *Sfi*I showed polymorphic sites. All five restriction enzymes, *Bam*HI, *Sfi*I, *Xba*I, *Hin*dIII and *Eco*RI showed polymorphic sites in the γ-casein locus. Four enzymes revealed polymorphic sites in the α and ε-loci while in the κ-casein locus three enzyme sites were polymorphic.

Restriction enzymes, *Bam*HI and *Sfi*I were used to construct the physical map of the casein locus (stain 129 of *Mus musculus*). Therefore, comparison of the fragments observed in the casein loci of the two species might help to understand the location of the sites of these enzymes in the two species.

α-β loci:

Two independent *Bam*HI fragments were observed for the α (25 kb) and β (20 kb)-casein loci of *M. spretus* while a common 40 kb fragment was revealed with both α and β-casein probes in addition to an independent 25 kb fragment found for α in *M. musculus*. Thus, the *Bam*HI fragments in *M. spretus* do not indicate any evidence for the linkage between the α and β casein genes. The enzyme, *Sfi*I produced 70 kb + 40 kb and 70 kb + 30 kb fragments in the α and β-casein loci of *M. spretus* respectively, where the smaller fragments together (40 + 30 = 70) constitute the size of the large fragment (70 kb), suggesting that the 70 kb fragment is a partially digested fragment and giving fairly good evidence for the physical linkage of the two genes in *M. spretus*. The low hybridisation intensity of the 70 kb fragment compared to the other fragments is also suggestive of partial digestion. This suggests that there is a *Sfi*I site between the α and β-loci in *M. spretus* with the genes contained in 40 and 30 kb *Sfi*I fragments respectively against a single 50 kb fragment encompassing both the genes in *M. musculus*. The

variation in sizes of *Bam*HI and *Sfi*I fragments between the two species propose the concurrent addition and deletion of these enzyme sites in *M. spretus*.

As per the physical map generated (*M. musculus*), it is evident that the α and β casein genes are located very close to each other with an intergenic region of about 8 kb. Comparing the sizes of the *Bam*HI fragments (α + β-loci), the fragments together constitute 45 kb and 65 kb in *M. spretus* and *M. musculus* respectively. Examinig the evolutionary divergence of satellite DNA sequences, revealed a marked reduction in the relative abundance of major satellite DNA, probably due to deletion, in *M. spretus* (~1%), as compared to *M. musculus* (~10%) (Brown and Dover, 1980; Matsuda and Chapman, 1991). It has also been shown that the reduction in the centromeric satellite DNA was not localized to a specific chromosome, but dispersed on several chromosomes (Matsuda and Chapman, 1991). Therefore, an alternate suggestion is, perhaps a deletion occurred in *M. spretus* DNA might have caused a shift in the enzyme sites across the casein locus, which account for the variation in the fragment sizes. This speculatation can be clarified only after detailed physical mapping studies in the casein locus of *M. spretus*.

Polymorphic sites for the frequent cutting enzymes, such as *Xba*I and *Eco*RI were also present in α-casein locus but all three enzymes, *Xba*I, *Hin*dIII and *Eco*RI showed monomorphic fragments for the β-casein locus.

γ, ε and κ-loci:

Small variations in size of *Bam*HI fragments of γ, ε and κ casein loci (γ- *M. spretus* ~23 and ~15 kb against 25 and 10kb in *M. musculus*; ε- ~25 kb in *M. spretus* against 30 kb in *M. musculus*; κ- ~25 kb in *M. spretus* against 20 kb in *M. musculus*) were also observed between *M. spretus* and *M. musculus* while the number of fragments were the same in both the species. But with the enzyme *Sfi*I,

the γ-casein locus of *M. spretus* showed two fragments (20 + 40 kb) against one fragment (60 kb) in *M. musculus*. This suggests the formation of an additional *Sfi*I site in the γ-locus of *M. spretus* or the loss of the site in *M. musculus*. The *Sfi*I fragment observed in ε-locus was larger in size (~80 kb) in *M. spretus* compared to the fragment in *M. musculus* (35 kb). The variations in sizes might be due to the appearance of new sites as well as deletion of existing sites during evolution.

The enzymes *Xba*I, *Hin*dIII and *Eco*RI also showed polymorphic sites in the γ-casein locus. Between the two species variations in the number and size of fragments were observed with *Xba*I. Restriction enzyme, *Hin*dIII produced fragments of varying size but of same number in the two species. *Eco*RI produced two fragments, of which one was of similar size in both the species. Polymorphic fragments were detected in ε-casein locus with *Hin*dIII and *Eco*RI. *Hin*dIII produced a similar sized fragment in the two species. *Xba*I and *Hin*dIII also showed variations in fragment sizes at the κ-casein locus. *Xba*I revealed one fragment of the same size at the κ-locus in both the species. Presence of similar sized fragments with an enzyme at the same locus of the two species suggest that, at least two sites for that enzyme are located at similar positions in that particular locus.

Restriction fragment length polymorphisms with various enzymes have also been identified in the casein loci of other species such as, bovines, ovines, caprine and porcines (Threadgill and Womack, 1990; Di Gregorio et al., 1991; Leroux et al., 1990; Archibald et al., 1994).

As described in the previous sections (chapter 3 and 4) the casein genes are physically linked and contained in a 250 kb genomic fragment in mouse genome. Similar findings have also been reported in cattle (Threadgill and Womack, 1990; Ferretti et al., 1990) and goat (Leroux and Martin, 1996). Lien et al. (1993) analysed 330 single sperm cells and observed no recombination in the

casein loci. A similar, lack of recombination was also observed between the casein loci of pigs (Archibald et al., 1994). All these findings indicate tight genetic linkage between the casein loci as would be predicted from the close physical proximity of the individual casein genes to one another.

Crosses between laboratory mice and the highly divergent wild-derived Mus species, *Mus spretus*, are a powerful tool for the analysis of genetic linkage (Avner et al., 1988). The polymorphic information obtained from the restriction analysis can be used to locate the casein gene cluster in the genetic map of mouse chromosome 5. The limited linkage data for the murine α and β casein loci places them 1 cM apart (Geissler et al., 1988). But in the physical map constructed, I have shown that the α and β casein genes are situated only ~8 kb apart. However, the genetic distance between the α and β genes has been estimated on the basis of a single/putative recombination event. Thus, the current apparent overestimate of the genetic distance between the α and β genes may be explained in several ways - chance observation of a rare event, a genotyping error or a real deviation from the average relationship between genetic and physical distances (1 cM = ~1.7 Mb in mice). In collaboration with Nancy Jenkins and Neal Copeland, Mammalian Genetics Laboratory, ABL-Basic Research programme, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA, the casein gene cluster has been added to their genetic (gene) map by linkage analysis using the κ-casein RFLP. As expected the κ-casein locus maps to chromosome 5 between Pdgfra and Alb1. The limited recombination close to the κ-casein marker would make it very difficult to orientate the casein gene cluster by linkage analysis in this population (Fig. 6.6). Their extensive gene map of the mouse is based on a *M. musculus*/ *M. spretus* cross (Dietrich et al., 1996).

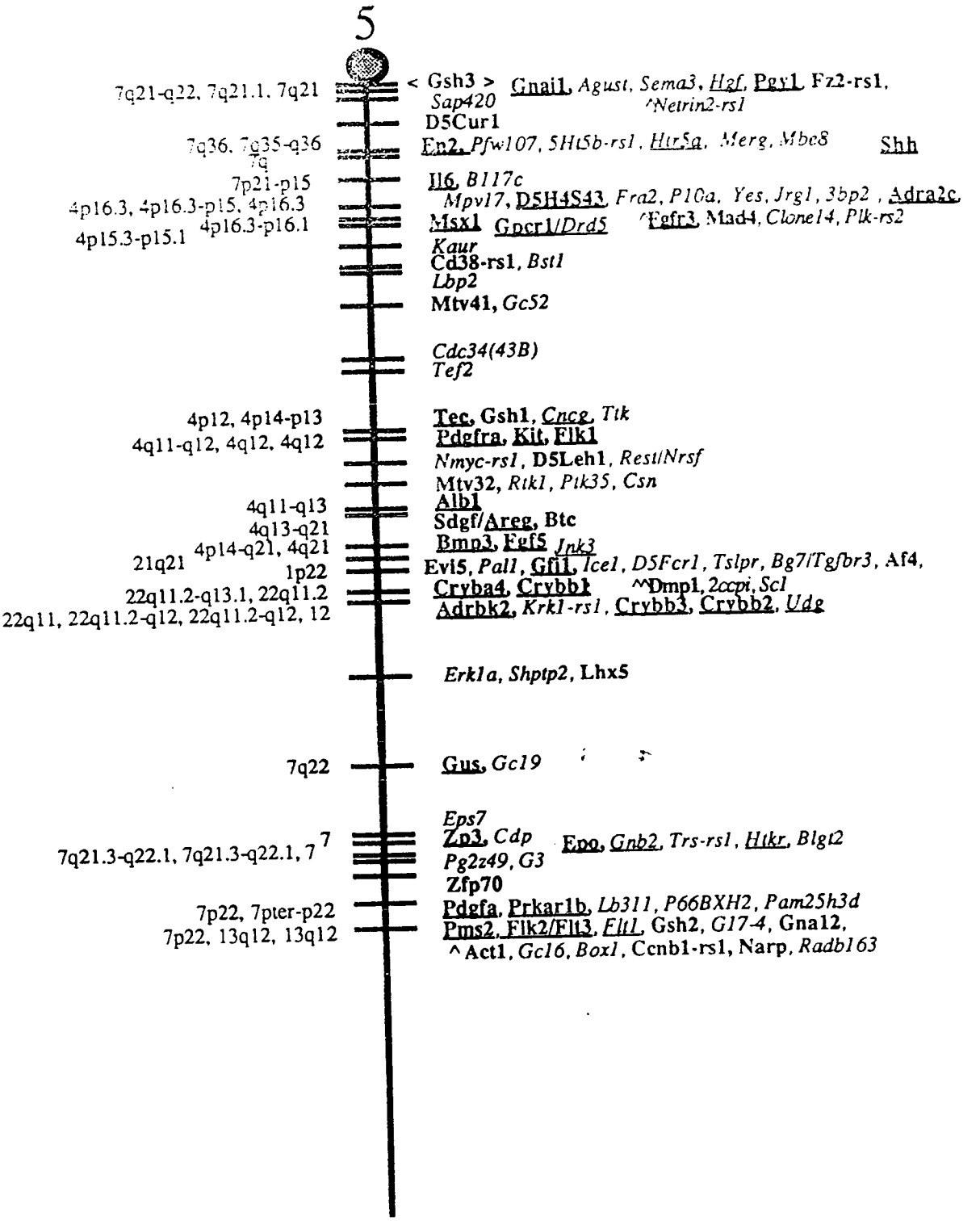


Fig. 6.6. Linkage map of mouse chromosome 5 showing the location of casein gene cluster between Pdgfra and Alb1. The cluster has been added to the genetic map by linkage analysis using the kappa-casein RFLP.

Chapter7

GENERAL DISCUSSION

The data obtained in the present study permit me to confirm the existence of five different casein genes and their position in the mouse genome. The whole casein locus is containined in a 250 kb genomic DNA fragment, in which the genes are arranged in the order α, β, γ, ε and κ. The α, γ and κ genes are oriented in the same direction whilst the β gene is in the opposite direction. The orientation of the ε gene remains unresolved.

7.1. Evolutionary analysis of casein genes

Knowledge of the locus organization allows me to draw conclusions regarding the evolution of casein genes. Nucleotide sequence analysis of a number of casein mRNA/cDNAs from different species reveals considerable species divergence. However, the hypothesis of a common ancestor for the α and β genes is supported by the presence of common sequence motifs in the proximal 5' flanking regions and the similar organization pattern of the first four exons (reviewed by Mercier and Vilotte 1993). Although the κ-casein gene is linked to the other casein genes, in mice (finding of the present research), cattle (Threadgill and Womack, 1990) and pigs (Archibald et al., 1994, 1995; Rohrer et al., 1994) it does not share any common sequence motifs with the other casein genes and is thought to be evolved from the fibrinogen gene family (Thompson et al., 1985). Interestingly, the fibrinogen gene cluster maps to the same chromosome as the casein genes in humans (HSA 4) and pigs (SSC 8) but not in cows (fibrinogen, BTA 17; casein BTA 6) or mice (fibrinogen, MMU 3; casein MMU 5). Comparison of the bovine/caprine and murine casein loci suggests that a further gene duplication has occurred in the murine lineage with murine γ and ε sharing a common ancestor with bovine/caprine αS2. The occurrence of a fifth casein gene

is not recorded in cattle or goats. Thus, the presumed homology between the genes encoding murine α, β and κ and bovine/caprine αS1, β and κ caseins is supported by sequence similarity and conserved gene order.

Thus, the caseins can be considered to be encoded by members of two small gene families - the casein gene family (α, β, γ and ε) and the fibrinogen gene family (κ). The functions of members of gene families range from those with extremely diverse functions, such as the protein kinase family (Hunter, 1987), those with very variable members such as immunoglobulin genes (Hood et al., 1975) and to those consisting of uniform members like ribosomal RNA genes (Dover, 1982). All existing gene families are the products of millions of years of evolution, and the organization and diversity of each family reflect the effects of natural selection. The simplest way to acquire new functions would be through duplication of a whole gene with subsequent differentiation by nucleotide substitution (Ohta, 1983a).

The clustering of the casein genes (α, β, γ, ε) can be explained by the gene duplication mechanism. The tight linkage between the κ-casein and fibrinogen genes in pigs supports the hypothesis that these are members of the same gene family, that have arisen by gene duplication. The proximity of the putative ancestral casein and fibrinogen genes is retained in pigs but has been progressively disrupted in other lineages. In humans the fibrinogen and casein genes are still on the same chromosome but are now much further apart possibly as a result of an intra-chromosomal inversion (Ellegren et al., 1993). The location of the κ-casein gene in humans is unknown.

7.2. Regulation of expression

Once the arrangement of the casein genes - a tightly linked cluster - was exposed it was of interest to look into their expression profiles. Close proximity

and co-ordinated expression of the casein genes, α, β and γ might imply the presence of a common regulatory region such as a locus control region (LCR). The map that I have developed will be of great help in unraveling the regulatory regions, independent or shared between the casein genes. Knowledge of such regulatory regions especially if shared elements such as a LCR can be identified would not only contribute to the understanding of the control of gene expression in gene clusters but would aid transgene construct design. For example, such information would allow the design of transgenes that are more likely to exhibit correct spacial and temporal regulation of expression in the mammary gland. Using my detailed map and the characterised contig/clones as a starting point and tools respectively, one of the post -doctoral scientists in our Division will scan the murine casein locus for a locus control region.

7.3. Application in transgenic studies

The characterisation of the murine casein locus has potentially significant applications in the field of transgenic studies pertaining to the production of novel proteins through milk as well as altering the composition of milk. The advent of transgenic animals and cloning vectors like YACs (yeast artificial chromosomes), BACs (bacterial artificial chromosomes) and PACs (P1 based artificial chromosomes) (Monaco and Larin, 1994) has made it feasible to manipulate large genes or even gene complexes in the laboratory before introducing them into the genome. It has been shown that such large cloned fragments can be introduced into the mouse germ line (Schedl et al., 1993) and transgenic mice carrying such exogenous DNA can be established. Thus, the BAC clones that I have characterised could allow transgenic approaches to study the entire casein locus.

Transgenic mice in which the transgene is delivered as a large fragment - from a YAC, PAC or BAC - may allow the gene of interest to be studied in

approximately the right context. However, the vagaries of random integration and the resulting position effects limit the conclusions that can be drawn from transgenic animals generated by microinjection.

Thus, increasingly genetic material is introduced into transgenic mice by homologous recombination in embryo-derived stem (ES) cells such that the site of integration is known (McFarlane and Wilson, 1996). As yet, such gene targetting technology has not been established in dairy animals such as cattle, sheep and goats. Nevertheless, for germ line manipulation studies of the casein locus of higher animals, the mouse can be used as an excellent model system for evaluating the effects of locus manipulation and can be achieved through ES cell manipulation. Correct modification of the endogenous genes is possible by homologous recombination techniques (Capecchi, 1989). Gene 'knockout' studies will help to understand the functions of specific genes *in vivo*. For example, no deleterious effect was detected in β-casein deficient mice either in fertility or in lactation and nurturing the young. However, in such mice the casein micelle size was significantly smaller, probably due to an increase in κ-casein content (Kumar et al., 1994). In contrast, to the β-casein deficient mice, mice developed with a knocked-out α-lactalbumin gene have problems during lactation due to their inability to synthesise lactose. The milk was highly concentrated and viscous and the pups were unable to suck it from the mammary gland (Stinnakre et al., 1994). Not only can genes be knocked out by homologous recombination, it is possible to replace a gene with another one (Stacey et al., 1994) and preciesly substitute a single base pair (Hasty et al., 1991) at a particular locus.

Appropriate well mapped cloned DNA for the gene(s) of interest is essential for genetic manipulation by homologous recombination. Moreover, there is clear evidence that the success or frequency of homologous recombination is increased by using DNA that is identical to or isogenic with the target locus

(teRiele et al., 1992). Most mouse ES cell lines are derived from the 129 strain of inbred mice. Thus, the clones (that comprise the casein locus contig described here, were derived from 129 mouse DNA) along with the map information will be invaluable for future genetic manipulation of the casein genes/locus in transgenic mice. Indeed, Dr. Satish Kumar, Centre for Cell and Molecular Biology, Hyderabad, India has initiated such experiments using my clones and map information.

The casein genes are found physically close to one another not only in mice (within a 250 kb fragment) but also in cattle (within a 200 kb fragment, Threadgill and Womack, 1990; Ferretti et al., 1990) and goat (within 250 kb fragment, Leroux and Martin, 1996). The casein genes are also very tightly linked in terms of genetic map or recombination distances. For example, no recombinants were observed between the bovine casein genes in a study of 330 single sperm cells from a bull heterozygous at the α, β and κ casein loci (Lien et al., 1993). Similarly an absence of recombinations between casein genes have also been reported in pigs (Archibald et al., 1994) and in sheep (Leveziel et al., 1991).

Although, the frequency of recombination within the casein locus is low many different casein haplotypes are observed in cattle (reviewed by Ng-Kwai-Hang and Grosclaude, 1992). Therefore, despite the lack of observed recombinants there is scope for selection of desirable casein haplotypes. Such marker assisted selection should be successful as the disruption by recombination of the selected haplotype should be rare. The lack of recombination, however, limits the scope for establishing new haplotypes by conventional breeding. Thus, the use of transgenic technology, once gene targetting methods are established in dairy animals, may be an efficient means for altering the composition of milk. Meanwhile such changes can be modelled in transgenic mice.

7.4. Application to enrich mouse genetic map

The lack of recombination within the casein locus will make it difficult to orient the casein gene cluster on the mouse genetic (linkage) map. However, by mapping one of the restriction fragment length polymorphisms described here the casein gene cluster can be added to the mouse gene and genetic maps. In the physical map I have generated, the α and β casein genes are situated only ~8 kb apart while the limited linkage data for the murine α and β casein loci places them 1 cM apart (Geissler et al., 1988), as estimated on the basis of a single/putative recombination event. Thus, the current apparent overestimate of the genetic distance between the α and β genes may be explained in several ways - chance observation of a rare event, a genotyping error or a real deviation from the average relationship between genetic and physical distances (1 cM = ~1.7 Mb in mice).

In collaboration with Nancy Jenkins and Neal Copeland, Mammalian Genetics Laboratory, ABL-Basic Research programme, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA, the casein gene cluster was assigned to chromosome 5 on their genetic (gene) map by linkage analysis using the κ-casein RFLP. As appropriate recombination events close to the κ-casein locus are not present in the Frederick mapping resource, it would not be worthwhile to score the α-casein RFLP in this population and orient the casein gene cluster. The limited number of meioses in the Frederick cross limits the scope for ordering genes that are physically very close to each other.

Although, at present, I will not be able to continue with developing the project, others have already started research on the mouse casein locus that depends on the resources and information that I have developed. But the training I received in molecular biology research methodology will certainly be useful to set up a molecular biology laboratory in the University where I am going to continue

my work (Kerala Agricultural University, India) as well as in my teaching and research career.

REFERENCES

Alexander, L.J and Beattie, C.W. (1992). The sequence of porcine beta casein cDNA. *Anim. Genet.* **23**: 369 - 371.

Alexander, L.J., Das Gupta, N.A., Beattie, C.W. (1992). The sequence of porcine alphaS2 casein cDNA. *Anim. Genet.* **23**: 365 - 367.

Alexander, L.J., Stewart, A.F., MacKinlay, A.G., Kapelinskaya, T.V., Tkach, T.M and Gorodetsky, S.I. (1988). Isolation and characterization of the bovine κ-casein gene. *Eur. J. Biochem.* **178**: 395 - 401.

Ali, S and Clark, A.J. (1988). Characterization of the gene encoding ovine β-lg similarity to the gene for retinol binding protein and other secretory proteins. *J. Mol. Biol.* **199**: 415 - 426.

Altiok, S and Groner, B (1994). β-casein messenger RNA sequesters a single stranded nucleic acid binding protein which negatively regulates the β-casein gene promoter. *Mol. Cell. Biol.* **14**: 6004 - 6012.

Andres, A.C., Schonenberger, C.A., Groner, B., Hennighausen, L., LeMeur, M and Gerlinger, P. (1987). Ha-ras oncogene expression directed by a milk protein gene promoter. Tissue specificity, hormonal regulation and tumor induction in transgenic mice. *Proc. Natl. Acad. Sci. USA.* **84**: 1299 - 1303.

Anema, S.G and Creamer, L.K. (1993). Effect of the A and B variants of both alphaS1 and kappa casein on bovine casein micelle solvation and kappa casein content. *J. Dairy. Sci.* **60**: 505 - 516.

Antequera, F and Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA.* **90**: 11995 - 11999.

Antequera, F., Boyes, J and Bird, A.P. (1990). High levels of *de novo* methylation and altered chromatin structure at CpG islands in cell lines. *Cell.* **62**: 503 - 514.

Antequera, F., Macleod, D and Bird, A.P. (1989). Specific protection of methylated CpGs in mammalian nuclei. *Cell.* **58**: 509 - 517.

Archibald, A.L; McClenaghan, M., Hornsey, V., Simons, J.P and Clark, A.J. (1990). High level expression of biologically active human α1-antitrypsin in the milk of transgenic mice. *Proc. Natl. Acad. Sci. USA.* **87**: 5178 - 5182.

Archibald, A.L., Couperwhite, S., Haley, C.S., Beattie, C.W and Alexander, L.J. (1994). RFLP and linkage analysis of the porcine casein loci - CASAS1, CASAS2, CASB and CASK. *Anim. Genet.* **25**: 349 - 351.

Archibald et al., (1995). The PiGMaP consortium linkage map of the pig (*Sus scrofa*). *Mammalian genome.* **6**: 157 - 175.

Arnheim, N. (1983). Concerted evolution of multigene families, In *Evolution of genes and proteins* (M. Nei and R.K. Koehn, eds), Sinauer, Sunderland, MA. pp. 38 - 61.

Aschaffenburg, R and Drewry, J. (1955). Occurrence of different β-lg in cow's milk. *Nature.* **176**: 218 - 219.

Ball, R.K., Friis, R.R., Schoenenberger, C.A., Doppler, W and Groner B. (1988). Prolactin regulation of beta casein gene expression and of a cytolytic 120kDa protein in a cloned mouse mammary epithelial cell line. *The EMBO. J.* **7**: 2089 - 2096.

Banerjee, M.R and Antoniou, M. (1985). Steroid and polypeptide hormone interaction in milk protein gene expression, In *Biochemical actions of hormones.* Vol XII, Academic Press, Inc. pp. 237 - 288.

Banerjee, M.R., Mehta, N.M., Ganguly, R., Majumder, P.K., Ganguly, N and Joshi, J. (1982). Selective gene expression in an isolated whole mammary organ *in vitro.* Cold Spring Harbor Conference on cell proliferation Book B, Growth of cells in hormonally defined media. 789 - 805.

Banerjee, M.R., Majumder, P.K., Antoniou, M and Joshi, J.B. (1983). Hormone inducible specific gene expression in an isolated whole mammary organ in serum free culture. Hormonally defined media: a tool in cell biology, lectures and posters presented at the 1st European conference on serum free cell culture, Heidelberg. 234 - 249.

Barlow, S.M., Morrison, P.J and Sullivan, F.M. (1974). Plasma corticosterone levels during pregnancy in the mouse: The relative contributions of the adrenal glands and foetoplacental units. *J. Endo.* **60**: 473 - 483.

Barrie, P.A., Jefreys, A.J and Scott, A.F. (1981). Evolution of the β-globin gene cluster in man and primates. *J. Mol. Biol.* **149**: 319 - 336.

Baruch, A., Shani, M., HurWitz, D.R and Barash, I. (1995). Developmental regulation of the ovine β-lactoglobulin/human serum albumin transgene is distinct from that of the β-lactoglobulin and the endogenous β-casein genes in the mammary gland of transgenic mice. *Develop. Genetics.* **16**: 241 - 252.

Bayna, E.M and Rosen, J.M. (1990). Tissue specific, high level expression of the rat whey acidic protein gene in transgenic mice. *Nucleic Acids Res.* **18**: 2977 - 2985.

Benvensity, N., Mencher, D., Meyuhas, O., Razin, A and Reshef, L. (1985). Sequential changes in DNA methylation patterns of the rat phosphoenolpyruvate carboxykinase gene during development. *Proc. Natl. Acad. Sci. USA.* **82**; 267 - 271.

Bhattacharjee, M and Vonderhaar, B.K. (1984). Thyroid hormone enhance the synthesis and secretion of α-lactalbumin by mouse mammary tissue *in vitro*. *Endo.* **115**: 1070 - 1077

Bird, A.P. (1986). CpG rich islands and the function of DNA methylation. *Nature.* **321**: 209 - 213.

Bird, A.P., Taggart, M.H., Nicholls, R.D and Higgs, D.R. (1987). Non-methylated CpG-rich islands at the human α-globin locus: implications for evolution of the α-globin pseudogene. *The EMBO. J.* **6**: 999 - 1004.

Bischoff, R., Degryse, E., Perraud, F., Dalemans, W., Ali-Hadji, D., Thepot, D., Devinoy, E., Houdebine, L.M and Pavirani, A. (1992). A 17.6 kbp region located upstream of the rabbit WAP gene directs high level expression of a functional human protein variant in transgenic mouse milk. *Fed. Eur. Biol. Soc. Lett.* **305**: 265 -

Blackburn, D.E., Hobbs, A.A and Rosen, J.M. (1982). Rat beta casein cDNA: Sequence analysis and evolutionary comparisons. *Nucleic Acids Res.* **10**: 2295 - 2307.

Bleck, G.T and Bremel, R.D. (1993). Variation in expression of a bovine α-lactalbumin trans gene in milk of transgenic mice. *J. Dairy. Sci.* **77**: 1897 - 1904.

Boisnard, M and Petrissant, G. (1985). Complete sequence of ovine alphaS2 casein mRNA. *Biochemie.* **67**: 1043 - 1051.

Bonifer, C., Vidal, M., Grosveld, F and Sippel, A.E. (1990). Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *The EMBO J.* **9**: 2843 - 2848.

Bonsing, J and Mackinlay, A.G. (1987). Recent studies on nucleotide sequences encoding the caseins. *J. Dairy. Res.* **54**: 447 - 461.

Bonsing, J., Ring, J.M., Stewart, A.F and Mackinlay, A.G. (1988). Complete nucleotide sequence of the bovine β-casein gene. *Aust. J. Biol. Sci.* **41**: 527 - 537.

Boulanger, A., Grosclaude, F and Mahe, M.F. (1984). Polymorphisme des caseines αS1 et αS2 de la chevre (Capra hircus). *Genet. Sel. Evol.* **16**: 157 - 175.*

Bovenhuis, H. (1992). Mapping QTL in dairy cattle using milk protein genes as genetic markers. *J. Dairy. Sci.* (Suppl.**1**), 285.

Bovenhuis, H. (1993). The relevance of milk protein polymorphisms for dairy cattle breeding. *Neth. Milk and Dairy J.* **47**: 46 - 47.

Bovenhuis, H., Van Arendonk, J.A and Krover, S. (1992). Association between milk protein polymorphisms and milk production traits. *J. Dairy. Sci.* **75**: 2549 - 2559.

Boyes, J and Bird, A.P. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell.* **64**: 1123 - 1134.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, S., Nemes, A., Temper, V., Razin, A and Cedar, H. (1994). Sp1 elements protect a CpG island from *de novo* methylation. 371: 435 - 438.

Breathnach, R and Chambon, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**: 349 - 383.

Brem, G., Hartl, P., Besenfelder, U., Wolf, E., Zinovieva, N and Pfaller, R. (1994). Expression of synthetic cDNA sequences encoding human insulin like growth factor I (IGFI) in the mammary glands of transgenic rabbits. *Gene.* **149**: 351 - 355.

Brew, K and Hill, R.L. (1975). Lactose biosynthesis. *Rev. Physiol. Biochem. Pharmacol.* **72**: 105 - 157.

Brew, K., Vanaman, T and Hill, R.L. (1967). Comparison of the amino acid sequence of bovine α-lactalbumin and hen egg white lysozyme. *J. Biol. chem.* **242**: 3747 - 3749.

Brignon, G., Chtourou, A and Ribadeau-Dumas, B. (1985). Preparation and amino acid sequence of human κ-casein. *Fed. Eur. Bioch. Soc. Lett.* **188**: 48 - 54.

Brignon, G., Mahe, M.F., Grosclaude, F and Ribadeau-Dumas, B. (1989). Sequence of caprine αS1-casein and characterization of those of its genetic variants which are synthesised at a high level, αS1-Cn A, B and C. *Protein Seq. Data. Anal.* **2**: 181 - 188.

Brignon, G., Mahe, M.F., Ribadeau-Dumas, B., Mercier, J.C and Grosclaude, F. (1990). Two of the three genetic variants of goat αS1-casein which are synthesised at a reduced level have an internal deletion possibly due to altered RNA splicing. *Eur. J. Biochem.* **193**: 237 - 241.

Brignon, G., Ribadeau-Dumas, B., Mercier, J.C and Pelissier, J.P. (!977). Complete amino acid equence of bovine αS2-casein. *Fed. Eur. Bioch. Soc. Lett.* **76**: 274 - 279.

Brodbeck, U and Ebner, K.E. (1966). Resolution of a soluble lactose synthetase into two protein components and solubilization of microsomal lactose synthetase. *J. Biol. Chem.* **241**: 762 - 764.

Brown, W.R.A and Bird, A.P. (1986). Long-range restriction site mapping of mammalian genomic DNA. *Nature.* **322**: 477 - 481.

Brown, S.D.M and Dover, G.A. (1980). Conservation of segmental variants of satellite DNA of *Mus musculus* in a related species: *Mus spretus*. *Nature.* **285**: 47 - 49.

Buhler Th.A., Bruyere Th., Went, D.F., Stranzinger, G and Burki, K. (1990). Rabbit beta casein promoter direct secretion of human interleukin2 into the milk of transgenic rabbits. *Bio/Technology.* **8**: 140 - 143.

Burdon, T., Sankaran, L., Wall, R.J., Spencer, M and Hennighausen, L. (1991). Expression of a whey acidic protein transgene during mammary development. *J. Biol. Chem.* **266**: 6909 - 6914.

Burdon, T., Wall, R.J., Shamay, A., Smith, G.H and Hennighausen, L. (1991). Over-expression of an endogenous milk protein gene in transgenic mice is associated with impaired mammary alveolar development and a milchlos phenotype. *Mech. Dev.* **36**: 67 - 74.

Busslinger, M., Hurst, J and Flavell, R.A. (1983). DNA methylation and the regulation of globin gene expression. *Cell.* **34**: 197 - 206.

Busslinger, M., DeBoer, E., Wright, S., Grosveld, F.G and Flavell, R.A. (1983). The sequence GGCmCGG is resistant to *Msp*I cleavage. *Nucleic Acids Res.* **11**: 3559 - 3569.

Campbell, S.M., Rosen, J.M., Hennighausen, L.G., Strech-Jurk, U and Sippel, A.E. (1984). Comparison of the whey acidic protein genes of rat and mouse. *Nucleic Acids Res.* **12**: 8685 - 8697.

Capecchi, M.R. (1989). Altering the genome by homologous recombination. *Science.* **244**: 1288 - 1292.

Carotti, D., Palitti, F., Lavia, P and Strom, R. (1989). *In vitro* methylation of CpG-rich islands. *Nucleic Acids Res.* **17**: 9219 - 9229.

Cartwright, T. (1987). Isolation and purification of products from animal cells. *Trends. Biotechnol.* **5**: 25 - 30.

Cedar, H. (1988). DNA methylation and gene activity. *Cell.* **53**: 3 - 4.

Cedar, H., Solage, A., Glaser, G and Razin, A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme *Msp*I. *Nucleic Acids Res.* **6**: 2125 - 2132.

Chen, L.H and Bissel, M.J. (1989). A novel regulatory mechanism for whey acidic protein gene expression. *Cell Regul.* **1**: 45 - 54.

Choi, Y.-C and Chae, C.-B. (1993). Demethylation of somatic and testis-specific histone H2A and H2B genes in F9 embryonal carcinoma cells. *Mol. Cell. Biol.* **13**: 5538 - 5548.

Chomczynski, P., Qasba, P and Topper, Y.J. (1986). Transcriptional and post transcriptional roles of glucocorticoid in the expression of the rat 25000 molecular weight casein gene. *Biochem. Biophys. Res. Commun.* **134**: 812 - 818.

Clark, A.J., Bessos, H., Bishop, J.O., Brown, P., Harris, S., Lathe, R., McClenaghan, M., Prowse, C., Simons, J.P., Whitelaw, C.B.A and Wilmut, I.

(1989). Expression of human anti haemophilic factor IX in the milk of transgenic sheep. *Bio/Technology.* **7**: 487 - 492.

Clarke, R.A., Sokol, D., Rigby, N., Ward, K., Murray, J.D and MacKinlay, A.G. (1994). Mammary gland specific expression of bovine αS1-casein derived transgenes in mice. *Transgenics.* **1**: 313 - 319.

Cleary, M.L., Schon, E.N and Lingrel, J.B. (1981). *Cell.* **26**: 181.

Collect, C., Joseph, R and Nicholas, K. (1990). Cloning, cDNA analysis and prolactin dependent expression of a marsupial α-lactalbumin. *Reprod. Fertil. Dev.* **2**: 693 - 701.

Collect, C., Joseph, R and Nicholas, K. (1991). A marsupial β-lactoglobulin gene: characterization and prolactin dependent expression. *J. Mol. Endocrinol.* **6**: 9 - 16.

Cooper, D.N., Taggart, M.H and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. *Nucleic Acids Res.* **11**: 647 - 658.

Cronin, M.A and Crocket, N. (1993). Kappa casein polymorphisms among cattle breeds and bison herds. *Anim. Genet.* **24**: 135 - 138.

Dale, T.C., Krnacik, M.J., Schmidhauser, C., Yang, C.L.Q., Bissel, M.J and Rosen, J.M. (1992). High level expression of the rat whey acidic protein gene is mediated by elements in the promoter and 3' untranslated region. *Mol. Cell. Biol.* **12**: 905 - 914.

Damiani, G., Ferretti, L., Rognoni, G and Sgaramella, V. (1989). Restriction fragment length polymorphism analyses of the kappa casein locus in cattle. *Anim. Genet.* **21**: 107 - 114.

Danielson, K.G., Oborn, C.J., Durban, E.M., Butal, J.S and Medina, D. (1984). Epithelial mouse mammary cell line exhibiting normal morphogenesis *in vivo* and functional differentiation *in vitro*. *Proc. Natl. Acad. Sci. USA.* **81**: 3756 - 3760.

David-Inouye, Y., Couch, C.C.H and Rosen, J.M. (1986). The isolation and transfection of the entire rat beta casein gene. Metabolic regulation: Application of recombinant DNA techniques. *Annals. NY. Acad.Sci.* **478**: 274 - 277.

Davis, J.A and Linzer, D.I.H. (1989). Expression of multiple forms of the prolactin receptor in mouse liver. *J. Mol. Endocrinol.* **3**: 674 - 679.

Dayhoff, M.O.(Ed). (1976). *Atlas of protein sequence and structure,* Vol. **5**, Suppl. 2, Natl. Biomed. Res. Found., Maryland.

de Jong, W.W., Hendriks, W., Mulders, J.W.M and Bloemendal, H. (1989). Evolution of eye lens crystallins: the stress connection. *Trends Biochem. Sci.* **14**: 365 - 368.

den Dunnen, J.T., Moormann, R.J.M., Lubsen, N.H and Schoenmakers, J.G.G. (1986). Concerted and divergent evolution within the rat γ-crystallin gene family. *J. Mol. Biol.* **189:** 37 - 46.

Devinoy, E,. Houdebine, L.M and Delouis, C. (1978). Role of prolactin and glucocorticoids in the expression of casein genes in rabbit mammary gland organ culture: Quantification of casein mRNA. *Biochem. Biophys. Acta.* **517**: 360 - 366.

Devinoy, E., Thepot, D., Stinnakre, M.G., Fontaine, M.L., Grabowski, H., Puissant, C., Pavirani, A and Houdebine, L.M. (1994). High level production of human growth hormone in the milk of transgenic mice; the upstream region of the rabbit whey acidic protein (WAP) gene targets transgenic expression to the mammary gland. *Transgenic Res.* **3**: 79 - 89.

Dietrich W.F., Miller, J., Steen, R and 18 others. (1996). A comprehensive genetic map of the mouse genome. *Nature.* **380**: 149 - 152.

Di-Gregorio P., Rando A., Pieragostini E., and Masina, P. (1991). DNA polymorphism at the casein loci in sheep. *Anim. Genet.* **22**: 21 - 30.

Dillon N and Grosveld, F. (1993). Transcriptional regulation of multigene loci: multilevel control. *Trend. genet.* **9**: 134 - 137.

DiTullio, P., Cheng, S.H., Marshall, J., Gregory, R.J., Ebert, K.M., Meade, H.M and Smith, A.E. (1992). Production of cystic fibrosis transmembrane conductance regulator in the milk of transgenic mice. *Bio/Technology.* **10**: 74 - 77.

Doppler, W., Groner, B and Ball, R.K. (1989). Prolactin and glucocorticoid hormones synergistically induce expression of transfected rat beta casein gene promoter constructs in a mammary epithelial cell line. *Proc. Natl. Acad. Sci. USA.* **86**: 104 - 108.

Doppler, W., Viilunger, A., Jennewein, P., Brduscha, K., Groner, B and Ball, R.K. (1991). Lactogenic hormone and cell type-specific control of the whey acidic protein gene promoter in transfected mouse cells. *Mol. Endo.* **5**: 1624 - 1632.

Dover, G and Coen, E. (1981). *Nature.* **290**: 731.

Dover,G.A. (1982). Molecular drive: A cohesive mode of species evolution. *Nature.* **299**: 111 - 117.

Drohan, W.N., Zhang, D., Paleyanda, R.K., Chang, R., Wroble, M., Velander, W and Lubon, H. (1994). Inefficient processing of human protein-C in the mouse mammary gland. *Transgenic Res.* **3**: 355 - 364.

Ebert, K.M., Selgrath, J.P., DiTullio, P., Denman, J., Smith, T.E., Memon, M.A., Schindler, J.E., Monastersky, G.M., Vitale, J.A and Gorden, K. (1991). Transgenic production of a variant of human tissue type plasminogen activator in goat milk: Generation of transgenic goats and analysis of expression. *Bio/Technology.* **9**: 835 - 838.

Edery, M., Jolicoeur, C., Levi-Meyrueis, C., Dusanter-Fourt, I., Petridou, B., Boutin, J.M., Lesueur, L., Kelly, P.A and Djiane, J. (1989). Identification and sequence analysis of a second form of prolactin receptor by molecular cloning of cDNA from rabbit mammary gland. *Proc. Natl. Acad. Sci. USA.* **86**: 2112 - 2116.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connel, C., Spritz, R.A., De Riel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E.,

Smithies, O., Baralle, F.E., Shoulders, C.C and Proudfoot, N.J. (1980). DNA sequence comparison of the human β-like globin genes. *Cell.* **21**: 653 - 668.

Ellegren, H; Fredholm, M; Edfors-Lilja, I; Wintero, A.K and Andersson, L. (1993). Conserved synteny between pig chromosome 8 and human chromosome 4 but rearranged and distorted linkage maps. *Genomics.* **17**: 599 - 603.

Evans, T., Felsenfield, G and Reitman, M. (1990). Control of globin gene transcription[1]. *Annu. Rev. Cell Biol.* **6**: 95 - 124.

Faruque, S.M., Antoniou, M and Skidmore, C.J (1987). [Communication]. The fine structure of the mouse α1 casein gene. *Biochem. Soc. Trans.* **16**: 4, 624

Federoff, N.V. (1979). On spacers. *Cell.* **16**: 697 - 710.

Feinberg, A.P and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6 - 13

Feinberg, A.P and Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**: 266 - 267.

Ferretti, L., Leone, P and Sgaramella, V. (1990). Long range restriction analysis of the bovine casein genes. *Nucleic Acids Res.* **18**: 6829 - 6833.

Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M and Kioussis, D. (1996). Locus control region function and heterochromatin -induced position effect variegation. *Science.* **271**: 1123 - 1125.

Fiat, A.M and Jolles, P. (1989) Caseins of various origins and biologically active casein peptides and oligosaccharides: Structural and physiological aspects. *Mol. Cell. Biochem.* **87**: 5 - 30.

Forrester, W.C., Takegawa, S., Papayanopoulou, T., Stamatoyannopoulos, G and Groudine, M. (1987). Evidence for a locus activation region: the formation of

developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res.* **15**: 10159 - 10177.

Forrester, W.C., Novac, K., Gelinas, R and Groudine, M. (1989). Molecular analysis of the human beta-globin locus activation region. *Proc. Natl. Acad. Sci.USA.* **86**: 5439 - 5443.

Foster, R.C. (1977). Changes in mouse mammary epithelial size during mammary gland development. *Cell. Differ.* **6**: 1 - 8.

Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A and Cedar, H. (1991). Demethylation of CpG islands in Embryonic cells. *Nature.* **351**: 239 - 241.

Gardiner-Garden, M and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**: 261 - 282.

Gartler, S.M and Riggs, A.D. (1983). Mammalian X-chromosome inactivation. *Ann. Rev. Genet.* **17**: 155 - 190.

Geissler, E.N., Cheng, S.V., Gusella, J.F and Housman, D.E. (1988). Genetic analysis of the dominant white spotting (*W*) region on mouse chromosome 5: Identification of cloned DNA markers near *W*. *Proc. Natl. Acad. Sci. USA.* **85**: 9635 - 9639.

Gellin, J., Echard, G., Yerle, M., Dalens, M., Chevalet, C and Gillois, M. (1985). Localization of the alpha and beta casein genes to the q24 region of chromosome 12 in the rabbit (Oryctolagus cuniculus L.) by *in situ* hybridization. *Cytogenet. Cell Genet.* **39**: 220 - 223.

Gessler, M., Poustka, A., Cavenee, W., Neve, R.L., Orkin, S.H and Bruns, G.A.P. (1990). Homozygous deletion in Wilm's tumours of a zinc finger gene identified by chromosome jumping. *Nature.* **343**: 774 - 778.

Ghali, M.I., Mehtar, N.S and Guerin, G. (1991). Sheep gene mapping: additional DNA markers included (CASB, CASK, LALBA, IGF-1 and AMH). *Anim. Genet.* **22**: 165 - 172.

Gilbert, W. (1978). Why genes in pieces?. *Nature.* **271**: 501.

Gilbert, W. (1985). Genes in pieces revisited. *Science.* **228**: 823 - 824.

Godovac Zimmermann, J., Conti, A., Liberatori, J and Braunitzer, G. (1985). Homology between the primary structures of β-lactoglobulin and human retinol-binding protein: evidence for a similar biological function. *Biol. Chem. Hoppe-Seyler.* **366**: 431 - 434.

Goldsmith, M.R and Kafatos, F.C. (1984). Developmentally regulated genes in silk moths. *Annu.Rev.Genet.* **18**: 443 - 487.

Gordon, K.E., Lee, J.A., Vitale, A.E., Smith, H., Westphal, H and Hennighausen, L. (1987). Production of human tissue plasminogen activator in transgenic mouse milk. *Bio/Technology.* **5**: 1183 - 1187.

Goodman, H.S and Rosen, J.M (1990). Transcriptional analysis of the mouse β-casein gene. *Mol. Endo.* **4**: 1661 - 1670.

Gorodetsky, S.I., Tkach, T.M and Kapelinskaya, T.V. (1988). Isolation and characterization of the *Bos taurus* β-casein gene. *Gene.* **66**: 87 - 96.

Goyal, N., Knox, J and Gronostajski, R.M. (1990). Analysis of multiple forms of nuclear factor 1 in human and murine cell lines. *Mol. Cell. Biol.* **10**: 1041 - 1048.

Greaves, D.R., Wilson, F.D., Lang, G and Kioussis, D. (1989). Human CD2 3' flanking sequences confer high level, T-cell specific, position independent gene expression in transgenic mice. *Cell.* **56**: 979 - 986.

Green, M.R and Pastewka, J.V (1976). Molecular weights of three mouse milk caseins by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and κ-like characteristics of a fourth casein. *J. Dairy Sci.* **59**: 1738 - 1745.

Greenberg, N.M., Anderson, J.W., Hsueh, A.J.W., Nishimori, K., Reeves, J.J., DeAvila. D.M., Ward, D.N and Rosen, J.M. (1991). Expression of biologically

active heterodimeric bovine follicle stimulating hormone in milk of transgenic mice. *Proc. Natl. Acad. Sci, USA.* **88**: 8327 - 8331.

Greenberg, R., Groves, M.L and Dower, H.J. (1984). Human β-casein. Amino acid sequence and identification of phosphorylation sites. *J. Biol. Chem.* **259**: 5132 - 5138.

Groenen, M.A.M., Dijkhof, R.J., Van Der Poel, J.J., Van Diggelen. R and Verstege, E. (1992). Multiple octamer binding sites in the promoter region of the bovine alphaS2 casein gene. *Nucleic Acids Res.* **20**: 4322 - 4318.

Groenen, M.A.M., Dijkhof, R.J.M., Verstege, A.J.M and Poel, J.J Van Der. (1993). The complete sequence of the gene encoding bovine alphaS2 casein. *Gene.* **123**: 187 - 193.

Grosclaude, F. (1988). Le polymorphisme genetique des principales lactoproteines bovines. *INRA Prod. Anim.* **1**: 5 - 17.*

Grosclaude, F., Mahe, M.F., Mercier, J.C and Ribadeau Dumas, B. (1972). Caracterisation des variants genetiques des caseines αS1 et β bovines. *Eur. J. Biochem.* **26**: 328 - 337.*

Grosclaude, F., Mahe, M.F., Brignon, G., Di Stasio, L and Jeunet, R. (1987). A Mendelian polymorphism underlying quantitative variations of goat αS1-casein. *Genet. Sel. Evol.* **19**: 399 - 412.

Grosclaude, F., Mercier, J.C and Ribadeau Dumas, B. (1973). Genetic aspects of cattle casein research. *Neth. Milk Dairy J.* **27**: 328 - 340.

Gross, D.S and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**: 159 - 197.

Grosveld, F., Assendelft, G.B.V., Greaves, D.R and Kollias, G. (1987). Position independent high level expression of human β-globin gene in transgenic mice. *Cell.* **51**: 975 - 985.

Grusby, M.J., Mitchell, S.C., Nabavi, N and Glimcher, L.H (1990). Casein expression in cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. USA.* **87**: 6897 - 6901.

Gunderson, G., Kolsto, A.B and Prydz, H. (1991). Differential methylation of a CpG island concatemer in hemi and homozygous transgenic mice. *Fed. Eur. Bio. Soc. Lett.* **295**: 214 - 218.

Gutierrez, A., Meade, H.M., Jimenz-Flores, R., Anderson, G.B., Murray, J.D and Medrano, J.F. (1995). Expression analysis of bovine k-casein in the mammary glands of transgenic mice. *Transgenic. Res. (In press)* (from Maga and Murray, 1996).

Gupta, P., Rosen, J.M., D' Eustachio, P and Ruddle, F.H. (1982). Localization of the casein gene family to a single mouse chromosome. *J. Cell. Biol.* **93**: 199 - 204.

Guyette, W.A., Matusik, R.J and Rosen, J.M. (1979). Prolactin mediated transcriptional and post transcriptional control of casein gene expression. *Cell.* **17**: 1013 - 1023.

Hall, L., Emery, D.C., Davies, M.S., Parker, D and Craig, R.K. (1987). Organization and sequence of the human α-lactalbumin gene. *Biochem. J.* **242**: 735 - 742

Hall, L., Laird, J.E and Craig, R.E. (1984a). Nucleotide sequence determination of guinea-pig casein B mRNA reveals homology with bovine and rat αS2 caseins and conservation of the non-coding regions of the mRNA. *Biochem. J.* **222**: 561 - 570.

Hall, L., Laird, J.E., Pascall, J.C and Craig, R.K (1984b). Guinea-pig casein A cDNA: Nucleotide sequence analysis and comparison of the deduced protein sequence with that of bovine αS2 casein . *Eur. J. Biochem.* **138**: 585 - 589.

Hamada, H., Seldman, M., Howrd, B.H and Gordon, C.M. (1984). Enhanced gene expression by the poly (dT-dG. poly (dC-dA)sequence. *Mol. Cell. Biol.* **4**: 2622 - 2630.

Hansen, R.S., Ellis, N.A and Gartler, S.M. (1988). Demethylation of specific sites in the 5' region of the inactive X-linked human phosphoglycerate kinase gene correlates with the appearance of nuclease sensitivity and gene expression. *Mol. Cell. Biol.* **8**: 4692 - 4699.

Hansson, L., Edlund, A., Johansson, T., Hernell, O., Stromqvist, M., Lindquist, S., Lonnerdal, B and Bergstrom, S. (1994). Structure of the human beta casein encoding gene. *Gene.* **139**: 193 - 199.

Hansson, L., Edlund, M., Edlund, A., Johansson, T., Marklund, S.L., Fromm, S., Stromqvist, M., and Tomell, J. (1994). Expression and characterisation of biologically active human extra-cellular super-oxide dismutase in milk of transgenic mice. *J. Biol. Chem.* **69**: 5358 - 5363.

Harris, S., Ali, S., Anderson, S., Archibald, A.L and Clark, A.J. (1988). Complete nucleotide sequence of the genomic ovine β-lg gene. *Nucleic Acids Res.* **16**: 10379 - 10380.

Harris, S., McClenaghan, M., Simons, J.P., Ali, S and Clark, A.J. (1991). Developmental regulation of the sheep β-lactoglobulin gene in the mammary gland of transgenic mice. *Devp. Genet.* **12**: 299 - 307.

Hasty, P., Ramirez-Solis, R., Krumlauf, R and Bradley, A. (1991). Introduction of a subtle mutation into the *Hox-2.6* locus in embryonic stem cells. *Nature.* **350**: 243 - 246.

Hayes, H., Petit, E.J., Bouniol, C and Popescu, P. (1993a). Localization of the alphaS2 casein gene (CASAS2) to the homologous cattle, sheep and goat chromosome 4 by *in situ* hybridisation. *Cytogenet. Cell. Genet.* **64**: 281 - 285.

Hayes, H., Petit, E.J., Lemieux, N and Dutrillaux. (1993b). Chromosomal localization of ovine β-casein gene by non isotopic *in situ* hybridisation and R-banding. *Cytogenet. Cell. Genet.* **61**: 286 - 288.

Hennighusen, L.G and Sippel, A.E. (1982a). Characterization and cloning of the mRNA specific for the lactating mouse mammary gland. *Eur. J. Biochem.* **125**: 131 -141.

Hennighusen, L.G and Sippel, A.E. (1982b). Mouse whey acidic protein is a novel member of the family of four " disulfide core" proteins. *Nucleic Acids Res.* **10**: 2677 - 2684.

Hennighausen, L.G., Sippel, A.E., Hobbs, A.A. and Rosen, J.M. (1982b). Comparative sequence analysis of the mRNAs coding for mouse and rat whey protein. *Nucleic Acids Res.* **10**: 3733 - 3744.

Hennighausen, L.G., Stendle, A and Sippel, A.E. (1982a). Nucleotide sequence of cloned cDNA coding for mouse epsilon casein. *Eur. J. Biochem.* **126**: 569 - 572.

Hennighausen, L., Westphal, C., Sankaran, L and Pittius, C.W. (1991). Regulation of expression of genes for milk proteins. *Biotechnology.* **16**: 65 - 74.

Hipp, N.J., Groves, M.L., Custer, J.H and McMeekin, T.L. (1952). Separation of α, β and γ caseins. *J. Dairy. Sci.* **35**: 272 - 281.

Hobbs, A.A and Rosen, J.M. (1982). Sequences of rat alpha and gamma casein mRNAs : Evolutionary comparison of the calcium dependent rat casein multigene family. *Nucleic Acids Res.* **10**: 8079 - 8098.

Hobbs, A.A., Richards, D.A., Kessler, D.J and Rosen, J.M. (1982). Complex hormonal regulation of casein gene expression. *J. Biol. Chem.* **257**: 3598 - 3605.

Holler, M., Westin, G., Jiricny, J and Shaffner, W. (1988). Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes and Dev.* **2**: 1127 - 1135.

Hood, L., Campbell, J.H and Elgin, S.C.R. (1975). The organization, expression and evolution of antibody genes and other multigene families. *Annu. Rev. Genet.* **9**: 305 - 353.

Houdebine, L.M. (1980). The role of prolactin, glucocorticoids and progesterone in the control of casein gene expression. *Hormones and Cell regulation.* **4**: 175 - 196.

Houdebine, L.M and Gaye. P. (1975). Regulation of casein synthesis in the rabbit mammary gland. Titration of mRNA activity for casein under prolactin and progesterone treatments. *Mol. Cell. Endocrin.* **3**: 37 - 55.

Hunter, T. (1987). A thousand and one protein kinases. *Cell.* **50**: 823 - 829.

Ihle, J.N and Kerr, I.M. (1995). Jaks and stats in signalling by the cytokine receptor superfamily. *Trends in genetics.* **11**: 69 - 74.

Jahn, G.A., Edery, M., Belair, L., Kelly, P.A and Djiane, J. (1991). Prolactin receptor gene expression in rat mammary gland and liver during pregnancy and lactation. *Endo.* **128**: 2976 - 2984.

Jarman, A.P and Higgs, D.R. (1988). Nuclear scaffold attachment sites in the human globin gene complexes. *The EMBO J.* **7**: 3337 - 3344.

Jeffreys, A and Harris, S. (1982). Process of gene duplication. Nature. 296: 9 - 10.

Jenness, R. (1973). Caseins and caseinate micelles of various species. *Neth. Milk Dairy J.* **27**: 251 - 257.

Jenness, R and Larson, B.L. (1985). Biochemical and nutritional aspects of milk and colostrum, In Lactation, Iowa State University Press, Ames, Iowa, USA. pp 164 - 197.

Jimenez-Flores, R and Richardson, T. (1988). Genetic engineering of the caseins to modify the behaviour of milk during processing : a review. *J. Dairy. Sci.* **71**: 2640 - 2654.

Johnson, P.F and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58**: 799 - 839.

Jolivet, G., Devinoy, E., Fontaine, M.L and Houdebine, L.M. (1992). Structure of the gene encoding rabbit αS1 casein. *Gene.* **113**: 257 - 262.

Jolles, J., Fiat, A.M., schoentgen, F., Alais, C and Jolles, P. (1974b)). The amino acid sequence of sheep κ-A casein. II sequence studies concerning the κ-A-caseinoglycopeptide and establishment of the complete primary structure of the protein. *Biochim. Biophys. Acta.* **365**: 335 - 343.

Jolles, P. and Henschen, A. (1982). Comparison between the clotting of blood and milk. *TIBS.* **7**: 325 - 328.

Jolles, P., Loucheux-Lefebbre, M.H and Henschen, A. (1978). Structural relatedness of κ-casein and fibrinogen γ-chain. *J. mol. Evol.* **11**: 271 - 277.

Jolles, J., Schoentgen, F., Hermann, J., Alais, C and Jolles, P. (1974a). The sequence of sheep κ-casein: primary structure of para κ-A-casein. *Eur. J. Biochem.* **46**: 127 - 132.

Jolles, J., Schoentgen, F., Alais, C., Fiat, A.M and Jolles, P. (1972). Studies on the primary structure of cow κ-casein. Structural features of para-κ-casein., N-terminal sequence of κ-caseinoglycopeptide studied with a sequencer. *Helv. Chim. Acta.* **55**: 2872 - 2883.*

Jones, W.K., Yu-Lee-Ly., Clift, S.M., Brown, T.L and Rosen, J.M. (1985). The rat casein multigene family. Fine structure and evolution of the beta casein gene. *J. Biol. Chem.* **260**: 7042 - 7050.

Jones, P., Wolkowickz, M., Harrington, M and Gonzales, F. (1990). Methylation and expression of the Myo D1 determination gene. *Phil. Trans. Roy. Soc. (Lond.)B* **326**: 277 - 284.

Keith, D.H., Singer-Sam, J and Riggs, A.D. (1986). Active X-chromosome DNA unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine rich island at the 5' end of the gene for phosphoglycerate kinase. *Mol. Cell. Biol.* **6**: 4122 - 4125.

Keshet, I., Lieman-Hurwitz, J and Cedar, H. (1986). DNA methylation affects the formation of active chromatin. *Cell.* **44**: 535 - 543.

Keshet, I., Yisraeli, J and Cedar, H. (1985). Effect of regional DNA methylation on gene expression. *Proc. Natl. Acad. Sci. USA.* **82**: 2560 - 2564.

Knight, C.H and Peaker, M. (1982). Development of the mammary gland. *J. Reprod. Fertil.* **65**: 521 - 536.

Koczan, D., Hobom, G and Seyfert, H.M. (1991). Genomic organization of the bovine alphaS1 casein gene. *Nucleic Acids Res.* **19**: 5591 - 5596.

Konieczny, S.F and Emerson Jr, C.P. (1984). 5-Azacytidine induction of stable mesodermal stem cell lineages from 10T1/2 cells: Evidence for regulatory genes controlling determination. *Cell.* **38**: 791 - 800.

Konkel, D.A., Maizal, J.V and Leder, P. (1979). The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. *Cell.* **18**: 865 - 873.

Krimpenfort, P., Rademakers, A., Eyestone, W., van der Schans, A., van den Broek, S., Kooiman, P., Koorwijk, E., Platenburg, G., Pieper, F., Strijker, R and de Boer, H. (1991). Generation of transgenic dairy cattle using *in vitro* embryo production. *Bio/Technology.* **9**: 844 - 847.

Kruczek, I and Doerfler, W. (1983). Expression of the chloramphenicol acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters: Effect of promoter methylation on gene expression. *Proc. Natl. Acad. Sci. USA.* **80**: 7586 - 7590.

Kuhn, N.J. (1983). The biochemistry of lactogenesis. *In Biochemistry of lactation.* T.B.Mepham (ed) Elsevier, Amsterdam, pp. 351 - 379.

Kumar, S., Clarke, A.R., Hooper, M.L., Horne, D.S., Law, A.J.R., Leaver, J., Springbett, A., Stevenson, E and Simons, P. (1994). Milk composition and lactation of β-casein deficient mice. *Proc. Natl. Acad. Sci.* **91**: 6138 - 6142.

Laird, J.E., Jack, L., Hall, L., Boulton, A., Parker, D and Craig, R.K. (1988). Structure and expression of the guinea-pig α-lactalbumin gene. *Biochem. J.* **254**: 85 - 94.

Larsen, F., Gundersen, G., Lopez, R and Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics.* **13**: 1095 - 1107.

Lee, K.F., Atiee, S.H., Henning, S.J and Rosen, J.M. (1989). Relative contribution of promoter and intragenic sequences in the hormonal regulation of rat beta casein transgene. *Mol. Endo.* **3**: 447 - 453.

Lee, C.S and Oka, T. (1992). A pregnancy specific mammary nuclear factor involved in the repression of the mouse beta casein gene transcription by progesterone. *J. Biol. Chem.* **267**: 5797 - 5801.

Lee, K.F., Anee, S.H and Rosen, J.M. (1989). Differential regulation of rat β-casein chloramphenicol acetyl transferase fusion gene expression in transgenic mice. *Mol. cell. Biol.* **9**: 560 - 565.

Lee, K.F., DeMayo, F.J., Atiee, S.H and Rosen, J.M. (1988). Tissue specific expression of the rat beta casein gene in transgenic mice. *Nucleic Acids Res.* **16**: 1027 - 1041.

Leroux, C and Martin, P. (1996). The caprine αS1 and β-casein genes are 12 kb apart and convergently transcribed. Proceedings of the XXVth International Conference on Animal Genetics. *Anim. Genet.* **27**: Suppl 2. abstract (in press).

Leroux, C., Martin, P., Mahe, M.F., Leveziel, H and Mercier, J.C. (1990). Restriction fragment length polymorphism identification of goat αS1-casein alleles. A potential tool in selection of individuals carrying alleles associated with a high level protein synthesis. *Anim. Genetics.* **21**: 341 - 351.

Leroux, C., Mazure, N and Martin, P. (1992). Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat αS1-casein transcripts: Structural organization of the relevant gene. *J. Biol. Chem.* **267**: 6147 - 6157.

Leveziel, H., Metenier, L., Guerin, G., Cullen, P., Provot, C., Bertand, M and Mercier, J. (1991). Restriction fragment length polymorphism of ovine casein genes: close linkage between alphaS1, alphaS2, beta and kappa casein loci. *Anim. Genet.* **22**: 1 - 10.

Leveziel, H., Rodellar, C., Leroux, C., Pepin, L., Grohs, C., Vaiman, D., Mahe, M.F., Martin, P and Grosclaude, F. (1994). A microsatellite with in the bovine kappa casein gene reveals a polymorphism correlating strongly with polymorphisms previously described at the protein as well as the DNA level. *Anim. Genet.* **25**: 223 - 228.

Levine, W.B., Alexander, L.J., Haganson, G.E and Beattie, C.W.C. (1992). Cloning and sequencing of the porcine kappa casein cDNA. *Anim. Genet.* **23**: 361 - 363.

Li, E., Beard, C and Jaenisch, R. (1993). Role of DNA methylation in genomic imprinting. *Nature.* **366**: 362 - 365.

Li, S and Rosen, J.M. (1995). NF1 and mammary gland factor (Stat5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. *Mol. Cell. Biol.* **15**: 2063 - 2070.

Li Cai., Taylor, J.F., Wing, R.A., Gallagher, D.S., Woo, S.-S and Davis, S.K. (1995). Construction and characterization of a bovine bacterial artificial chromosome library. *Genomics.* **29**: 413 - 425.

Lien, S., Alestrom, P., Klungland, H and Rogne, S. (1992). Detection of multiple beta casein (CASB) alleles by amplification created restriction sites (ACRS). *Anim. Genet.* **23**: 333 - 338.

Lien, S., Kaminski, S., Alestrom, P and Ronge, S. (1993). A simple and powerful method for linkage analysis by amplification of DNA from single sperm cells. *Genomics.* **16**: 41 - 44.

Liskay, M.R. and Evans, R.J. (1980). Inactive X-chromosome DNA does not function in DNA-mediated cell transformation for the hypoxanthine phosphoribosyl transferase gene. *Proc. Natl. Acad. Sci. USA.* **77**: 4895 - 4898.

Lock, L.F., Takagi, N and Martin, G.R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell.* **48**: 39 - 46.

Lubon, H and Hennighausen, L. (1987). Nuclear proteins from lactating mammary glands bind to the promoter of a milk protein gene. *Nucleic Acids Res.* **15**: 2103 - 2121.

Lui, X., Robinson, G.W., Gouilleux, F., Groner, B and Hennighausen, L. (1995). Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue. *Proc. Natl. Acad. Sci. USA.* 92: 8831 - 8835.

.

MacKinlay, A.C and Wake, R.G. (1971). κ-casein and its attack by rennin (chymosin), In *Milk proteins: Chemistry and Molecular Biology II* (H.A. Mckenzie, ed), Academic Press, New York, pp. 175 - 215.

Macleod, D., Charlton, J., Mullins, J and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes and Dev.* **8**: 2282 - 2292.

Maga, E.A., Anderson, G.B., Huang, M.C and Murray, J.D. (1994). Expression of human lysozyme mRNA in the mammary gland of transgenic mice. *Transgenic Res.* **3**: 36 - 42.

Maga, E.A and Murray, J.D. (1995). Mammary gland expression of transgenes and the potential for altering the properties of milk. *Bio/technology.* **13**: 1452 - 1457.

Mahe, M.F and Grosclaude, F. (1989). αS1-cnD, another allele associated with a decreased synthesis rate at the caprine αS1 casein locus. *Genet. Sel. Evol.* **21**: 127 - 129.

Majumder, P.K., Joshi, J.B and Banerjee, M.R. (1983). Correlation between nuclear glucocorticoid receptor levels and casein gene expression in murine mammary gland *in vitro. J. Biol. Chem.* **258**: 6793 - 6798.

Maki, M., Tibbetts, C., Nagao, M., Sasaki, R and Chiba H (1985). Nucleotide sequence comparison of mRNAs coding for major calcium sensitive caseins between cow and rat. *Agril. Biol. Chem.* **49**: 1099 - 1109.

Maniatis, T., Fritsch, E. F., Lauer, J and Lawn, R.M. (1980). The molecular genetics of human haemoglobins. *Ann. Rev. Genet.* **14**: 145 - 178.

Maniatis, T., Fritsch, E. F and Sambrook, J. (1982). *Molecular Cloning. A laboratory mannual.* Cold-Spring Harbour Laboratory. Cold-Spring Harbour, NY.

Markert, C.L. (1987). Isozymes and the regulatory structure of the genome. *Current.Top. Biol. Med. Res.* **14**: 1 - 17.

Markoff, E and Talamantes, F. (1980). The lactogenic response of mouse mammary explants to mouse prolactin and growth hormone. *Endocri. Res. Commun.* **7**: 269 - 278.

Martin, P. (1993) Genetic polymorphism of goat milk proteins. *Lait.* **73**: 511 - 532.

Martin, P and Grosclaude, F. (1993). Improvement of milk protein quality by gene technology. *Livestock Production. Sci.* **35**: 95 - 115.

Martin, P and Leroux, C. (1992). Exon skipping is responsible for the 9 amino acid residue deletion occuring near the N-terminal of human beta casein. *Biochem. Biophys. Res. Commun.* **183**: 750 - 757.

Maschio, A., Brickell, P.M., Kioussis, D., Mellor, A.L., Katz, D and Graig, R.K. (1991). Transgenic mice carrying the guinea pig α-lactalbumin gene transcribe milk protein genes in their sebacious glands during lactation. *Biochem. J.* **275**: 459 - 467.

Matsuda, Y and Chapman, V.M. (1991). *In situ* analysis of centromeric satellite DNA segregating in *Mus* species crosses. *Mammalian genome.* **1**: 71 - 77.

Matusik, R.J and Rosen, J.M. (1978). Prolactin induction of casein mRNA in organ culture. A model system for studying peptide hormone regulation of gene expression. *J. Biol. Chem.* **253**: 2343 - 2347.

McFarlane, M and Wilson, J.B. (1996). A model for the mechanism of precise integration of a microinjected transgene. *Transgenic Res.* **5**: 171 - 177.

Meada, N and Smithies, O. (1986). The evolution of multigene families: human haptoglobin genes. *Annu. Rev. Genet.* **20**: 81 - 108.

Meade, H., Gates, L., Lacy, E and Lonberg, N. (1990). Bovine α-casein gene sequences direct high level expression of active human urokinase in mouse milk. *Bio/Technology.* **8**: 443 - 446.

Meehan, R.R., Lewis, J.D., Mckay, S., Kleiner, E.L and Bird, A.P. (1989). Idetification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell.* **58**: 499 - 507.

Mehta, N.M., EI-Gewely, M.R., Joshi J., Helling, R.B and Banerjee, M.R. (1981). Cloning of mouse beta casein gene sequences. *Gene.* **15**: 285 - 288.

Mehtali, M., LeMeur, M and Lathe, R. (1990). The methylation free status of a house-keeping transgene is lost at high copy number. *Gene.* **91**: 179 - 184.

Meier, V.S and Groner, B. (1994). The nuclear factor YY1 participates in repression of the beta casein gene promoter in mammary epithelial cells and is counteracted by mammary gland factor during lactogenic hormone induction. *Mol. Cell. Biol.* **14**: 128 - 137.

Meier, H., Hoag, W.G and McBurney, J.J. (1965). Chemical characterization of inbred-strain mouse milk. I. Gross comparison and amino acid analysis. *J. Nutr.* **85**: 305 - 308.

Menon, R.S., Chang, Y.F., Jeffers, K.F and Ham, R.G. (1992). Exon skipping in human beta casein. *Genomics.* **12**: 13 - 17.

Mepham, T.B. (1987). *Physiology of lactation.* Open University Press, Milton Keynes and Philadelphia.

Mepham, T.B., Gaye, P., Martin, P and Mercier, J.C. (1992) Biosynthesis of milk protein, In *Advanced dairy chemistry, I. Proteins*, (P.F. Fox, ed), Elseiver applied science, London and NY. pp 491 - 540.

Mercier, J.C. (1986). Genetic engineering applied to milk producing animals. some expectations In *Exploiting New Technologies in Animal Breeding- Genetic Developments.* Smith C; King JWB; Mckay JC, (Ed). Oxford University Press, Oxford, England. pp 122

Mercier. J.-C. (1992). Structure and functions of milk protein genes. *J. Dairy. Sci.* **75**: (Suppl 1) 90.

Mercier, J.-C., Grosclaude, R and Ribadeau-Dumas, B. (1971). Structure primerie de la caseine αS1-bovine. Sequence complete. *Eur. J. Biochem.* **23**: 41 - 51.*

Mercier, J.-C., Chobert, J.M and Addeo, F. (1976). Comparative study of the amino acid sequences of the caseinomacropeptides from seven species. *Fed. Eur. Biol. Soc. Lett.* **72**: 208 - 214.

Mercier. J.-C., Gaye. P., Soulier. S., Hue-Delahaie, D and Vilotte, J.L. (1985). Construction and identification of recombinant plasmids carrying cDNAs coding for ovine αS1, αS2, β, κ casein and β-lactoglobulin. Nucleotide sequence of αS1-casein cDNA. *Biochemie.* **65**: 959 - 971.

Mercier, J.-C and Vilotte, J.L. (1993). Structure and function of milk protein genes. *J. Dairy. Sci.* **76**: 3079- 3098.

Miller, L.K . (1988). Baculoviruses for foreign gene expression in insect cells. In *vectors : a survey of molecular cloning vectors and their uses.* pp 457 - 465. Rodriguez RL and DEnhardt DT (eds), Butterworth, Boston.

Monaco, A.P and Larin, Z. (1994). YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *TIBTECH.* **12**: 280 - 286.

Mummaneni, P., Bishop, P.L and Turker, M.S. (1993). A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenosine phosphoribosyltransferase gene. *J. Biol. Chem.* **268**: 552 - 558.

Myers, R.M., Tilly, K and Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. *Science.* **232**: 613 - 618.

Nagao, M., Maki, M., Sasaki, R and Chiba, H. (1984). Isolation and sequence analysis of bovine alphaS1-casein cDNA clone. *Agril. Biol. Chem.* **48**: 1663 -- 1667.

Nagaiah, K., Bolander, F.F Jr., Nicholas, K.R., Takemoto, T and Topper, Y.J. (1981). Prolactin induced accumulation of casein mRNA in mouse mammary explants: A selective role of glucocorticoid. *Biochem. Biophys. Res. Commun.* **98**: 380 - 387.

Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.N and Numa, S. (1979). Nucleotide sequence of cloned cDNA for bovine corticotropin-β-lipotropin precursor. *Nature.* **278**: 423 - 427.

Nakhasi, H.L., Grantham, F.H and Gullino, P.M. (1984). Expression of kappa casein in normal and neoplastic rat mammary gland is under the control of prolactin. *J. Biol. Chem.* **259**: 14894 - 14898.

Nakhasi, H.L and Qasba, P.K. (1979). Quantitation of milk proteins and their mRNAs in rat mammary gland at various stages of gestation and lactation. *J. Biol. Chem.* **254**: 6016 - 6025.

Ng-Kwai-Hang, K.F and Grosclaude, F. (1992). Genetic polymorphism of milk proteins. In *Advanced Dairy chemistry-I: Proteins* (P.F. Fox ed), Elsevier Science publishers LTD, Crown house, Linton Road, Barking, Essex IG11 8JU, Englnd. pp 405 - 455.

Ng-Kwai-Hang, K.F., Zadworny, D., Hayes, J.F and Kuhnlein, U. (1991). Identification of kappa casein genotypes in Holstein sires. A comparison between analysis of milk samples from daughters and direct analysis of semen samples from sires by PCR. *J. Dairy. Sci.* **74**: 2410 - 2415.

Nicoll, C.S and Tucker, H.A. (1965). Estimates of parenchymal, stromal and lymphnode DNA in mammary glands of C3H/Cgrl/2 mice. *Life. Sci.* **4**: 993 - 1001.

Nizetic, D., Drmanac, R and Lehrach, H.(1991). An improved bacterial colony lysis procedure enables direct DNA hybridisation using short (10, 11 bases) oligonucleotides to cosmids. *Nucleic Acids Res.* **19**: 182.

Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Hirose, T., Inayama, S., Nakanishi, S and Numa, S. (1982). Cloning and sequencing analysis of cDNA for bovine adrenal preproenkephalin. *Nature.* **295**: 202 - 206.

Ohno, S. (1970). Evolution by gene duplication. Springer-Verlag, NewYork.

Ohta, T. (1980). *Evolution and variation of multigene families.* Lecture notes in biomathematics. Vol. **37**. Springer-Verlag, Berlin. pp 107 - 121.

Ohta, T. (1983a). On the evolution of multigene families. *Theor. Pop. Biol.* **23**: 216 - 240.

Ohta, T. (1983b). Time until fixation of a mutant belonging to a multigene family. *Genet. Res.* **41**: 47 - 55.

Ohta, T. (1989). Gene families. In *Evolution and Animal Breeding.* (W.G. Hill and T.F.C. Mackay, eds), C.A.B. International, Wallingford, Oxon OX10 8DE, U.K. pp 61 - 66.

Ohta, T. (1991). Multigene families and their implications for evolutionary theory In *New Aspects of the Genetics of Molecular evolution.* (M. Kimura and Takahata, N (eds), Japan. Sci. Soc. Press. Tokyo/Springer-Verlag, Berlin. pp. 15 - 25.

Ono, M and Oka, T. (1980). α-lactalbumin-casein induction in virgin mouse mammary explants: dose dependent differential action of cortisol. *Science.* **207**, 1367 - 1369.

Orkin, S.H. (1990). Globin gene regulation and switching: Circa 1990. *Cell.* **63**: 665 - 672.

Pearse, M.J., linklater, p.M., Hall, R.J and MacKinlay, A.G. (1986). The effect of casein composition and casein dephosphorylation on the coagulation and syneresis of artificial micelle milk. *J. Dairy. Res.* **53**: 381 - 390.

Persuy, M.A., Stinnakre, M.G., Printz, C., Mahe, M.F and Mercier, J.-C. (1992). High expression of the caprine beta casein gene in transgenic mice. *European J. Biochem.* **205**: 887 - 893.

Perutz, M.F. (1983). The haemoglobin molecule In *The Molecular basis of life*. pp 39 - 51.

Pervais, S and Brew, K. (1985). Homology of β-lactoglobulin, serum retinol binding protein and protein HC. *Science.* **228**: 335 - 337.

Pfeifer, G.P and Riggs, A.D. (1991). Chromatin differences between active and inactive X-chromosomes revealed by genomic footprinting of permeabilised cells using DNase I and ligation-mediated PCR. *Genes and Dev.* **5**: 1102 - 1113.

Pfiefer, G.P., Tanguay, R.L., Steigerwald, S.D and Riggs, A.D. (1990). *In vivo* footprint and methylation analysis by PCR-aided genomic sequencing: Comparison of active and inactive X-chromosomal DNA at the CpG island and promoter of human Pgk-1. *Genes and Dev.* **4**: 1277 - 1287.

Pierce, J.C., Sternberg, N and Sauer, B. (1992). A mouse genomic library in the bacteriphage P1 cloning system: Organization and characterisation. *Mammalian Genome.* **3**: 550 - 558.

Pierre, S., Jolivet, G., Devinoy, E., Theron, M.C., Malienou N- Gassa, R., Puissant, C and Houdebine, L.M. (1992). A distal region enhances the prolactin induced promoter activity of the rabbit alphaS1 casein gene. *Mol. Cell. Endocrin.* **87**: 147 - 156.

Piletz, J.E and Ganschow, R.E. (1981). Genetic variation of milk proteins in mice. *Biochem. Genetics.* **19**: 1023 - 1030.

Pinder, S.J., Perry, B.N., Skidmore, C.J and Savva, D. (1991). Analysis of polymorphism in the bovine casein genes by use of the polymerase chain reaction. *Anim. Genet.* **22**: 11 - 20.

Pittius, C.W., Sankaran, L., Topper, Y.J and Hennighausen, L. (1988). Comparison of the regulation of the whey acidic protein gene with that of a hybrid gene containing the whey acidic protein gene promoter in transgenic mice. *Mol. Endo.* **2**: 1027 - 1032.

Pittius, C.W., Hennighausen, L., Lee, E., Westphal, H., Nicholas, E., Vitale, J and Gordon, K. (1988). A milk protein gene promoter directs the expression of human tissue plasminogen activator cloned DNA to the mammary gland in transgenic mice. *Proc. Natl. Acad. Sci. USA*. **85**: 5874 - 5878

Platenberg, G.J., Koorwijk, E.P.A., Kooiman, P.M., Woloshuk, S.L., Nuijens, J.H., Krimpenfort, P.J.A., Piepper, F.R., deBoer, H.A and Strijker, R. (1994). Expression of human lactoferrinin milk of transgenic mice. *Transgenic Res.* **3**: 99 - 108.

Poyet, P., Henning, S.J and Rosen, J.M. (1989). Hormone dependent beta casein mRNA stabilization requires on going protein synthesis. *Mol. Endo.* **3**: 1961 - 1968.

Primrose, S.B. (1987). Modern Biotechnology. Blackwell Scientific Publications, Oxford.

Provot, C., Persuy, M.-A and Mercier, J.-C. (1995). Complete sequence of the ovine β-casein encoding gene and interspecies comparison. *Gene.* **154**: 259 - 263.

Qasba, P.K and Safaya, S.K. (1984). Similarity of the nucleotide sequence of α-lactalbumin and chicken lysozyme genes. *Nature.* **308**: 377 - 380.

Rando, A., Di Gregorio, P and Masin, P. (1987). Identification of bovine kappa casein genotype at the DNA level. *Anim. Genet.* **19**: 51 - 54.

Raught, B., Liao, W.S.-L and Rosen, J.M. (1995). Developmentally and hormonally regulated CCAAT/enhancer-binding protein isoforms influence β-casein gene expression. *Mol. Endo.* **9**: 1223 - 1232.

Reitman, M and Felsenfeld, G. (1990). Developmental regulation of topoisomerase II sites and DNase I hypersensitive sites in the chicken β-globin locus. *Mol. Cell. Biol.* **10**: 2774 - 2786.

Ribadeau-Dumas, B., Brignon, G., Grosclaude, F and Mercier, J.-C. (1972). Structure primaire de la caseine β-bovine. Sequence complete. *Eur. J. Biochem.* **25**: 505 - 514.*

Rijnkels, M., Kooiman, P.M., Krimpenfort, P.J.A., De Boer, H.A and Pieper, F.A. (1995). Expression anlysis of the individual bovine β-, αS2- and κ- casein genes in transgenic mice. *Biochem. J.* **311**: 929 - 937.

Roberts, B., DiTullio, P., Vitale, J., Hehir, K and Gordon, K. (1992). Cloning of the goat beta casein encoding gene and expression in transgenic mice. *Gene*.**121**: 255 - 262.

Rohrer, G.A., Alexander, L.J., Keele, J.W., Smith, T.P and Beattie, C.W (1994). A microsatellite linkage map of the porcine genome. *Genetics.* **136**: 231 - 245.

Rosen, J.M.(1987). Milk protein gene structure and expression. In *The mammary gland: development, regulation and function.* (M.C. Neville and C.W. Daniel, eds) Plenum press, NewYork. pp. 301 - 322.

Rosen, J.M and Schields, D. (1980). Post translational modifications of the rat mammary gland caseins: *in vitro* synthesis, processing and segregation. In *Testicular Developments structure and function.* (A. Steinberger and E. Steinberger, eds). Raven Press, NY.

Rosen, J.M., Jones, W.K and Lee, L.Y.Y. (1984). Hormonl regulation of casein gene expression in normal and transformed cells. Proceedings of the eleventh annual meeting of the International Foundation for Biochemical Endocrinlogy. 385 - 396.

Rosen, J.M., Jones, W.K., Campbell, S.M., Bisbee. C.A and Lee, L.Y.Y. (1985). Structure and regulation of peptide hormone responsive genes. UCLA- Symposia on Molecular and Cellular Biology. 23: 385 - 396.

Rosen, J.M., Jones, W.K., Rodgers, J.R.,Compton, J.R., Bisbee. C.A and Lee, L.Y.Y. (1986). Regulatory sequences involved in the hormonal control of casein gene expression. *Ann. NY. Acad. Sci.* **464**: 87 - 99.

Rosen, J.M., O'Neal, D.L., McHugh, J.E., Comstock, J.P. (1978). Progesterone-mediated inhibition of casein mRNA and polysomal casein synthesis in the rat mammary gland during pregnancy. *Biochemistry.* **17**: 290 - 297.

Rosen, J.M., Rodgers, J.R., Couch, C.H; Bicbee, C.A., David-Inouye, Y., Campbell, S.M and Yu-Lee, L.Y. (1986). Multihormonal regulation of milk protein gene expression. *Annals. N.Y. Acad. Sci.* **478**: 63 - 76.

Rosen, J.M., Woo, S.L.C and Comstock, J.P. (1975). Regulation of casein mRNA during the development of the rat mammary gland. *Biochemistry.* **14**: 2895 - 2903.

Saiki, R.K., Gelfand, D.H and Stoffel, S. (1988). Primer directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science.* **239**: 487 - 491.

Sambrooke, J., Fritsch, E.F and Maniatis, T. (1989). Molecular cloning: A laboratory mannual, 2nd Edition, Cold spring harbor university press, Cold spring harbor.

Sankaran, L., Qasba, P and Topper, J.J. (1984). Effects of oestrogen depletion on rat casein gene expression. *Biochem. Biophys. Res. Commun.* **125**: 682 - 689.

Sarmeintos, P., Duchesne, M., Denefle, P., Boiziau, J., Fromage, N., Delporte, N., Parker, F., Lelievre, Y., Mayaux, J.F and Cartwright, T. (1989). Synthesis and purification of active human tissue plasminogen activator from *E.coli*. *Biotechnology.* **7**: 495 - 501.

Sasaki, T., Sasaki, M and Enami, J. (1992). Mouse gamma-casein cDNA: PCR cloning and sequence analysis. *Zool. Sci.* **10**,:65 - 72.

Schedel, A., Montoliu, L., Kelsey, G and Schutz, G. (1993). A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice. *Nature.* **362**: 258 - 261.

Scheller, R.H., Jackson, J.F., McAllister, L.B., Schwartz, J.H., Kandel, E.R and Axel, R. (1982). A family of genes that code for ELH, a neuropeptide eliciting a stereotyped pattern of behaviour in *Aplysia. Cell.* **28**: 707 - 719.

Schmidhauser, C., Bissell, M.J., Myers, C.A and Casperson, G.F. (1990). Extracellular matrix and hormones transcriptionally regulate bovine beta casein 5' sequences in stably transfected mouse mammary cells. *Proc. Natl. Acad. Sci. USA.* **87**: 9118 - 9122.

Schmidt, D.G. (1982). Association of caseins and casein micelle structure. In *Developments in Dairy Chemistry* - I. (P.F. Fox, ed), Applied Science Publishers, London. pp. 61 - 86.

Schmitt-Ney, M., Happ, B., Hofer, P., Hynes, N.E., and Groner B. (1992a). Mammary gland specific nuclear factor activity is positively regulated by lactogenic hormones and negatively by milk stasis. *Mol. Endo.* **6**: 1988 - 1997.

Schmitt-Ney, M., Doppler, W., Ball, R.K and Groner, B. (1991). Beta casein gene promoter activity is regulated by the hormone mediated relief of transcriptional repression and a mammary gland specific nuclear factor. *Mol. Cell. Biol.* **7**: 3745 - 3755.

Schmitt-Ney, M., Happ, B; Ball, R.K and Groner B.(1992b). Developmental and environmental regulation of a mammary gland specific nuclear factor essential for transcription of the gene encoding beta casein. *Proc. Nat. Acad. Sci. USA.* **89**: 3130 - 3134.

Schnenberger, C.A., Zuk, A., Groner, B., Jones, W and Andres, A.C. (1990). Induction of the endogenous whey acidic protein (WAP) gene and a WAP-*myc* hybrid gene in primary murine mammary organoids. *Dev. Biol.* **139**: 327 - 337.

Shamay, A., Pursel, V.G., Wall, R.J and Hennighausen, L. (1992). Induction of lactogenesis in transgenic virgin pigs: evidence for gene and integration site-specific hormonal regulation. *Mol. Endo.* **6**: 191 - 197.

Shani, M., Barash, I., Nathan, M., Ricca, G., Searfoss, G.H., Dekel, I., Faerman, A., Givol, D and Hurwitz, D.R. (1992). Expression of human serum albumin in the milk of transgenic mice. *Transgenic Res.* **1**: 195 - 208.

Shapira, S.K and Finnerty, V.G. (1986). The use of genetic complimentation in the study of eukaryotic macromolecular evolution: rate of spontaneous gene duplication at two loci of *Drosophila melanogaster. J. Mol. Evol.* **23**: 159 - 167.

Sharp, P.A. (1985). On the origin of RNA splicing and introns. *Cell.* **42**: 397 - 400.

Sherbon, I.W., Ledford, R.A., Regenstein, I and Thompson, M.P. (1967). Variants of milk proteins and their possible relation to milk properties. *J. Dairy. Sci.* **50**: 951.

Shizuya, H., Birren, B., Ung-Jin Kim., Mancino, V., Slepak, T., Tachiiri, Y and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor based vector. *Proc. Natl. Acad. Sci. USA.* **89**: 8794 - 8797.

Silva, M.C., Wong, D.W.S and Batt, C.A. (1990). Cloning and partial nucleotide sequence of the genomic bovine β-lactoglobulin gene. *Nucleic Acids Res.* **18**: 3051.

Simons, J.P., McClenaghan, M and Clark, A.J. (1987). Alteration of the quality of milk by expression of sheep beta lactoglobulin in transgenic mice. *Nature.* **328**: 530 - 532.

Simpson, K.J., Bird, P., Shaw, D and Nicholas, K. (1996). Isolation, characterisation and hormone-dependent expression of the porcine whey acidic protein. *Bioch. Soc. Transactions.* **24**: 367S.

Sleigh, R.W., Sculley, T.B and MacKinlay, A.G. (1979). The binding of b-casein to hydroxyapatite: the effect of phosphate content and lactation. J. Dairy. Res. 46: 337 - 342.

Solage, A and Cedar, H. (1978). Organisation of 5-methylcytosine in chromosomal DNA. *Biochemistry.* **17**: 2934 - 2938.

Soulier, S., Vilotte, J.L., Stinnakre, M.G and Mercier, J.-C. (1992). Expression anlysis of ruminant α-lactalbumin in transgenic mice: developmental regulation and general location of important cis-regulatory elements. *Fed. Eur. Biol. Soc. Lett.* **297**: 13 - 18.

Stacey, A., Schnieke, A., McWhir, J., Cooper, J., Colman, A and Melton, D.W. (1994). Use of double-replacement gene targetting to replace the murine α-lactalbumin gene with its human counterpart in embryonic stem cells and mice. *Mol. Cell. Biol.* **14**: 1009 - 1016.

Stapleton, G., Somma, M.P and Lavia, P. (1993). Cell-type-specific interactions of transcription factors with a house keeping promoter *in vivo*. *Nucleic Acids Res.* **21**: 2465 - 2471.

Stevenson, E.M and Leaver, J. (1994). Chromatographic separation of the proteins of mouse milk. *Int. dairy. J.* **4**: 205 - 220.

Stewart, A.F and MacKinlay, A.G. (1984). Evolution of casein gene complex: insights provided by comparison of casein mRNAs. *Proc. Aust. Bioch. Soc.* **16**: S48.

Stewart, A.F., Willis, I.M and MacKinlay, A.G. (1984). Nucleotide sequences of bovine alphaS1 and kappa casein cDNAs. *Nucleic Acids Res.* **12**: 3895 - 3907.

Stewart, A.F., Bonsing, J., Beattie, C.W., Shah, F., Willis, I.M and MacKnlay, A.G. (1987). Complete nucleotide sequence of bovine alphaS2 and beta casein cDNAs: comparison with related sequences in other species. *Mol. Biol. Evol.* **4**: 231 - 241.

Stinnakre, M.G., Vilotte, J.L., Soulier, S., L'Haridon, R., Charlier, M., Gaye, P and Mercier, J._C. (1991). The bovine α-lactalbumin promoter directs expression of ovine trophoblast interferon in the mammary gland of transgenic mice. *Fed. Eur. Biol. Soc. Lett.* **284**: 19 -

Stinnakre, M.G., Vilotte, J.L., Soulier, S and Mercier J.-C. (1994). Creation and phenotypic analysis of a-lactalbumin-deficient mice. *Proc. Natl. Acad. Sci. USA.* **91**: 6544 - 6548.

Sulimova, G.E., Sokolova, S.S., Semikozova, O.P., Nguet, L.M and Berberov, E.M .(1992). Analysis of DNA polymorphism of cluster genes in cattle : casein genes and major histocompatibility complex (MHC) genes. *Tsitologiya-i-Genetika.* **26**: 18 - 26.*

Szyf, M., Tanigawa, G and McCarthy, P.L.J. (1990). A DNA signal from the thy-1 gene defines *de novo* methylation patterns in embryonic stem cells. *Mol. Cell. Biol.* **10**: 4396 - 4400.

Tamame, M., Antequera, F., Villanueva, J.R and Santos, T. (1983). High frequency conversion to a 'fluffy' developmental phenotype in *Aspergillus* spp. by 5-azacytidine treatment: evidence for involvement of a single nuclear gene. *Mol. Cell. Biol.* **3**: 2287 - 2297.

Terada, N., Leiderman, L.J and Oka, T. (1983). The interaction of cortisol and prostaglandins on the phenotypic expression of the α-lactalbumin gene in the mouse mammary gland in culture. *Biochem. Biophys. Res. Commun.* **111**: 1059 -

te Riele, H., Maandag, E.R and Berns, A. (1992). Highly efficient gene targetting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. USA.* **89**: 5128 - 5132.

Terry, P.M., Banerjee, M.R and Lui, R.M. (1977). Hormone inducible casein messenger RNA in a serum free organ culture of whole mammary gland. *Proc. Natl. Acad. Sci. USA.* **74**: 2441 - 2445.

Teyssot, B and Houdebine, L.M .(1980). Role of prolactin in the transcription of the beta casein and 28S ribosomal genes in the rabbit mammay gland. *Euro. J. Biochem.* **110**: 263 - 272.

Thepot, D., Devinoy, E., Fontaine, M.L., Hubert, C and Houdebine, L.M. (1990). Complete sequence of the rabbit whey acidic protein gene. *Nucleic Acids Res.* **18**: 3641.

Thepot, D., Devinoy, E., Fontaine, M.L and Houdebine, L.M. (1991). Structure of the gene encoding rabbit beta casein. *Gene.* **97**: 301 - 306.

Thompson, M.D., Dave, J.R and Nakhasi, H.L. (1985). Molecular cloning of mouse mammary gland κ-casein: Comparison with rat κ-casein and rat and human γ-fibrinogen. *DNA.* **4**: 263 - 271.

Thordarson, G and Talamantes, F. (1987). Role of the placenta in mammary gland development and function. In *The mammary gland: development, regulation and function.* (M.C. Neville and C.W. Daniel, eds.) Plenum press, NewYork. pp. 459 - 498.

Threadgill, D.W and Womack, J.E. (1990). Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Res.* **18**: 6935 - 6942.

Topper, Y.J. (1970). Multiple hormone interactions in the development of mammary gland in vitro. *Rec. Prog. Horm. Res.* **26**: 287 - 308.

Tuan, D., Solomon, W., Li, Q., London, I.M. (1985). The "beta-like-globin" gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. USA.* **82**: 6384 - 6388.

Tuan, D.Y.H., Solomon, W.B., London, I.M and Lee, D.P. (1989). An erythroid specific, developmental stage independent enhancer far upstream of the human 'beta-like globin' genes. *Proc. Natl. Acad. Sci. USA.* **36**: 2554 - 2558.

vanAssendelft, G.B., Hanscombe, O., Grossveld, F and Greaves, D.R. (1989). The beta globin dominant control region activates homologous and heterologous promoters in a tissue specific manner. *Cell.* **56**: 969 - 977.

Velander, W.H., Page, R.L., Morcol, T., Russel, C.G., Canseco, R., Young, J.M., Drohan, W.N., Gwazdauskas, F.C., Wilkins, T.D and Johnson, J.L. (1992). Production of biologically active protein-C in the milk of transgenic mice. *Ann. N. Y. Acad. Sci.* **665**: 391 - 403.

Vilotte, J.L and Soulier, S. (1992). Isolation and characterization of the mouse α-lactalbumin encoding gene: interspecies comparison, tissue and stage-specific expression. *Gene.* **119**: 287 - 292.

Vilotte, J.L., Soulier, S., Mercier, J.C., Gaye, P., Hue-Delahale, D and Furet, J.P. (1987). Complete nucleotide sequence of bovine alpha lactalbumin gene. Comparison with its rat counter part. *Biochimie.* **69**: 609 - 620.

Vilotte, J.L., Soulier, S., Printz, C and Mercier, J.-C. (1991). Sequence of the goat alpha lactalbumin encoding gene: Comparison with the bovine gene and evidence of related sequences in the goat genome. *Gene.* **98**: 271 - 276.

Vilotte, J.L., Soulier, S., Stinnarkre, M.G., Massoud, M and Mercier, J.-C. (1989). Efficient expression of bovine α-lactalbumin in transgenic mice. *Eur. J. Biochem.* **186**: 43 - 48.

Vonderhaar, B.K. (1984). Hormones and growth factors in mammary gland development. In *Control of cell growth and proliferation*, ed. C.M. Venziale, NY: Van Vostrand Reinhold. pp. 11 - 33.

Vonderhaar, B.K and Nakhasi, H.L. (1986). Bifunctional activity of EGF on α and κ casein gene expression in rodent mammary gland *in vitro*. *Endo.* **119**: 1178 - 1184.

Vonderhaar, B.K and Ziska, S.E. (1989). Hormonal regulation of milk protein gene expression. *Annu. Rev. Physiol.* **51**: 641 - 652.

Wakao, H., Schmitt-Ney, M and Groner, B. (1992). Mamaryd specific nuclear factor is present in lactating rodent and bovine mammary tissue and composed of a single polypeptide of 89kDa. *J. Biol. Chem.* **267**: 16365 - 16371.

Wall, R.J., Pursel, V.G., Shamay, A., McKnight, R.A., Pittius, C.W and Hennighausen, L. (1991). High level synthesis of a heterologous milk protein in the mammary glands of transgenic swine. *Proc. Natl. Acad. Sci. USA.* **88**: 1696 - 1700.

Watson, C.J., Gordon, K.E., Robertson, M and Clark, A.J. (1991). Interaction of DNA binding proteins with a milk protein gene promoter *in vitro*: Identification of a mammary gland specific factor. *Nucleic Acids Res.* **19**: 6603 - 6610.

Waugh, D.F. (1971). Formation and structure of casein micelles, In *Milk proteins: Chemistry and Molecular Biology II* (H.A. Mckenzie, ed). Academic Press, New York, pp. 3 - 85.

Welte, T., Philipp, S., Cairns, C., Gustaffson, J.A and Doppler, W. (1993). Glucocorticoid receptor binding sites in the promoter region of milk protein genes. *J. Steroid. Biochem. Mol. Biol.* **47**: 75 - 81.

Whitelaw, C.B.A., Archibald, A.L., Harris, S., McCleaghan, M., Simons, J.P and Clark, A.J. (1991). Targetting expression to the mammary gland: intronic sequences can enhance the efficiency of gene expression in transgenic mice. *Transgenic Res.* **1**: 3 - 13.

Whitelaw, C.B.A., Harris, S., McCleaghan, M., Simons, J.P and Clark, A.J. (1992). Position-independent expression of the ovine beta lactoglobulin gene in transgenic mice. *Biochem. J.* **286**: 31 - 39.

Wilde, C.J., Clark, A.J., Keer, M.A., Knight, C.H., McClenaghan, M and Simons, J.P. (1992). Mammary development and milk secretion in transgenic mice expressing the sheep β-lactoglobulin gene. *Biochem. J.* **284**: 717 - 720.

Willis, I.M., Stewart, A.F., Caputo, A., Thompson, A.R and MacKinlay, A.G. (1982). Construction and identification by partial nucleotide sequence analysis of bovine casein and β-lactoglobulin cDNA clones. *DNA.* **1**: 375 - 386.

Wilson, B.J. (1977). Growth curves: Their analysis and use. In *Growth and poultry meat production.* (K.N. Boorman and Wilson, B.J. eds), British poultry science Ltd., Great Briton pp 90 - 115.

Wistow, G.J., Mulders, J.W.M and de Jong, W.W. (1987). The enzyme lactate dehydrogenase as a structural protein in avian and crocodilian lenses. *Nature.* **326**: 622 - 624.

Wistow, G.J., Lietman, T., Williams, L.A., Stapel, S.O., de Jong, W.W., Horwitz, J and Piatigorsky, J. (1988). γ-crystallin/α-enolase: one gene encodes both an enzyme and a lens structural protein. *J. Cell. Biol.* **107**: 2729 - 2736.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedman, T and Migeon, B.R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X chromosome inactivation. *Proc. Natl. Acad. Sci. USA.* **81**: 2806 - 2810.

Womack, J.E., Threadgill, D.W., Moll, Y.D., Faber, L.K., Foreman, M.L., Dietz, A.B., Tobin, T.C., Skow, L.C., Zneimer, S.M., Gallagher, D.S and Rogers, D.S. (1989). Syntenic mapping of 37 loci in cattle. Chromosomal conservation with mouse and man. *Cytogenet. Cell Genet.* **51**, 1109.

Woo, S-S., Jiang, J., Gill, B.S., Paterson A.H., and Wing, R.A. (1994). Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor. Nucleic Acids Res.* **22**: 4922 - 4931.

Wright, G., Carver, A., Cottom, D., Reeves, D., Scott, A., Simons, P., Wilmut, I., Garner, I and Colman, A (1991). High level expression of active human α1-antitrypsin in the milk of transgenic sheep. *Bio/Technology.* **9**: 830 - 834.

Wright, S., Rosenthal, A., Flavell, R.A and Grosveld, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. *Cell.* **38**: 265 - 273.

Yen, P.H., Patel, P., Chinault, A.C., Mohandas, T and Shapiro, L. J. (1984). Differential methylation of hypoxanthine phosphoribosyl transferase genes on

active and inactive human X chromosomes. *Proc. Natl. Acad. Sci. USA.* **81**: 1759 - 1763.

Yisraeli, J., Adelstein, R.S., Melloul, D., Nudel, U., Yaffe, D and Cedar, H. (1986). Muscle specific activation of a methylated chimeric actin gene. *Cell.* **46**: 409 - 416.

Yisraeli, J., Frank, D., Razin, A and Cedar, H. (1988). Effect of *in vitro* DNA methylation on β-globin gene expression. *Proc. Natl. Acad. Sci. USA.* **85**: 4638 - 4642.

Yisraeli, J and Szyf, M. (1984). *DNA methylation: Biochemistry and Biological Significance*. (A. Razin; H. Cedar and A.d. Riggs. eds). Springer, NewYork. pp 353 - 378.

Yoshimura, M., Banerjee, M.R and Oka T (1986). Nucleotide sequence of a cDNA encoding mouse beta casein. *Nucleic Acids Res.* **14**: 8224.

Yoshimura, M and Oka, T. (1989). Isolation and structural analysis of the mouse beta casein gene. *Gene.* **78**: 267 - 275.

Yoshimura, M and Oka, T. (1990). Hormonal induction of β-casein gene expression: requirement of ongoing protein synthesis for transcription. *Endo.* **126**: 427 - 433.

Yu-Lee, L-Y and Rosen, J.M. (1983). The rat casein multfigene family. I. Fine structure of the γ-casein gene. *J. Biol. Chem.* **258**: 10794 - 10804.

Yu-Lee, L-Y., Richter-Mann, L., Couch, C.H., Stewart, A.F., Mackinlay, A.G and Rosen, J.M. (1986). Evolution of the casein multigene family: conserved sequences in the 5' flanking and exon regions. *Nucleic Acids Res.* **14**: 1883 - 1902

Zadworny, D and Kuhnlein, U. (1990). The identification of kappa casein genotype in holstein dairy cattle using the polymerase chain reaction. *Theoretical and Applied Genetics.* **80**: 631 - 634.

Zadworny, D., Kuhnlein, U and Ng-Kwai-Hang, K.F. (1990). Determination of kappa casein alleles in Holstein dairy cows and bulls using polymerase chain reaction. Proceedings of the 4th world congress on genetics applied to livestock production Edinburgh, Dairy cattle genetics and breeding, adaptation, conservation. 251 - 254.

*Original articles not consulted.