# ALTERNATIVE EXPRESSION FORMS OF THE BOVINE $\alpha_{S1}$-CASEIN GENE


The University of Waikato
*Te Whare Wānanga o Waikato*

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy in Biological Sciences
at the University of Waikato
by

**Tao Xie**

June 1999

# ABSTRACT

The $\alpha_{S1}$-casein A variant protein results from a rare polymorphism in the bovine $\alpha_{S1}$-casein gene. It is dramatically different from other $\alpha_{S1}$-caseins in that it has a 13 amino acid (encoded by the entire exon 4 of the $\alpha_{S1}$-casein gene) deletion, which significantly alters the physico-chemical properties of the protein. Previous work suggested that the deletion is caused by exon 4 skipping during RNA processing and that a T→A transversion mutation at position +6 in the intron 4 splice donor site is responsible for this skipping (Mohr *et al.* 1994). The current investigation initially focused on the genetic aetiology of variant A animals in an New Zealand herd; it was found that a single base (adenine) deletion, which is different from the previously reported mutation in the $\alpha_{S1}$-casein gene, occurs at position +4 of the intron 4 splice donor site in these animals. An amplification created restriction site (ACRS) based method has been developed to genotype this variant of calves and bulls. Later work centred on the investigation of aberrant $\alpha_{S1}$-casein A variant RNA and protein arising from RNA processing perturbations rather than mutations in the coding sequences of the gene *per se*. Interestingly, it was found that $\alpha_{S1}$-casein A variant mRNA and protein is actually expressed by all normal cows examined, at 1-5 percent of the level of the normal gene product! In addition, the identification of low levels of exon 17 and exon 4&5 aberrant skipped transcripts beside that of the exon 4 skipping suggests that, under normal circumstances, bovine $\alpha_{S1}$-casein gene mRNA undergoes multi-exon-skipping during processing. Thus, it is possible that this gene might also express a spectrum of other hitherto unsuspected, truncated, $\alpha_{S1}$-casein proteins. These studies utilised the sloughed off mammary epithelia cells in milk to characterise the $\alpha_{S1}$-casein mRNAs. This technique was found to provide an ideal non-invasive approach for both gene expression studies on the mammary gland and as a screening technique for milk protein gene polymorphisms.

# ACKNOWLEDGMENTS

During the course of the research presented in this thesis, a number of people provided valuable contributions, which I would like to acknowledge.

First to my supervisor Dr. Dick Wilkins, whose encouragement, wisdom, advice and patience throughout has been pivotal in the completion of this thesis.

I would also like to express my deepest gratitude to Dr. Dave Musgrave, Dr. Ray Cursons and his wife Dr. Gwyneth Verkerk, and Ms. Raewyn Towers, for their kind support in and out of the research laboratory.

To my "lab mates": Bridget, Linda, the two Gregs, Anandan, Peter, Stefan, Gune, Marcel, Quanah, Julie, Ivy, Jo and Seumas for their assistance, understanding, friendship and humour that made my work that much more enjoyable.

I want to extend a warm thanks to Dr. H. Davey, Dr. C. Coker, Dr. D. Otter as well as the other people, here unnamed, from AgResearch, DRI, DRC, LIC and the University of Waikato who assisted with helpful technical counselling, administrative support and generous supply of experimental materials. During the course of this work, founding and resources from Marsden grant UOW 602 contributed to the research.

Finally, to Xi, to Kaiyi, to my Mum and Dad. I could not complete this work without your support and contributions. This thesis is for you.

# TABLE OF CONTENTS

*Abstract* ..... ii
*Acknowledgments* ..... iii
*Table of contents* ..... iv
*List of figures* ..... viii
*List of tables* ..... x
*Abbreviations* ..... xi

*Chapter I*
**LITERATURE REVIEW**

**1.1 Bovine milk and its proteins:** ..... 2

1.1.1 Milk composition ..... 2
1.1.2 Major protein components ..... 3
1.1.3 The genetic variants of major milk proteins ..... 6
1.1.4 Production, composition & processing consequences of genetic variations ..... 9

**1.2 The structural characteristics of caseins** ..... 11

1.2.1 The primary structures ..... 11
1.2.2 The Secondary and tertiary structures ..... 13
1.2.3 Casein micelle models ..... 13

**1.3 The structure and expression of casein genes** ..... 15

1.3.1 Chromosomal location ..... 15
1.3.2 Genomic organisation of casein genes ..... 16
1.3.3 The evolution of casein genes ..... 19
1.3.4 Expression regulation: *cis*-elements and transcription factors ..... 21
1.3.5 Gene expression in the mammary gland epithelial cells ..... 23

**1.4 The splicing and aberrant splicing of pre-mRNA** ..... 25

1.4.1 The conserved splice sites, spliceosome and exon definition ..... 25
1.4.2 Pre-mRNA intron removal and exon joining ..... 29
1.4.3 Alternative splicing and its common patterns ..... 31
1.4.4 Mutations at splice sites and their effects on splicing ..... 33
1.4.4.1 5' splice site mutations and their effect on pre-mRNA splicing ..... 33
1.4.4.2 3' splice site mutations and their effect on pre-mRNA splicing ..... 34
1.4.4.3 The phenotypic consequences of splice site mutation ..... 34
1.4.4.4 The mechanism of exon skipping ..... 35

**1.5 Alternative splicing in milk protein genes** ..... 36

1.5.1 The $\alpha_{S1}$-casein gene ..... 36
1.5.2 Exon skipping and $\alpha_{S1}$-casein gene ..... 37
1.5.3 Alternative splicing in other milk protein genes ..... 37

**1.6 The aim of this project** ..... 40

*Chapter II*
MATERIALS AND METHODS

**2.1 Techniques, reagents and sample sources** ..... 43

**2.2 Sample preparation** ..... 46

2.2.1 DNA isolation from blood samples ..... 46
2.2.2 RNA preparation from mammalian cells ..... 47
2.2.3 cDNA preparation ..... 48
2.2.4 Casein isolation ..... 48

**2.3 Nucleic acid manipulations** ..... 49

2.3.1 Polymerase chain reaction ..... 49
2.3.2 PCR products purification and band isolation ..... 50
2.3.3 Radioactive labelling ..... 50
2.3.4 Denaturing PAGE for labelled PCR fragments ..... 51
2.3.5 Southern hybridisation ..... 51
2.3.6 PCR product cloning ..... 52

2.3.6.1 Preparation of pBluescript® II SK vector ..... 52
2.3.6.2 Preparation of competent cells ..... 53
2.3.6.3 Ligation ..... 53
2.3.6.4 Transformation ..... 54
2.3.6.5 The identification of recombinant plasmids ..... 54
2.3.6.6 Single strand DNA rescue ..... 55

2.3.7 DNA sequencing ..... 56
2.3.8 GenScan analysis ..... 56

**2.4 Protein manipulations** ..... 56

2.4.1 Urea-PAGE for casein proteins ..... 56
2.4.2 Western blotting ..... 57
2.4.3 Fast protein liquid chromatography ..... 58

*Chapter III*
THE DRI $\alpha_{S1}$-CN A VARIANT AND IDENTIFICATION OF THE GENE MUTATION

**3.1 Introduction** ..... 61

**3.2 The DNA mutation in DRI $\alpha_{S1}$-casein A cows** ..... 63

**3.3 The design of a PCR test for the DRI $\alpha_{S1}$-casein A mutation** ..... 66

**3.4 Discussion** ..... 72

3.4.1 The relationship between $\alpha_{S1}$-casein A gene mutation and aberrant splicing.... 72
3.4.2 PCR based testing for $\alpha_{S1}$-casein A mutations ..... 75

*Chapter IV*
A VARIANT PROTEIN EXPRESSION OF THE NORMAL $\alpha_{S1}$-CN GENE

**4.1 Introduction** ..... 80

**4.2 The identification of the exon 4 skipped product** ..... 83

4.2.1 RT-PCR analysis ..... 83

4.2.2 Oligonucleotide hybridisation analyses ..... 85
4.2.3 DNA sequencing of the exon 4 skipped product ..... 88

**4.3 Quantitative analysis of the exon 4 skipping** ..... 90

4.3.1 Quantitation by RT-PCR – Southern hybridisation ..... 91
4.3.2 Quantitation by GeneScan analysis ..... 93

**4.4 Identification of $\alpha_{S1}$-CN A protein in normal milk** ..... 97

4.4.1 $\alpha_{S1}$-CN A protein detection by urea-PAGE ..... 97
4.4.2 Western blot for identification of $\alpha_{S1}$-CN A protein ..... 100
4.4.3 Isolation of $\alpha_{S1}$-CN variant A protein from normal milk ..... 103

**4.5 Discussion** ..... 105

4.5.1 The validity of sampling RNAs from milk ..... 105
4.5.2 The methods for quantitative analysis ..... 107
4.5.2.1 Quantitative RT-PCR ..... 107
4.5.2.2 GeneScan analysis ..... 109
4.5.3 "Leaky" exon skipping of the $\alpha_{S1}$-casein gene ..... 110

*Chapter V*
**SCANNING THE $\alpha_{S1}$-CASEIN GENE FOR OTHER EXON SKIPPING EVENTS AND CRYPTIC MUTATIONS**

**5.1 Introduction** ..... 113

**5.2 Additional anomalous RT-PCR products in the exon 4 region of the $\alpha_{S1}$-CN mRNA (exon 4&5 skipping)** ..... 113

5.2.1 Sequence analysis of the 192 bp RT-PCR product ..... 114
5.2.2 The cloning and sequencing of the 192 bp product ..... 116
5.2.3 DNA sequencing of flanking regions of the genomic DNA ..... 117

**5.3 $\alpha_{S1}$-Casein exon 17 skipping** ..... 118

5.3.1 RT-PCR analysis of the full length coding region of $\alpha_{S1}$-CN mRNA ..... 119
5.3.2 Characterisation of the minor band ..... 120
5.3.2.1 *Pvu* II digestion ..... 120
5.3.2.2 Sequencing results ..... 121

**5.4 Screening of the full length coding region** ..... 122

**5.5 Discussion** ..... 124

5.5.1 Exon skipping events in the region of exon 4 ..... 124
5.5.2 The putative protein translated from exon 17 skipped mRNA ..... 125

*Chapter VI*
**CONCLUSIONS**

6.1 **RNA sampling from milk permits large scale gene expression studies** ..... 129

6.1.1 Practical and ethical advantages ..... 129
6.1.2 Basic research advantages ..... 130

**6.2 The $\alpha_{S1}$-casein A variant** ..... 131

6.2.1 What range of mutations could give rise to the A variant? ... 131
6.2.2 PCR tests ... 132

**6.3 Synthesis of small amounts of A variant $\alpha_{s1}$-CN protein in normal animals** ... 133

**6.4 Other mRNA transcripts arising via alternative splicing from the bovine $\alpha_{s1}$-casein gene** ... 135

**6.5 Do exon skipping events occur in all $\alpha_{s1}$-casein genes?** ... 136

**6.6 A new approach for detecting milk protein genetic variants** ... 137

**6.7 Future work** ... 140

6.7.1 Exon skipping events in $\alpha_{s1}$-CN and other milk protein genes ... 140
6.7.2 Qualitative and quantitative analysis of putative protein products produced from exon skipped mRNAs ... 140
6.7.3 The dynamics of exon skipping ... 141
6.7.4 The mechanism of leaky exon skipping ... 142

***Appendices*** ... 144

I Average conc. of some constituents in the milk of a number of species ... 144
II Organisation of milk protein genes ... 144
III The full sequence of the *B, taurus* gene for alpha-s1-casein ... 145
IV Exon skipping reported in major milk protein genes ... 150
V Single base-pair substitutions in the 5' splice sites of human genes ... 151

***List of publications / presentations*** ... 154

***References*** ... 156

# LIST OF FIGURES

**Figure 1.2.3** Two models of the casein micelle ..... 15

**Figure 1.3.1** Physical map of the bovine casein locus ..... 17

**Figure 1.3.2** Organisation of the bovine casein genes ..... 18

**Figure 1.3.3** Hypothetical scheme for the evolution of $\alpha_{S2}$-casein and β-casein from a common ancestor ..... 20

**Figure 1.3.5** Diagram of the transcriptional unit and cognate mRNA encoding $\alpha_{S1}$-casein ..... 24

**Figure 1.4.1.1** Splice sites consensus of vertebrate genes ..... 26

**Figure 1.4.1.2** Early steps of splice site recognition ..... 28

**Figure 1.4.2** A schematic representation of two steps of pre-mRNA splicing ..... 30

**Figure 1.4.3** Alternative splicing patterns ..... 32

**Figure 3.1.1** A schematic diagram shows the occurrence of the fourth exon skipping of the German Red $\alpha_{S1}$-casein A variant ..... 63

**Figure 3.2.1** PCR amplification of DNA from $\alpha_{s1}$-casein A/A and B/B animals ..... 65

**Figure 3.2.2** The 5' splice junction of intron 4 of $\alpha_{S1}$-CN gene ..... 65

**Figure 3.3.1** Designing of amplification created restriction site (ACRS) primers ..... 68

**Figure 3.3.2** The PCR products amplified by ACRS primers Spe and Tru ..... 69

**Figure 3.3.3** The two-step ACRS-PCR method to check the $\alpha_{S1}$-casein DRI A variant.. 70

**Figure 3.3.4** Genotyping newborn calves by the two-step ACRS PCR method ..... 71

**Figure 4.1.1.1** A typical FPLC separation of caseins on a Mono Q column ..... 82

**Figure 4.2.1.1** The schematic diagram showing two RT-PCR reactions capable of detecting $\alpha_{S1}$-casein exon 4 skipping ..... 84

**Figure 4.2.1.2** RT-PCR amplification of $\alpha_{S1}$-casein mRNA using primers A6/A7 ..... 85

**Figure 4.2.2.1** Oligonucleotide probes designed to detect normal and exon 4 skipped RT-PCR products ..... 86

**Figure 4.2.2.2** Electrophoresis of triplicate RT-PCR products for blotting ..... 87

**Figure 4.2.2.3** The results of oligonucleotide hybridisations ..... 88

**Figure 2.2.3.1** Isolation of the 216 bp RT-PCR Product ..... 89

**Figure 4.2.3.2** Sequencing results of 216 bp and 255 bp PT-PCR fragments ..... 90

**Figure 4.3.1.1** Southern hybridisation of RT-PCR products amplified by A6/A7 primers. 92

**Figure 4.3.2 A–E** Examples of electropherograms in the GeneScan analyses ..... 96

**Figure 4.4.1.1** Urea-PAGE analysis of whole casein and standards ..... 99

**Figure 4.4.1.2** Casein from fresh milk samples electrophoresed by urea-PAGE ..... 100

**Figure 4.4.2.1** Flow diagram for ECL immunodetection ..... 102

**Figure 4.4.2.2** Western blotting analysis using an anti-$\alpha_{S1}$-casein antibody ..... 103

**Figure 4.4.3.1** Bovine caseins separated by FPLC ..... 104

**Figure 4.4.3.2** Urea-PAGE analysis of the $\alpha_{S0}$-CN fraction prepared by FPLC ..... 105

**Figure 5.2.1.1** The isolated 192 bp product ..... 115

**Figure 5.2.1.2** Sequencing result of the 192 bp product ..... 115

**Figure 5.2.1.3** The sequence of the 255 bp product from the cow with the putative heterozygotic exon 6 ..... 115

**Figure 5.2.2.1** The sequence of cloned 192 bp product with clean exon 4 and 5 skipping. ..... 116

**Figure 5.2.2.2** The sequence of cloned 192 bp product with exon 5 skipping and exon 4 and 6 partial deletion ..... 116

**Figure 5.2.3.1** A schematic diagram of amplification of the exon 4 - 6 region ..... 117

**Figure 5.2.3.2** The PCR products amplified for sequencing ..... 118

**Figure 5.2.3.3** The sequence of the PCR fragment amplified using A8/A10 primers ..... 118

**Figure 5.3.1.1** The primers used to amplify the full length $\alpha_{S1}$-casein coding region ..... 119

**Figure 5.3.1.2** RT-PCR result using primers A13/A7 ..... 120

**Figure 5.3.2.1** Restriction map for *Pvu* II digestion ..... 121

**Figure 5.3.2.2** The result of *Pvu* II digestion of full length and truncated $\alpha_{S1}$-casein cDNA ..... 121

**Figure 5.3.2.3** The sequence of the 521 bp minor product amplified by A7/A13 ..... 122

**Figure 5.4.1** The RT-PCR product for $\alpha_{S1}$-casein cryptic variant scanning ..... 123

**Figure 5.4.2** The alignment of $\alpha_{S1}$-CN coding sequences from 9 normal cows ..... 124

**Figure 5.5.2** Sequences alignment of the normal $\alpha_{S1}$-CN and the putative protein generated by exon 17 skipping ..... 127

# LIST OF TABLES

**Table 1.1.1** Average composition of bovine milk ..... 2

**Table 1.1.2** Major protein components of bovine milk ..... 3

**Table 1.1.3** Location of amino acid substitution in genetic variants of milk proteins identified in *Bos taurus* ..... 8

**Table 1.1.4** Effects of genetic variants in bovine milk ..... 10

**Table 1.2.1** The primary structure of caseins ..... 12

**Table 1.3.4** Potential binding sites for nuclear factors in the 5'-flanking regions of bovine casein genes ..... 22

**Table 2.1.1** DNA samples ..... 44

**Table 2.1.2** RNA samples ..... 44

**Table 2.1.3** Oligonucleotides for PCR, RT-PCR and Southern hybridisation analysis ..... 45

**Table 2.3.1** PCR and RT-PCR analyses ..... 49

**Table 2.4.1** Composition of urea-PAGE gels ..... 57

**Table 3.4.1** Variations from consensus in the splice sites of the bovine $\alpha_{S1}$-CN gene ..... 74

**Table 3.4.2** Relative amplification efficiencies of 3'-terminal mismatches in the Presence of 800 μM dNTPs ..... 77

**Table 4.5.3.1** Complete exon skipping in major milk protein genes ..... 111

# LIST OF ABBREVIATIONS

| | |
|---|---|
| α | alpha |
| A | adenine or adenosine |
| ADP | adenosine diphosphate |
| APS | ammonium persulfate |
| β | beta |
| ME | β-mercaptoethanol |
| BCIP | 5-bromo-4-chloro-3-indolyl phosphate |
| bp | base pairs |
| C | cytosine or cytidine |
| $Ca^{2+}$ | calcium |
| $CaCl_2$ | calcium chloride |
| cDNA | complementary DNA |
| cm | centimeter |
| cpm | counts per minite |
| DNA | deoxyribonucleic acid |
| DNase | deoxyribonuclease |
| DTT | dithiothreitol |
| EDTA | ethylenediamine tetraacetic acid |
| EtOH | ethanol |
| FPLC | fast protein liquid chromatography |
| γ | gamma |
| g | gram |
| G | guanine or guanosine |
| HPLC | high performance liquid chromatography |
| HnRNPs | heterogeneous ribonucleoprotein particles |
| IPTG | β-D-isopropyl-thiogalactopyranoside |
| kb | kilobase |
| kDa | kilodalton |
| λ | lambda |
| l | liter |
| M | molar |
| mg | milligram |
| $Mg^{2+}$ | magnesium |
| $MgCl_2$ | magnesium chloride |

| | |
|---|---|
| ml | milliliter |
| mm | millimeter |
| mM | millimolar |
| mRNA | messenger RNA |
| MW | molecular weight |
| NaCl | sodium chloride |
| nmole | nanomole |
| nt | nucleotides |
| -P | phosphate group |
| PAGE | polyacrylamide gel electrophoresis |
| pg | picomole |
| PMN | polymorphonuclear leucocytes |
| PMSF | phenyl methyl sulfonyl fluoride |
| RNA | ribonucleic acid |
| RNase | ribonuclease |
| SDS | sodium dodecyl sulfate |
| snRNPs | small nuclear ribonucleoprotein particles |
| T | thymine or thymidine |
| TE | Tris-EDTA buffer |
| tRNA | transfer RNA |
| μ | micro |
| U | Uracil or uridine |
| mg | milligram |
| ml | milliliter |
| mM | millimolar |
| (v/v) | volume:volume ratio |
| (w/v) | weight:volume ratio |
| X-Gal | 5-bromo-4-chloro-3-indolyl-β galactopyranoside |

# *CHAPTER I*

# LITERATURE REVIEW

*Bovine milk and its proteins*
*The structural characteristics of caseins*
*The structure and expression of casein genes*
*The aberrant splicing of pre-mRNA*
*Alternative splicing in milk protein genes*
*The aim of this project*

# 1.1 Bovine Milk and Its Proteins

## 1.1.1 Milk composition

Milk is a polyphasic secretion of the mammary gland. The major constituents consist of proteins, fats, sugars, water and the minor constituents include a variety of biochemicals, vitamins and mineral ions. The average composition of bovine milk is shown in Table 1.11. These basic components of milk are found in all mammalian species, but considerable quantitative differences are found in the concentrations of fats, proteins and carbohydrates (as there is in the precise nature of these substances). Indeed, even within a species such as bovine, considerable variations are observed between breeds such as Holstein and Jersey and, in addition, milk varies between individuals and at different lactation stages (Hartmann and Prosser 1984).

**Table 1.1.1 Average composition of bovine milk**

| *Component* | *g/100 ml* |
|---|---|
| Lipids | ~3.7 |
| proteins | ~3.3 |
| Lactose | ~4.6 |
| minerals | ~0.8 |
| water | 87.3 |

The milk proteins provide a complete source of amino acids, especially the essential amino acids, plus nitrogen for the neonate and have a number of other non-nutritional functions. Milk fats are also a very important component which contributes over 50 percent of the gross energy supplied by milk. Moreover, lactose acts as an energy source and is a significant component of the milks produced by many mammals, e.g. humans, cows, sheeps, goats, pigs and mice. Interestingly, it is also the major osmole and thus largely determines the water content of the milk, which is regulated so that the milk and blood plasma are near isotonic. In some mammals such as seals and whales, lactose levels are very low, which means that mineral ions and higher molecular weight components are effectively the only osmoles, with the consequence the milk is very concentrated with very little water being required to make it isotonic with blood plasma (see Mepham 1987 for a general review).

## 1.1.2 Major protein components

In bovine milk, there are two major protein types: caseins and whey. Caseins were defined originally as those phosphoproteins that precipitated from raw skim milk when it was either acidified to pH 4.6 at 20°C or digested with the gastric enzyme chymosin (also called rennin) (Jenness *et al.* 1956). These proteins are synthesised in the mammary gland and comprise approximately 80% of the total milk proteins. Four casein types, $\alpha_{S1}$, $\alpha_{S2}$, β and κ are found in bovine milk. Whey proteins are those which remain in solution at the pH of casein precipitation, and these consist of both mammary synthesised proteins such as β-lactoglobulin, α-lactalbumin and lactoferrin and a number of minor proteins originating from the blood, notably, serum albumin, immunoglobulins and β-microglobulin. At different stages of lactation and in different disease conditions, the amounts of lactoferrin and some of the blood derived proteins can vary markedly.

**Table 1.1.2 Major protein components of bovine milk**

| | Quantity (g/l) | | Proportion (%) | |
|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** |
| ***Caseins*** | ***26.92*** | ***1.54*** | ***82.2*** | ***0.6*** |
| $\alpha_{S1}$-Casein | 10.25 | 0.57 | 31.3 | 0.5 |
| β-Casein | 9.60 | 0.50 | 29.3 | 0.6 |
| κ-Casein | 3.45 | 0.32 | 10.5 | 0.5 |
| $\alpha_{S2}$-Casein | 2.74 | 0.21 | 8.4 | 0.5 |
| γ-Casein | 0.88 | 0.15 | 2.7 | 0.4 |
| ***Whey Proteins*** | ***5.79*** | ***0.32*** | ***17.8*** | ***0.6*** |
| β-Lactoglobulin | 3.14 | 0.19 | 9.6 | 0.4 |
| α-Lactalbumin | 1.23 | 0.09 | 3.8 | 0.3 |
| IPL | 0.97 | 0.10 | 3.0 | 0.2 |
| Serum albumin | 0.45 | 0.04 | 1.4 | 0.1 |
| ***Total*** | 32.71 | 1.80 | | |

*(Data source: Davies et al. 1983. SD, standard deviation; IPL, immunoglobulin, proteose-peptone components and lactoferrin)*

The classification of milk proteins is somewhat confusing as in the early days of milk protein chemistry, technical limitations resulted in a number of casein products being incorrectly identified as distinct proteins rather than variations of one basic protein. The American Dairy Science Association (ADSA) Committee on the nomenclature and methodology of milk proteins differentiated caseins according to their relative

electrophoretic mobility in alkaline polyacrylamide or starch gels containing urea with or without β-mercaptoethanol (Whitney *et al.* 1976). However, such differentiation was not especially easy because of various and variable post-translational phosphorylations, glycosylations and genetic polymorphisms. Later it was possible to classify the caseins more accurately using protein sequencing techniques which enabled comparisons of homologies to be made at the primary structure level (amino acid sequences). The whey proteins were somewhat easier to study as they were more easily resolved by size alone by a number of electrophoretic techniques. Indeed, polyacrylamide or starch gel electrophoresis still can be used to characterize and identify individual members of each family.

## Caseins

$\alpha_{S1}$-Casein, $\alpha_{S2}$-casein and β-casein are called ***calcium sensitive caseins*** because of their ability to bind metal ions, in particular calcium, strongly. Indeed, this high capacity for calcium is extremely important as it enables milk to deliver high levels of calcium to the neonate. Substantial amounts of other ions such as zinc are also chelated by the phosphate residues of caseins. By themselves, these three caseins would precipitate but κ-casein plays a crucial role by interacting with and forming micelles with these three calcium sensitive caseins.

For each of the casein families, several distinct protein species can be found because of post-translational modifications and genetic polymorphisms. Based on the recommendation of the ASDA committee, the post-translational variations, such as phosphorylation, are designated by an Arabic number and a letter after the Latin letter that designates the genetic variant. Thus, for example, the major genetic variant (B) of $\alpha_{S1}$-caseins in *Bos taurus* would be designated $\alpha_{S1}$-CN B-8P. However, this full nomenclature is not always used.

Electrophoretically, the ***$\alpha_{S1}$-caseins*** ($\alpha_{S1}$-CN) can be resolved into one major and one minor component. The major component contains 8 phosphorylated amino acid residues and is designated $\alpha_{S1}$-CN B-8P. The minor component, historically classified as $\alpha_{S0}$-casein, actually has the same amino acid sequence as $\alpha_{S1}$-casein but contains one

additional phosphorylated residue at position 41. So the term $\alpha_{S0}$-casein really should be dropped in favour of $\alpha_{S1}$-CN B-9P.

The major component of the ***$\alpha_{S2}$-caseins*** is $\alpha_{S2}$-CN A-13P, and the minor components are $\alpha_{S2}$-CN A-12P, $\alpha_{S2}$-CN A-11P and $\alpha_{S2}$-CN A-10P. In earlier literature, these last three minor products were classified as $\alpha_{S2}$, $\alpha_{S3}$, $\alpha_{S4}$. A fifth product named $\alpha_{S5}$ in the earlier literature is actually a dimer consisting of $\alpha_{S3}$- $\alpha_{S4}$-casein molecules linked together by a disulfide bond.

In the case of ***β-caseins*** (β-CN), considerable confusion arose in early work because three distinct proteins could be identified in bovine milk, designated γ1-, γ2- and γ3-casein, which primary structure information later revealed to be degradation products of β-casein itself. Plasmin, a protease in milk is responsible for the proteolysis of β-casein, which leads to these products.

***κ-Casein*** (κ-CN) is the only glycosylated member of the casein group of proteins. As isolated from milk, κ-CN occurs in polymeric forms linked via intermolecular disulfide bonds into multimers ranging in molecular weight from 60 to 600 kDa. In alkaline gel electrophoresis, in the presence of β-mercaptoethanol and urea, κ-CN electrophoreses as a monomer. However, distinct electrophoretic bands are observed in different milk samples; these correspond to variant forms of the protein with the A variant possessing the greatest mobility.

## Whey proteins

***a-Lactalbumin*** (α-LA) is a small whey protein molecule, MW 14.6 KDa, and it has considerable homology with lysozyme. It is unusual amongst the milk proteins in that it has an essential cofactor role in lactose synthesis, binding to the lactose synthetase complex within mammary epithelial cells and increasing the $K_m$ of the complex for glucose some 100-fold. In the absence of α-lactalbumin, lactose synthesis in the mammary gland is almost non-existent, as observed in α-lactalbumin-gene-disrupted mice (Stinnakre *et al.* 1994); in lactating (-/-) female mice, very low levels of α-lactalbumin result in very low lactose level which are ineffective as an osmole with the result that the milk is a viscous concentrate which does not secrete properly and can not support the pups. This is obviously an extreme case of a more general phenomenon,

namely that for a range of mammals, the concentration of α-lactalbumin and lactose in milk are correlated (Mepham 1987).

***β–Lactoglobulin*** is a globular protein and the major whey protein (7-12% of total protein in skim milk) in bovine. However, it is not found in all mammals, e.g. in the mouse which, instead, possesses a whey acidic protein (WAP). β-Lactoglobulin has been extensively studied by a variety of techniques but its biological function remains uncertain. Possibly β-lactoglobulin has a role as a carrier of small hydrophobic molecules.

A large number of other proteins and peptides are found in milk at low concentrations. Some of these have potential non-nutritional functions, e.g. the immunoglobulins, enzymes and enzyme inhibitors, binding or carrier proteins, growth factors and antibacterial agents. Among them, the immunoglobulins are not unique to lacteal secretions. Except for secretory component (SC) which forms secretory immunoglobulin A (SIgA), with immunoglobulin A (IgA), all immunoglobulins in milk are also normal components of serum. Lactotransferrins are a group of evolutionarily related specific metal-binding (especially iron binding) proteins. Serum Albumin (SA) prepared from milk is physically and immunologically identical to blood SA. For some factors such as the insulin growth factors and growth hormone, which were once thought to originate from serum, local production in the mammary gland may also occur.

### 1.1.3 The genetic variants of major milk proteins

Genetic variants (or genetic polymorphisms) have been found in dairy cattle populations for all seven major milk proteins (Ng-Kwai-Hang and Grosclaude 1992; Eigel *et al.* 1984). These variant forms usually involve either substitutions of amino acids or the deletions of specific amino acids sequences within the peptide chain and result from mutations in the corresponding genes. Historically, most of the variants were detected electrophoretically as the result of changes in either the net electrical charge, or in a few cases, the conformation of the protein. As many authors have pointed out, such electrophoretic techniques are relatively insensitive and would not be expected to reveal all of amino acid substitutions or micro-deletions. Thus, one might

reasonably expect that more comprehensive investigations, involving protein and DNA sequencing will reveal a number of hitherto unknown variants.

- For $\alpha_{S1}$-casein, six genetic variants are known and are designated A, B, C, D, E and F. The first five variants (A to D) are designated in order of decreasing relative electrophoretic mobilities in alkaline gels containing urea (Peterson 1963). The B variant is predominant in *Bos taurus* and the C variant is also relatively common. Unlike other single amino acid substitution variants, the A variant contains a 13 amino acid deletion (residues 14-26). The A, D and F variants are rare in *Bos taurus* (Prinzenberg *et al.* 1998). No E variants are found in Western dairy herds.
- For $\alpha_{S2}$-casein there are four recognised genetic variants. The A and D variants are observed in European breeds (*Bos taurus*) with D in the Vosgienne and Montbeliarde breeds. Variant B is found in *Bos taurus* and *Bos indicus* (Nepalese), and variant C is observed specifically in yaks (*Bos grunniens*).
- Seven genetic variants of β-casein are known, $A^1$, $A^2$, $A^3$, B, C, D and E. Their differentiation by gel electrophoresis is more complicated than for the other caseins. $A^1$, $A^2$ and B are the more common variants.
- Six genetic variants of κ-casein have been reported. The A variant tends to be predominant in most breeds and the B is another common variant in Western cattle.
- Three genetic variants A, B and C are known to occur in α-lactalbumin. Only the B variant has been observed in milk from Western cattle.
- The positions of amino acid substitutions have been determined for seven β-lactoglobulin variants, A and B are common in *Bos taurus*. C is not uncommon while E, F and G are not found in Western cattle.

The known genetic variants and their amino acid changes are summarised in Table 1.1.3. So far, the most common tool for exploring genetic variation is protein electrophoresis. It seems likely, on a probability basis, that more variants will be discovered in dairy cattle. The existence of such variants may have been overlooked because of the complexity of the casein electrophoretic patterns resulting from the variable post-translational modifications. There are more uncharged than charged

amino acids in proteins and thus there are probably even more uncharged than charged genetic variants.

**Table 1.1.3 Location of amino acid substitution in genetic variants of milk proteins identified in *Bos Taurus***

| *Proteins* | *Variants* | *Mol. W.* | *Position and amino acid in the protein* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha_{S1}$-CN | | | 14-26 | | | 53 | | 59 | 66 | 192 | |
| (199) | A | 22,068 | Del. | | | | | | | | |
| | B | 23,614 | | | | Ala | | Gln | SerP | Glu | |
| | C | 23,542 | | | | | | | | Gly | |
| | D | 23,724 | | | | ThrP | | | | | |
| | E | 23,542 | | | | | | Lys | | Gly | |
| | F | | | | | | | | Leu | | |
| $\alpha_{S2}$-CN | | | 33 | | 47 | | 50-58 | | 130 | | |
| (207) | A | | GLu | | Ala | | | | Thr | | |
| | B | | | | | | | | | | |
| | C | | Gly | | Thr | | | | Ile | | |
| | D | | | | | | Del. | | | | |
| β-CN | | | 18 | 35 | 36 | 37 | 67 | 106 | 122 | | |
| (209) | A1 | 24,023 | | | | | His | | | | |
| | A2 | 23,983 | SerP | SerP | Glu | Glu | Pro | His | Ser | | |
| | A3 | 23,974 | | | | | | Gln | | | |
| | B | 24,092 | | | | | His | | Arg | | |
| | C | 23,944 | | Ser | | Lys | His | | | | |
| | D | 23,944 | Lys | | | | | | | | |
| | E | 23,982 | | | Lys | | | | | | |
| κ-CN | | | 97 | | 136 | | | 148 | | 155 | |
| (169) | A | 19,039 | | | Thr | | | Asp | | Ser | |
| | B | 19,007 | Arg | | Ile | | | Ala | | | |
| | C | | His | | | | | | | | |
| | E | | | | | | | | | Gly | |
| α-LA | | | | | | 10 | | | ? | | |
| (123) | A | 14,147 | | | | Gln | | | | | |
| | B | 14,175 | | | | Arg | | | Asp | | |
| | C | | | | | | | | Asn | | |
| β-LG | | | 45 | 50 | 59 | 64 | 78 | 118 | 129/130 | 130 | 158 |
| (162) | A | 18,363 | | | | Asp | | Val | | | |
| | B | 18,277 | Glu | Pro | Gln | Gly | Ile | Ala | Asp | Asp | Glu |
| | C | 18,286 | | | His | | | | | | |
| | D | 18,276 | Gln | | | | | | | | |
| | E | 18,205 | | | | | | | | | Gly |
| | F | 18,243 | | Ser | | | | | Tyr | | Gly |
| | G | 18,233 | | | | | Met | | | | Gly |

*Number in parentheses indicated the total number of amino acid residues in the protein; Del.: deletion (adapted and modified from Ng-Kwai-Hang and Grosclaude 1992)*

### 1.1.4 Production, composition and processing consequences of genetic variations

In the past 15 years, several groups of researchers around the world have indicated that there are possible associations between some genetic variants of milk proteins and production traits, in particular milk composition (Hayes *et al.* 1983; Mclean *et al.* 1984; Bovenhuis *et al.* 1992; Winkelman and Wickham 1997). The reports relating milk protein variants with milk production are inconclusive and, at times, contradictory. This is mainly due to factors such as relatively small animal numbers in some trials, different breeds of cattle and different methods of determining milk yields. In addition, the rigour of statistical analysis differ among the various studies, for instance, when correcting for major factors, such as breed of cows, stage of lactation, season, health status, age and genetic status of the animal or other protein variant which could contribute to milk production. However, it appears that reports on the associations of genetic variants with different milk components are more consistent than those for milk production. The best examples are the κ-casein B variant and the β-lactoglobulin A variant, both of which are associated with higher milk protein concentrations. The ratio of casein to whey protein is different for the different variants of κ-casein and β-lactoglobulin (Ng-Kwai-Hang 1997). High ratios are generally regarded as advantageous for manufacturing a range of products from milk .

The processing properties of milk are largely influenced by both the absolute amounts and the relative proportions of each of the milk constituents. The constitution of the protein fraction is of particular importance in this regard. In recent years, attention has focussed even more specifically on the actual variants of milk proteins present in the milk. *A priori* one might expect the differences in amino acid composition and sequence to lead to a series of modifications including net charge, hydrophobicity, degree of phosphorylation and glycosylation, all of which contribute to the behaviour of milk proteins and hence to the overall manufacturing properties of the milk. An example is the $\alpha_{S1}$-casein A variant protein which form micellar aggregates with κ-casein which differ from those formed by the $\alpha_{S1}$-casein B and C variants (Thompson *et al.* 1969; Anema and Creamer 1993). The milk and products containing $\alpha_{S1}$-casein A protein behave differently from those containing $\alpha_{S1}$-casein B or C; for example in the

viscosity of the whole casein. This phenomena has been related to the decreased hydrophobic character of $\alpha_{S1}$-casein A which can be linked to the loss of a region that has a significant concentration of hydrophobic residues. More recently, Coker *et al.* (1997) have suggested that $\alpha_{S1}$-casein A affects the normal maturation of cheese because the major chymosin-sensitive $Phe^{23}$-$Phe^{24}$ bond of $\alpha_{S1}$-casein B and C is in the amino acid segment that is deleted in $\alpha_{S1}$-casein A.

**Table 1.1.4 Effects of genetic variants in bovine milk**

| *Protein* | *Effect on milk composition* | *Effect on cheese processing* | *Effect on conc.ed. milk* | *Possible reasons* |
|---|---|---|---|---|
| $\alpha_{s1}$-casein A | | --yields a softer curd and cheese body (Creamer *et al.* 1988) | | Loss of a hydrophobic domain and of a phosphorylation site |
| $\alpha_{s1}$-casein B | --Correlates with higher milk yield (Ng-Kwai-Hang *et al.* 1984) | | --No significant differences among variants (Mclean *et al.* 1984) | |
| $\alpha_{s1}$-casein C | --Lower concentration of k-casein (Mclean *et al.* 1984) | | | |
| β-casein A1<br>β-casein A2 | | | --No significant differences among variants (Mclean *et al.* 1984) | |
| β-casein B | --Significant increase in fat (Maclean *et al.* 1984) | --Gives faster and firmer coagulum | | An increased quantity of the B variant in milk and /or increase in positive charge of the B variant may enhance casein interactions |
| κ-casein A | | --gives a longer clotting time | | |
| κ-casein B | --Significant increase in concentration of κ-casein and decrease in citrate (Mclean *et al.* 1984) | --Gives faster and firmer coagulation, and better yield | confers greater heat stability (Maclean *et al.* 1984) | The larger amounts of κ-casein may result in smaller micelles, along with differences in composition |

| | | | | |
|---|---|---|---|---|
| β-LG B (for the variant A/B) | --Significant increase in fat, total solids, casein and a decrease in whey protein (Mclean *et al.* 1984) | --Significant increase in yield when κ-casein B is present | --confers greater heat stability (Mclean *et al.* 1984) | A marked decrease in the quantity of variant B may result in less interaction with other proteins upon heating |

*(adapted and modified from Richardson et al 1992)*

## 1.2 The Structural Characteristics of Caseins

### 1.2.1 The primary structures

The complete amino acid sequences of all four major caseins and their variants are summarised in Table 1.2.1 and Table 1.1.3 respectively. The casein sequences are rather different from those of most other proteins (Sawyer and Holt 1993), with a large number of proline residues (which prevent the formation of certain types of secondary structure) and acidic regions that include a number of phosphoserine residues. The cystinyl and cysteinyl residues in κ-casein and $\alpha_{S2}$-casein allow these proteins to participate in sulfhydryl-disulfide interchange and the absence of cysteinyl residues in $\alpha_{S1}$-casein and β-casein indicate these proteins can not participate in these crosslinking reactions.

Special properties are conferred on caseins as a result of their post-translational modifications. All caseins are phosphorylated to varying extents at seryl and, occasionally, threonyl residues. Genetic variants of κ-, $\alpha_{S1}$-, and β-caseins usually contain a characteristic number of phosphoseryl residues, viz. 1-P, 8-P or 9-P and 4-P or 5-P, respectively. The $\alpha_{S2}$-casein family exhibits the greatest variability in the extent of phosphorylation; for example, $\alpha_{S2}$-CN A has 10-13 phosphoseryl residues (Brignon *et al.* 1977; Whitney 1976). At present, it is not known whether this variability results from different extents of phosphorylation by casein kinases or from varying degree of dephosphorylation by phosphatases. In addition to post-translational phosphorylation, κ-casein is also glycosylated. These modifications cause extensive heterogeneity in the properties of this family of proteins.

**Table 1.2.1 The primary structure of caseins**

| *Proteins* | *Amino acid sequence* | | | | |
|---|---|---|---|---|---|
| $\alpha_{s1}$-casein | RPKHPIKHQG | LPQEVLNENL | LRFFVAPFPE | VFGKEKVNEL | SKDIG**S**E**S**TE |
| | DQAMEDIKQM | EAE**S**I**SSS**EE | IVPN**S**VEQKH | IQKEDVPSER | YLGYLEQLLR |
| | LKKYKVPQLE | IVPN**S**AEERL | HSMKEGIHAQ | QKEPMIGVNQ | ELAYFYPELF |
| | RQFYQLDAYP | SGAWYYVPLG | TQYTDAPSFS | DIPNPIGSEN | SEKTTMPLW |
| $\alpha_{s2}$-casein | KNTMEHV**SSS** | EESIISQETY | KQEKNMAINP | SKENLCSTFC | KEVVRNANEE |
| | EYSIG**SSS**EE | **S**AEVATEEVK | ITVDDKHYQK | ALNEINQFYQ | KFPQYLQYLY |
| | QGPIVLNPWD | QVKRNAVPIT | PTLNREQL**S**T | **S**EENSKKTVD | ME**S**TEVFTKK |
| | TKLTEEEKNR | LNFLKKISQR | YQKFALPQYL | KTVYQHQKAM | KPWIQPKTKV |
| | IPYVRYL | | | | |
| κ-casein | **Q**EQNQEQPIR | CEKDERFFSD | KIAKYIPIQY | VLSRYPSYGL | NYYQQKPVAL |
| | INNQFLPYPY | YAKPAAVRSP | AQILQWQVLS | NTVPAKSCQA | QPTTMARHPH |
| | PHLSFMAIPP | KKNQDKTEIP | TINTIASGEP | TSTPTTEAVE | STVATLEDSP |
| | EVIESPPEIN | TVQVTSTAV | | | |

*Phosphorylation sites are underlined. Date sources: Mercier et al. 1971; Brignon et al. 1977; Ribadeau-Dumas et al. 1972 and Mercier et al. 1973.*

Hydrophobic and charged resides are not uniformly distributed in the casein amino acid sequences. For example, three hydrophobic regions are discernible in $\alpha_{S1}$-casein, broadly including residues 1-44, 90-113 and 132-199. However, clustering of the phosphoseryl residues is perhaps the most unique feature of the primary structure of the calcium-sensitive caseins. The cluster sequence EpSXpSpSpSEE, where X is I or L and pS is SerP, occurs in all known variants of $\alpha_{S1}$- and β-casein. Two similar sequences, pSpSpSEE and pSpSpSEEpSAE, occur in $\alpha_{S2}$-caseins. All the phosphorylation cluster sites are located in the N-terminal region. Furthermore, almost all the sites are on or close to a flexible loop region formed by a β-turn structure (Holt and Sawyer 1988). These structures, most likely to be responsible for the caseins unique functional properties, suggest that their tertiary structures are organised into polar and hydrophobic domains.

In contrast to the calcium-sensitive caseins, κ-casein has few potential Ser residues for phosphorylation, these are in the C-terminal segment, and are not found in clusters. Furthermore, in contrast to the calcium-sensitive caseins in which the polar domains contain anionic clusters comprising phosphoseryl residues, the polar domain of the κ-

casein lack these residues, although they are strongly anionic. In addition to conferring solubility in the presence of calcium ions, these domains are important in micelle formation.

### 1.2.2 The secondary and tertiary structures

Because caseins have proved impossible to crystallise, their exact secondary structures cannot be derived by X-ray crystallography. Various spectral methods for determination of casein structures, and algorithms that predict secondary structures from a knowledge of the primary structure, have yielded some information; however, these techniques are far from perfect.

In general, caseins appear to have an open conformation rather than the compact structures of globular proteins, to have relatively little regular secondary structure and to be characterised by a large fraction of random coil conformation.

Although the exact three-dimensional structures of the caseins are not known, some attempts have been made to predict the tertiary structures from their primary structures by molecular modelling (Kumosinski *et al.* 1991; Sawyer and Hot 1993). Physicochemical properties indicate that the tertiary structures of these proteins are more open and flexible than those of typical globular proteins. The high frequency of prolyl residues in caseins may provide an architectural stiffness, yielding an overall open structure with typical flexibility around individual residues and a rapidly fluctuating secondary structure. Thus, caseins probably have a higher structural motility than globular proteins.

### 1.2.3 Casein micelle models

In milk, the caseins are aggregated into casein micelles in which the $\alpha_{S1}$-, $\alpha_{S2}$- and $\beta$-caseins bind calcium and phosphate. In the absence of these minerals, the casein molecules are water soluble, but nevertheless associate with one another to form aggregates.

Casein micelles range in diameter from 500 to 3000 Å with a mean size of about 800 Å and a MW of about 2.5 x $10^8$ Da. The micelles comprise 93% (W/W) caseins with the

$\alpha_{S1}$- to $\alpha_{S2}$- to β- to κ-caseins in the proportion of 3:1:3:1 weight ratio. The remaining 7% consists of inorganic calcium (2.87%), phosphate (2.89%), citrate (0.40%) and small amounts of magnesium, sodium and potassium (Schmidt 1979).

The internal structure of the casein micelle has been the subject of much speculation and investigations and basic features are fairly well agreed upon (Wong *et al.* 1996). A casein micelle is a core consisting of a mixture of $\alpha_S$- and β-caseins stabilised by a surface coat of κ-casein (Slattery 1976). However, the exact structure of the casein micelle and the forces responsible for its stability are still the subject of speculation and a number of models of the micelle have been proposed (Creamer and MacGibbon 1996).

Model A showed in Figure 1.2.3 suggested that micelle composed of submicelles, small subunits consisting of $\alpha_{S1}$-, $\alpha_{S2}$- β- and κ-caseins arranged in the same manner as mentioned above, in which the hydrophobic regions of the casein molecules are oriented in the interior and the hydrophilic regions are located on the surface. The κ-caseins, due to self-association, are restricted to one area on the surface of the submicelle. Submicelles aggregate to form micelles via calcium or calcium phosphate cross-links between phosphoserine residues on the outer surface of the submicelle.

In another model (Holt 1992), Figure 1.2.3 B, the micelle is postulated to coalesce gradually from the components, and the final form has a discontinuous distribution of proteins and of calcium phosphate, although there is a tendency for the κ-casein to be on the outside and for the minerals to be associated with the phosphoserine residues of the caseins and to form in clumps.

Regardless of which model, if either is correct, the undisputed functional role of the micelle, and the casein components, is as a carrier of high level of calcium which would otherwise precipitate out of milk.

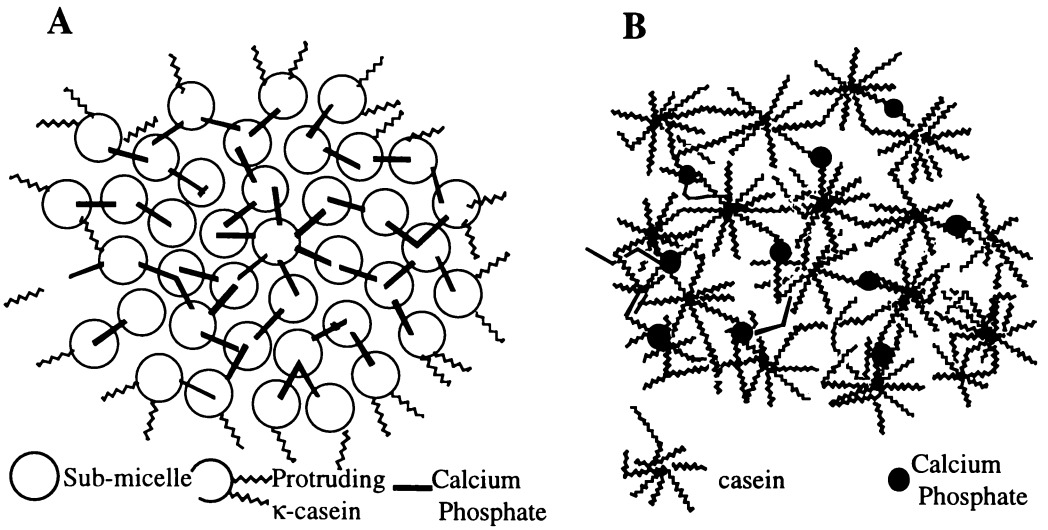



**Figure 1.2.3 Two models of the casein micelle.** *Model A is centred on the concept of sub-micelles as distinct entities, held together by calcium phosphate linkages and with the κ-casein oriented towards the surface. Model B is a more dynamic system of interacting casein molecules with occlusions of calcium and phosphate complexes, with regions of varying ionic, protein and dielectric constants within the micelle. (based on Creamer and MacGibbon 1996)*

## 1.3 The Structure and Expression of Casein Genes

### 1.3.1 Chromosomal location

Each casein gene occurs as a single functional copy per haploid chromosome set with no related pseudogene sequences being reported. Early genetic studies showed a tight linkage between the four bovine casein genes (Groslaude *et al.* 1978). Similarly tight linkages were also demonstrated in some other species (Di Gregorio *et al.* 1991; Leveziel *et al.* 1991) and it is a reasonable assumption that this will prove to be the case for all mammals. The chromosome location of the bovine gene cluster has been assigned to chromosome 6q31 by various locus mapping method (Gallagher *et al.* 1994). Long range restriction analysis gave the casein locus cluster order: $\alpha_{S1}$-β-$\alpha_{S2}$-κ within a range of 200 Kb to 300 Kb (Ferretti *et al.* 1990; Threadgill and Womack

1990). More recently, Rijnkels *et al.* (1997), by combining the approaches of the long-range mapping and analysis of cosmid or phage clones encompassing the casein locus, established a more accurate physical map which is shown in Figure 1.3.1. The order of the casein genes, $\alpha_{S1}$-β-$\alpha_{S2}$-κ, is confirmed as described previously (Ferretti *et al.* 1990; Threadgill and Womack 1990). The $\alpha_{S1}$- and β-casein genes are about 20 kb apart and are convergently transcribed (that is, towards each other). The $\alpha_{S2}$-casein gene is located about 70 kb upstream of the β-casein gene with a divergent transcriptional direction relative to the β-casein gene. The κ-casein gene is located in a region 95-120 kb downstream of the $\alpha_{S2}$-casein gene, about 200 kb from the $\alpha_{S1}$-, β-casein gene region. The transcriptional orientation with respect to the $\alpha_{S2}$-casein gene is most likely identical. The total size of the locus is estimated to be 250 kb.

### 1.3.2 Genomic organisation of casein genes

The bovine genes coding for the four caseins, as well as the major whey genes, have been cloned and sequenced. The bovine major milk protein genes are relatively small (all are < 20 kb) and they have a relatively simple structure. However, the $\alpha_S$- and β-casein genes are somewhat larger than the whey protein genes and have a more complex intron-exon structure. Most exons are quite short, but a few extend to several hundred base pairs. The existence of such long and short exon stretches may be due to ongoing evolution. No codon is disrupted by an intron, a fact which is of crucial importance when exon skipping occurs (see section 1.5). All casein genes contain repeated elements, mostly *Alu*-like sequences of retroposons. For example, repetitive sequences represent around 14% of the bovine $\alpha_{S2}$-casein gene. Retroposon inserts in casein genes increase the intron size and thus alter the exon surrounds. Figure 1.3.2 shows the structures of the bovine milk protein genes.

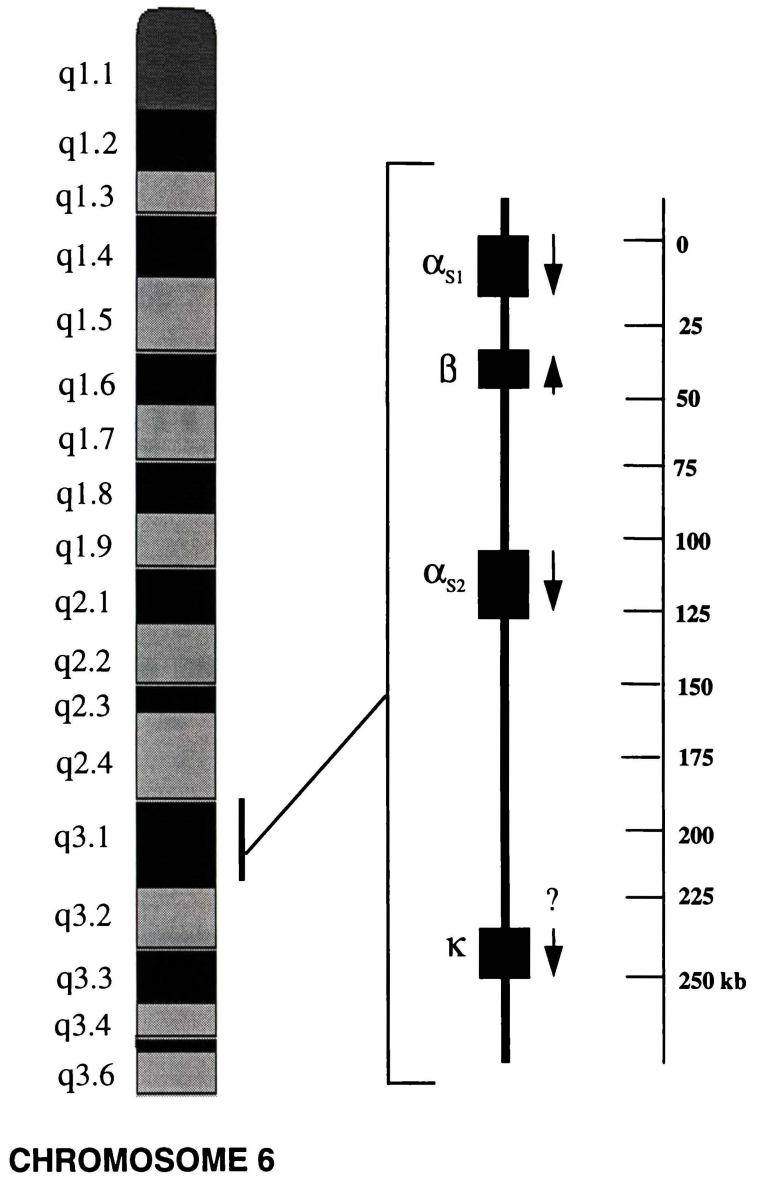



**Figure 1.3.1 Physical map of the bovine casein locus.** *Four casein genes (closed boxes) are clustered on chromosome 6q3.1 in a range of 250 Kb. Their orientation are indicated by arrows. (not to scale). (after Rijnkels et al. 1997).*

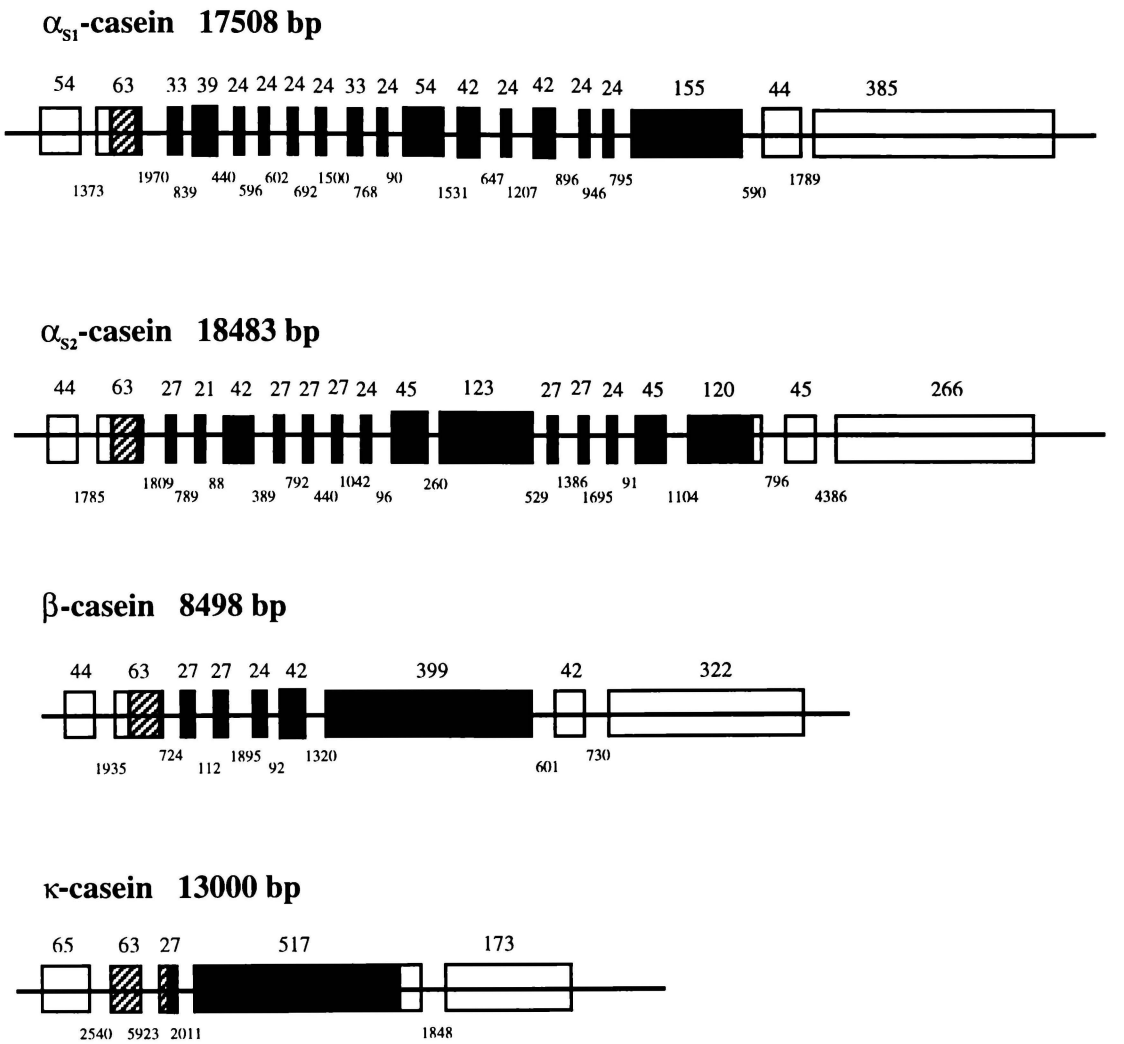



**Figure 1.3.2 Organisation of the bovine casein genes.** *Number above the exons indicate the number of base pairs (bp). The number below the exons is the length (bp) of introns. Open boxes, 5' and 3' non-coding regions; Hatched box, region coding for the signal peptide; closed boxes, regions coding for the protein. (adapted from Groenen and van der Poel 1994).*

### 1.3.3 The evolution of casein genes

Alignment of casein genes from bovine and other mammalian species suggests that the overall organisation of these genes has been conserved during the evolution of mammals (Groenen and van der Poel 1994). Indeed, from the sequences of the $Ca^{2+}$-sensitive caseins it is clear that these genes probably evolved from one ancestral gene (Bonsing and Mackinlay 1987; Yu-Lee *et al.* 1986). In these genes, the sequence coding for the leader peptide and the promoter region are both highly conserved. Based upon the gene organisation (e.g. the number and length of exons and introns) it is likely that the $\alpha_{S1}$- and $\alpha_{S2}$-casein genes are more closely related to each other than either is to the β-casein gene, which only contains nine exons. However, actual sequence comparisons between the first exons of these three genes indicate that $\alpha_{S2}$-casein and β-casein genes are more closely related to each other than to $\alpha_{S1}$-casein gene. Both genes could have evolved from a common ancestor by a small number of internal duplications as is shown in Figure 1.3.3. The large duplication in $\alpha_{S2}$-casein, which has resulted in the formation of exons 12 to 16, is reflected in both the amino acid and mRNA sequences. At the nucleic acid sequence level, the duplications mentioned above are only apparent in some of the exons, i.e., the region containing exons 9 and 10 shows 66% sequence similarity to the region containing exons 14 and 15. Furthermore, exon 14 of $\alpha_{S2}$-casein and a small part of intron 13 has significant similarity with exon 5 and a small part of intron 4 of bovine β-casein genes. Finally, within $\alpha_{S2}$-, a stretch of 51 nt containing the 27 bp exon 3 shows 75% sequence similarity with the region around exon 12.

Despite the fact that the κ-casein gene is located close to the other three casein genes, it is functionally distinct and conceivably may represent a different earlier branch in the evolution of the casein genes (Alexander *et al.* 1988).

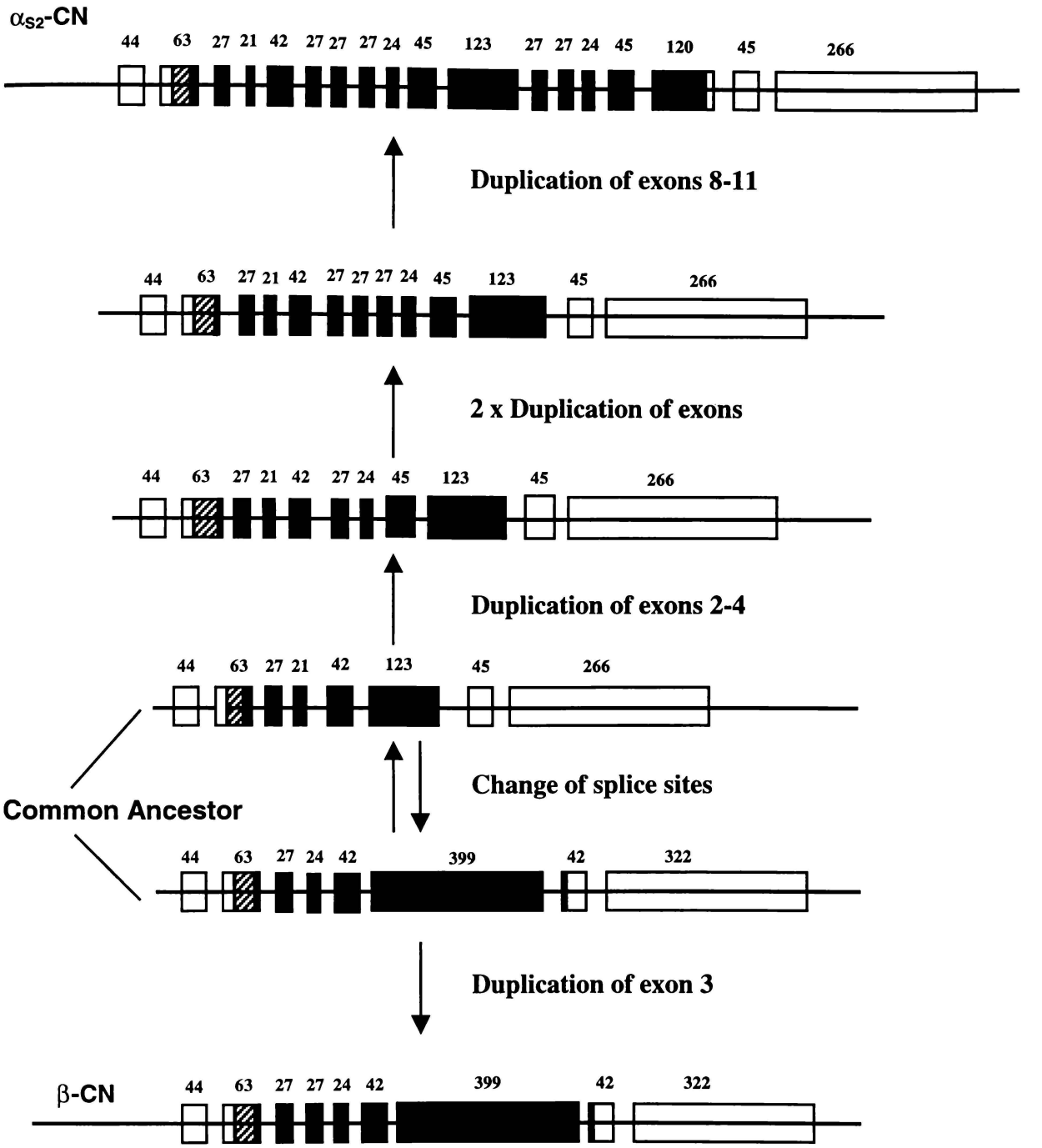



**Figure 1.3.3 Hypothetical scheme for the evolution of $\alpha_{s2}$-casein and β-casein from a common ancestor.** *The order of some of the steps is chosen arbitrarily. Open boxes indicate the 5' and 3' untranslated regions of the mRNAs, the hatched box indicates the sequences that code for the conserved signal peptide and the solid boxes indicate the sequences that codes for proteins (CDS). ( adapted from Groenen et al. 1993)*

### 1.3.4 Expression regulation: *cis*-elements and transcription factors

The 5'-flanking regions of $\alpha_{S1}$-, $\alpha_{S2}$-, β- and κ- casein genes have been examined by a number of authors (Popov 1996). The first 150-200 bp in the 5'-flanking regions contain the essential promoter regions, with sequences identical or similar for all casein genes (Laird *et al.* 1988; Vilotte *et al.* 1991). All milk protein genes have the transcription initiation start--TATA box. The gene for β-casein has a TATATATAAA sequence between position -35 to -26 (Bonsing *et al.* 1988). The TATA box of $\alpha_{S2}$-casein gene is between -30 to -24 with a sequence of TATAAATA (Groenen *et al.* 1993). In $\alpha_{S1}$-casein gene this element is somewhat altered and it has TTTAAATA between -29 to -22, the same sequence is also found at positions -89 to -82 (Koczan *et al.* 1991). The κ-casein gene's TATA box is TTAATTA at -32 to -25 (Alexander *et al.* 1988). There are also other conserved sequences, some of which have shown to bind nuclear factors (discussed below). Comparison of different milk protein genes revealed a 30 bp sequence showing a 70-80% sequence similarity in both caseins and whey protein genes which was termed the "milk box" (Laird *et al.* 1988) although its functional significance still remains unclear.

It is known that *cis*-elements influence transcription by interacting with *trans*-acting factors. DNA-binding nuclear factors regulating differential gene expression have been found in many cell and tissue types. Beside some common nuclear transcription factors such as NF1, Oct-1 and PEA3, several factors which are thought to be a specific in mammary gland cells and which act by binding with 5'-flanking (and possibly other) regions of milk protein genes have been investigated (Groenen and van der Poel 1994) e.g. the milk protein binding factor (MPBF), pregnancy-specific mammary factor (PMF) and mammary gland-specific factor (MGF). Sequence comparison of the MPBF sites indicated that this factor might play a substantial part in regulating the expression of all milk protein genes (Watson *et al.* 1991). This MPBF is related, if not identical to, the MGF which was first identified as a protein which became activated by hormonal signals, and then played a crucial role in relaying these signals to the β-casein gene in cultured mammary cells (Schmitt-Ney *et al.* 1991); *in vivo* experiments also showed that its level of activation increased during gland development, peaking during lactation. The MGF protein was actually found to be a new member of the Signal

Transducer and Activator Transcription (STAT) families of proteins, and since has been renamed STAT5. Even more recently, it has been further designated as STAT5a because a closely related gene produces a protein, STAT5b, with very similar functions, although STAT5a is the predominant form in the mammary gland (Liu *et al.* 1995). Experiments in which STAT5 gene disrupted mice were produced, demonstrate that STAT 5a (-/-) female mice do not lactate properly and STAT5b (-/- ) mice also have disrupted mammary gland development and lactation (Udy *et al.* 1997).

**Table 1.3.4 Potential binding sites for nuclear factors in 5'-flanking regions of bovine casein genes**

| *Factor* | *Binding Site* | *Site Position* | | | |
|---|---|---|---|---|---|
| | | *$\alpha_{s1}$-CN* | *$\alpha_{s2}$-CN* | *β-CN* | *κ-CN* |
| (Y1, NF-E1, δ) | 5-GCCATNTT-3 | 201, 352, 1697 | 2384 | 413, 506, 747, 893, 1427 | 96, 415 |
| PEA3 | 5-TGACTCA-3 | 297, 795, 940, 1254, 1425, 1663 | 41, 149, 422, 765, 2263 | 497, 1668 | 69, 431 |
| Ap1 | 5-TGACTCA-3 | 117, 285, 405, 615, 2007 | 2324 | 536, 607, 931 | 53 |
| C/EBP | 5-CCAAT-3 | 731, 801, 1470 | 14, 851, 2272 | ND | ND |
| Oct1 | 5-ACTTGGCATFT-3 | ND | 480 | ND | ND |
| | 5-GTATTAAAAT-3 | 322, 1023, 2004 | 267 | 331, 1322 | ND |
| | 5-TCATGACATAT-3 | 1938 | 218 | ND | ND |
| | 5-ATTACCATAT-3 | 359 | 56 | 55, 1320 | ND |
| NF1 | 5-TGGN$_{(6-8)}$CCA-3 | 74(N$_{11}$), 718, 1531 | 680(N$_{10}$), 2204(N$_{12}$), 2395(N$_{9}$) | 259 | ND |
| MGF | 5-TTCTTAGAATT-3 | 97 | 97 | 102 | ND |
| MPBF | 5-CCAGGAACCG-3 | 24, 346 | 355 | 955, 1540 | ND |
| PMF | 5-TGAT/ATCA-3 5-ATCA/TGAT-3 | 1753 | 189 | 9, 442 | ND |

*numeration begins in the region preceding the transcription start; (ND) not detected (adapted from Popov 1996)*

Considerable information on promoter sequences has come from animal experiments employing transgenes utilising milk protein gene regulatory sequences. These demonstrate that the κ-casein 5'-flanking region (-552→ +64), in contrast to that in the other milk protein genes, is a poor mammary gland promoter. Thus, the potential binding sites for nuclear factors located therein (Table 1.3.4) appear to be insufficient to regulate the expression of this gene. Possibly the full tissue-specific regulation of this gene depends on *cis*-elements located in the introns, in distal 5' regions or in the 3'-

flanking region, and possibly the 5'-flank has binding site(s) for a yet unidentified negatively acting factors. Indeed, enhancer elements have been identified and postulated for various milk protein genes (Schmidhauser *et al.* 1992; Pierre *et al.* 1992; Groenen and van der Poel 1994).

### 1.3.5 Gene expression in the mammary gland epithelial cells

The daily average mammary gland production of milk across a wide number of species is 50-120 ml/kg body weight. At the cellular level, expression levels are high with each mammary epithelial cell capable of daily synthesising and secreting an equivalent of its own weight of milk proteins and some other milk contents (Delouis and Richard 1993).

The specific patterns of expression of milk protein genes in epithelial cells of the mammary gland during lactation are under complex multifactoral control. The minimal hormonal requirement for lactogenesis are the increased secretion of prolactin, glucocorticoids and estradiol-17β and decreased secretion of progesterone. Many of these hormones are also involved in the development of the mammary gland during gestation. A cascade of events occurs in the endocrine system during the third trimester of gestation that prepares the mammary gland for the secretion of milk. Indeed, at this stage significant expression of mRNA for milk protein occurs in the mammary gland, but the protein produced is not secreted. Furthermore, in addition to hormones, interactions between the epithelial cells and the extracellular matrix (ECM) plays a vital role in the expression of the milk protein genes (Aggeler *et al.* 1988). Very recent evidence suggests that, for some genes there may be a switch of promoter at parturition, with the mRNA produced during pregnancy being far less translatable than that produced during lactation. In the β-1,4-galactosyltransferase gene, this is achieved through eliminating a 5' UTR region which has extensive secondary structure (Charron *et al.* 1998). It is not yet known if similar mechanisms operate with other milk protein genes.

The transcripts of milk protein genes share the general organisation of mRNA encoding secretory proteins. 1) a M7Gppp cap, which protects the 5' untranslated region against enzyme degradation and facilitates the binding of the ribosomal 40S subunit-Met-

$tRNA^{Met}$-initiation factors complex; 2) a coding frame delineated by the initiation and stop codons; and 3) an untranslated 3' region with the recognition signal for polyadenylation located 13 to 20 nucleotides up-stream from poly(A) tail. From mid-pregnancy to lactation, those mRNA encoding major milk proteins can account for up to 60 to 80% of the mRNA in mammary epithelial cells (but, as stated above, in the light of recent findings there is a need to establish whether or not these transcripts are identical, and as translatable, as those produced after parturition). Rat studies show that casein mRNAs are present in very high levels once lactation starts, at about 90,000 mRNA/cell, due an increase in both the transcription rate and stabilisation of these mRNAs by hormonal influences (Rosen *et al.* 1981). Rodent species exhibit particular high levels of casein mRNA expression and, on a mammary unit volume basis, expression is some 5-10 fold higher than in other species, e.g. bovine and ovine (R.J. Wilkins, pers. comm.), even though, one would still classify these latter expression levels as high.

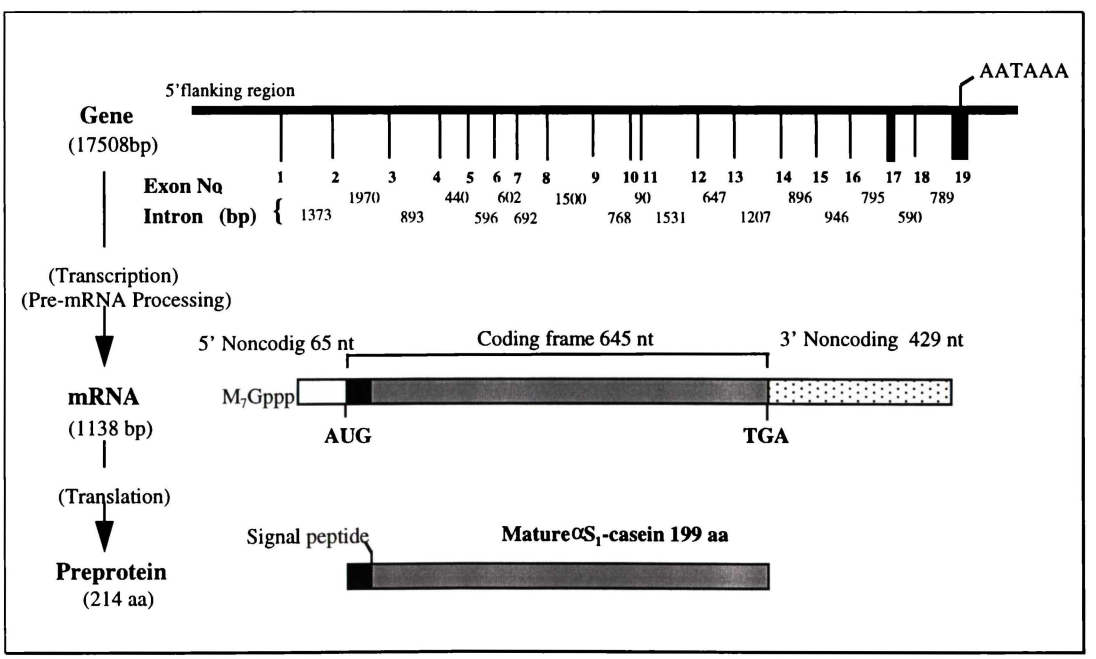



**Figure 1.3.5 Diagram of the transcriptional unit and cognate mRNA encoding $\alpha_{S1}$-CN.**

## 1.4 The Aberrant Splicing of Pre-mRNA

### 1.4.1 The conserved splice sites, spliceosome and exon definition

The splicing process in higher eukaryotes, which converts pre-mRNA into mature mRNA species, is characterised by the precise excision of introns (which in extreme cases can be longer than 50,000 bases) from the pre-mRNA and the concomitant joining of the now juxtaposed exons, each of which is rarely over 300 bases in length. This joining of exons separated by introns (some large, some small) implies that the molecular mechanisms employed allow: 1) the accurate recognition of the exons embedded in an excess of intronic sequences, and 2) the precise joining of the exons with the concomitant release of introns.

The most distinctive common features associated with the exon-intron junctions (or splice site) are the 5' (upstream or donor) and 3' (downstream or acceptor) sequences of introns. These sequences at each exon-intron junction are conserved in virtually all nuclear mRNA genes of almost all species that have been studied (see Figure 1.4.1.1 below). Note the invariant GT at the 5' end of the intron and the invariant AG at the 3' end; moreover, reasonably high consensus is conserved several nucleotides into the intron. In contrast, the exon sequence at the boundary is hardly conserved at all, although the 5' splice site is frequently flanked by (C/A)RG (where R is a purine) and a G often precedes the 3' splice sites. In general, however, there can be wide variations in the exon sequences bordering introns, with mutations at these locations usually not abolishing splicing, although they may affect the rate thereof. One practical, and frustrating, consequence of this wide range of flanking sequences is that it is virtually impossible to predict the exon components of a cDNA by inspection of mRNA sequences alone.

As mentioned above, the most invariant structural features associated with splicing lie within the intron. Padgett *et al.* (1986) scanned some 400 vertebrate genes and derived the following consensus sequences:

5' (donor) splice site consensus sequence

| -4 | -3 | -2 | -1 | | +1 | +2 | +3 | +4 | +5 | +6 |
|---|---|---|---|---|---|---|---|---|---|---|
| $C_{29}/A_{34}$ | $C_{38}/A_{35}$ | $A_{62}$ | $G_{77}$ | | $G_{100}$ | $T_{100}$ | $A_{60}$ | $A_{74}$ | $G_{84}$ | $T_{50}$ |
| | | | Exon | | Intron | | | | | |

3'(acceptor) splice site consensus sequence

| -14 | -13 | -12 | -11 | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | | +1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y_{78}$ | $Y_{81}$ | $Y_{83}$ | $Y_{89}$ | $Y_{85}$ | $Y_{82}$ | $Y_{81}$ | $Y_{86}$ | $Y_{91}$ | $Y_{87}$ | N | $C_{78}$ | $A_{100}$ | $G_{100}$ | | $G_{55}$ |
| | | | | | | Intron | | | | | | | | | Exon |

**Figure 1.4.1.1 Splice sites consensus of vertebrate genes.** *[Y = pyrimidine, N = any base, subscript numerals refer to percentage frequency of occurrence]*

The first two nucleotides at the 5' end of the intron in DNA are virtually always GT (GU in RNA); the next four are not invariant, although the sequence AAGT appears to be the consensus. Mutations of either the G or T residue at the junction generally abolish splicing and changes at the next positions can affect splicing to varying extents (Technically, it can be difficult to establish the importance of various mutations, because *in vitro* experiments utilising cultured cells transfected with *in situ* mutagenised constructs do not always accurately reflect splicing events *in vivo*; Lewin 1997). The six nucleotides at the 5' end of the intron are usually sufficient for 5' splice site function. The 3' terminus of the intron invariably ends with AG, most often preceded in mammalian introns by a pyrimidine-rich stretch that may extend far into the intron (Oshima and Gotoh 1987). Much of the reliable data on the *in vivo* effects of various mutations on splicing, actually comes from studies of human genetic disease; some 15% of all point mutations causing human genetic disease have been found to cause mRNA splicing defects (Krawczak *et al.* 1992).

The branchpoint sequences YNCUGAC in mammals (where the underlined adenosine is the residue that forms the 2' to 5' link) are usually found in the intron, 17-40 nucleotides (nt) upstream of the 3' splice site (Langford *et al.* 1984; Nelson and Green 1989). In contrast to this, branchpoints located as far as 177 nts upstream from the 3' splice junction are used efficiently in several alternatively spliced pre-mRNAs (Helfman and Ricci 1989; Smith and Nanal-Ginard 1989). This sequence plays a role in forming a branch with the 5' terminal of the intron but it appears that the consensus

is relatively weak (Krainer and Maniatis 1988). Smith *et al.* (1989) proposed that the branchpoint association with a polypyrimidine tract is more important than its proximity to the 3' splice site AG. Furthermore, they suggest that the 3' splice site chosen is the first AG downstream of the branchpoint-polypyrimidine tract. The AG must be in a governable context, that is, at least 12 nt distal to the branchpoint and not part of a hairpin (Smith *et al.* 1993).

Splicing occurs within the spliceosome, a complex assembly of small nuclear ribonucleoprotein particles (snRNPs), heterogeneous ribonucleoprotein particles (hnRNPs), and various splicing factors (Guthrie and Patterson 1988; Guthrie 1991; Wassarman and Steitz 1991). Exon recognition is accomplished by the interactions of certain components of the spliceosomal complex with the 5' and 3' splice sites. Formation of a commitment complex between a pair of splice sites triggers the sequential interaction of splicing factors, which assemble into a spliceosome to carry out intron removal. Critical to this process is U1 snRNP base pairing with conserved sequences of the 5' splice site and $U2AF^{65}$ recognition of 3' splice site consensus. (Manley and Tacke 1996; Wu and Maniatis 1993). Commitment between the pairs of splice sites was proposed to be mediated by serine-arginine-rich (SR) proteins through an interaction with $U2AF^{35}$ and U1 snRNP 70K protein. The protein-protein contacts involved in the 70K-SR-$U2AF^{35}$ bridging interaction are mediated by serine-arginine-rich domains (RS).

Although the existence of a consensus motif can be thought of as defining a putative intron/exon boundary, such an isolated conserved sequence cannot be considered by itself to allow the precise recognition of exons. Rather, in the exon definition model, it is believed that both 5' and 3' splice sites of an exon must be recognized as pairs by the splicing apparatus and they only define an exon when there is a certain proximity between these two sites (Robberson *et al.* 1990; Berget 1995). Terminal exons, that is, the first and the last exons in a mRNA molecule, require special mechanisms for their recognition and these appear to involve cap and poly (A) binding proteins. Moreover, the last exon is generally longer than internal exons, although, interestingly this is not true for β and κ casein (see Figure 1.3.2). There is also strong experimental evidence demonstrating that certain conformations within exon sequences can affect splicing.

Experimentally, testing the strength and effect of various intronic and exonic sequences is difficult because, as in the case of consensus sequences, the splicing events *in vitro* and in cell culture (using introduced test constructs) do not always reflect what occurs *in vivo* (Lewin 1997). Indeed the splicing machinery is extremely complex and factors which modulate its efficacy and selectivity are not well understood.

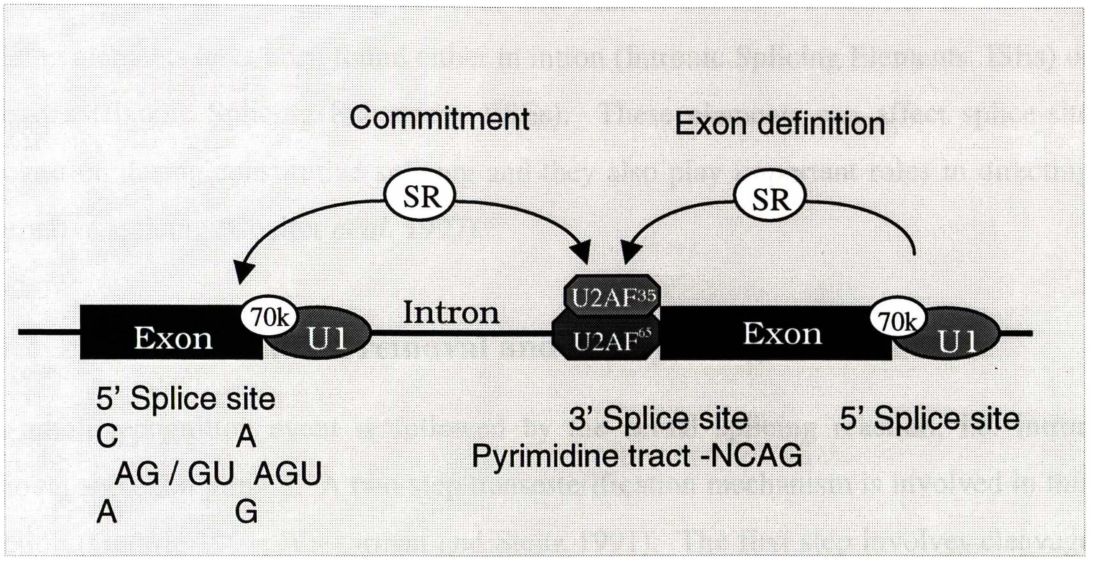



**Figure 1.4.1.2 Early steps of splice site recognition.** *5' splice site sequences are initially recognised by virtue of their complementarity to the 5' end of U1 snRNP. 3' splice site sequences are recognised by U2AF$^{65}$, which exists in association with U2AF$^{35}$. The strength of splice sites is defined by their match to the consensus sequences. SR proteins are proposed to mediate commitment between pairs of splice sites through intron-bridging interactions. SR proteins can interact with each other and so intron-bridging interactions may involve combinations of SR proteins. Alternatively, because SR proteins can interact simultaneously with the U1 snRNP 70k protein and U2AF$^{35}$, only SR protein might be required. SR proteins have also been implicated in exon-bridging interactions (exon definition model) (Adapted from Chabot 1996)*

While the 5' splice sites of constitutive introns are usually well conserved, 3' splicing signals show much greater sequence variation. To ensure efficient U2AF binding and to avoid aberrant splicing, a U1-bound 5' splice site can stimulate U2AF binding to the upstream 3' splice site and that requires SR proteins (Figure 1.4.1.2). The abundant examples of exon skipping induced by naturally occurring 5' splice site mutations (Talerico and Berget 1990) support this model since the failure to recognise the 5' splice site of an exon usually prevents inclusion of that exon in the mRNA.

Experimental support for this exon definition model is also provided by studies which demonstrate that assembly of spliceosome components upon a splicing substrate lacking an upstream 5' splice site is greatly enhanced by the presence of a downstream 5' splice site (Robberson *et al.* 1990).

Recent studies indicated that distinct sequence elements distant from the splice sites are also needed for normal splicing (Cooper and Mattox 1997). A number of auxiliary splicing elements have been found either in intron (Intronic Splicing Elements, ISEs) or in exons (Exonic Splicing Enhancers, ESEs). These elements can affect splice-site recognition during constitutive splicing and they also play important roles in directing alternative splicing (Coulter *et al.* 1997).

## 1.4.2 Pre-mRNA intron removal and exon joining

The exon recognition event is followed by the actual splicing reaction, i.e. intron removal and exon joining. A two-step transesterification mechanism is involved in this reaction (Guthrie 1991; Wassarman and Steitz 1991). The first step involves cleavage at the 5' end of the intron (the 5' splice site) and formation of a 2'-5' phosphodiester bound between the 5' terminal guanine of the intron and the 2' hydroxyl of a conserved adenosine located in the intron (known as the branchpoint). The intermediates generated during the first step of splicing consist of a free "exon 1" plus the intron with its 5' end joined to the branchpoint and its 3' end still joined to "exon 2". The latter is known as a lariat structure due to its configuration. In the second step of splicing, cleavage at the 3' end of the intron (the 3' splice site) and concomitant ligation of the two exons occurs. As the reaction proceeds, two products accumulate; the correctly spliced, ligated exons and the released complete intron-containing lariat structure (Figure 1.4.2). These enzymatic steps are mediated by a catalytic centre that is formed by base pair interactions between conserved residues of U2 and U6 snRNPs (Nilsen 1994; Newman 1997). In addition, several other components of the spliceosomal complex including SR and hnRNP proteins are also involved in the mechanisms that bring together defined exons (Berget 1995).

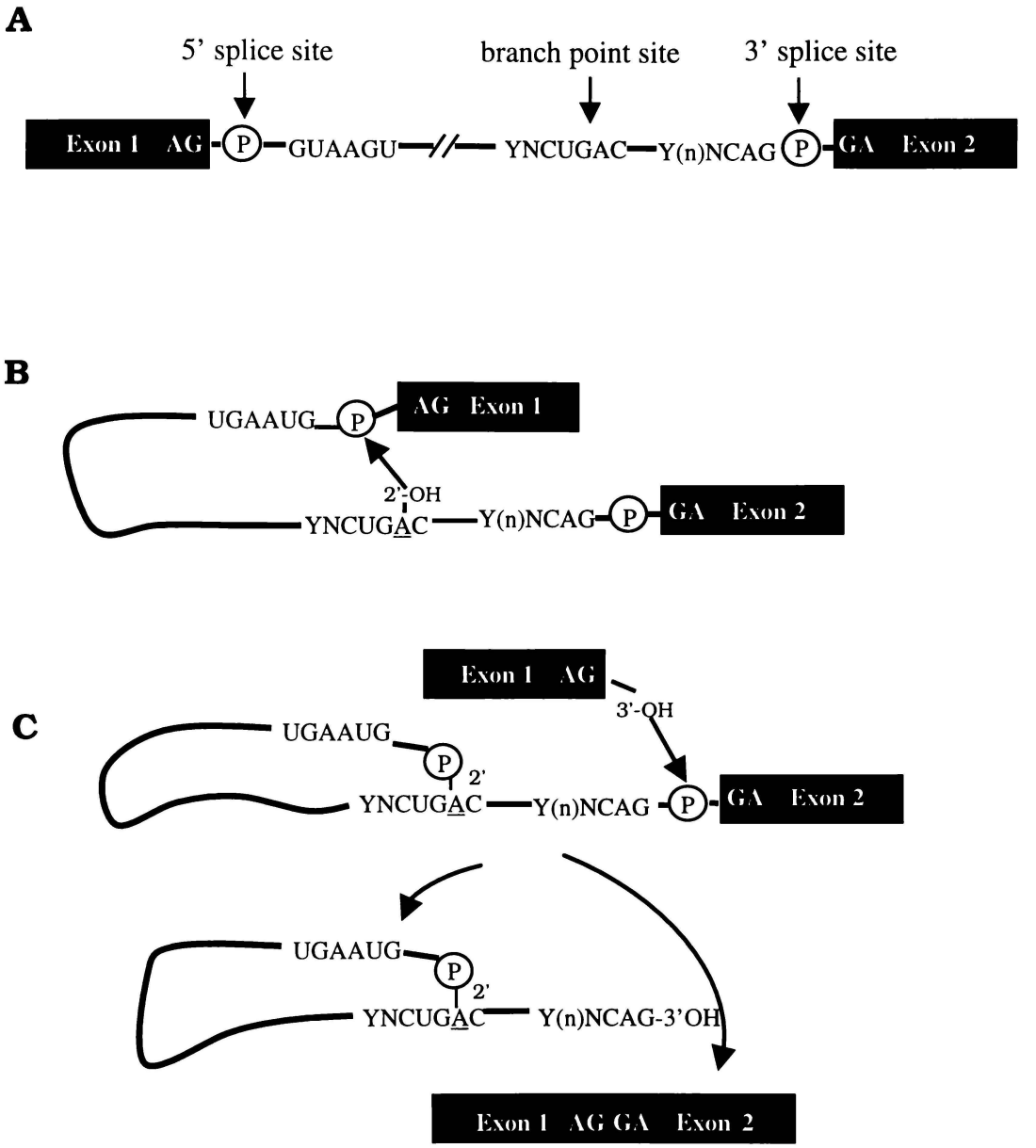



**Figure 1.4.2 A schematic representation of two steps of pre-mRNA splicing.** *A) The intron is represented as a line, flanked by two boxes representing exon 1 and 2. Step one (B) involves joining of the branch point adenosine to the 5' end of the intron via a 5'-2' phosphodiester bond. This results in free exon 1 and a lariat structure consisting of the intron attached to exon 2. The second step (C) involves cleavage at the 3' end of the intron and ligation of exon 1 to the 5' end of exon 2. This results in exon 1 and 2 being joined by a phosphodiester bond and release of the intron lariat. (re-drawn from Mepham 1992)*

### 1.4.3 Alternative splicing and its common patterns

Most pre-mRNAs are spliced by a pathway that excises each intron at its respective 5' and 3' splice site. Consequently, each one of the transcript's exons is conserved in its original order as a continuous sequence in a single mature mRNA (constitutive splicing). However, some pre-mRNAs undergoing a special "alternative splicing" process. In alternative splicing, the pre-mRNA are spliced in more than one way, thereby yielding more than one structurally related mRNAs, each with a subset of exons and each putatively encoding one member of a family of protein isoforms. An increasing number of examples of alternative splicing of genes are being documented, occurring in species as diverse as Drosophila and humans.

Alternative splicing provides the means to diversify the output from a single gene without altering its genomic organisation. It represents an efficient way to generate a variety of mRNAs encoding structurally related proteins in which the shared amino acid sequences are specified by the common exons and the individually distinctive sequences derive from the alternatively spliced exons. The alternatively spliced mRNAs are, in some instances, produced concurrently and/or the their relative abundance is developmentally regulated. A well documented case of alternative splicing involves sex determination in Drosophila embryogenesis. Thus, the different activities of regulatory genes in males and females are due largely to sex-specific differences in RNA splicing that lead to the production of functionally different transcripts in the two sexes (Hodges and Bernstein 1994). The activities of individual genes in this regulatory hierarchy are controlled both at the level of splicing and by the specificity of splicing. As the various mRNAs resulting from alternative splicing often encode proteins with distinct functions, the splicing process must be stringently regulated during both development and differentiation.

The regulation of alternative splicing is still poorly defined. The interactions of SR and hnRNP proteins with target sequences within pre-mRNAs are expected to play a key role in alternative splicing. A large variety of splicing decisions may be managed by combining cell specific variations in the abundance of individual SR and hnRNP proteins with pre-mRNA-specific differences in the strength of splice sites and the

position of modulating elements. Splicing unit regulation may require the participation of additional tissue-, sex- or developmental-specific factors.

Patterns of alternative splicing can be very complex and can involve alternative introns and exons as well as variations in the position of individual splice junctions. Some common alternative splicing patterns are illustrated in Figure 1.4.3. A and B are shown that exons can be lengthened or shortened by utilising alternative 5' or 3' splice sites. When there is an alternative 5' splice site and an alternative 3' site within an exon and both are activated, an internal intron can be eliminated (C). Under some circumstances exon skipping occurs, as shown in D, the entire exon(s) can be skipped during the splicing process. Exons can even be mutually exclusive, in that one or the other is present in the final transcript (E). Alternative transcriptional promoters (F) or polyadenylation sites (G) may influence use of alternative splice junctions, simply by providing a template containing or lacking specific splice sites.

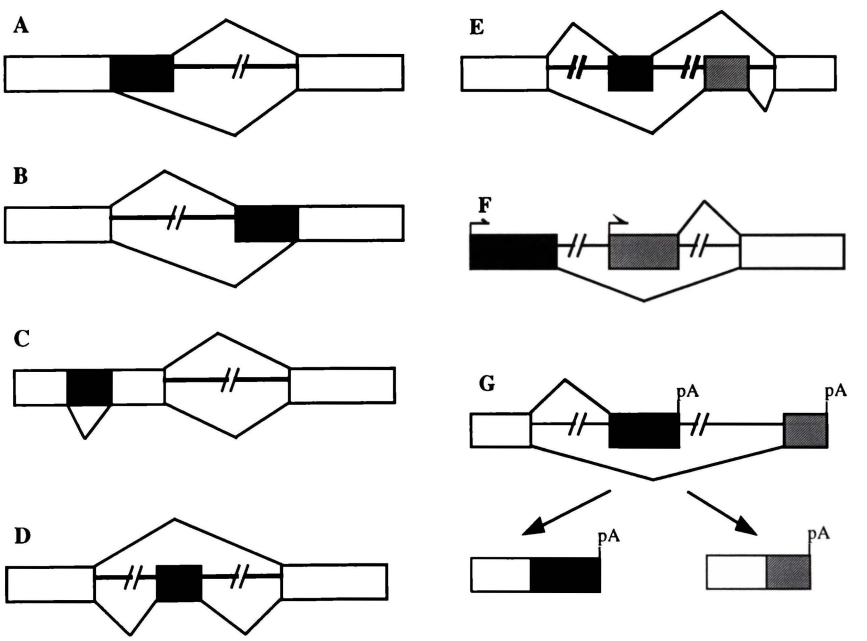



**Figure 1.4.3 Alternative splicing patterns.** *Constitutive exons are represented by open boxes and alternative exons by black or cross-hatched boxes. Introns are indicated by bold horizontal lines. Lines above and below each gene structure indicate alternative modes of splicing. In (F), two potential transcriptional start sites are indicated by arrows. In (G), alternative polyadenylation sites are indicated by pA. (adapted from Hodges and Berstein 1994)*

## 1.4.4 Mutations at splice sites and their effects on splicing

Defective RNA splicing is now recognised as the cause of a number of genetic diseases. The molecular lesions abolish or, at the very least, reduce the amount of mature mRNA generated and/or activate cryptic splice sites in the vicinity of the wild-type sites. Thus, an exon may no longer be recognised as such, thereby excluding it from the mature mRNA transcript (exon skipping). Alternatively, utilisation of cryptic splice sites results in the production of mRNAs that either lack a portion of the coding sequence (cryptic splice site activation) or that instead contain an additional sequence of intronic origin (intron retention) (Krawczak *et al.* 1992). The net result may be the production of a non-functional or dysfunctional protein or no protein at all.

Nakai and Sakamoto (1994) constructed an aberrant splicing database containing an extensive collection (90 genes, 209 mutations) of mammalian genetic disease mutations. Their database showed that more than 90% of mutations either destroy or create the splice site consensus sequences. In addition to mutations within splice sites, it is known that the accuracy and efficiency of splicing is affected by the nucleotide sequence in the immediate vicinity of the splice junction, e.g. the length and /or sequence composition of exons, perhaps as a result of secondary structure consideration (Somasekhar and Mertz 1985; Reed and Maniatis 1986; Matsuo *et al.* 1991). Nevertheless, the vast majority of known genetic lesions that affect splicing are within the "essential" 5' and 3' splice sites. This is probably because (1) they are the most readily detectable and (2) they are most likely to result in a severe clinical phenotype.

### 1.4.4.1 5' Splice site mutations and their effect on pre-mRNA splicing

The pattern of point mutations affecting different 5' splice sites deviates significantly from that expected on the basis of dinucleotide mutabilities in gene coding sequences. The number of mutations mapped at individual residues in the splice site regions roughly correlates with degree of conservation in the consensus sequences. The majority of mutations (60% in the samples of Krawczak 1992) thus occur in the invariant GT dinucleotide, comprising a significant excess at both +1 and +2. G→A transitions significantly outrank G→C and G→T transversions at position +1. This may reflect the interference of non-homologous purines with U1 snRNP binding. It was suggested that 5' splice sites that are less than optimal in terms of their similarity

to the consensus sequence are prone to the deleterious effected of mutation, perhaps because U1 snRNP binding is more readily perturbed. So it appears likely that the observed non-randomness of mutation in 5' splice site is a reflection of the relative phenotypic severity (and hence detection bias) rather than any intrinsic difference in the underlying frequency of mutation. If this is indeed the case, one would predict the existence of a low level of silent mutations in these splice sites which have minimal phenotypic effects.

### 1.4.4.2 3' Splice site mutations and their effect on pre-mRNA splicing

The effects of mutations at the 3' splice site are less well understood than at 5' splice site, in part because of the higher redundancy of the 3' splice site consensus sequence. In the screening studies of Krawczak *et al.* (1992), the significantly skewed distribution of mutations amongst 3' splice sites involves the A residue at position -2. Mutations at other positions do not occur at frequencies significantly different from expectation. Purines are introduced by mutation significantly more often than pyrimidines. This was also observed in mutations at 5' splice sites and was attributed to disturbance of the association with the U1 snRNP

*In vitro* studies demonstrated that the adjacent pyrimidine tract play an important role in the splicing of the 3' end (Reed 1989). If the track is short (e.g. 14 bases), then an adjacent AG appears to be essential for efficient splicing. If, however, it is longer (e.g. 26 bases), then an adjacent AG does not seem to be required. A few *in vivo* mutations have already been detected associated with pyrimidine tracts in 3' splice sites (Krawczak *et al.* 1992).

### 1.4.4.3 The phenotypic consequences of splice site mutations

Three classes of splice site mutations are known to have significant phenotypic effects, namely, failure to remove an intron, cryptic splice site activation and exon skipping.

Intron retention has occasionally been reported to result from a splice junction mutation. Ohno and Suzuki (1988) showed in humans, intron 13 can be retained in transcripts of the β-hexosaminidase gene due to a G to C mutation at site +1 of the intron. This mutation results in Tay-Sachs disease. Another example of intron

retention arises from a mutation at a 5' splice site of the P53 gene (Sameshima *et al.* 1990). The resulting absence of the P53 protein can be correlated to some types of carcinoma.

Mutation of splice junctions sometimes leads to activation of cryptic splice sites. Thalassemias are one of the diseases resulting from this kind of aberrant splicing. A 3' splice site which results in activation of a cryptic 3' splice site within the adjacent exon and some mutations of a 5' splice site which results in activation of several cryptic 5' splice sites have been observed (Aebi *et al.* 1986). Another more complicated example occurs when a mutation creates a 5' splice junction in the intron of the β-globin gene (Treisman *et al.* 1983). This new splicing site is used as well as a cryptic site located upstream in the intron, thus creating a new exon.

In genetic diseases, exon skipping occurs more frequently than cryptic splice site usage (Appendix V). The results of a series of observations and the comparative studies on vertebrate splice sites mutations suggests that exon skipping is the most common phenotype observed as a result of 5' splice site mutations (Talerico and Berget 1990). When mutation of the 5' splice site led to exon skipping, it was always the upstream exon immediately preceding the lesion that was removed from the subsequent mRNA transcript (Kuivaniemi *et al.* 1991). In the a few occasions that the mutation occurred at the 3' splice site, the skipped exon was that immediately downstream of the lesion.

### 1.4.4.4 The mechanism of exon skipping

Several studies have be aimed at elucidating the mechanism of exon skipping *in vitro*. Talerico and Berget (1990) have shown that splicing of an internal exon to its flanking exons is strongly inhibited by mutations at the 5' splice site of the internal exon. Kuo *et al.* (1991) and Grabowski *et al.* (1991) demonstrated that enhanced U1 binding to the 5' splice site of an internal alternative exon of a rat preprotachykinin transcript dramatically increases inclusion of that exon in the processed transcript. In agreement with the exon definition model (see in section 1.4.1), removal of both the upstream and downstream introns is dependent upon strong U1 binding at the 5' splice site of this exon. Strong binding of U1 and the 70-kDa U1 protein might allow interactions with SR proteins and $U2AF^{35}$. This could bring the associated $U2AF^{65}$ subunit into contact

with the upstream 3' splice site and initiate splice site recognition by U2 followed by upstream intron removal (Wu and Maniatis 1993).

Masking of a single splice junction of an exon can clearly result in its exclusion from the mature mRNA. Thus secondary structures that inhibit recognition of a splice site (Libri *et al.* 1991), *trans*-acting factors that block a particular splice site, or perhaps even post-transcriptional modification of splicing signals (Feagin *et al.* 1987; Bass and Weintraub 1988), or other base conversions (Powell *et al.* 1987) may be implicated in exon skipping during alternative RNA splicing.

## 1.5 Alternative Splicing in Milk Protein Genes

### 1.5.1 The $\alpha_{s1}$-casein gene

The complete sequence and structure of the $\alpha_{S1}$-casein gene was determined from a 'deutsche Schwarzbunte' cow's genomic DNA library by Koczan *et al.* in 1991. $\alpha_{S1}$-casein is the largest and most complicated of the milk protein genes, being 17508 bp in size (without including promoter sequences), consisting of 1138 bp of exon and over 16300 bp of intron DNA sequences (Appendix III). The gene is split into 19 exons, ranging in size from 24 bp to 385 bp and 18 introns from 90 to 1967 bp (Figure 1.3.2). The first exon of 53 bp is untranslated, but the entire leader-peptide as well as the first two amino acids of the mature protein are encoded by exon 2, spanning 63 bp, exactly as in bovine and the mouse β-casein genes and similar to that in all rat casein genes (Yashimura and Oka 1989; Yu-Lee *et al.* 1986). As mentioned in section 1.3.2, no coding triplet of the reading frame is disrupted by any of the splice junctions. The only exception is that the translation stop codon UGA is created by splicing the final nucleotides UG of exon 17 onto the first nucleotide A of exon 18. All splice junctions follow the 5' GT/3'AG splice rule. Nine out of 16 coding exons begin with a 'GAX' sequence. An intron-exon-intron stretch of 154 bp is precisely duplicated, with 97.4% homology and only 4 C/T base transitions. These regions encompass, respectively, intron 9 - exon 10 - intron 10 and intron 12 - exon13 - intron 13, i.e. two of the short 24

bp exons, together with their flanking intron regions (Appendix III, nucleotide positions +9101 to +9254 and +11489 to +11642).

## 1.5.2 Exon skipping and $\alpha_{s1}$-casein gene

The $\alpha_{S1}$-casein A variant (Larsen and Thymann 1966) has been found only in Holstein Friesian-, Red Holstein- and German Red cattle. Grosclaude (1988) described the occurrence of the $\alpha_{S1}$-casein A variant in one out of 23 different breeds investigated, with a allele frequency lower than 0.01; Erhardt (1993) found $\alpha_{S1}$-casein A variant in a German Red cattle breed with an allele frequency of 0.001. Because of this low frequency, homozygous individuals were not found. From protein sequencing data it was found that amino acids 14 to 26 were missing from the bovine $\alpha_{S1}$-casein A variant (Grosclaude *et al.* 1972). In accordance with these data, McKnight *et al.* (1989) observed a 39-bp deletion in the cDNA of the variant A. It is notable that this $\alpha_{S1}$-casein A cDNA was from the mammary tissue of a homozygous B variant cow! Based on the known genomic exon-intron structure of $\alpha_{S1}$-casein gene, it is clear that an exact deletion must have occurred of the entire exon 4 in $\alpha_{S1}$-casein A cDNA.

Mohr *et al.* (1994) studied the genomic DNA of two $\alpha_{S1}$-casein A/B German Red animals and demonstrated that the $\alpha_{S1}$-casein A is not the result of a genomic deletion across the exon 4 region. The only relevant mutation they could find in the $\alpha_{S1}$-casein gene was a single point mutation at position +6 in the splice donor sequence distal of exon 4, and they postulate that this is responsible for upstream exon skipping during the serial splice reactions of the A allele $\alpha_{S1}$-casein pre-mRNA. They postulated that the A allele specific mutation at position +6 is able to interrupt the perfect complementarity of the intron-4 splice donor signal (positions one to eight) with U1-snRNP, which may then no longer be able to compensate for a rather weak exon 4 upstream splice acceptor sequence in facilitating the initial binding of $U2AF^{65}$ to that polypyrimidine tract.

## 1.5.3 Alternative splicing in other milk protein genes

**Bovine $\alpha_{s2}$-casein D** Among the identified genetic variants of bovine milk protein genes, $\alpha_{S2}$-casein D differs from the common type A by the deletion of a stretch of 9 amino acids. Bouniol *et al.* (1993) analysed the sequences of cloned PCR-amplified

genomic DNA from three homozygous cows. Two unrelated females carrying the $\alpha_{S2}$-CN D allele and one carrying the $\alpha_{s2}$-CN A allele, did not reveal any deletion and showed two identical nucleotide (nt) substitutions in both $\alpha_{S2}$-CN D alleles from the cDNA encoding $\alpha_{S2}$-CN A. One C→A transition (allele A→D) observed within intron 7, but as it was 66 nt downstream from exon 7, it is probably harmless. In contrast, the second substitution was a G→T transversion (allele A→D), which affects the last nt of exon 8 (aa 51→59), i.e., the 5' consensus splicing site, (C/A)AGgtaagt in vertebrates. This data strongly suggests that the deleted bovine $\alpha_{S2}$-CN D casein arises from exon 8 skipping during processing of the $\alpha_{S2}$-CN encoding pre-mRNA and that the G→T transversion affecting the last nt of that exon might be responsible for the exon skipping in allele D.

**Human $\alpha_{S1}$-casein** Originally it was thought that $\alpha_{S1}$-casein was absent from human milk. Recently though, the presence of a human counterpart of $\alpha_{S1}$-casein has been confirmed by the isolation and amino acid sequencing of such a protein from mature (first month of lactation) milk. Human $\alpha_{S1}$-casein is smaller than its ruminant counterparts (170 vs 199 amino acid residues) and displays weak similarities with the mammalian $\alpha_{S1}$-caseins so far characterised. Two research groups Johnsen *et al.* (1995) and Martin *et al.* (1996) have reported independently the cloning and sequencing of the three types of human $\alpha_{S1}$-casein transcripts from a human breast λgt-11 cDNA library and two mRNA samples prepared from human mammary epithelial cells. The type-1 transcript was the largest and most abundant and contained an open reading frame encoding 185 amino acids (the first 15 amino acids are signal peptide). Type 2 differed from type 1 by an internal deletion of three nucleotides CAG. Type 3 different from type 1 by an internal deletion of 24 nucleotides and is less abundant, and comprised only two of the 13 analysed $\alpha_{S1}$-casein clones. The fact that multiple forms of $\alpha_{S1}$-casein mRNA were identified from individual samples strongly suggests that differential splicing occurring during pre-mRNA processing. According to the multiple-protein alignment and the identify of neighbouring sequences, the deletion in type-3 transcripts are thought to correspond to skipping of exon 8 (encoding DTRNESTQ). For the type-2 transcript, the single codon deletion is located at an exon boundary, as would be expected if the deletion was due to aberrant processing.

**Porcine $\alpha_{S1}$-casein** Three classes of $\alpha_{S1}$-casein cDNA were found during screening of a cDNA library constructed from the lactating porcine mammary gland. Compared with the full length porcine $\alpha_{S1}$-casein cDNA, the other two classes lacked 18 bp and 60 bp respectively. These difference appear to be due to altered RNA splicing (Alexander and Beattie 1992)

**Caprine $\alpha_{S1}$-casein** Extensive variations have been observed in $\alpha_{S1}$-casein fractions from goats milk, some of which involve large quantitative differences in absolute yields. $\alpha_{S1}$-casein A. B. C and E variants differ from one another by amino acid substitutions, while $\alpha_{S1}$-casein D, F and G are internally deleted (Brignon *et al.* 1990). It was demonstrated that the deletion occurring within variant F arises from the out splicing of three exons (9, 10 and 11) during the pre-mRNA processing, most likely due to a single nucleotide deletion within the first removed exon. Furthermore, the $\alpha_{S1}$-casein F allele was shown to yield at least nine alternatively spliced transcripts (Leroux *et al.* 1992). More recently, multiple forms of caprine $\alpha_{S1}$-casein in variant A. B. C have been determined by Ferranti *et al.* (1997). Mature goat $\alpha_{S1}$-casein in these variants exists as a mixture of at least four molecular species which differ in peptide chain length. The shorter forms originated from skipping events at exons 13 and 16, and from the presence of a cryptic splice site within exon 11 affecting primary transcript processing. The findings of these splicing abnormalities in the three common variants A, B and C suggests that this is a general feature of $\alpha_{S1}$-casein in goat. One general comment that should be made about the goat studies, is that many such animals originate from isolated herds which have been inbred for many centuries with no scientific attempts at herd improvement.

## 1.6 The Aim of This Project

$\alpha_{S1}$-Casein is the most abundant protein content in bovine milk. The A variant which results from skipping of exon 4 gives rise to a truncated A variant protein which has different physiochemical properties from the B variant and thus has potential as a protein with different and valuable processing properties for the dairy industry. There is only one published study pinpointing an underlying gene mutation in $\alpha_{S1}$-casein A variant animals. This was in a German Red animal and involved a point mutation in the splicing site following exon 4. The authors postulated that this point mutation was sufficient to depress the splicing process which normally results in the inclusion of exon 4 in the mature mRNA. However, theoretical considerations render such an explanation somewhat dubious.

In this thesis, the first part of the investigation was to detect the molecular lesion in a New Zealand herd which also gives rise to $\alpha_{S1}$-casein A protein in milk. The aim was both to see if the mutation was the same as reported for the German Reds, and also to enable simple DNA tests to be designed, utilising the polymerase chain reaction (PCR), so that other animals in the New Zealand herd could be genotyped.

The second part of the investigation was aimed at investigating the nature of the exon 4 skipping of the $\alpha_{S1}$-casein gene in more detail. Several lines of evidence suggested that this might yield interesting data. Thus, it was known that:

1) McKnight *et al.* (1989) isolated one of the variant A cDNA from a mammalian gland cDNA library of $\alpha_{S1}$-casein B/B homologous animal;
2) There is an extremely weak homology for the splice acceptor sequence in intron 3 of this gene (Mohr *et al.* 1994);
3) Alternative splicing exists in some other genes which are evolutionary related to the bovine $\alpha s_1$-casein gene.

This thesis aimed to explore exactly how the bovine $\alpha_{S1}$-casein gene was expressed in "normal" animals. Two important outcomes were seen. First, a better understanding of the factors which influence milk protein gene expression, both at the mRNA

(processing) and protein level, and second, to make use of the unique experimental advantages of the mammary gland for examining more basic aspects of exon skipping.

Finally, to achieve the above aims, and enable the large-scale collection of mammary mRNA samples, the feasibility of developing a non-invasive sampling method was investigated. Without this, it would not be possible to set up large scale of experiments so as to examine individual variability (both due to genetic and environmental factors) because of problems involving animal ethics, economic expense which arise when collecting large numbers of samples by biopsy and/or autopsy.

# *CHAPTER II*

# MATERIALS AND METHODS

*Techniques, reagents and sample sources*
*Sample preparation*
*Nucleic acid manipulations*
*Protein manipulations*

## 2.1 Techniques, Reagents and Sample Sources

Standard molecular biology methods, including electrophoresis, culture of *E. coli* clones, nucleic acid precipitation with ethanol or isopropanol, restriction digestion, Southern blotting and hybridisation etc. were performed essentially as described by Sambrook *et al.* (1989) and Ausubel *et al.* (1987). Milli Q filtered water (Millipore Corp.) was used for all molecular biology experiments. Diethylpyrocarbonate (DEPC) treated water was used for work with RNA. All processes were carried out at room temperature in C.203 Genetic laboratory in The University of Waikato unless otherwise stated. Common laboratory reagents were obtained from BDH Chemicals NZ, Ltd. and Sigma Chemical Co. Molecular biology reagents were purchased from Life Technologies, Boehringer Mannheim, Promega, Amersham Life Science etc.

N.Z. Dairy Research Institute (DRI) kindly provided the rabbit anti-bovine $\alpha_{S1}$-casein antibody for immuno-blotting experiments and the purified $\alpha_{S1}$-casein A, B and C variant proteins for electrophoresis standards.

Samples were collected for DNA isolation from the herds maintained by the Dairying Research Corporation (DRC) and the New Zealand Dairy Research Institute (DRI). The DRI and Livestock Improvement Corporation Inc. (LIC) also routinely collect blood samples from dairy farmers. These have been used in some experiments. DNA was also isolated from frozen semen supplied by the LIC (Table 2.1.1).

The RNA and casein protein samples (Table 2.1.2) were isolated from milk which was collected from cows of the DRC by Dr. Gwyneth Verkerk.

The oligonucleotide primers used in this research were synthesised by Life Technologies Inc. Sequences and other details of these oligonucleotides are summarised in Table 2.1.3.

**Table 2.1.1 DNA samples**

| *DNA No.* | *Cow's Industry No.* | *DNA No.* | *Cow's Industry No.* |
|---|---|---|---|
| 1 | FWXP-91-40 | 24 | DDMH-91-4 (No 337) |
| 2 | FWXP-91-67 | 25 | CRGF-91-11(336) |
| 3 | FWXP-91-66 | 26 | CRGF-91-11(336) |
| 4 | FWXP-91-82 | 27 | DDMH-91-46 (No 299) |
| 5 | FWXP-91-59 (No 322) | 28 | DDMH-91-13 ( No 207) |
| 6 | FWXP-91-35 (No 298) | 29 | 9512 |
| 7 | FWXP-91-82 (No 271) | 30 | 9540 |
| 8 | FWXP-91-75 (No192) | 31 | 9530 |
| 9 | FRGR-91-22 (No 52) | 32 | 9541 |
| 10 | DDWL-91-95 (No 338) | 33 | 9532 |
| 11 | JHF-91-15 (No 15) | 34 | 9539 |
| 12 | JHF-91-4 ( No 242) | 35 | 9616 |
| 13 | HF-91-16( No 216) | 36 | 9534 |
| 14 | CRBF-91-4( No 119) | 37 | 9533 |
| 15 | YFH-91-96 (No 141) | 38 | 967 |
| 16 | YFH-91-16 (No 29) | 39 | 9626 |
| 17 | FRVN-91-33 (No 31) | 40 | 9602 |
| 18 | FRGR-91-19 (No 178) | 41 | 9335 |
| 19 | FWXP-91-83 (No 63) | 42 | 8940 |
| 20 | FRGR-91-13 (No 175) | 43 | 9744 |
| 21 | YFP-91-95 (No 194) | 44 | 832 |
| 22 | DDMH-91-21 (No 324) | 45 | 871 |
| 23 | DDMH-91-49 (No 112) | 46 | 5950 |

**Table 2.1.2 RNA samples**

| *RNA No.* | *Cow's Industry No.* | *RNA No.* | *Cow's Industry No.* | *RNA No.* | *Cow's Industry No.* |
|---|---|---|---|---|---|
| 1 | Pooled milk | 28 | 1323 | #10 | 669 |
| 2 | 2771 | 29 | 4754 | #11 | 6229 |
| 3 | 9744 | 30 | 4768 | #12 | 6277 |
| 4 | 1343 | 31 | 2811 | #13 | 636 |
| 5 | 9744 | 32 | AB | #14 | 616 |
| 6 | 9751 | 33 | 5514 | #15 | 6230 |
| 7 | 1245 | 34 | 9751 | #16 | 6278 |
| 8 | 1323 | 35 | 4765 | #17 | 615 |
| 9 | 2801 | 36 | 0814 | #19 | 0505 |
| 10 | 3802 | 37 | 5765 | #20 | 2519 |
| 11 | 2538 | 38 | 5791 | #21 | 2534 |
| 12 | 3775 | 39 | 5293 | #22 | 3513 |
| 13 | 2811 | 40 | 2815 | #23 | 3515 |
| 14 | 4765 | 41 | 0709 | #24 | 3522 |
| 15 | 4506 | 42 | 4769 | #25 | 4503 |

| | | | | | |
|---|---|---|---|---|---|
| 16 | 5531 | 43 | 0708 | #26 | 4522 |
| 17 | 5511 | 44 | 9751 | #27 | 4527 |
| 18 | 4515 | 45 | 4756 | #28 | 4528 |
| 19 | 3507 | #1 | 655 | #29 | 5504 |
| 20 | 4525 | #2 | 6285 | #30 | 5505 |
| 21 | 9751 | #3 | 635 | #31 | 5507 |
| 22 | 5528 | #4 | 6262 | #32 | 5514 |
| 23 | 9415 | #5 | 656 | #33 | 5529 |
| 24 | 8509 | #6 | 693 | #34 | 5534 |
| 25 | 3772 | #7 | 6261 | | |
| 26 | 3774 | #8 | 646 | | |
| 27 | 2758 | #9 | 645 | | |

**Table 2.1.3 Oligonucleotides for PCR, RT-PCR and Southern hybridisation analysis**

| *Name* | *Sequences 5'→3'* | *Locations in $\alpha_{s1}$-CN gene** | *Notes* |
|---|---|---|---|
| alpha-B | cctgagtaaattttcattgagg | 6520←6541 (22mer) | on antisense chain, in exon 4 |
| alpha-A | ctggaaaaggttgagggagt | 6505←6514:6554←6563 (20 mer) | on antisense chain, sequence from exon 3 jumped to exon 5 |
| Spe | aatttactcaggtttttgtggcacta | 6530→6556 (27mer) | nt 6554 g was replaced by c |
| Tru: | aatttactcaggtttttgtggcatta | 6530→6556 (27mer) | nt 6554 g was replaced by t |
| EXS | cttgctcagttcattgaccttctc | 7659←7682 (24mer) | on antisense chain. in exon 6 |
| A1 | ctccttttctgactgtgtttttcac | 6461→6485 (25mer) | on sense chain, in intron 3 |
| A2 | gtggttgcttgggtgagtaaatga | 6653←6676 (24mer) | on antisense chain. in intron 4 |
| A3 | atcctatcaagcaccaaggactcc | 5593→5616 (24mer) | on sense chain, in exon 3 |
| A6 | cctcacttgacgaaatgctttcag | 10535←10558 (24mer) | on antisense chain, in exon 9 |
| A7 | ccatgaaacttctcatccttacctg | 3567→3591 (25mer) | on sense chain, in exon 2 |
| A8 | gcacaagcatttttgacataccac | 7815←7839 (25mer) | on antisense chain, in intron 6 |
| A9 | ggtaaagtggagaaagctgtgcag | 7531→7554 (24mer) | on sense chain, in intron 5 |
| A10 | ccagacagagagatgaagataact | 7148→7171 (24mer) | on sense chain, in intron 5 |
| A11 | ttcatcttgtttaagtcatagttctg | 7172←7197 (26mer) | on antisense chain, in intron 5 |
| A12 | ataactgtggagtccctcagaattc | 18429←18453 (25mer) | on antisense chain, in exon 18 |
| A13 | tggagtccctcagaattcacttgac | 18422←18446 (25mer) | in exon 18, partly overlapped with A12 |

** The location is referenced to the full-length sequence of the B. taurus gene for $\alpha_{s1}$- casein submitted to Genbank by Koczen et al. (Genbank Accession X59856; Appendix III).*

## 2.2 Sample Preparation

### 2.2.1 DNA isolation from blood samples

Peripheral blood samples were stored in a -20°C freezer before extracting the DNA (DNA could be successfully isolated from blood stored frozen for at least a year and samples could be thawed at least twice).

The frozen blood sample was thawed gradually at room temperature. Five ml of the thawed blood sample was transferred to a 50 ml Falcon centrifuge tube containing 45 ml of cold lysis buffer (10 mM Tris.Cl pH 7.5, 0.32 M sucrose, 1% Triton X-100, 5 mM $MgCl_2$). The blood was mixed into the lysis solution immediately and gently by inverting the tube a few times. The lysed solution was then centrifuged at 2,000 rev/min (1000g) in a swing out rotor for 8 min at room temperature. The supernatant was aspirated and the pellet of cell nuclei drained by inverting the tube on paper towel for 2 min and then wiping out the inside of the tube thoroughly with tissues. Two ml of digestion buffer (50 mM Tris.Cl pH 8.0, 100 mM EDTA, 0.5% SDS and proteinase K to a final concentration of 75 µg/ml) was added to the pellet. The tube was tightly capped and incubation carried out in a Hybaid oven with rotating overnight at 55°C. The digested nuclear lysate was extracted with an equal volume of phenol (Tris.Cl pH 8.0 saturated)/chloroform/isoamyl achohol (1:1:1/25) by gently shaking for 10 min. The phase separation was performed by centrifuging 5 min at 2000 rev/min. The lower organic phase was pipetted out as thoroughly as possible. Two volumes of ethanol were then added to the upper aqueous phase and a glass rod with a small hook used to spool the DNA out. The DNA was washed three times with 70% ethanol and once with TE buffer. The glass rod was then placed into a 1.5 ml microcentrifuge tube. The tip with DNA adhered was then broken off and 400 µl TE buffer was added. The sample was shaken at room temperature until the DNA dissolved (usually overnight), then stored at 4°C. Normally, at least 100 µg of DNA was obtained from 5 ml of blood.

## 2.2.2 RNA preparation from mammalian cells

Total RNA was prepared from the mammalian cells sloughed off into milk. The milk was collected directly into sterile polypropylene centrifuge tube from the teat of a cow during routine milking and stored at 4°C for no more than two hours before RNA extraction. Ideally, 50 ml of milk was collected from each cow. The milk was centrifuged at 2,000 rev/min (1000g) in a swing out rotor for 10 min at room temperature. The fat was removed with a plastic tip, the supernatant aspirated and the pellet drained by inverting the tube for 2 min. The pellet was then processed by either of the two following methods:

1) Lysis solution (750 µl) containing 5 M guanidinium thiocyanate, 0.25% β-mercaptoethanol and 25 mM sodium citrate was added to the cell pellet. The mixture was left for about 10 min, pipetting periodically to aid homogenisation. Homogenised solution was transferred to a 1.5 ml Eppendorf tube and centrifuged at 7,500 rev/min for 2 min to sediment undissolved materials. The 700 µl supernatant was transferred to a new microcentrifuge tube, 70 µl 2 M sodium acetate (pH 4.5) and 730 µl isopropanol were added, and the mixture stored at –20°C for 2 hours. The sample was then centrifuged at 15,000g for 15 min. The supernatant was carefully removed and the small pellet air dried. The pellet was resuspended with 20 µl RNase free water and stored at -70°C.

2) The cell pellet was resuspended in 750 µl of TRIzol®LS reagent (Life Technologies); 200 µl $H_2O$ was then added. The suspension was left at room temperature for 5 min with periodic pipetting to ensure homogenisation and dissociation. Then 200 µl of chloroform was added to the sample and vortexed twice for 10 secs with a 5-10 min interval. To separate the phases the sample was centrifuged at 12,000g for 15 min in a cold room. The upper aqueous phase, which contained the RNA, was transferred to a new 1.5 ml tube. The RNA was pelleted by adding 0.5 ml of 100% isopropanol, incubating for 10 min at room temperature and centrifuging at 12,000g for 10 min in a cold room. The supernatant was removed and the pellet air dried for 2 min and then dissolved in 20 µl RNase free water and stored at -70°C.

## 2.2.3 cDNA synthesis

To prepare cDNA, 2 µl of the RNA sample prepared as described in section 2.2.2 was make up to 11 µl with water and 2 pmol of an oligonucleotide fragment A6, EXE or A12 (Table 2.1.3; these are reverse primers located within exon 6, 9, 18 of $\alpha_{S1}$-casein gene, respectively). This reaction mixture was placed into a 0.2 ml PCR tube, heated to 70°C for 10 min and then quickly chilled on ice. Four µl of 5 x first strand buffer (250 mM Tris.Cl, 375 mM KCl, 15 mM $MgCl_2$), 2 µl of 0.1 M dithiothreitol and 2 µl of 10 mM dNTP mix (2.5 mM each of dATP, dGTP, dCTP and dTTP) were added to the tube and mixed, and then 1 µl of Superscript II (Life Technologies; 200 units/µl) added. After incubation for 2-3 hours at 42°C in a thermocycler, the reaction was inactivated by heating to 95°C for 5 min. The completed reactions were stored at -20°C and used as templates for PCR.

## 2.2.4 Casein protein isolation

Casein proteins was prepared by an acid precipitation method (Thompson and Kiddy 1964; Annan and Manson 1969). Fresh milk was collected directly into sterile Falcon tubes from individual cows at DRC during routine milking. The milk was centrifuged for 10 min at 2500 rev/min (1200g) in a swing out rotor. The fat layer was removed and the skim milk was transferred to a clean Falcon tube. 1 M HCl was gradually dropped into the skim milk until the pH decreased to 4.6-4.8. After the casein precipitated out, the sample was centrifuged at 800g for 5 min to pellet the caseins. The casein precipitate was washed three times by suspension in distilled water before being finally dissolved by addition of 1 M NaOH to increase the pH to 7.5.

An alternative procedure was also performed to increase the proportion of $\alpha_S$-casein in the casein precipitate. The casein was dissolved in 6.6 M urea and briefly centrifuged to remove the undissolved materials. The $\alpha_S$-casein complex was recovered by diluting the 6.6 M urea casein to 3.3 M urea at pH 4.6-4.8 (Thompson and Kiddy 1964; Thompson *et al.* 1969).

## 2.3 Nucleic Acid Manipulations

### 2.3.1 Polymerase chain reaction

One µl aliquot of DNA (approximately 10-20 ng) or the cDNA product (see section 2.2.3) were added to a 20 µl reaction mix containing 10 mM Tris.Cl (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, 200 nM of each dNTP and 0.5 U of Taq polymerase (Boehringer Mannheim) plus the appropriate primers (Table 2.1.3) at a final concentration of 1 µM. In the reverse transcription - polymerase chain reactions (RT-PCR), the reverse primer was the same as, or nested to, that used in synthesis of cDNA. The reactions were performed by denaturing at 94°C for 15 secs, annealing for 50 secs and elongating at 72°C for 1 min. The annealing temperatures are given in Table 2.3.1. The number of cycles varied from 15-29 in different experiments.

**Table 2.3.1 PCR and RT-PCR analyses**

| *For. primer* | *Rev. primer* | *Templete* | *annealing Tm.* | *length of product* |
|---|---|---|---|---|
| A3 | EXS | cDNA | 60 °C | 116 bp |
| A1 | A2 | DNA | 57 °C | 216 bp |
| A7 | A6 | cDNA | 61 °C | 255 bp |
| Spe | A2 | A1/A2 PCR product | 57 °C | 147 bp |
| Tru | A2 | A1/A2 PCR product | 57 °C | 147 bp |
| A7 | A13 | cDNA | 62 °C | 676 bp |
| A1 | A11 | DNA | 57 °C | 736 bp |
| A10 | A8 | DNA | 57 °C | 695 bp |
| A9 | A8 | DNA | 57 °C | 309 bp |

Products from PCR reactions were analysed by electrophoresis either directly or after being digested by restriction enzymes. All the restriction enzymes and related buffers used in this research were from Boehringer Mannheim. The gels for electrophoresis were either agarose (FMC BioProducts), high resolution Metaphor agarose (FMC BioProducts) or denaturing polyacrylamide. Both the restriction enzyme digestion and

the electrophoresis were carried out using standard molecular techniques (Sambrook *et al.* 1989; Ausubel *et al.* 1987) and following manufacturer's instructions unless otherwise stated.

### 2.3.2 PCR product purification and band isolation

Isolation of PCR products, either for sequencing or cloning, was done using either Wizard PCR-Preps Kit (Promega) or High Pure PCR Product Purification Kit (Boehringer Mannheim), following the manufacturer's instructions.

A band stab method (Bjourson and Cooper 1992) was used for isolation of bands of interest from a mixture of PCR products. The bands from the PCR reactions were separated by agarose gel electrophoresis in TAE/EB buffer system and visualised on an UV transilluminator. To prevent UV damage of the DNA, a piece of Mylar Glad oven bag was placed between the gel and UV transilluminator (Dick Wilkins, pers. comm.). The desired band was stabbed with a clean Pipetman tip and the small amount of gel in the tip, which contained the desired DNA, was used as template of another 20-25 cycles of PCR under the same conditions as the original PCR.

### 2.3.3 Radioactive labelling of DNA

Oligonucleotides, primers and PCR fragments labelled with radioactive isotope were used as hybridisation probes, or for DNA analysis on denaturing gels. Oligo probes were labelled by a 5'-end-labelling method and PCR fragments were labelled either by amplifying the fragments with a $^{33}P$ end labelled primer or by, after the PCR reaction, random primer labelling method. For the end labelling, 2 µl [$\gamma$-$^{33}P$]-ATP, 1 µl 10 x phosphorylation buffer, 2 µl oligonucleotides primer (20 µM), 1 µl polynucleotide kinase (10 units/µl) and 4 µl $H_2O$ were mixed and incubated at 37°C for 30 min. Both the enzyme and the buffer were purchased from Boehringer Mannheim. For random primer labelling, the Rediprime Kit was used and according to the manufacturer's instructions (Amershan RPN 1633). [$\alpha$-$^{32}P$]-dCTP was used for random primer labelling. Both [$\gamma$-$^{33}P$]-ATP and [$\alpha$-$^{32}P$]-dCTP were product of NEN™ Life Science with the specific activity 2000 Ci/mmol and 3000 Ci/mmol, respectively.

## 2.3.4 Denaturing PAGE for labelled PCR fragments

The PCR products labelled with [$\gamma$-$^{33}$P]-ATP 5'-end labelled primer were analysed by denaturing polyacrylamide gel electrophoresis (PAGE), using the Model S2 Sequencing Gel Electrophoresis System (Life Technologies) according to the manufacturer's instructions. To prepare 50 ml of solution for a 8% (W/V) polyacrylamide gel, 10 ml of 40% (w/v) 19:1 (w/v) acrylamide/bisacrylamide (BioRad; prepared in 50% urea, deionised by BioRad AG® 501-X8 (D) Resin and filtered through Watman No 1 paper), 5 ml 10 x TBE, 35 ml of 50% urea (deionised and filtered) were mixed. Immediately before pouring the gel, 300 µl of 10% ammonium persulfate (Sigma) and 35 µl of TEMED (Life Technologies) were added for polymerisation. Gels were allowed to polymerise for at least 30 min at room temperature before using.

The gel was pre-electrophoresed in 1 x TBE buffer for approximately 20 minutes at 90W constant power (using a BioRad Model 3000xi power supply). The mixture of labelled PCR products and loading buffer (10 mg xylene cyanol-FF, 10 mg bromophenol blue, 20 µl 0.5 M EDTA in 10 ml of formamide) were heated at 95°C for 5 minutes before loading them. The gel was electrophoresed at 65W constant power for approximately 1.5 hours. Finally, the gel was dried using a gel dryer (BioRad Model 583) for two hours at 80°C under vacuum and then autoradiographed.

## 2.3.5 Southern hybridisation

Blotting was performed using modifications of the Amersham protocols (Hybond$^{TM}$-N$^{+}$ Version 2.0 Amersham). After electrophoresis in TAE buffer, the gel was soaked in three changes of denaturing buffer (0.5 M NaOH, 1.5 M NaCl) for total 30 minutes while rocking at room temperature. The gel was then placed upside down on 5 sheets Gel-Blotting-Paper (Schleicher & Schuell GB 002; pre-wetted in denaturing buffer) on a glass plate. Hybond$^{TM}$-N$^{+}$ nylon membrane was cut to the appropriate gel size, soaked in denaturing solution and carefully placed on top of the gel to avoid air bubbles. The pieces of pre-wetted Gel-Blotting-Paper were placed on top of the membrane followed by a 10 cm wad of paper towels and the stack was compressed lightly with a plate of glass. The blot was left overnight at room temperature to allow

maximum capillary transfer of DNA onto the membrane. The membrane was then washed in 3 x SSC and used for hybridisation.

The hybridisation were performed following the modified Church and Gilbert protocol (Church and Gilbert 1984). After blotting, the membrane was placed in a Hybaid bottle that contained 10 ml of Church & Gilbert (C&G) hybridisation solution (0.5 M sodium phosphate buffer pH 7.0, 1 mM EDTA, 7% SDS). The bottle was rotated in the mini MK II hybridisation oven at 65°C for one hour. To denature $^{32}$P-DNA probes labelled by random primer labelling, freshly made 4 M NaOH was added to the radioactive probe to a final concentration of 1 M and left at room temperature for 5 minutes. The C&G solution was removed from the bottle and replaced with 4 ml of fresh C&G and the denatured radiolabelled probe was added directly to the hybridisation solution, ensuring that the probe solution did not contact the membrane until diluted. The hybridisation tube was rotated overnight in the mini MK II hybridisation oven at 65°C. For oligonucleotide probes, the hybridisation was carried out at 37°C.

After hybridisation the probe was transferred to a 50 ml Falcon tube and stored at 4°C for future use. The membrane was then washed 3 times with 3 x SSC (0.45 M NaCl, 45 mM sodium citrate) containing 0.1% SDS, at 65°C or room temperature (for oligonucleotide hybridisation). A Geiger counter (Survey Meter model II) was used to monitor the strength of signal while washing. When using $^{32}$P-probes, the washed membrane was sealed in a thin plastic bag and in a darkroom, a piece of X-ray film (Kodak X-OMAT AR film) placed on top. Then both were sandwiched between two intensifying screens in a light tight X-ray cassette. The cassette was placed in a -70°C freezer for exposing and then the X-ray film removed and developed. For $^{33}$P-probes, the membrane was dried and film placed directly against the membrane in a cassette which was exposed at room temperature.

## 2.3.6 PCR products cloning

### 2.3.6.1 Preparation of Bluescript® II SK vector

The vector used for the PCR product cloning in this project was T-tailed pBluescript® II SK$^+$ plasmid (Stratagene). This vector was prepared by digesting the pBluescript® II SK$^+$ plasmid with *Eco*R V to create blunt ends and then attaching a single thymine

residue to the 3' ends of the both strands. This procedure allows the T-tailed vector to ligate efficiently with a PCR product, as Taq DNA polymerase places an additional adenine (A) at the 3' end of the DNA fragment during PCR (Clark 1988).

The following method is essentially that described in Ausubel *et al.* (1987). Ten µg pBluescript® II $SK^+$ plasmid was digested overnight at 37°C with 100 units *Eco*R V. The digestion product was extracted with an equal volume of Tris-equilibrated phenol. The aqueous phase was then extracted with an equal volume of chloroform. The DNA was precipitated with 0.1 volume of 7.5 M ammonium acetate and 2.5 volumes 100% ethanol and stored at -20°C for 15 minutes. The DNA was pelleted by centrifuging for 15 minutes, 12,000g at room temperature. The DNA precipitate was resuspended in 10 µl 10 x PCR buffer (Boeringer Mannheim), 1 µl 100 mM dTTP, 2.5 µl Taq DNA polymerase (1u/µl) and 86.5 µl $H_2O$. The DNA sample was incubated overnight at 73°C. The T-tailed DNA was then electrophoresed on a 0.7% low melting point (LMP) agarose gel containing EB at a final concentration of 0.5 mg/ml. The plasmid band was identified using a long wave UV transilluminator, cut out, melted and stored at 4°C.

### 2.3.6.2 Preparation of competent cells

One ml of an overnight culture of *E. coli* cells that had been inoculated from a single clone was transferred into prewarmed 100 ml LB broth (Bacto-Tryptone 1%, Bacto-Yeast extract 0.5% and NaCl 1%) with 0.2% glucose and 10 mM $MgSO_4$. The cells were incubated for ~2.5 hours at 37°C with shaking at 250 rev/min and then pelleted in a Beckman J2-21M centrifuge for 10 min at 3000 rev/min (1000g) at 4°C. The cells were gently resuspended at 4°C with 2.5 ml TSS solution (Chung *et al.* 1989), which is LB broth containing 10% PEG (mw 3350) and 50 mM $MgSO_4$. One hundred µl aliquots of the cell suspension were transferred to 0.5 ml pre-cooled microcentrifuge tubes and stored at -70°C.

### 2.3.6.3 Ligation

Before use, the T-tailed plasmid in LMP agarose was melted at 65°C for 10 minutes. For a 20 µl ligation reaction mixture, 3 µl of melted LMP plasmid was mixed with 7 µl of purified PCR product, 2 µl of 10 x T4 DNA ligase buffer (660 mM Tris.Cl, 50 mM

$MgCl_2$, 10 mM dithioerythritol, 10 mM ATP, pH 7.5), 2 units T4 DNA Ligase (Boehringer Mannheim) and 6 µl $H_2O$. The mixture was incubated overnight at room temperature to allow the ligation of the PCR product and vector and then heat inactivated at 65°C for 10 min.

### 2.3.6.4 Transformation

An aliquot of competent *E. coli* cells was thawed on ice. Ten µl of ligation product was added to 100 µl of competent cells. Each tube was gently swirled with a Pipetman tip and incubated on ice for 30 min, heat shocked for 90 sec in a 42°C water bath and then chilled on ice for 2 min. Fifty µl of the transformed cells was added to a sterile 0.5 ml Eppendorf tube containing 450 µl of LB broth which contained 0.4% glucose pre-warmed to 37°C. The cells were grown for 1 hour at 37°C to allow the bacteria to recover and express the antibiotic resistance.

Approximately 200 µl of the cells were plated out onto a pre-warmed LB agar plate ($\phi$=9 cm) on which had been previously spread solutions containing 250 µg Ampicillin, 80 µg X-gal and 80 µg IPTG using a sterile glass rod. The plates were left at room temperature until the liquid had been absorbed. The plates were then inverted and grown at 37°C for 12-18 hours.

### 2.3.6.5 The identification of recombinant plasmids

Single white, putatively positive, colonies were inoculated into 5 mls of LB broth containing Ampicillin at a final concentration of 75 µg/ml. The cells were grown overnight at 37°C with vigorous shaking at 250 rpm. The overnight culture was transferred into a 10 ml centrifuge tube and centrifuged at 2500g for 10 min to pellet the cells. The cell pellet was then resuspended in 700 µl of STET buffer (8% sucrose, 50 mM EDTA, 50 mM Tris pH 8.0 and 0.5% Triton X-100), transferred to a 1.5 ml Eppendorf tube, and after adding 50 µl lysozyme (10 mg/ml in STET) to the suspension, vortexed for 5 secs and then immediately heated in a boiling water for 80 secs. The boiled sample was centrifuged at 15,000g for 10 min and the supernatant transferred to a new 1.5 ml microcentrifuge tube. An equal volume of -20°C cold isopropanol was added to the supernatant, mixed well, stood at room temperature for 5

min, then microfuged at 15,000g for 5 min to pellet the DNA, which was resuspended in 50 µl TE.

To identify the cloned DNA insert, a restriction enzyme digestion was performed. *Bam* HI and *Hin*d III, located on either side of *Eco*R V, were chosen to digest the plasmid DNA and the resulting products were electrophoresed and analysed on 1% agarose gel. Sometimes the cloned fragment was also identified by PCR by using the isolated plasmid DNA as the template for amplification.

### 2.3.6.6 Single stranded DNA rescue

The following method was used to make single stranded DNA from recombinant plasmids for DNA sequencing.

A single colony was inoculated into 2 ml of 2 x YT broth (Trypton 1.6%, Yeast extract 1% and NaCl 0.5%, pH 7.0) containing 50 µg/ml of Ampicillin and 20 µl $>10^9$ pfu/ml VCSM13 helper phage (Stratagene). The culture was grown for 3 hrs at 37°C with vigorous shaking and then 6 µl of 25 mg/ml Kanamycin added to select phage infected cells. The culture was grown overnight, at 37°C, with vigorous shaking, to maximize the yield of phage.

These overnight cultures were centrifuged at 10,000g for 5 minutes at room temperature. 1.3 ml of supernatant was transferred to a 1.5 ml microcentrifuge tube and 195 µl of a solution containing 20% PEG and 2.5 M NaCl was then added with careful mixing. The mixture was left on ice for 15 min to allow the phage particles to precipitate out.

The sample was centrifuged again at 15,000g for 15 min to pellet the phage particles. Care was taken to remove the viscuos supernatant completely and the pellet resuspended in 100 µl TE (pH 8.0). The sample was vortexed for 30 seconds to resuspend the phage, which were then extracted with 50 µl of Tris-saturated phenol, vortexed for 30 seconds and then 50 µl chloroform added following by further vortexing for 30 seconds. The sample was centrifuged at 12,000g for 2 min to separate the solvent phases. The upper aqueous phase (about 90 µl) was transferred to a new 1.5

ml microcentrifuge tube, 1/10 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of 100% ethanol added, left on ice for 5 minutes then DNA was pelleted by centrifugation at 15,000g for 10 minutes at room temperature. The DNA pellet was resuspended in 12 µl TE (pH 8) and stored at -20°C.

### 2.3.7 DNA sequencing

DNA samples were sequenced by the Waikato DNA Sequencing Facility using an ABI Prism 377 Automated DNA Sequencer with dye terminator chemistry. The PCR products to be sequenced were purified by Wizard® PCR-Preps DNA Purification System (Promega) following the manufacturer's instructions. For the ssDNAs, no further purification procedure was needed after the sample had been prepared by PEG pptn protocol. DNA concentration in all samples for sequencing were evaluated by mini-agarose gel electrophoresis to ensure to satisfy the concentration requirement of the ABI Prism 377 sequencing, namely, ~50 ng/µl for ssDNA, ~5 ng/100bp for PCR product (The primer concentration was 0.8–1.6 pmol/µl).

### 2.3.8 GeneScan analysis

The PCR fragments for GeneScan analysis were prepared by PCR reactions carried out the same way as being described in section 2.3.1 except for addition of fluorescent [R6G]-dUTP (Perkin-Elmer P/N 402174) to the PCR reaction mixture to a final concentration of 1 µM. Reactions were normally carried out for less than 16-20 cycles and then the product purified using a MicroCon-30 (Perkin-Elmer) column before loading onto an Applied Biosystem Sequencer. The data on the DNA fragments collected by the sequencer were analysed by GeneScan™ software (Perkin-Elmer).

## 2.4 Protein Manipulations

### 2.4.1 Urea-PAGE for casein proteins

Milk protein samples, consisting of either lyophilised proteins or skim milk, were dissolved in 40% urea, then 20 µl of β-mercaptoethanol and 20 µl of tracking dye (0.1% bromophenol blue) were added per 100 µl of sample. The mixture was left for more than 45 min at room temperature.

Urea polyacrylamide gel electrophoresis was performed essentially as described by Medrano and Sharrow (1989). The electrophoresis was carried out either in large format gels (16 x 20 cm, 1.5 mm thick, Bio-Rad Protean™ II Slab Cell) or in mini format gels (9 x 10 cm, 1 mm thick, Hoefer Might Small™ SE 250 /SE 260 apparatus). All the gels were prepared in a discontinuous form (Laemmli 1970) with the gel contents as described below.

**Table 2.4.1 Composition of urea-PAGE gels**

| *Chemical* | *Stacking gel* | *Dissolving gel* |
|---|---|---|
| Acrylamide (30%T; 3.3%C in 50% urea) | 4.5 % | 8 % |
| 50% Urea | 4 M | 4 M |
| 0.5M Tris.Cl pH6.8 | 0.125 M | |
| 1.5M Tris.Cl pH8.9 | | 0.375 M |
| 10% APS | 0.1% | 0.04% |
| TEMED | 0.075% | 0.05% |

The running buffer contained 25 mM Tris and 0.19 M glycine at pH 8.3. A constant current was applied for electrophoresis. For the Protean™ II Slab Cell gel, the electrophoresis was at 20 mA for 16 hours. The small Hoefer Might Small™ gel was electrophoresed for 4 hours at 9 mA.

For Coomassie staining, the gel was removed from the glass plates immediately after electrophoresis and soaked in a fix-stain solution (0.1% Coomassie Blue R250 in 40% methanol and 10% glacial acetic acid) for 1 hour with rotary mixing. The gel was destained by rocking the gel in a solution containing 30% methanol and 10% glacial acetic acid overnight, or until the background was essentially clear.

## 2.4.2 Western blotting

The electrophoretically separated proteins were transferred from the polyacrylamide gels onto Hybond™-P: PVDF or Hybond™-C Super (Amersham; a supported pure nitrocellulose membrane) by semi-dry blotting in a LKB Electrophoretic Transfer Kit. A constant current of 80 mA/cm$^2$ was applied for 2 hours. The transfer buffer

contained 25 mM Tris, 192 mM glycine and 20% methanol (Towbin *et al.* 1979). Immediately following the blotting, the membrane was briefly washed twice with phosphate buffered saline-Tween (PBS-T, 0.1% Tween-20, 80 mM $Na_2HPO_4$, 20 mM $NaH_2PO_4$ and 100 mM NaCl). The membrane was then placed in a Hybaid bottle and the non-specific binding sites were blocked by incubating in blocking buffer (PBS-T solution containing 3% bovine serum albumin) for 1 hour with rotation. The blot was then washed three times with PBS-T. Each wash was 5 minutes with rotary shaking.

The primary antibody, anti-$\alpha_{S1}$-casein, was diluted 1:10,000 times with PBS-T. The membrane was incubated in the antibody solution for 1 hour at room temperature. The membrane was then washed with PBS-T for 10 minutes once and 5 minutes twice. The second antibody anti-rabbit-IgG (whole molecule) Peroxidase Conjugate (Sigma Product No. A-6154) was diluted 1:10,000 with PBS-T and incubated with the membrane for 1 hour. Before being further developed, the blot was thoroughly washed 10 minutes twice and 5 minutes twice with a large volume of PBS-T.

The bound second antibody was detected using enhanced chemiluminescence (ECL) reagents (ECL™ Western blotting kit, Amersham). ECL was carried out according to the manufacturer's instructions. The combined reagent 1 and 2 (ratio=1/1) were poured over the membrane so that it was completely covered. The membrane was left to stand for one minute without agitation, then removed from the solution and placed in a plastic bag. The membrane was immediately placed against Kodak X-OMAT AR film in a X-ray cassette. The exposure time varied from 15 sec to a few minutes.

### 2.4.3 Fast protein liquid chromatography (FPLC)

For fractionation of caseins, the fast protein liquid chromatography (FPLC) was carried out on a Mono Q HR 5/5 (Pharmacia) anion-exchange column which was fitted onto a Phamacia liquid chromatography controller LCC-500 system (Andrews *et al.* 1985).

The caseins precipitate prepared by the method described in section 2.2.4 was redissolved in FPLC buffer I (20 mM Tris.Cl, 8 M urea, 10 mM β-mercaptoethanol) at a concentration of about 10 mg/ml. Undissolved particles were removed by filtration

(0.22μm filter, Millipore Corp.) just before the sample was loaded onto the FPLC column.

A 500 μl casein sample was loaded in each run. The flow rate of the mobile phase was 1 ml/min and the two elution buffers were salt free buffer I (20 mM Tris.Cl pH 9, 8 M urea, 10 mM β-mercaptoethanol) and buffer II, which was the same as buffer I except for containing 0.35 M NaCl. After a 4 min wash using salt free buffer I, the next 5 min elution was performed using a gradient of sodium chloride concentration that was achieved by increasing the buffer II component from 0-75% of the mobile phase over the next 5 min. A more shallow increase in the salt gradient was then applied for the next 20 min so that the buffer II portion in the mobile phase reached 100% at the end of this time. A two-channel chart recorder (Pharmacia REC-482) was used to monitor the eluted protein peaks. The protein fractions were collected every half minute by a Pharmacia Frac-100 fraction collector.

## *CHAPTER III*

# THE DRI $\alpha_{S1}$-CN A VARIANT AND IDENTIFICATION OF THE GENE MUTATION

*The DNA mutation in DRI $\alpha_{S1}$-casein A cows*
*The design of a PCR test for the DRI $\alpha_{S1}$-casein A mutation*
*Discussion*
*The relationship between $\alpha_{S1}$-CN A gene mutation and aberrant splicing*
*PCR based testing for $\alpha_{S1}$-casein A mutations*

## 3.1 Introduction

The $\alpha_{S1}$-casein A variant is one of the six known protein polymorphisms in the bovine $\alpha_{S1}$-casein family. Originally, it was identified in the American Holstein (Thompson *et al.* 1962). Later, it was found in Red Danish, German Red and some other dairy cows (Larsen and Thymann 1966 a,b; Farrell and Thompson 1971). The $\alpha_{S1}$-casein A variant is rare, both in terms of breed distribution and frequency. Grosclaude (1988) described the occurrence of variant A in one out of 23 different breeds investigated, with an allele frequency lower than 0.01. Erhardt (1993) reported $\alpha_{S1}$-casein A in a German Red cattle breed with an allele frequency of 0.001.

The $\alpha_{S1}$-casein A protein is dramatically different from other $\alpha_{S1}$-casein members in that it has a 13 amino acid deletion which significantly alters the properties of the protein (Grosclaude *et al.* 1972). Protein sequencing data shows that the deleted amino acid residues of this variant are His-Gln-Gly-Leu-Pro-Gln-Pro-Phe-Pro-Glu, located at aa residues 14 to 26 in the N-terminal region of the protein. In accordance with these data, McKnight *et al.* (1989) observed a cDNA form of $\alpha_{S1}$-casein in which a 39 bp stretch encoding the deleted 13 amino acids of $\alpha_{S1}$-casein A variant was absent. Based on the exon-intron structure of the $\alpha_{S1}$-casein gene, it was clear that the entire exon 4 was missing in the mRNA of the $\alpha_{S1}$-casein A variant.

Loss of an entire exon may be caused either by a deletion in genomic DNA or by an alteration during mRNA splicing. Mohr *et al.* (1994) reported that in a German Red animal with a heterozygotic $\alpha_{S1}$-casein A allele, the lack of exon 4 of $\alpha_{S1}$-casein A mRNA is not the result of a gross genomic deletion. The German Red A allele genomic fragment extending across the exon 4 segment was identical in size with the fragment of the common B allele. Furthermore, sequencing analysis showed that both exon 4 and intron 3 sequences in the German Red $\alpha_{S1}$-casein A allele were entirely unchanged, and the only obvious change was a single point mutation (T→A) at position +6 in the intron 4 splice donor sequence which occurred in one half of the clones obtained from this heterozygotic animal. The authors stated that this T→A transversion mutation is responsible for the skipping of exon 4 during RNA processing (Figure 3.1.1).

In spite of its rare occurrence, the $\alpha_{S1}$-casein A protein is of interest to scientists in both the basic sciences and the dairy industry because its physico-chemical properties are different from other $\alpha_{S1}$-casein variants. Due to the deletion of the 13 amino acid residues, one of the major chymosin cleavage sites is absent, also the hydrophilic N-terminal section can closely contact to the distal phosphate-rich section (Kumonski *et al.* 1991). These changes in the $\alpha_{S1}$-casein A protein reduce the hydrophobic nature of the protein compared to $\alpha_{S1}$-casein B (Creamer *et al.* 1982), leading to a change in the rate of proteolysis during cheese ripening. Casein adhesives manufactured from casein containing $\alpha_{S1}$-casein A are less viscous and have better adhesive strength than those containing either the B or C variants (Mountfort 1984). This desirable attribute of $\alpha_{S1}$-casein A protein has prompted researchers in the New Zealand Dairy Research Institute (DRI) to establish a small herd of such A variant animals based on a founder A variant Jersey cow first identified in a Massey University herd.

In this part of the project, a study at the genomic DNA level was carried on the DRI (NZ) $\alpha_{S1}$-casein A herd with two aims:

- Determining the DNA sequence in the exon 4 region of the DRI $\alpha_{S1}$-casein A gene to see if was the same as in the German Red and

- To develop a DNA test which would enable new calves to be genotyped.

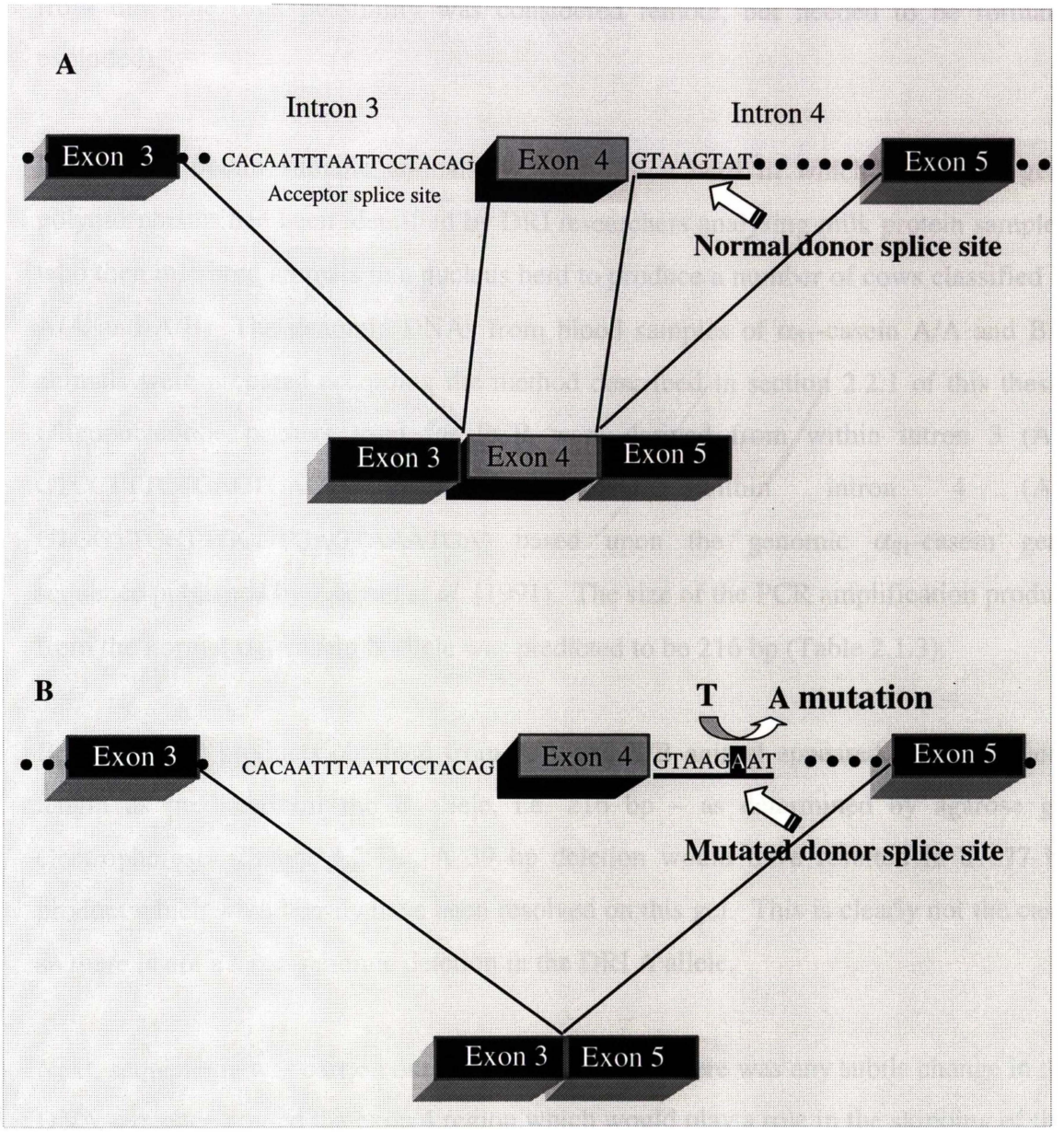



**Figure 3.1.1 A schematic diagram shows the occurrence of the exon 4 skipping of German Red $\alpha_{S1}$-casein A variant.** *A). Exon 4 is flanked by normal acceptor and donor splice sequences and included in mature mRNA.* ***B).*** *In German Red A variant, the 6$^{th}$ nucleotide in intron 4 is mutated from T to A, and this is believed to cause the skipping of the exon 4 in the mature mRNA.*

## 3.2 The DNA Mutation in DRI $\alpha_{S1}$-Casein A Cows

To explore the DNA mutation in the DRI $\alpha_{S1}$-casein A allele, the DNA samples of DRI $\alpha_{S1}$-casein A variant cows were first analysed by a polymerase chain reaction to see if 39 base pairs, corresponding to the deletion of 13 amino acid residues, were deleted

from the gene (this possibility was considered remote, but needed to be formally excluded).

In order to obtain suitable animals, we sampled a herd in which $\alpha_{S1}$-casein gene polymorphisms had been identified by DRI researchers analysing milk protein samples, who then interbred animals in a nucleus herd to produce a number of cows classified as A/A and A/B. The genomic DNAs from blood samples of $\alpha_{S1}$-casein A/A and B/B animals were prepared according the method described in section 2.2.1 of this thesis. Oligonucleotide primers used for PCR were derived from within intron 3 (A1: CTCCTTTCTGACTGACTGTGTTTTTCAC) and within intron 4 (A2: GTGGTTGCTTGGGTGAGTAAATGA) based upon the genomic $\alpha_{S1}$-casein gene sequence published by Koczan *et al.* (1991). The size of the PCR amplification product from the normal $\alpha_{S1}$-casein B allele was predicted to be 216 bp (Table 2.1.3).

Both the PCR products obtained from A/A and B/B animal appeared to be the same length as predicted for the B allele, i.e. 216 bp – as determined by agarose gel electrophoresis (Figure 3.2.1). A 39 bp deletion would have resulted in a 177 bp product which would easily have been resolved on this gel. This is clearly not the case, so there is not a large genomic deletion in the DRI A allele.

Next, sequencing was carried out in order to detect if there was any subtle change in the DNA sequence around the exon 4 region which would play a role in the skipping of this exon during mRNA maturation. The PCR product, which encompassed the 4th exon region, was directly sequenced using an ABI Prism 377 sequencer. Interestingly, a point deletion was found in the sequence from the $\alpha_{S1}$-casein A/A animal, whereas the sequence of the B/B control animal perfectly matched the published sequence (Genbank Accession X59856). The mutation in the DRI $\alpha_{S1}$-casein A allele is a single base (A) deletion which occurs just four (three ?)* bases inside intron 4. (Figure 3.2.2). This is quite different from the mutation in the German Red A allele identified by Mohr *et al.* (1994); in their case it was a base transversion in position 6 of intron 4.

---

* *Formally, it is not possible to state whether it is nucleotide 4 and 3 which has been deleted. It is stated as 4 in this thesis because it is the fourth base that is different from the normal sequence.*

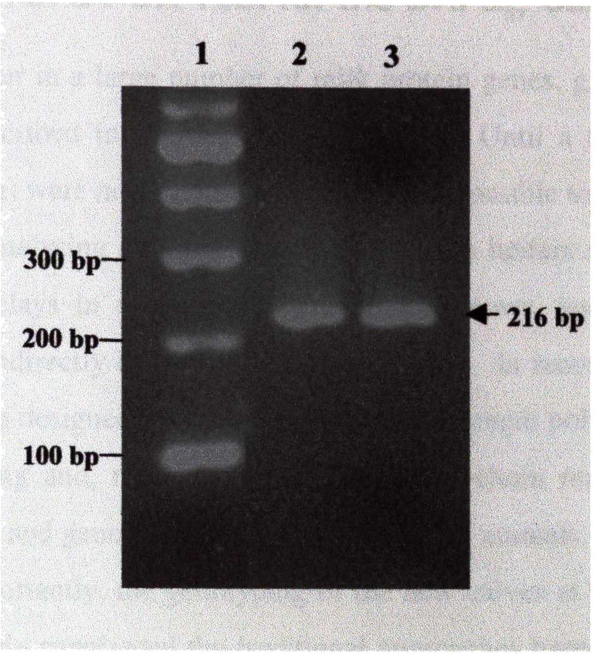



**Figure 3.2.1 PCR amplification of DNA from $\alpha_{S1}$-casein A/A and B/B animals.** *Primers A1 and A2 were used for amplification. The amplified products were electrophoresed on 2.5% agarose gels. Lanes: 1. 100 bp Molecule marker. 2. $\alpha_{S1}$-casein A/A animal. 3. $\alpha_{S1}$-casein B/B animal.*

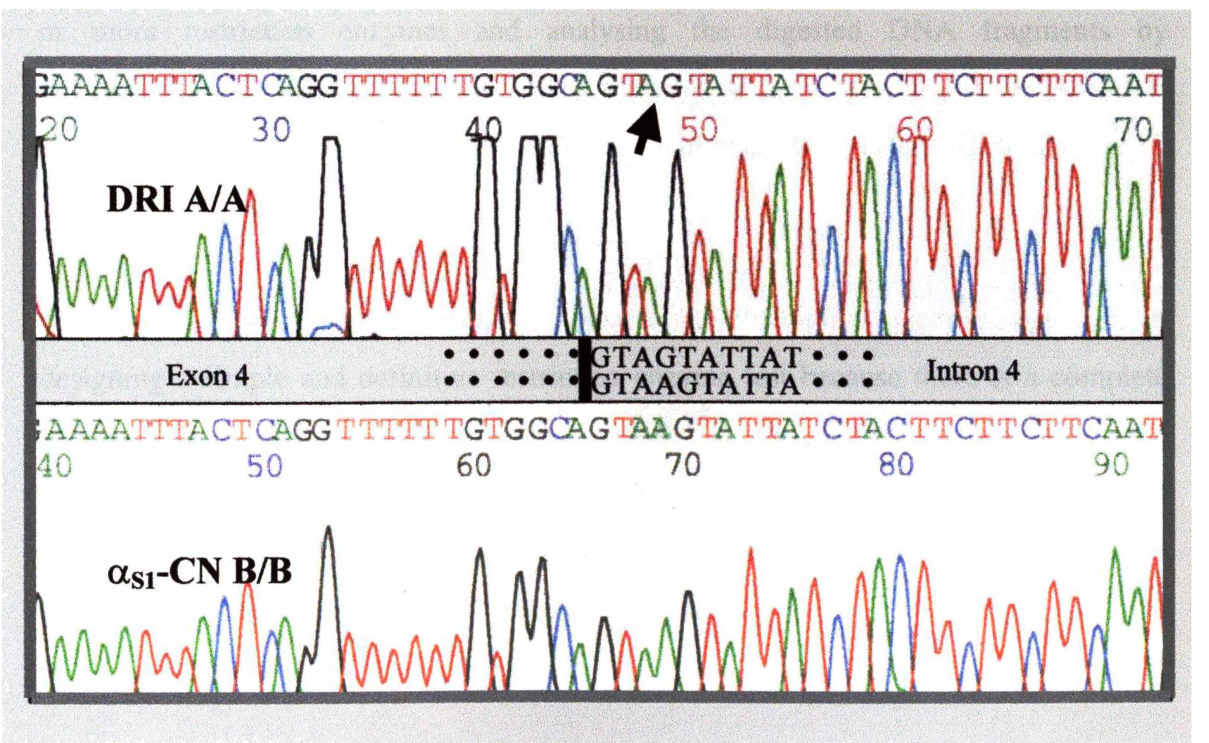



**Figure 3.2.2 The 5' splice junction of intron 4 of $\alpha_{S1}$-CN gene.** *The black arrow indicates the position of a single base (A) deletion in DRI $\alpha_{S1}$-casein A/A homologous cow (upper sequence).*

## 3.3 The Design of a PCR Test for the DRI $\alpha_{S1}$-Casein A Mutation

Point mutations occur in a large number of milk protein genes, giving rise to the milk protein variants described in chapter 1 of this thesis. Until a few years ago, DNA analyses (genotyping) were not feasible and it was only possible to phenotype suspected variant animals by analysing milk proteins obtained from heifers at their first lactation. This led to long delays in establishing genotypes of cows, and uncertainties often remained over the indirectly deduced genotypes of sires. In recent years, a number of DNA tests have been designed, first utilising restriction length polymorphisms (RFLPs) and Southern blotting and, more recently, polymerase chain reactions (PCR) which enable the screening and genotyping of a large number of animals (both dams and sires) and, even more importantly, the genotyping of the new calves at birth. This approach has almost completely supplanted the traditional approaches based on the examination of the milk proteins themselves. The most simple PCR tests are those in which the DNA mutations result in the loss or gain of a restriction enzyme recognition site. That is, the milk protein polymorphisms are reflected in restriction fragment length polymorphisms (RFLPs) which can be detected by digesting the PCR product with one or more restriction enzymes and analysing the digested DNA fragments by electrophoresis on an agarose gel. Such methods work very well for the A and B variants of both κ-casein and β-lactoglobulin. (Schlieben *et al.* 1991; Prinzenberg *et al.* 1996; Wilkins and Kuys 1992).

For the $\alpha_{S1}$-casein A DRI mutation, considerable difficulties were experienced in designing a simple and definitive restriction enzyme test because there is a complete absence of suitable restriction enzyme sites in the DNA region where the "polymorphism"* occurs. Instead, a method was designed with two steps of PCR amplification creating a diagnostic restriction site in the mutation site of the A variant or in the "normal" sequence of other $\alpha_{S1}$-casein alleles.

---

* *Strictly speaking, the A variant occurs at such a low frequency that they should formally be considered mutations not polymorphisms*

In this method, DNAs were first amplified by forward A1 and reverse A2 primers (Fiqure 3.3.1 and Figure 3.3.3). The first step PCR product was then used as template for the next amplification. The second step PCR reactions were carried out on each DNA sample using one or other of the two specially designed forward primers, Spe and Tru, and a common reverse primer A2, to create the diagnostic restriction sites. This Amplification Created Restriction Site (ACRS) method (see, for example, Lien *et al.* 1992) relies on a purposeful mismatch in the 3' region of the forward primer creating a restriction site. Thus, the Spe and Tru primers hybridise immediately adjacent to the point deletion site in the $\alpha_{S1}$-casein gene but are designed to contain slightly mismatched sequences which create restriction sites consisting of part primer sequence and part genome sequence in the amplified PCR product (Figure 3.3.1). Spe was designed so that the restriction enzyme *Spe* I only cuts the A variant PCR product (ACTAGT), and the second ACRS primer (Tru) only introduces a *Tru* 91 site (TTAA) in the PCR product from the $\alpha_{S1}$-casein B and not in the variant A allele (Figure 3.3.2). Although one ACRS reaction would suffice for diagnosing the $\alpha_{S1}$-casein A variant, the second reaction is useful for validating the results, especially as the consequences of misdiagnosing an A variant in the nucleus DRI herd are rather costly in both time and money.

As can be seen from Figure 3.3.3, cows with the three phenotypes (A/A, A/B and B/B) were tested using this method. The restriction fragment length patterns (147 bp and 120 bp for, respectively, uncut and cut fragment) occurred exactly in the manner predicted for the presumed genotypes. Note that one disadvantage of the ACRS method is that the cleaved products only differ from the uncleaved products by approximately the length of the primer, so it is necessary to use Metaphor gels to achieve good resolution.

Several calves in the DRI herd have been genotyped using this PCR test method. As shown in Figure 3.3.4, all the results were clear and unquestionable.

In an effort to speed up the PCR analysis, some analyses were also performed without using the initial PCR reaction to "pre-amplify" a larger $\alpha_{S1}$-casein gene fragment containing the mutated site. When the two ACRS primers were used directly to PCR

genomic DNA, the products could still be analysed using the *Tru* 91 and *Spe* I restriction enzymes, but unequal amounts of the two products were obtained and the results were not quite as simple to interpret (data not shown) as when using the two step procedure.

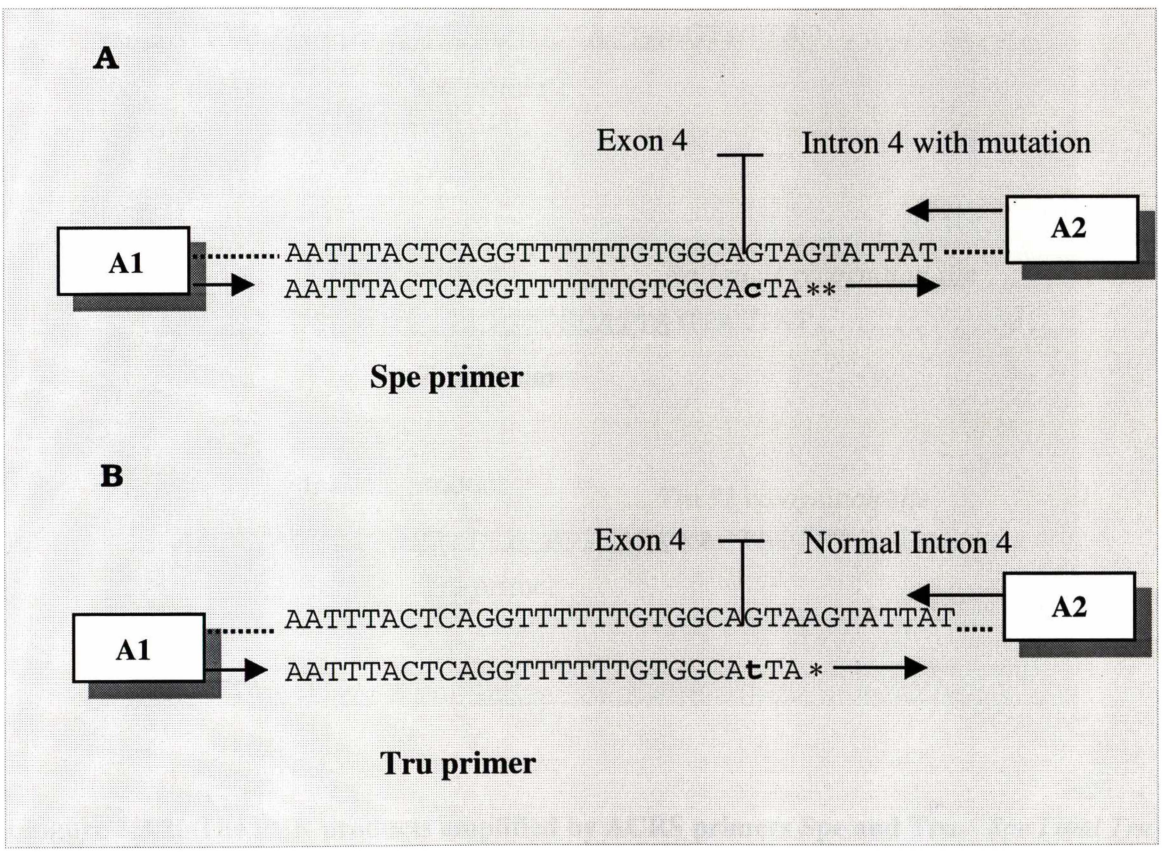



**Figure 3.3.1 Design of amplification created restriction site (ACRS) primers.** *Primer Spe and Tru hybridise to exon 4 and the donor splicing sites. **A**). A mismatching nucleotide at the third position from the 3' end (G→C, bold lowercase) was introduced to the Spe primer. **B**). A mismatching nucleotide (G→T, bold lowercase) at the 3rd position from the 3' end was introduced to the Tru primer.*

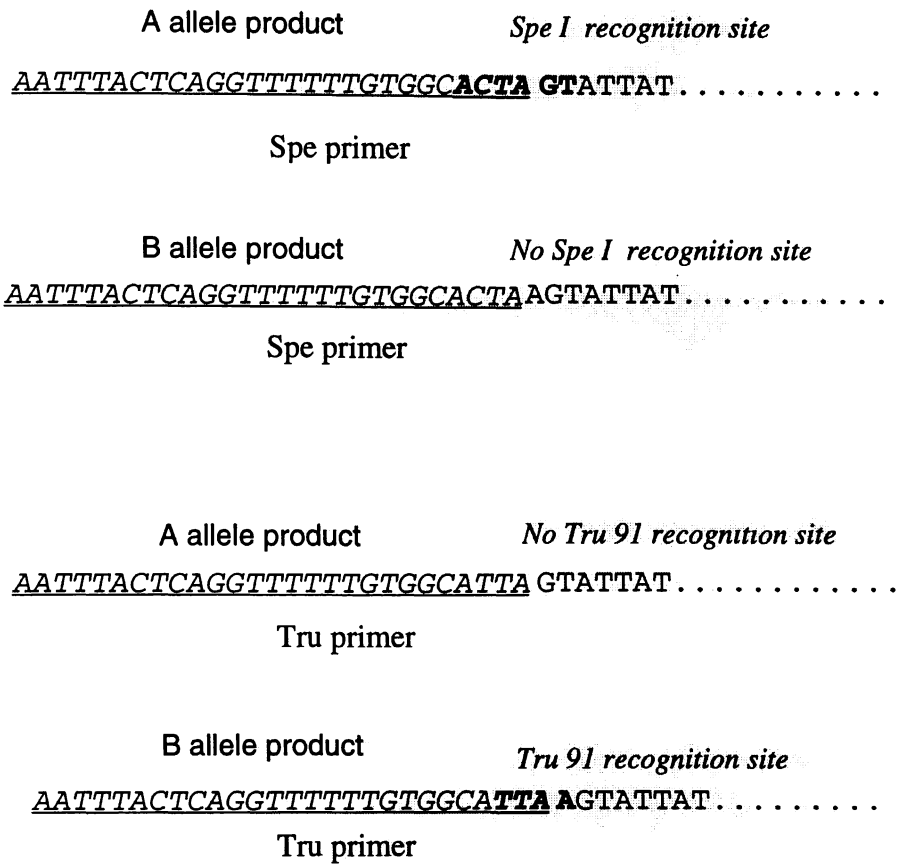



**Figure 3.3.2 The PCR products amplified by ACRS primers Spe and Tru.** *Spe I and Tru 91 restriction enzyme recognition sites (bold sequences) are created for, respectively, mutated A allele (Spe I: ACTAGT) and normal B allele (Tru 91: TTAA).*

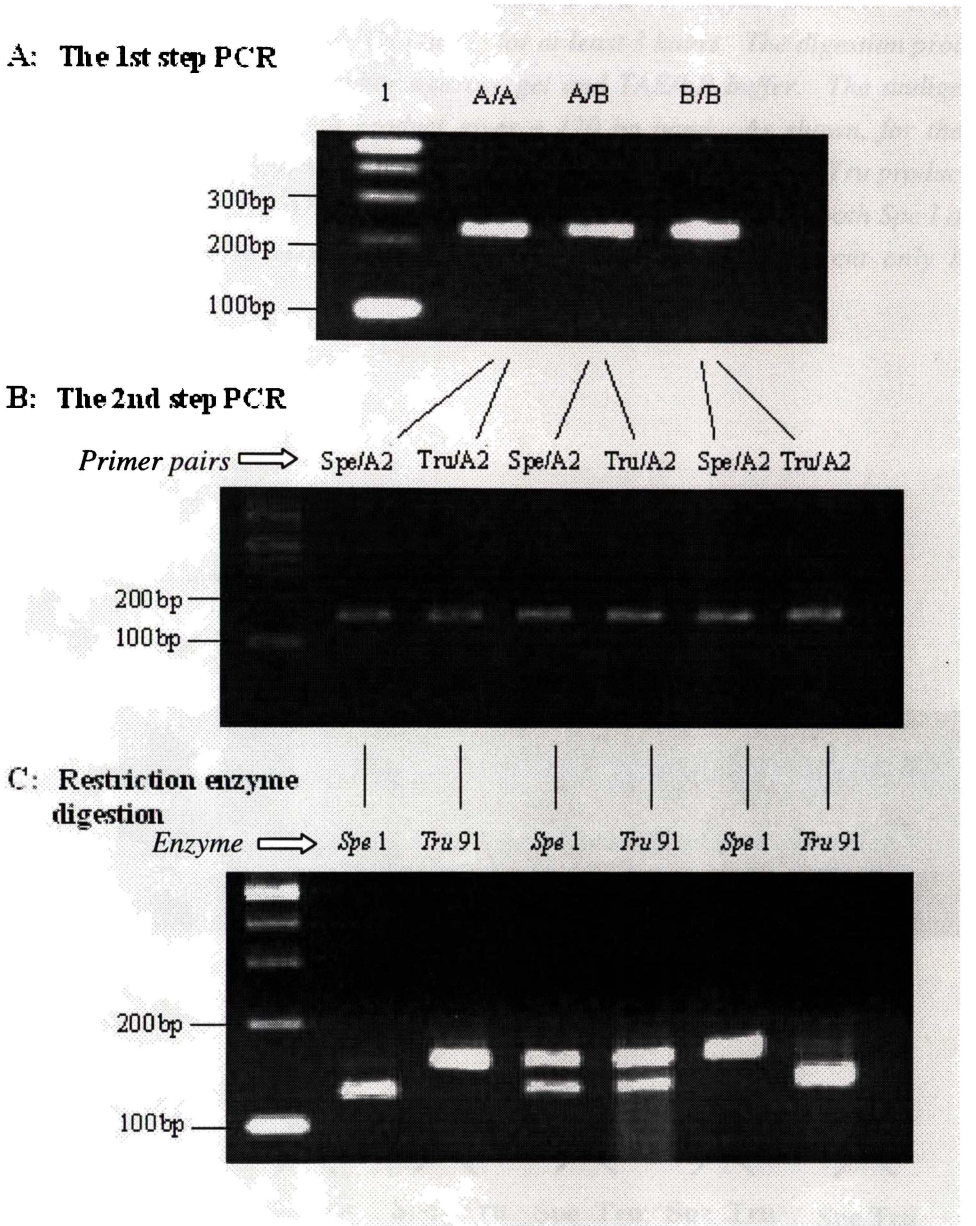



**Figure 3.3.3 The two-step ACRS PCR for analysing the $\alpha_{s1}$-casein DRI A variant.** *A). The first step PCR: DNAs from $\alpha_{s1}$-casein A/A, A/B and B/B animals were amplified by A1/A2 primers as previously described. A 216 bp product was obtained.* ***B****). The second step ACRS amplification: Two 20 µl PCR reactions were used for each animal. One reaction with primers Spe (*AATTTACTCAGGTTTTTTGTGGCACTA*) and A2 (*GTGGTTGCTTGGGTGAGTAAATGA*), another reaction with primers Tru (*AATTTACTCAGGTTTTTTGTGGCATTA*) and A2. The products of the first step PCR were diluted 1:50 with $H_2O$ and 1 µl of the dilution was used as template for each 20 µl second step PCR reaction. PCR was performed for 15 cycles in a PROGENE Thermocycler and a 147 bp product was obtained.* ***C****). The restriction enzyme analysis: 10 µl enzyme mixtures were prepared (1 µl 10 x B buffer, 3 µl Spe 1 or Tru 91 enzyme (10 U/µl), 0.5 µl 1 M $MgCl_2$ and 5.5 µl $H_2O$). For each 10 µl of the PCR reaction*

*amplified using Spe primer, 3 μl of the Spe 1 enzyme mixture was added. A similar process was used for Tru amplified PCR products using a Tru 91 enzyme mixture. Digestion was performed at 37℃ (Spe I) and 65℃ (Tru 91) for at least 3 hours. The digestion products were electrophoresed using 3% Metaphor agarose gel and TAE/EB buffer. The undigested DNA fragment is 147 bp and the cut product gives a 120 bp band. As shown, for the DRI A/A animal the Spe product (Spe/A2) was digested completely by Spe I but its Tru product (Tru/A2) was not cut by Tru 91; The A/B heterozygote animal product was cut by both Spe I and Tru 91 (for Spe/A2 and Tru/A2 respectively); The B/B animal product was cut only by Tru 91 (Tru/A2).*

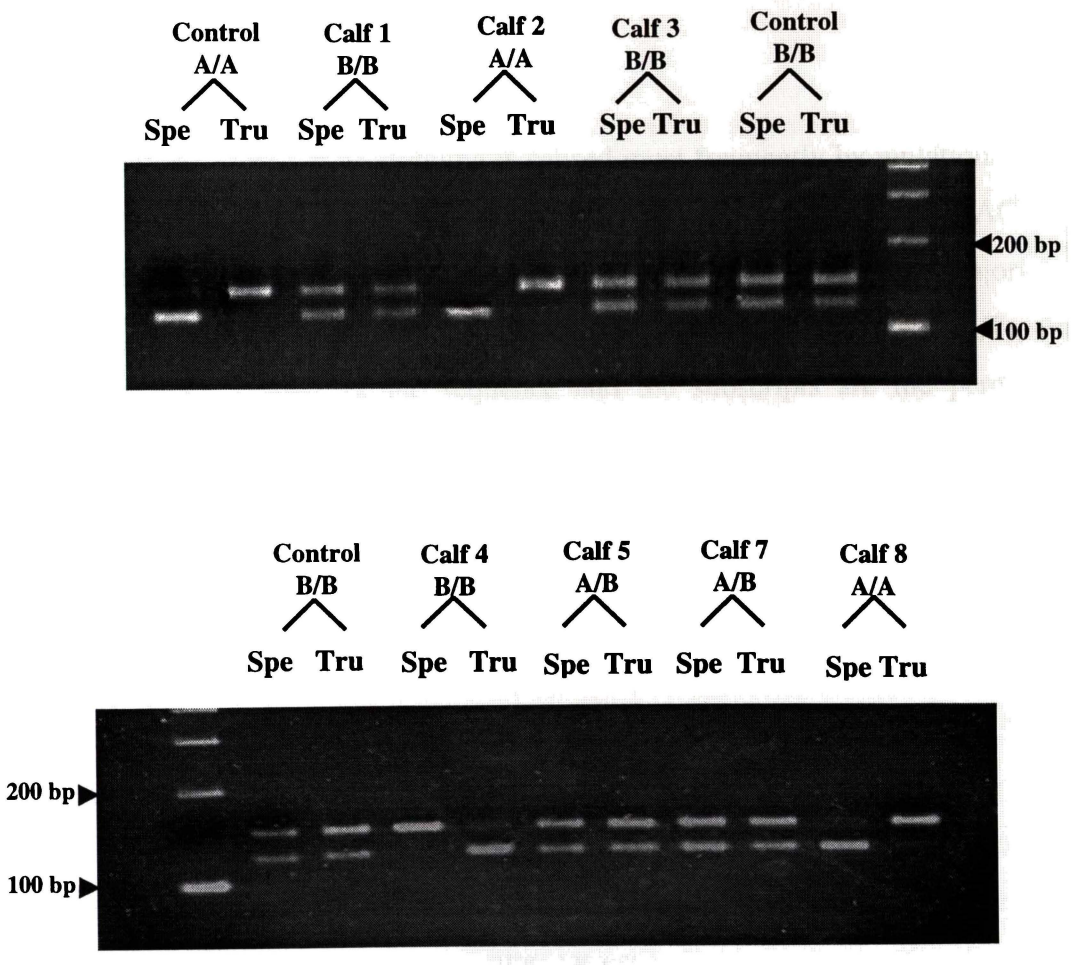



**Figure 3.3.4 Genotyping of $\alpha_{S1}$-CN gene in newborn calves by the two-step ACRS PCR method.**

## 3.4 Discussion

### 3.4.1 The relationship between $\alpha_{S1}$-CN A gene mutation and aberrant splicing

The $\alpha_{S1}$-casein A variant protein clearly arises as a result of the exon 4 skipping during $\alpha_{S1}$-casein RNA processing. At least two distinct gene mutations can give rise to this exon skipping. In the work presented here, a deletion of a single A in intron 4 has the effect of disrupting the 5' consensus splice sequence from the fourth base onwards by putting the distal portion of the consensus out of register by one base. The German Red A mutation (Mohr *et al.* 1994) is much more subtle, involving a T→A transversion at position 6 of this consensus sequence, but Mohr *et al.* also believed this is responsible for the aberrant splicing they observed.

It is well documented that exon skipping is caused most frequently by mutations in the 5' splice site region. The normal sequence of the 5' donor splice site (GTAAGT) in intron 4 of the $\alpha_{S1}$-casein gene perfectly matches the consensus given in the majority of publications e.g. Padgett *et al.* (1986) which quotes $\underline{A}_{62}\underline{G}_{77}/G_{100}T_{100}A_{60}A_{74}G_{84}T_{50}$ (the underlined capitals are the exonic sequence and the subscripts are the percentage frequency of the consensus base). However, the two $\alpha_{S1}$-casein gene mutations documented to date disrupt this consensus to quite different extents. Thus, it is relatively easy to accept that the DRI A mutation could result in exon 4 skipping because the deletion disrupts this consensus in the splice site very strongly, not only by replacing the 4$^{th}$ A with a G, but the 5$^{th}$ base with a T and the 6$^{th}$ with a A. But, why should exon 4 skipping be caused by a transversion mutation at position +6, which even Mohr *et al.* (1994) accept as being weakly consensual and not expected, on *a priori* considerations, to have such a profound effect on exon skipping? They suggest that the 3' splice acceptor preceding the 4$^{th}$ exon is very weak and requires a strong splice donor sequence at the start of intron 4 to ensure correct splicing, i.e., via effective initial binding of U1-snRNP, in order to assist binding of U2AF$^{65}$ to that upstream acceptor signal. Thus, they argue that even a mutation which subtly weakens this receptor sequence could cause exon skipping, so that selection pressure may have forced the system into conservation of that donor sequence (in intron 4) up to nt +8. One does

wonder, however, using the same argument, why selection pressure has not strengthened the 3' splice acceptor sequence!

Given the questions that arise from this type of argument, it is worth making simple comparisons between the acceptor and donor sites flanking exon 4 of the German Red A allele with the other splice sites in the bovine $\alpha_{S1}$-casein gene. Clearly, when one makes these comparisons (Table 3.4.1) one cannot convincingly argue that the 3' acceptor splice site of exon 4 is weaker than others in the $\alpha_{S1}$-casein gene. Moreover, if one examines a number of acceptor and donor site pairs, e.g. those flanking exons 11, 12 and 14, these are even weaker than the mutated splice site of the German Red A exon 4 allele but do not result in obvious exon skipping (though one does wonder if careful studies might reveal some skipping).

Although it is instructive to make comparisons such as those shown in Table 3.4.1, one must bear in mind that, besides the splice sites themselves, a number of other *cis*- and non *cis*-element factors (elements) also contribute to correct splicing. Thus, it is not feasible to accurately evaluate the strength of a given splice site by simply comparing it with the consensus sequence.

Numerous sophisticated methods have been proposed for analysing and making predictions about consensus sequences. For donor splice sites, methods have included weight (consensus) matrices (Shapiro and Senapathy 1987), discrimination energy (Penotti 1991), perceptron (Nakata *et al.* 1985), κ-tuple statistics (Solovyev *et al.* 1994), logitlinear models (Kleffe *et al.* 1996), and division into subclasses (Kudo *et al.* 1992). All these methods assume that nucleotide positions in the consensus sites are independent and fail to reveal the influence of more distant correlations (Zhang and Marr 1993; Rogozin and Milanesi 1997). These correlations in the site positions seem quite promising for revealing either three-dimensional structures of nucleic acid or interactions between biological macromolecules. More recently, Rogozin and Milanesi (1997) reported their new donor splice site consensus sequence – AG/GT(A,G)NNN, NG/GTNAGN, NN/GT(A,G)AGN, NN/GT(A,G)NGT, and NG/GTANNN, obtained by applying a new algorithm which considered the correlations between the nucleotide composition of the sites. These new consensus sequences reflect the different

nucleotide composition combinations (both exonic and intronic nucleotide) at exon/intron junctions and are thought more accurate than those achieved by other methods. Interestingly, in contrast to the slight disruption of the classical consensus, though at a very weak consensual +6 position, which is argued by Mohr *et al.* as being causative of the German Red A exon skipping, the application of those newer consensual analysis rules i.e. /GT(A,G)AG would not predict exon skipping though obviously the DRI mutation would still be predicted to be disruptive.

**Table 3.4.1 Variations from consensus in the splice sites of the bovine $\alpha_{S1}$-casein gene**

| *3'acceptor splice site* | *Location* | *5'donor splice site* |
|---|---|---|
| Consensus<br>YYYYYYYYYYNCAG/*G* | | Consensus<br>GT(A/G)AGT |
| **a**.....**a**..**t**../*A* | $\alpha_{S1}$-CN Exon 2 | ..... |
| .....**a**..../*A* | $\alpha_{S1}$-CN Exon 3 | ..... |
| ...**aa**...../*G* | $\alpha_{S1}$-CN Exon 4 | ..... |
| .....**t**../*C* | $\alpha_{S1}$-CN Exon 5 | .....**a** |
| .....**g**.....**a**../*G* | $\alpha_{S1}$-CN Exon 6 | .....**g** |
| **aa**.....**g**..../*G* | $\alpha_{S1}$-CN Exon 7 | .....**a** |
| .....**a**../*G* | $\alpha_{S1}$-CN Exon 8 | .....**a** |
| .....**t**../*C* | $\alpha_{S1}$-CN Exon 9 | ....**ac** |
| ....**a**.....**a**../*G* | $\alpha_{S1}$-CN Exon 10 | .....**a** |
| ...**gaa**.....**t**../*C* | $\alpha_{S1}$-CN Exon 11 | ....**aa** |
| .....**ga**.**t**../*G* | $\alpha_{S1}$-CN Exon 12 | ....**ta** |
| ....**a**.....**a**../*G* | $\alpha_{S1}$-CN Exon 13 | .....**a** |
| .....**a**..**t**../*G* | $\alpha_{S1}$-CN Exon 14 | ....**ta** |
| ...../*A* | $\alpha_{S1}$-CN Exon 15 | ..... |
| .....**a**../*G* | $\alpha_{S1}$-CN Exon 16 | ....**a**. |
| .....**t**../*C* | $\alpha_{S1}$-CN Exon 17 | ..... |
| ....**g**...../*A* | $\alpha_{S1}$-CN Exon 18 | ..... |
| ...**aa**...../*G* | DRI A Exon 4 | ...**gta** |
| ...**aa**...../*G* | German Red A Exon 4 | .....**a** |

*Italic capital case indicates the exonic consensus nucleotides and regular capital case indicates the intronic consensus nucleotides. The bold lower case indicates the nucleotides that are not in accordance with these consensus sequences. The small dot indicates that the nucleotide is the same as the consensus.*

A number of auxiliary splicing elements, either Intronic Splicing Elements (ISEs) or Exonic Splicing Enhancers (ESEs), play important roles in splice site recognition (Cooper and Mattox 1997). These elements direct the specific recognition of splice sites during constitutive and alternative splicing via regulatory factors (Ryan and

Cooper 1996) or SR proteins (Manley and Tacke 1996). It is reasonable to assume that the diversity (variation) of these elements, plus other unknown *cis* elements, present in the exons and introns would affect the strength of the related splice sites. Thus, it could be that in the German Red A mutation, these *cis* elements provide relatively weak signals and fail to preserve the correct splicing.

Whatever the correct explanation, it would appear certain that $\alpha_{S1}$-casein exon 4 splicing in the normal wild type allele is not a robust part of the overall RNA processing reactions.

### 3.4.2 PCR based testing for $\alpha_{s1}$-casein A mutation

Traditionally, most polymorphisms in milk proteins have been detected by the direct examination of the proteins themselves using either chemical or physical techniques, with the exact nature of the polymorphism being confirmed by sequencing the variant protein. More recently, it has been possible to sequence the actual milk protein genes and to pinpoint the DNA variations which give rise to these polymorphisms.

A variety of procedures have been employed for identifying DNA mutations at previously determined sites. In a surprisingly large number of cases, at least for milk protein genes, restriction fragment length polymorphisms (RFLPs) have been found associated with the nucleic acid sequences coding for the amino acid substitutions in the variant proteins. In such cases it is usually possible to design a simple PCR based mutation detection method which uses the relevant enzyme to digest the PCR fragment and analyse the RFLPs of the digestion products. However, when there is a complete absence of suitable restriction enzyme sites at the DNA mutation site, more complicated methods need to be applied. A number of approaches, such as allele-specific oligonucleotide hybridization (Studencki and Wallace 1984), allele-specific PCR (Okayama *et al.* 1989), allele-specific amplification through ligase chain reactions (Troutt *et al.* 1992), and the methods that create or destroy restriction sites at a point mutation (Davidow 1992) can be used. Generally, for overcoming the problem that occurs at the DRI mutation of $\alpha_{S1}$ casein A variant gene (namely, lack of a restriction enzyme recognition sequence associated with the mutation site), two PCR based

methods, allele-specific amplification (ASP) and amplification created restriction site (ACRS) should be considered first.

Allele-specific amplification (ASP) is a method through use of PCR with primers matching the mutated sequence at their 3' end (Ugozzoli and Wallace 1991). The discrimination of different PCR targets is based on the fact that Taq DNA polymerase lacks a 3'→5' exonuclease activity (Lawyer *et al.* 1989) and that mismatched 3' termini are extended at a lower rate than matched termini. However, not all 3'-terminal mismatches inhibit PCR equally. Kwok *et al.* (1990) demonstrated that in a system designed to tolerate mismatches, 3'-terminal mismatches involving A:G, G:A, C:C and G:G reduced product yield by 100-fold and A:A by 20-fold, whereas all other mismatches had little effect as shown in Table 3.4.2. It is important to note that sequence context can also significantly alter the properties of these mismatches, and as much as a 10-fold difference among sites has been reported (Huang *et al.* 1992). Sometimes, a single-base mismatch at the 3' terminus is insufficient to achieve the desired level of discrimination. An additional mismatch either 1, 2, or 3 bases from the 3' terminus may need to be generated, beside the 3'-terminal mismatch, to gain sufficient discrimination. These technical problems can affect the reliability of allele-specific amplification (ASP) and, in addition, we have found that translating a successful ASP protocol from one themocycler to another can be a problem. For this reason, the ASP method was not chosen for $\alpha_{S1}$-casein A allele detection.

Checking $\alpha_{S1}$-casein A variants by creating an ACRS is slightly more time consuming and expensive than ASP because three PCR reactions and 2 enzyme digestions are required for each sample. But, obviously, it is more reliable and the result can be evaluated accurately. As single-base mismatches can significantly lower the melting temperature of the primer-target duplex, the two ACRS primers (Spe and Tru) used in this study are designed to be slightly longer (27 mer) than normal (Tm 72°C, estimated by % C+G method) to overcome the mismatch while preserving a reasonable annealing temperature. In addition, the mismatched nucleotides were designed at the third position from the 3' end to enable amplification to occur without obvious compromise. As stated before, the other advantage of a longer ACRS primer is that it increases the difference in size of full length and cleaved PCR products.

**Table 3.4.2 Relative amplification efficiencies of 3'-terminal mismatches in the presence of 800 µM dNTP**

| | | *Primer 3' base* | | | |
|---|---|---|---|---|---|
| | | T | C | G | A |
| *Corresponding template base* | T | 1.0 | 1.0 | 1.0 | 1.0 |
| | C | 1.0 | ≤0.01 | 1.0 | 1.0 |
| | G | 1.0 | 1.0 | 1.0 | ≤0.1 |
| | A | 1.0 | 1.0 | ≤0.01 | 0.5 |

*Product yields were normalised to be the perfect match (1.0).*
*Reprinted from Kwok et al. 1990.*

Although it possible to amplify bovine DNA directly using the ACRS and reverse primers to give PCR products, better results were obtained if an upstream forward primer was used in an initial round of PCR and then a second round carried out using the ACRS primer in the nested reaction. This two-step reaction ensured that both $\alpha_{S1}$-casein B and DRI A alleles were equally amplified. In the one-step reaction, this was not always so. Moreover, although both hetero- and homozygotes for the $\alpha_{S1}$-casein DRI A variant could be diagnosed using *Spe* I alone, small amounts of undigestible product were observed in some reactions so it was felt advisable to use the additional *Tru* 91 digestion to confirm absolutely whether or not was present a heterozygote.

A final comment about the ACRS approach is that there is virtually no mutation that we have come across to date, real or predicted, that is not capable of being detected by careful design of an ACRS primer. One has enormous flexibility, because one can juxtapose the primer in the forward or reverse direction and can set it right next to the base under question, or back several bases, and one can also mismatch any one of several bases right up to the penultimate 3' base or, in some cases, even mismatch two bases. Given this range of possibilities, it is virtually always possible to find one, and often two or more restriction enzymes which will provide diagnostic cleavage.

Two concluding comments should be made about other possible mutations in the 5' splice region of intron 4 of the $\alpha_{S1}$-casein gene. It is possible that the other rare variant

A animals which have been reported around the world actually arise from mutations which are different from both the DRI and German Red mutations. Indeed, given the extreme effect on the relatively weak consensual German Red mutation, one would predict that a whole range of mutations from +1 to +6 could result in exon 4 skipping. Unfortunately, it is not possible to design one simple PCR test that could detect all such mutations. The other possibility, which is perhaps of more fundamental than practical interest, is that some mutations may occur in this splice region which do not cause exon skipping (e.g. in positions +3 to +6). It would be feasible to design 4 ACRS primers which would detect all such variations from the B sequence for these 4 bases, but screening large numbers of animals in this way would be laborious and one would have to question the worth of such an investigation!

*CHAPTER IV*

# A VARIANT PROTEIN EXPRESSION BY NORMAL $\alpha_{S1}$-CN GENE

*The identification of the exon 4 skipped product*
*Quantitative analyses of the exon 4 skipping*
*Identification of $\alpha_{S1}$-CN A variant protein in normal milk*
*Discussion*
*The validity of sampling RNAs from milk*
*Methods of quantitative analysis*
*"leaky" exon skipping of the $\alpha_{S1}$-CN gene*

## 4.1 Introduction

The surprising aspect of McKnight and co-workers (1989) work was that they isolated the $\alpha_{S1}$-casein A variant cDNA from a cDNA library of a B/B homologous cow (In this thesis, we will refer to this as a "normal" cow and the common B allele as a "normal" allele because of its predominant occurrence in Western dairying herds). They did not really comment on this paradoxical result in their paper. Moreover, in the later work of Mohr *et al.* (1994) the full implications of a very subtle mutation in the German Red A variant (see chapter 3), 6 bases away from the exon–intron boundary, causing exon 4 skipping are not discussed. The clear message from the results of Mohr *et al.* is that the RNA processing signals resulting in the inclusion of exon 4 of the bovine $\alpha_{S1}$-casein mRNA in normal animals must be weak, resulting in marginal splicing. Given the observations in these two earlier papers, it is reasonable to assume that some skipping of exon 4, yielding A variant mRNA might occur, at least at the gene transcription level, in non-mutated "normal" animal. In other words, the normal bovine $\alpha_{S1}$-casein gene may express significant levels of an exon 4 skipped product, the so-called A variant, along with its normal expression product. We will refer to this as "leaky" exon skipping. Whether this leaky expression is restricted to the mRNA or actually results in authentic variant A protein in the milk is one of the possibilities being addressed in this chapter. Such phenomena would be extremely interesting from the basic science point of view and, of course, may have practical significance as the minor milk components may affect the processing properties of milk and could also introduce different peptide sequences into milk products with possible health consequences. This chapter describes experiments which were conducted to detect, first, the predicted $\alpha_{S1}$-casein A variant mRNA product and second any traces of variant A protein in the milk of normal animal ($\alpha_{S1}$-casein B/B or B/C animals).

To look in more detail at the expression of $\alpha_{S1}$-casein gene, it is essential to collect a large number of RNA samples from lactating mammary gland. Normally, RNA samples are obtained from the mammary tissue of lactating cows at slaughter, but obtaining a large number of samples in this way is obviously neither simple nor convenient, is limited to a sub-group of animals and does not allow for repeated sampling. One other possibility is biopsy, but this is a reasonably impractical

procedure. The mammary gland of a lactating cow is relatively easy to biopsy either surgically, or with a specialised tool (Farr *et al.* 1996), but while the latter procedure is preferable, it is still inconvenient and reasonably invasive so that animal welfare and ethics considerations would limit its use for mass or repeated sampling. Moreover, as a very localised region of the gland is sampled and somewhat less than a gram of tissue is recovered and it may not be representative of the other 99.9% plus of the gland which is not sampled.

One alternative approach is to sample exfoliated cells from the mammary gland. While a number of studies have reported the isolation of DNA from the somatic cells in milk (Lipkin *et al.* 1993, Amills *et al.* 1997), there have been relatively few reports concerning RNA recovery (Lindquist 1994). It is well known that secretory cells can be observed in the vacoules of alveoli of healthy lactating cows. The epithelial cells of mammary gland origin can be observed in milk, and that some of these can be cultured (Buehring 1990). Thus one might predict that mammary specific mRNAs which were representative of gene expression levels in the mammary gland, could be recovered from the secretory cells exfoliated in milk, and that by exploiting the very high sensitivity of PCR techniques, these tissue-specific RNA could be applied for PCR-based gene expression studies. What is less clear, is the extent to which degradation of mRNA in stored milk would negate these analyses, and the extent to which other non-mammary specific mRNA from the somatic cells in the milk would complicate the analyses.

While obviously sampling milk does not present the same problem as sampling mammary RNA, there are a number of technical problems in resolving variant forms of $\alpha_{S1}$-casein. The analysis of caseins by SDS-PAGE is problematical (Creen and Pastewka 1976; Basch *et al.* 1985). Preparative separation of caseins can be even more difficult and many methods depend on differential solubilities or selective proteolytic cleavage.

Thus, a number of methods have been developed to preferentially isolate the A variant protein from the milk of $\alpha_{S1}$-casein A heterozygous animals (A/B), e.g. based on differential solubilities in $CaCl_2$ solution (Thompson *et al.* 1967; 1969), or selective degradation of $\alpha_{S1}$-casein B by pepsin or rennet followed by isolation of intact $\alpha_{S1}$-

casein A protein using DEAE-cellulose chromatography (Creamer 1973; Thompson 1966). These methods were successfully used with mixtures containing around 50% of $\alpha_{S1}$-casein A variant protein but it was felt that their use in the present work (with possibly less than 5% $\alpha_{S1}$-casein A) might be problematical.

In the standard laboratory, an attractive approach is fast protein liquid chromatography (FPLC) and high performance liquid chromatography (HPLC) of caseins on strong anion-exchange columns, i.e., Mono Q 50 mm x 5 mm I.D., which is a refinement of DEAE liquid chromatography and yields better resolutions. A typical result obtained by Andrews *et al.* (1985) is shown in Figure 4.1.1.1 with the four major casein components of bovine caseins ($\alpha_{S1}$, $\alpha_{S2}$, β and κ) as well as the minor γ-caseins being generally well separated by FPLC on the Mono Q column. The major resolution problems in this method occur in the $\alpha_S$-casein zones. The principal peak is $\alpha_{S1}$-casein, but there is considerable overlap between $\alpha_{S1}$- and $\alpha_{S0}$-casein. In addition, $\alpha_{S1}$-casein is not always well resolved from the $\alpha_{S2}$-casein.

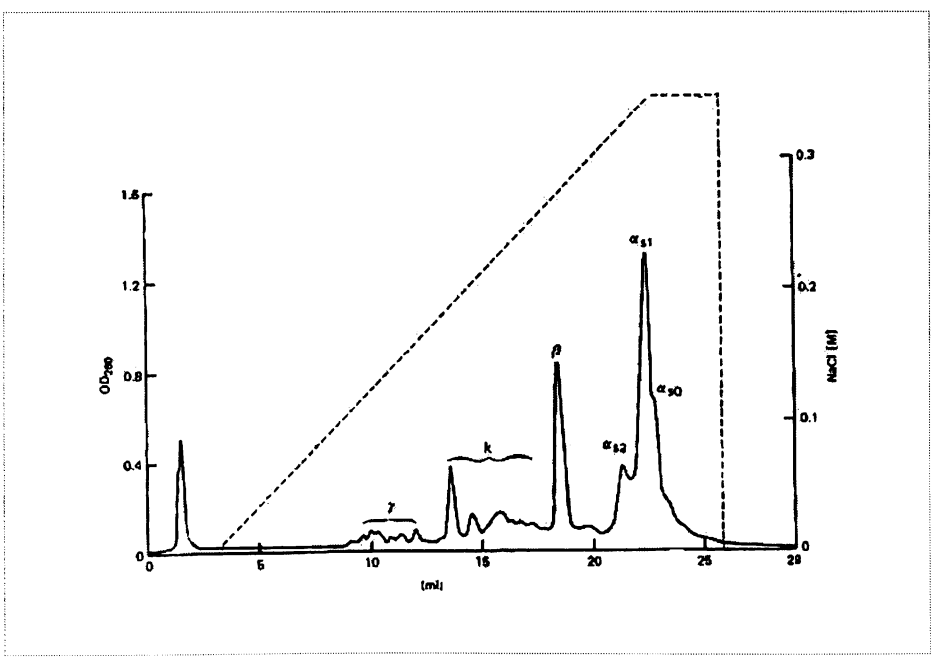



**Figure 4.1.1.1 A typical FPLC separation of caseins on a Mono Q column.** *A 200 μl sample of total casein (10 mg/ml) dissolved in 20 mM Tris-HCl buffer pH7.0 containing 4.5 M urea and 10 mM 2-mercaptoethanol was applied and the column eluted at 1 ml/min using a 0–0.35 M NaCl gradient in this same buffer. Peaks: γ= γ-caseins; κ= κ-caseins; β = β-caseins; $\alpha_{S2}$ = $\alpha_{S2}$-casein; $\alpha_{S1}$ = $\alpha_{S1}$-casein; $\alpha_{S0}$ = $\alpha_{S0}$-casein. (from Andrews et al. 1985).*

## 4.2 The Identification of Exon 4 Skipped Product

Fresh milk samples were collected from individual cows or pooled milk from the herds in the Dairying Research Corporation (Table 2.1.2). Two RNA isolation protocols were designed and applied to these milk samples (detailed in section 2.2.2 of this thesis). Both of the methods were successful and the quality of the RNA recovered from most milk samples was satisfactory for further applications. However, the method 2 which use commercial available solution TRIzol® LS reagent (Life Technologies, Cat No: 10296) to homogenise the cells was preferable because it was less time consuming than using the standard guanidium isothiocynate solutions. It enabled RNA to be obtained from milk samples in just one hour and there appeared to be less debris in the final RNA solution than when using Method 1 (section 2.2.1).

### 4.2.1 RT-PCR analysis

The RNA samples prepared from milk were used as templates for cDNA synthesis. The cDNA was prepared as described in section 2.2.3. Primers located downstream of the exon 4 sequence in the $\alpha_{S1}$-casein mRNA were used to reverse transcribe the "milk" RNA, and then this primer and a second forward primer upstream from exon 4 was used in a PCR reaction to obtain RT-PCR products.

Figure 4.2.1.1 shows schematically the position of the primers and the products expected obtained using two primer sets in RT-PCR experiments. The first utilises a reverse primer (A6: 5'CCTCACTTGACGAAATGCTTTCAG 3') located across exon 9 and 10 for cDNA synthesis followed by a PCR amplification with forward primer A7 (5'CCATGAAACTTCTCATCCTTACCTG 3') within exon 2. Thus, one would expect exon 4 to be contained in the amplification product. This would give a predicted PCR product of 255 bp, but, if exon 4 was skipped, the product would only be 216 bp in size.

The second utilises a reverse primer (EXS 5'CTTGCTCAGTTCATTGACCTTCTC3') located in exon 6 for cDNA synthesis followed by a PCR amplification with forward primer A3 (5'ATCCTATCAAGCACCAAGGACTCC3') in exon 3. The full length RT-PCR product should be 116 bp and the 4th exon skipped product should be 77 bp.

The products obtained from a RT-PCR reaction using A6/A7 primer pairs is shown in Figure 4.2.1.2. The predominant product was 255 bp long, which is the expected size for the normal, non-exon-skipped, $\alpha_{S1}$-casein mRNA. However, in all cases a much weaker band of approximately 216 bp length, corresponding to the size of exon 4 skipped $\alpha_{S1}$-casein mRNA, can be observed on the gel, but does not show up clearly on the photograph.

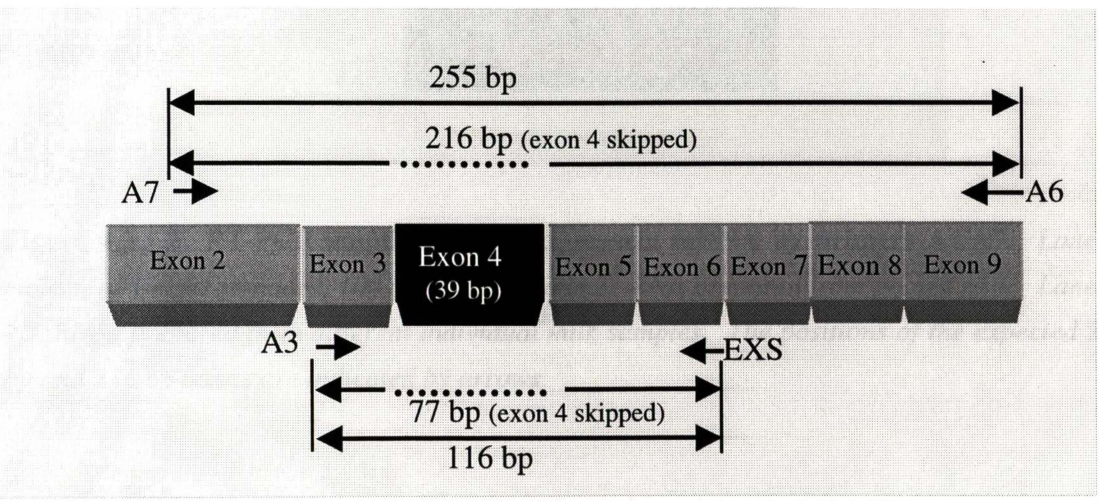



**Figure 4.2.1.1 Schematic diagram showing two RT-PCR reactions capable of detecting $\alpha_{S1}$-casein exon 4 skipping.**

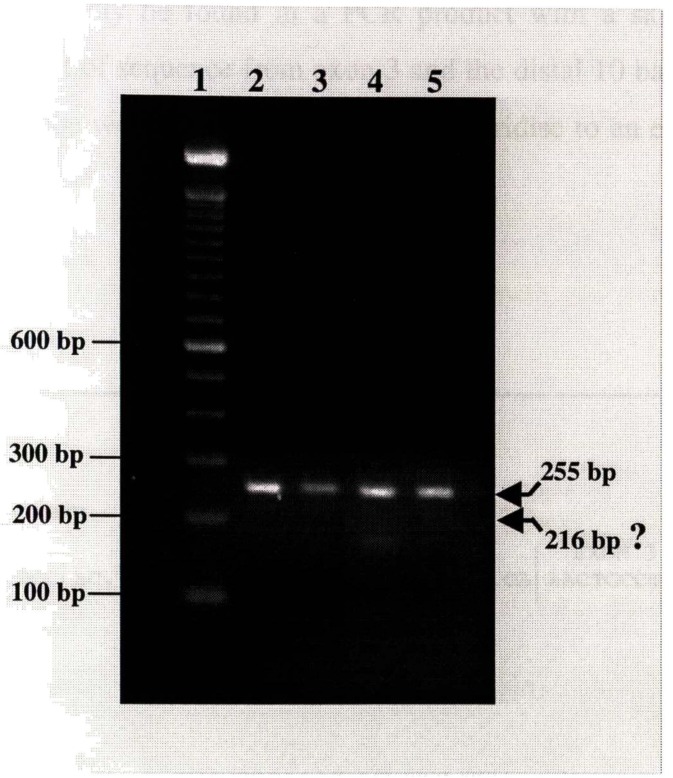



**Figure 4.2.1.2 RT-PCR amplification of $\alpha_{s1}$-casein mRNA by primers A6/A7.** *Lane 1, molecular weight standard, 100 bp ladder; Lane 2, RNA prepared from pooled milk; Lanes 3 - 5, RNAs prepared from different individual milk samples. The positions of the expected 255 bp and 216 bp bands are indicated by arrows.*

## 4.2.2 The results of oligonucleotide hybridisation

The second set of primers, EXS/A3, was used to further examine the weak product. The RT-PCR products of this primer pair would be much shorter (116 and 77 bp for, respectively, the normal full length and exon 4 skipped products) than those created by the A6/A7 primers, thus increasing the relative spacing of the putative "skipped" and normal full length products on agarose gels. For the qualitative analysis of the weak band, the products of RT-PCR were separated on an agarose gel, and then blotted to a Hybond-$N^+$ membrane which was then probed by specially designed oligonucleotides.

The two oligonucleotide probes were designed for the hybridisation (Figure 4.2.2.1). A 22-mer Alpha-B (5'CCTGAGTAAATTTTCATTGAGG 3') which corresponds to sequence within exon 4 was used to identify the normal full length (116 bp) product on the blot.

Alpha-A (5'CTGGAAAAGGTTGAGGGAGT 3') was 20 nucleotides long corresponding to a sequence which would only be found in a PCR product with a skipped exon 4; namely, the proximal 10 nt of sequence from exon 3 and the distal 10 bases from exon 5. Thus the Alpha-A probe would only be expected to hybridise to an exon 4 skipped product.

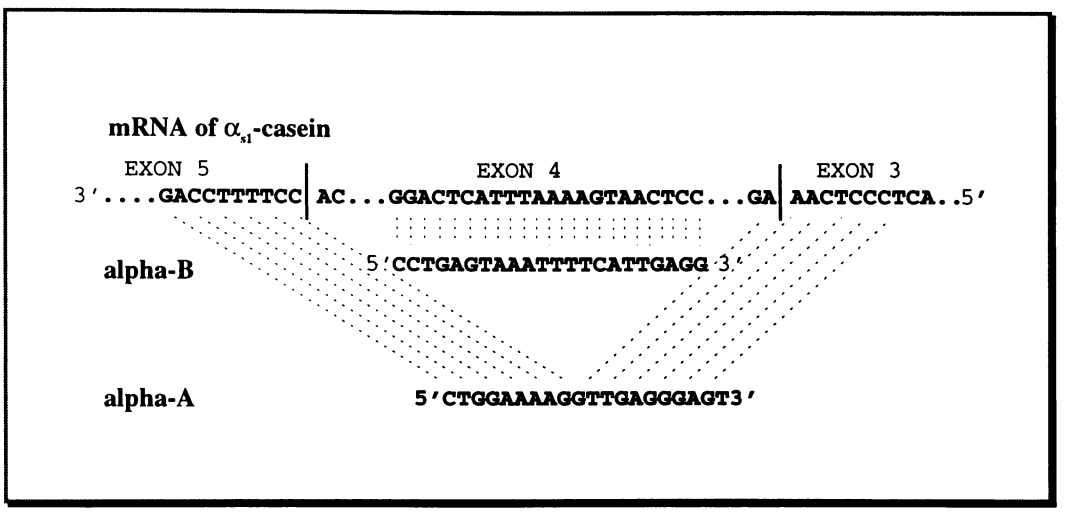



**Figure 4.2.2.1 Oligonucleotide probes designed to detect normal and exon 4 skipped RT-PCR products.** *The alpha B sequence is wholly within exon 4 while alpha A is a chimeric sequence comprised of the distal part of exon3 and proximal part of exon 5. The vertical bars indicate the junction of exons. The dotted line indicates the complementary sequences between the probes and the mRNA of $\alpha_{s1}$-casein.*

The RNA isolated from one pooled and two individual milk samples were used as templates for RT-PCR using the primer pair EXS and A3. The RT-PCR products were loaded in triplicate groups on an agarose gel (Figure 4.2.2.2) so that after electrophoresis and transfer to Hybond-$N^+$, the membrane could be cut to yield three identical blots.

Each of the three membranes were hybridisation with a $\gamma$-$^{33}$P ATP end labelled probe (section 2.3.3), either alpha-B, alpha-A or a mixture of alpha B + A. The autoradiographic result of this hybridisation experiment are shown in Figure 4.2.2.3.

The results reveal that, in spite of some cross hybridisation on all of the three membranes, the alpha-A probe (membrane A) does give a positive signal with the 77 bp band for all three samples (pooled and individual animals). This results suggested strongly that exon 4 skipped mRNA exists in all the animals, albeit at a reasonably low levels. Clearly, the hybridisation of the probes is not completely definitive, probably because the hybridisation temperature (37°C) is too low and there are some background problems. Nevertheless, the results are clear cut and, in any case, were confirmed in the next set of experiments using DNA sequencing.

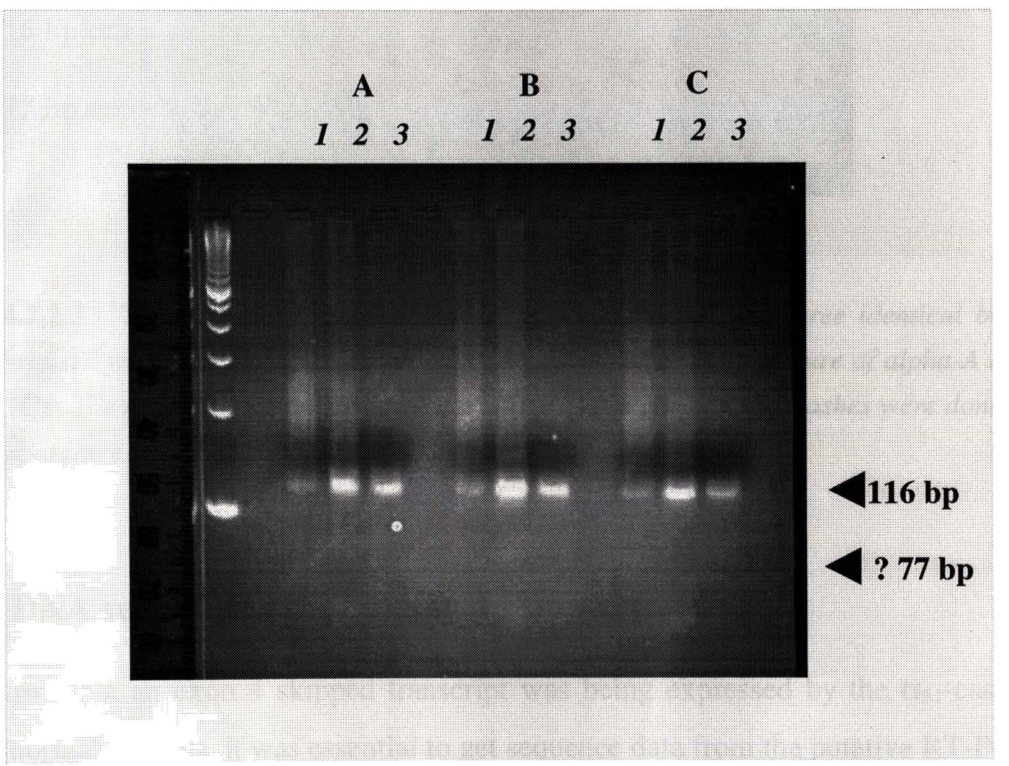



**Figure 4.2.2.2 Electrophoresis of triplicate RT-PCR products for blotting.** *cDNA samples were PCR'd by the EXS/A3 primer pair in a 50 ul reaction volume. The PCR products were loaded onto 4% agarose gel (containing 1% LMP agarose) as three identical group marked **A**, **B** and **C**. 15 µl of each PCR samples were loaded into each well. Lane 1-3, the RT-PCR products from a bulk milk sample and two individual animals. The arrows indicate positions of, respectively, the full length (116 bp) and putative exon 4 skipped product (77 bp).*

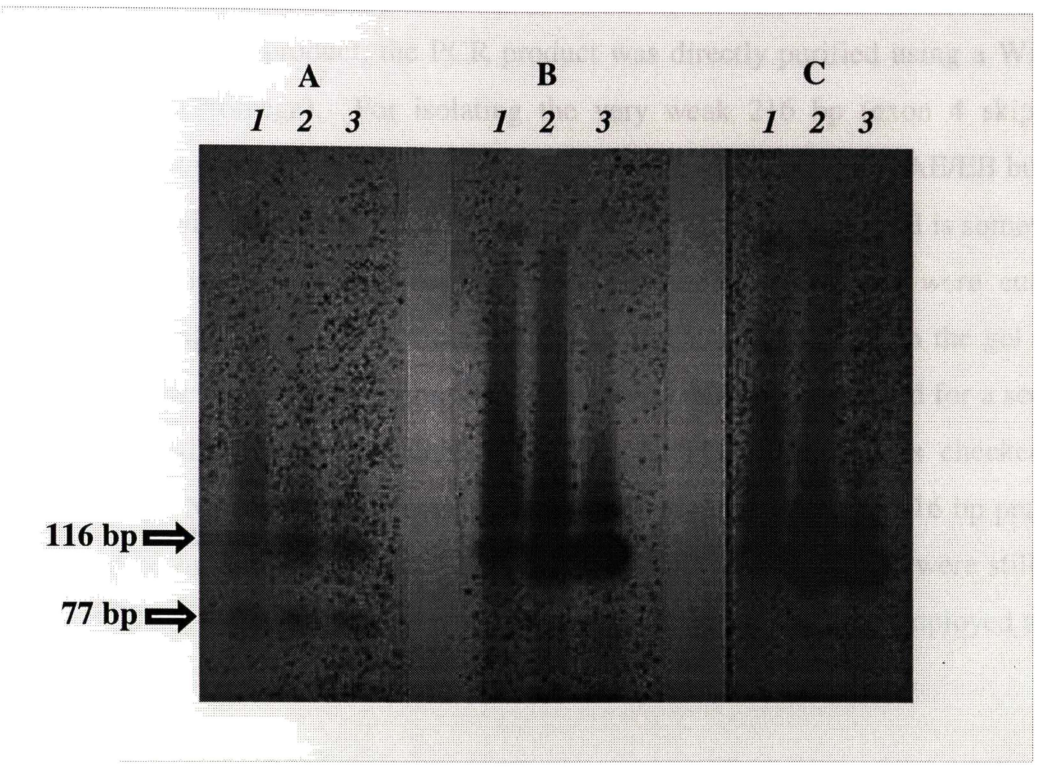



**Figure 4.2.2.3 The results of oligonucleotide hybridisations.** *The three identical blots were hybridised with $\gamma^{33}P$ end labelled alpha-A (**A**), alpha-B (**B**) or a mixture of alpha-A and alpha-B (**C**). The hybridisations were carried at 37 °C overnight. All the washes were done at room temperature.*

## 4.2.3 DNA sequencing of the exon 4 skipped product

To confirm that an exon 4 skipped transcript was being expressed by the $\alpha_{S1}$-casein gene in normal animals, it was essential to get sequence data from the putative RT-PCR product. However, this involved some technical hurdles because the difference between two products is just an internal deletion, so it is not possible to amplify just one of the two products specifically. Moreover, the normal full length RT-PCR product always constitutes at least 95% of the total product and causes the reaction to plateau if one attempts to build up the amount of the minor product by increasing the number of PCR cycles.

Two types of PCR reactions were employed in an attempt to obtain sufficient of the putative exon skipped product for sequence analysis. In the first approach, the RNAs of

normal animals were used for cDNA synthesis using the A6 reverse primer and then PCR'd using the A6/A7 primer pair (as described in section 4.2.1). For sequencing of the full length normal product, the PCR product was directly purified using a Wizard PCR-Preps Kit (Promega). For isolating the very weak 216 bp (exon 4 skipped) product, the PCR products were electrophoresed in a 3% agarose gel in TAE/EB buffer. The bands were visualised on an UV transilluminator. As the 216 bp band is somewhat difficult to localise (too weak), the gel slices encompassing 216 bp were cut off carefully, avoiding contamination from the 255 main band. The DNA in the gel slice was recovered by Wizard PCR-Preps. This recovered DNA was then used for a second PCR reaction using the same primers. The second PCR product was checked by electrophoresis. Normally, this second round of PCR yielded a specific 216 bp product which could be further purified for sequencing (Figure 2.2.3.1). If there were still any trace of a 255 bp, the second-round band isolation procedure would be employed again to obtain a "clean" 216 bp band.

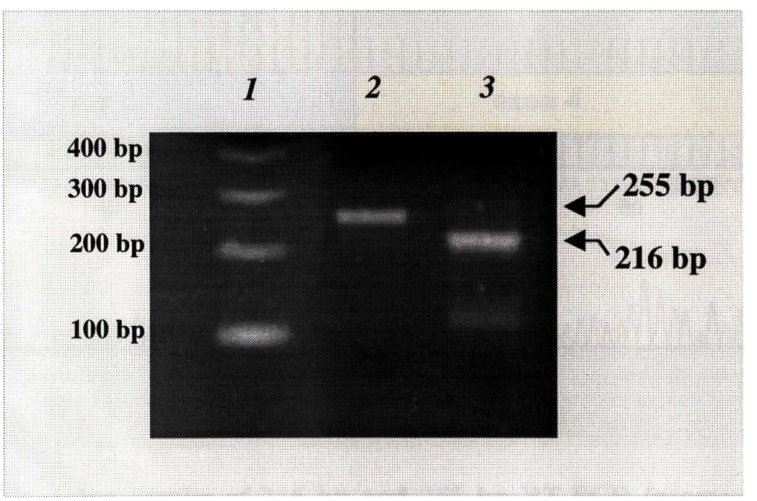



**Figure 2.2.3.1 Isolation of the 216 bp RT-PCR Product.** *Lane 1: Molecular weight standard, 100 bp ladder; Lane 2, the predominant 255 bp band of the first PCR amplification; lane 3, The re-amplified 216 bp product produced from lane 2.*

The 255 bp and 216 bp PCR products that had been purified by Wizard PCR Preps Kit were sequenced on an ABI Prism 377 Automated DNA Sequencer (see section 2.3.7) using both A6 and A7 sequencing primers. The results obtained with A6 are shown in Figure 2.4.3.2. These clearly demonstrate that the 255 bp sequence is identical to its counterpart in Genbank (Genbank Accession X59856) while the 216 bp fragment

matches exactly in the exon 3 and exon 5 regions but has a deleted 39 bases stretch, corresponding to the entire exon 4. These results definitively proved that besides the normal full-length product, the $\alpha_{S1}$-casein gene also, in normal cows, expresses the A variant mRNA albeit at low levels.

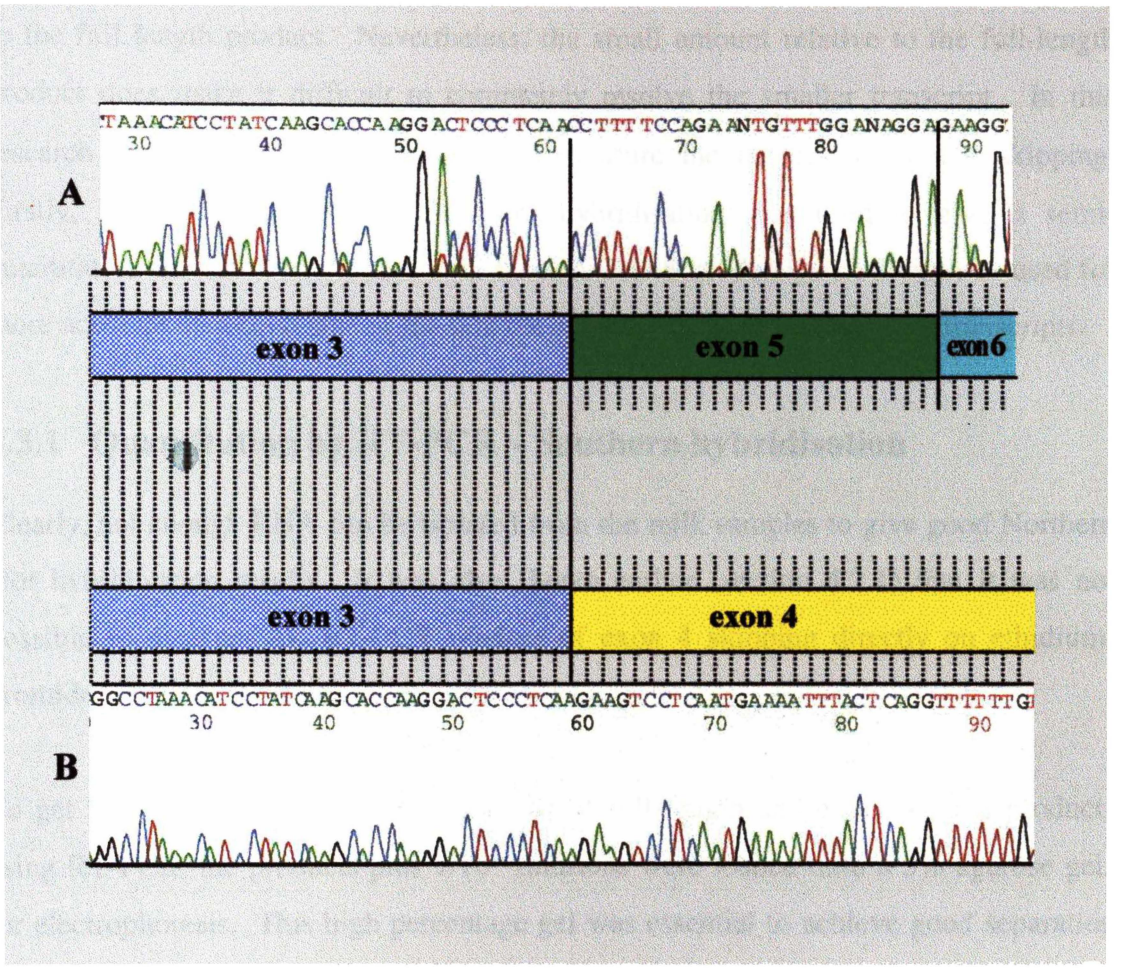



**Figure 4.2.3.2 Sequencing results of 216 bp and 255 bp RT-PCR fragments.** ***A. the sequence of the 216 bp fragment. B. the sequence of the 255 bp full length band. The result indicate that the exon 4 (39 bp) is missing from the 216 bp fragment.***

## 4.3 Quantitative Analyses of the Exon 4 Skipping

The next step after proving the existence of $\alpha_{S1}$-casein exon 4 skipping in normal cows, was to quantitate the relative amounts of full length and skipped transcripts. Generally, quantitation of a specific transcript of a gene in absolute terms, requires an accurate

measurement of that product e.g. on a Northern blot, using RNAse protection assays or quantitative RT-PCR using carefully standardised amounts of sample, internal controls etc. These methods are very error prone and additional errors can also arise when estimating the intensity of the X-ray film by densitometry or of fluorescence from ethidium bromide stained DNA bands. Fortunately, in the case of the $\alpha_{S1}$-casein gene, we are only interested in the level of expression of the exon 4 skipped mRNA relative to the full length product. Nevertheless, the small amount relative to the full-length product does make it difficult to completely resolve the smaller transcript. In this research two methods were employed to measure the $\alpha_{S1}$-casein exon 4 skipping. Firstly, RT-PCR coupled with Southern hybridisation was used to give a semi-quantitative estimate. Later, an assay based on GeneScan Analysis System was used for more accurate measurements of the relative expression levels of these two transcripts.

### 4.3.1 Quantitation by RT-PCR – Southern hybridisation

Clearly, not enough RNA can be isolated from the milk samples to give good Northern blot hybridisation results. It was also shown earlier (section 4.2.1) that it was not possible to measure the RT-PCR product of exon 4 skipping directly on ethidium-bromide stained agarose gel.

To get an estimate of the relative amounts of full length and exon skipped products using RT-PCR, the products plus $1/10^{th}$ dilutions were loaded onto a 3% agarose gels for electrophoresis. This high percentage gel was essential to achieve good separation of the 255 and 216 bp products; however, very poor capillary transfer to Hybond $N^+$ membrane was obtained, so it was found necessary to electrophorese the samples approximately half way down the gel then from this portion of the gel electrophorese the products out into a 1% gel so that they could be efficiently blotted! The membrane was probed with a full length 255 bp probe, and then the hybridisation intensities of the two species compared using a laser densitometer (LKB 2202 Ultroscan).

The results of this experiment (Figure 4.3.1.1) show a strong 255 bp product in all pooled and individual samples which has $1/10^{th}$ the intensity in diluted samples. By comparing the intensity of this band with the 216 bp band in the undiluted samples, it is

clear, in all cases, that the amount of 216 bp product relative to 255 bp non-skipped product is somewhat less than 10%. The exact fraction of 216 bp product varied somewhat from sample to sample, and was generally around the 1% – 5% level. Possibly, the signal strength from the 216 bp band is slightly depressed due to mismatches with the full length probe.

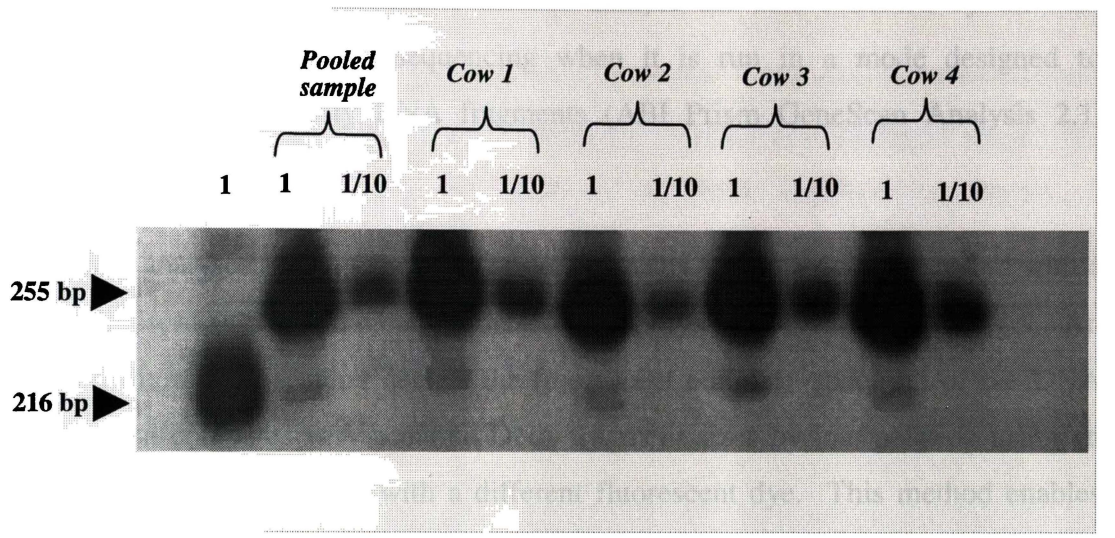



**Figure 4.3.1.1 Southern hybridisation of RT-PCR products amplified by A6/A7 primers.** *Lane 1, purified 216 bp RT-PCR product (previously confirmed by sequencing to correspond to the exon 4 skipped product); Pooled sample and Cow 1 – Cow 4 represent 10 µl RT-PCR products either loaded directly (10ul) or after 1/10 dilution.*

***Electrophoresis*** *The samples were loaded as mentioned above on a 3% agarose gel which was made in TAE buffer. The gel was electrophoresed at 100 volts in a Horizon 11.14 submarine gel apparatus, until the 255 bp band migrated to the distal half of the gel. At this time, electrophoresis was interrupted, the gel, in its holder removed from the buffer tank and placed in a Horizon 11.14 gel former, the first 7 cm (first half) of the 3% gel out of which the amplified products had electrophoresed was cut away parallel to the loading wells, the distal part of the gel slid up to the end, and the same volume of 1% agarose gel was added to the distal end of the gel by pouring in approximately 50 ml of molten 1% agarose in the same buffer. This discontinuous gel was re-placed in the electrophoresis tank and electrophoresis carried on at 100 volts until both the 255 bp and 216 bp DNA products had electrophoresed from the 3% into 1% agarose.*

***Southern hybridisation*** *The post-electrophoresis handings and gel blotting were done as described in section 2.3.5 of this thesis. The probe for hybridisation was made by isolating, on a 1% low melting point agarose gel, the 255 bp DNA which had been amplified by A6/A7 primers. The 255 bp band was purified by Wizard PCR Preps kit and then this DNA was labelled with $^{32}$P-dCTP using a Amersham Rediprime kit. The hybridisation was carried out at*

*60 °C for 16 hours. After hybridisation, the membrane was washed with four changes of 3x SSC and 0.1% SDS and exposed in a cassette to Kodak X-OMAT AR film.*

### 4.3.2 Quantification by GeneScan analysis

The GeneScan™ analysis software is used to analyse the data collected by the ABI Prism™ 377 automated DNA sequencing when it is run in a mode designed to accurately size and quantify DNA fragments (ABI Prism GeneScan Analysis 2.1, User's Manual, 1996).

While being analysed by the GeneScan system, the DNA fragments are labelled with a coloured fluorescent dye and electrophoresed on a polyacrylamide sequencing gel. An automated fluorescent scanner detects the fluorescent emission spectrum of the DNA fragments and computes very accurate DNA fragment sizes by interpolation using an internal size standard labelled with a different fluorescent dye. This method enables precise size calling without the problems often encountered using other techniques, such as band-shift artefacts and run-to-run variation. The GeneScan software also integrates the area under each band peak to give an estimate of relative amounts.

There are two techniques can be used to fluorescently label DNA fragments during the amplification steps of the PCR process: 5'-end-labelling and fluorescent dUTP labelling. In this study the fluorescent dUTP labelling technique was used. This introduced a rhodamine dye [F] into the DNA during the PCR amplification process. This method, rather than end-labelling, was used because the incorporation of multiple fluorophores gives increased signal strength and allows the use of smaller PCR reaction volumes and fewer amplification cycles. This is desirable, because reliable quantitative analyses based on PCR amplification requires that the reaction does not proceed beyond the linear part of the reaction curve. One possible risk, however, is that the PCR product may be detected as double peaks due to different mobilities of the two strands on the denaturing sequencing gel.

In this study, the fluorescent dUTP labelling technique was applied to label the RT-PCR products amplified by the A6/A7 primers. The fluorescent dye used in this

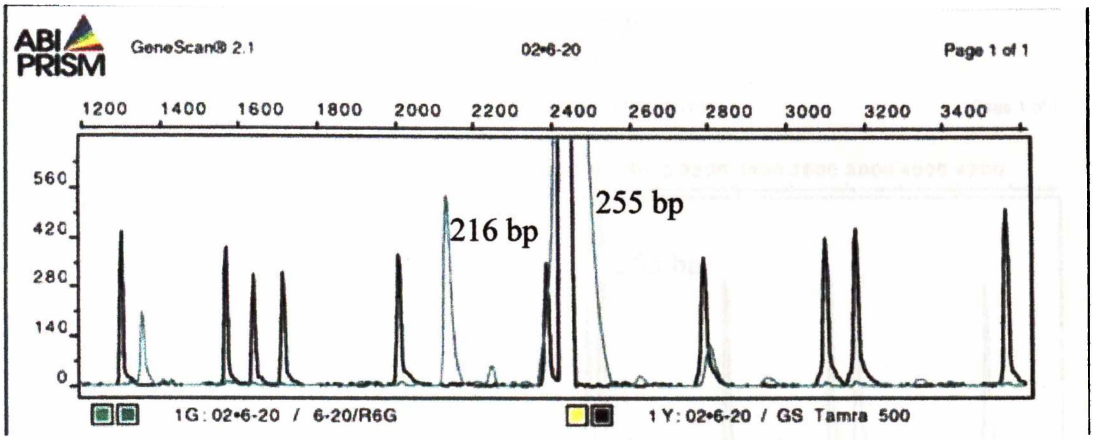
research was a dUTP analogue, 2'-deoxyuridine 5'-triphosphate labelled with rhodamine dye [R6G]. The [R6G]-dUTP was added directly to the PCR reaction. TAMRA labelled GeneScan-500, manufactured by the Perkin-Elmer Corporation, was used as the internal size standard.

Twenty GeneScan analyses in total were done on 15 different milk samples. Parts of the electropherograms are shown in Figure 4.3.2. A–E. These results indicate that the degree of exon 4 skipping varies from sample to sample. The percent of the mRNA exhibiting exon skipping varied from around 1% to 4%, and averaged about 2%. Unfortunately, experiments in which analyses on RNA samples were repeated, indicated that part of this variation depended on the number of cycles used in the PCR reaction, and the absolute size of the minor peak. This probably reflects artefacts arising from non-linear PCR reactions and also the background cut-off value which was used in the GeneScan program – with the minor peak much more prone to over or under estimation that the major peak.

It is noticeable that besides the expected bands of size 255 and 216 nt, extra bands occurred in some samples, e.g. a 192 nt band in Figure 4.3.2 E. The peak area of this 192 nt band was 14298 (fluorescence) units, representing about 7% of the non skipped 255 nt product, which had a peak area of 195,866 units. Although initially these addition bands were thought to be spurious background noise, further investigations indicated that they represent more complicated RNA processing aberrations. These are discussed in chapter V of this thesis.

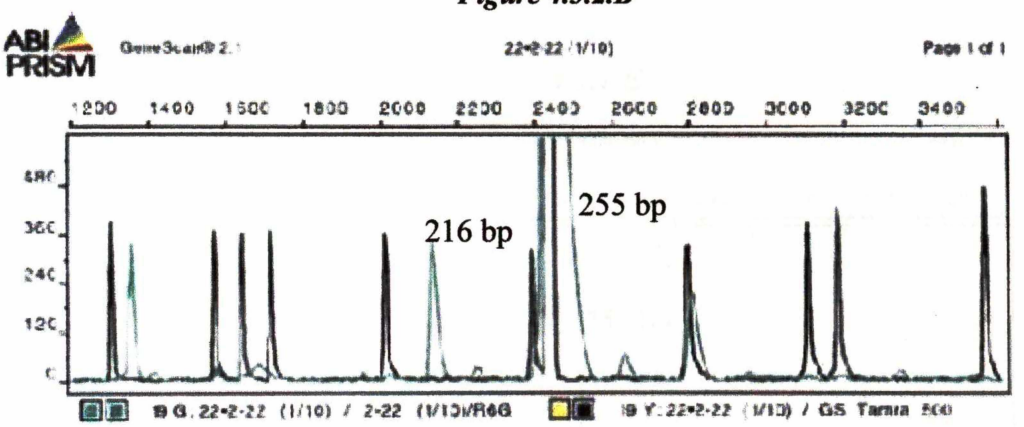
Overall then, the GeneScan analyses confirm the earlier observations using cruder methods (agarose gel electrophoresis and Southern blotting). The quantitation part of the analyses was not particularly robust, but the accurate estimation of band sizes proved invaluable, especially for accurately estimating (to the exact nt) the size of the minor additional bands seen in some samples.

***Figure 4.3.2 A***



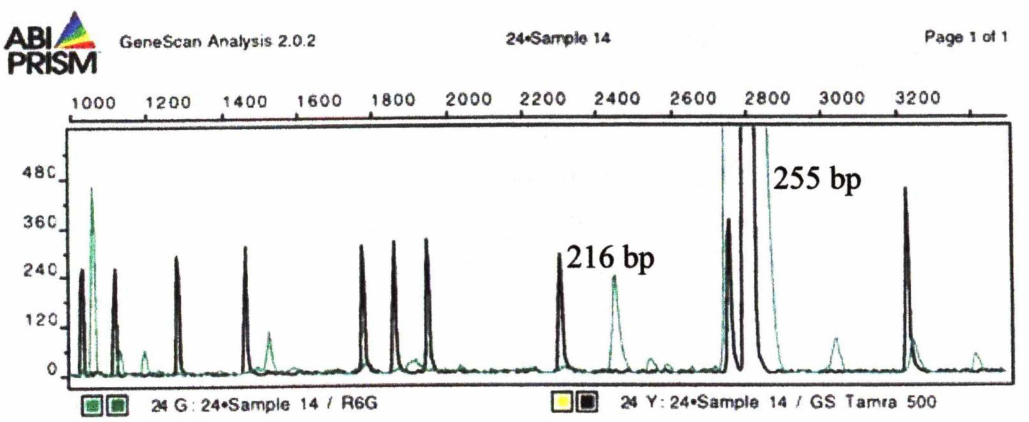

*GeneScan electropherogram sample A. 216 bp peak area is 10147 and 255 bp peak area is 320126. The amount of 216 bp exon 4 skipped transcript is 3% of the 255 bp transcript.*

***Figure 4.3.2.B***



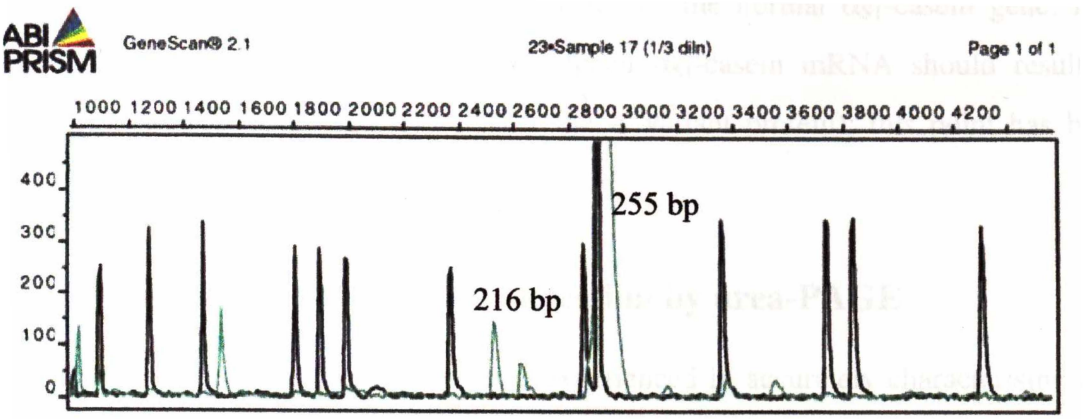

*GeneScan electropherogram sample B. 216 bp peak area is 6815 and 255 bp peak area is 195866. The ratio of 216 bp exon 4 skipping is about 4% of normal non skipped transcript.*

***Figure 4.3.2 C***



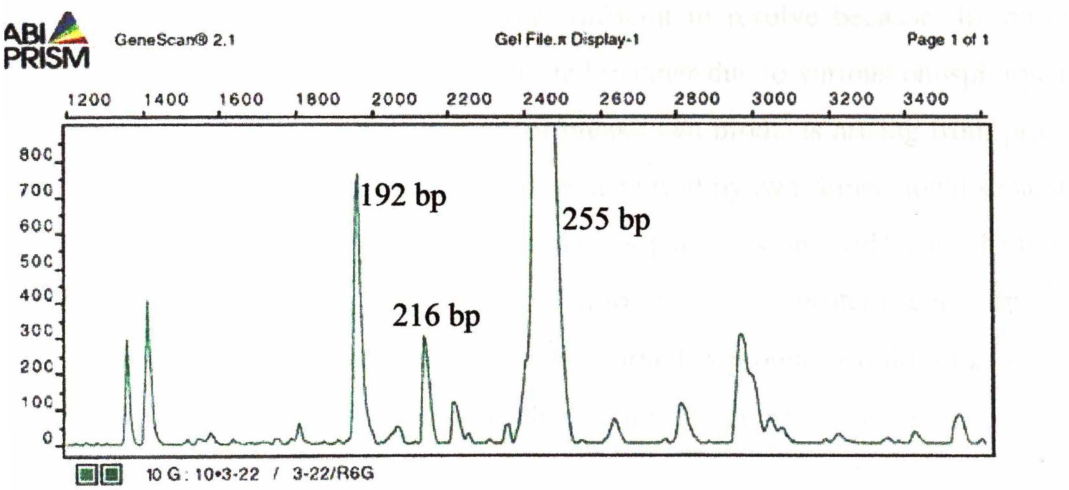

*GeneScan electropherogram sample C. 216 bp peak area is 4479 and 255 bp peak area is 327562. The ratio of 216 bp exon 4 skipping is about 1% of normal non skipped transcript.*

***Figure 4.3.2.D***




*GeneScan electropherogram sample D. 216 bp peak area is 2776 and 255 bp peak area is 144237. The ratio of 216 bp exon 4 skipping is 2% of normal non skipped transcript.*

***Figure 4.3.2 E***




*GeneScan electropherogram sample E. 216 bp peak area is 5224 and 255 bp peak area is 198829. The ratio of 216 bp exon 4 skipping is about 3% of normal non skipped transcript.*

**Figure 4.3.2 A–E Examples of electropherogram of the GeneScan analysis.** *Green peaks are [R6G] labelled PCR product, black peaks are [TAMRA] labelled internal DNA size standard. [R6G]-dUTP was added to PCR reaction mixture to a final concentration of 1 μM without adjusting the concentration of dNTP. The PCR parameters were the same as described before except the cycle numbers decreased from 30 to 16 or 20.*

## 4.4 Identification of $\alpha_{s1}$-CN Variant A Protein in Normal Milk

As $\alpha_{S1}$-casein A mRNA can been transcribed by the normal $\alpha_{S1}$-casein gene, it is reasonable to assume that the exon 4 skipped $\alpha_{S1}$-casein mRNA should result in truncated $\alpha_{S1}$-casein protein product (i.e. the A variant protein); this point has been tested experimentally in this section.

### 4.4.1 $\alpha_{s1}$-CN variant A protein detection by urea-PAGE

Historically, considerable problems were experienced in accurately characterising and resolving the various caseins in milk. Early methods for the analysis and preparation of milk proteins relied upon salt and organic solvent fractionation. For analytical purposes, these gave way to much higher resolution starch gel techniques and then to polyacrylamide gel electrophoresis (PAGE). Nevertheless, minor protein products in milk, especially those for caseins, are very difficult to resolve because the multiple protein species electrophorese in a complicated manner due to various phosphorylation and glycosylation states and the presence of breakdown products arising from protease action. Although the highest resolution can be achieved by two-dimensional separation or isoelectric focusing, simple single-dimension separations are widely used for milk protein analysis, with the exact method tailored to the specific proteins being analysed. Thus, although a minor band of, say $\alpha_{S1}$-casein variant A protein, would, in general, be hard to distinguish from other protein bands, conditions can be chosen which make the task easier.

SDS-PAGE and urea-PAGE have been widely used for casein analysis (Strange *et al.* 1992). Unfortunately, although SDS-PAGE has been applied successfully in a wide range of protein analyses, casein proteins behave somewhat anomalously e.g. retardation of $\alpha_{S1}$-casein relative to the larger β-casein. Despite these problems, SDS-PAGE was attempted is some early experiments as it was felt that a system which sized proteins would be invaluable, especially as the predicted MWs of the exon 4 (variant A) and other predicable truncated proteins could be accurately estimated. However, as earlier reported (Creen and Pastewka 1976; Basch *et al.* 1985), all four caseins appear to migrate in a narrow region near or above 30 kDa (data not shown), despite the fact that their molecular weights are in the range of 19 to 25 kDa. The reasons for these

anomalous mobilities are not completely clear. An explanation has been proposed by Creamer and Richardson (1984) who suggested that both $\alpha_{S1}$- and β-caseins bound the same 1.3 g of SDS per gram of protein, but that $\alpha_{S1}$-casein had an unexpectedly large hydrodynamic size. These kinds of complications become even more of a problem when one is attempting to distinguish variant forms of each casein. Indeed, even when using purified $\alpha_{S1}$-casein, including A, B and C standards, resolution was not achieved (data not shown). Thus, SDS-PAGE was not suitable for these investigations.

Urea-PAGE was chosen in this work because it had been reported to enable the separation of the three common variants of $\alpha_{S1}$-casein A, B and C and was used in early studies to detect rare A variant animals. The addition of urea to the PAGE gel prevents aggregation and precipitation of caseins (if the electrophoresis is at acid pH conditions). Urea is not considered a denaturant because the tertiary structures of caseins are less ordered than typical globular proteins. In urea-PAGE the proteins are negatively or positively charged, depending on buffer conditions (principally pH) and migrate with mobilities related to the charge/mass ratio of each individual protein. Thus, with care, caseins can be well separated and genetic variants and differences in the degree of phosphorylation can be detected. In the case of $\alpha_{S1}$-casein, alkaline pHs (around 8.0) are optimal. One disadvantage with these gels is that the proteins are still in the native form and their molecular weights cannot be determined as simply as in SDS gels (at least in theory!); standards of isolated purified casein(s) must be included in the gel. The standards of $\alpha_{S1}$-CN A, $\alpha_{S1}$-CN B and $\alpha_{S1}$-CN C used in this research were kindly supplied by Dr. Christine Coker of N.Z.D.R.I.

Figure 4.4.1.1 shows that under alkaline urea conditions, the $\alpha_S$-, β- and κ-casein families are well separated. The $\alpha_{S2}$-CN could not be clearly distinguished from $\alpha_{S1}$-CN because of their similar charge/mass rate ($\alpha_{S2}$-CN running slight slower than $\alpha_{S1}$-CN) (Van Hekken and Thompson 1992). However, the $\alpha_{S1}$-CN A variant mobility is distinctly different from that of the normal $\alpha_{S1}$-CN B protein and therefore is readily detectable.

In order to see if traces of $\alpha_{S1}$-casein A variant protein could be detected in normal milk, 40 casein samples from individual milk samples obtained from DRC cows were

isolated by the method described in section 2.2.4. These caseins, along with $\alpha_{S1}$-CN A and B standards, were electrophoresed as described above. The results of urea-PAGE analyses do indeed suggest that the exon 4 skipped $\alpha_{S1}$-casein mRNA is translated into $\alpha_{S1}$-casein A variant protein. Thus, all the casein samples had a weak band at the position of $\alpha_{S1}$-casein A (Figure 4.4.1.2). This weak band constituted a few percent of the total $\alpha_{S1}$-casein protein on the gel, roughly the same as that occurred in the mRNA.

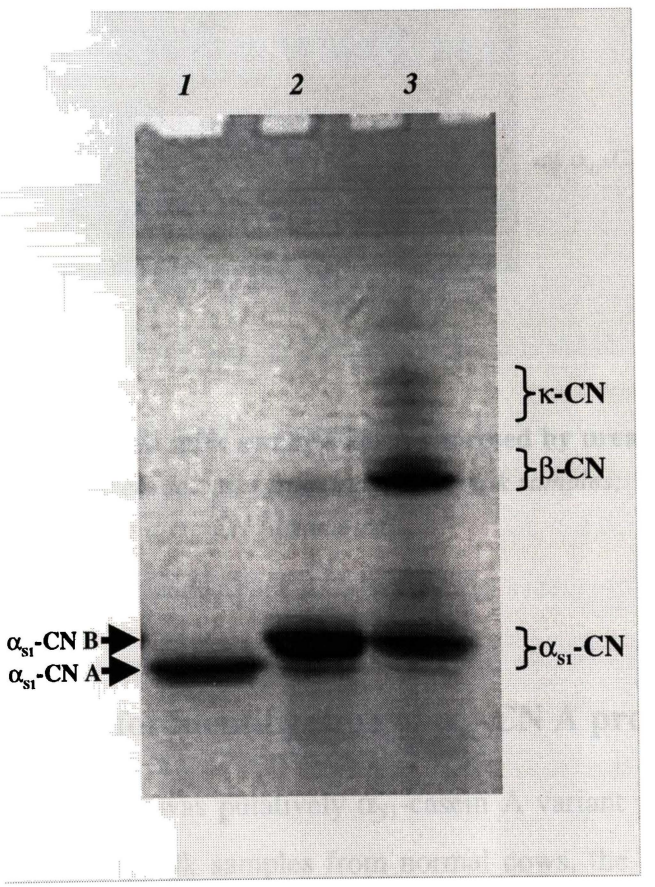



**Figure 4.4.1.1 Urea-PAGE analysis of whole casein and standards.** *Lane 1 $\alpha_{s1}$-CN A standard; Lane 2 $\alpha_{s1}$-CN B standard; Lane 3 Whole caseins isolated from fresh milk. The mini format discontinuous gel (4M urea – 8%PAGE resolving gel) was electrophoresed for 4 hours at 9 mA in Tris-glycine running buffer (pH 8).*

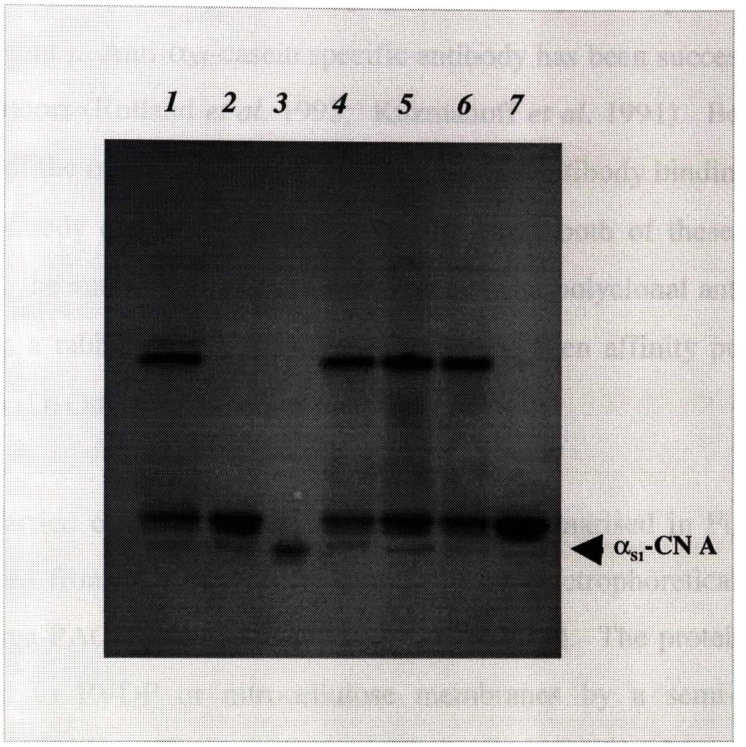



**Figure 4.4.1.2 Caseins from fresh milk samples electophoresed by urea-PAGE.** *Lane 1, lane 4, lane 5 and lane 6 are casein samples from individual milk samples; Lane 2 and lane 7 are $\alpha_{s1}$-CN B standards; Lane 3 is a $\alpha_{s1}$-CN A standard.*

## 4.4.2 Western blotting for identification of $\alpha_{s1}$-CN A protein

Given that a weak band, which was putatively $\alpha_{S1}$-casein A variant protein, could be seen on urea-PAGE in all the milk samples from normal cows, the next step was to confirm that this was indeed $\alpha_{S1}$-casein. As mentioned before, bovine milk protein patterns on gels are very complicated because of the post-translational modifications, proteolysis, etc. so one cannot simply assume that the minor band is $\alpha_{S1}$-casein. Western blotting was conducted to confirm the nature of the band.

Immunological techniques such as Western blotting are very specific and sensitive for the detection of specific proteins. Initially, as the $\alpha_{S1}$-casein variant A specific antibody was not available, it was planned to try to produce an antibody which could be specific to the 12 junction amino acid residues which span (6 on each side) the 13 amino acid region which is missing from the $\alpha_{S1}$-casein A protein. However, the creation of such a

specific anti-$\alpha_{S1}$-casein A antibody is, although possible, risky and expensive (Research Genetics, pers. Comm.). Anti-$\alpha_{S1}$-casein specific antibody has been successfully raised by a number of authors (Rolland *et al.* 1993; Kuzmanoff *et al.* 1991). Because of the same antigenicity of the $\alpha_{S1}$-casein B and A proteins in the antibody binding region, the anti-$\alpha_{S1}$-casein antibody can be used to specifically detect both of these proteins. In these experiments, the antibody, supplied by NZDRI, was a polyclonal antibody which had been raised in a rabbit using purified $\alpha_{S1}$-casein and then affinity purified on an $\alpha_{S1}$-casein column (D. Otter, pers. comm.).

The experiment carried out in this section was briefly summarised in Figure 4.4.2.1. The caseins isolated from individual milk samples were electrophoretically separated on large format urea-PAGE gels (16 x 20 cm, 1.5 mm thick). The proteins in the gel were then blotted to PVDF or nitrocellulose membranes by a semi-dry blotting procedure (section 2.4.2). The membrane was incubated with the anti-$\alpha_{s1}$-casein polyclonal antibody (rabbit) and the bound antibody was recognised by a goat anti-rabbit IgG antibody-peroxidase conjugate. Amersham ECL, a chemi-luminescence system that uses a substrate that emits light after reacting with the peroxidase enzyme, was applied to detect the bound second antibody.

Western blotting results are shown in Figure 4.4.2.2. The first point to make is that the antibody is extremely specific, as the same binding pattern is seen in lanes containing whole casein fractions and those containing purified $\alpha_{S1}$-caseins. In particular, no cross-reaction occurred between the antibody and β- or κ-caseins; possibly there was some affinity for $\alpha_{S2}$-casein as this band has a mobility similar to $\alpha_{S1}$-casein B. It is clear that the putative $\alpha_{S1}$-casein A band which is observed in all the casein samples, is specifically bound by anti-$\alpha_{S1}$-casein antibody. It is most interesting that a similar small component of $\alpha_{S1}$-casein A protein is present in the $\alpha_{S1}$-casein B standard purified by DRI! One would conclude that the casein fractionating techniques such as ion-exchange chromatography on DEAE-cellulose columns used for preparing the $\alpha_{S1}$-casein standard from B variant animals, also co-purifies the small amount $\alpha_{S1}$-casein A that I have demonstrated to be in the milk of B variant animals!

It proved extremely difficult to quantitate the relative expression of the two products as the ECL development conditions which enabled the minor band to be adequately exposed, completely over exposed the major band. Conversely, if conditions were adjusted to give good exposure of the major band, the minor band could not be seen! This problem could not be overcome by altering exposure times or the concentrations of the antibody components (data not shown)

Hence, the Western blotting data confirms that $\alpha_{S1}$-casein A variant protein is indeed expressed at low levels from the $\alpha_{S1}$-casein gene of normal cows. The exact level of expression as a percentage of total expression is difficult to estimate but is similar to the percent of exon 4 skipped mRNA in the mammary gland, about 1-5%.

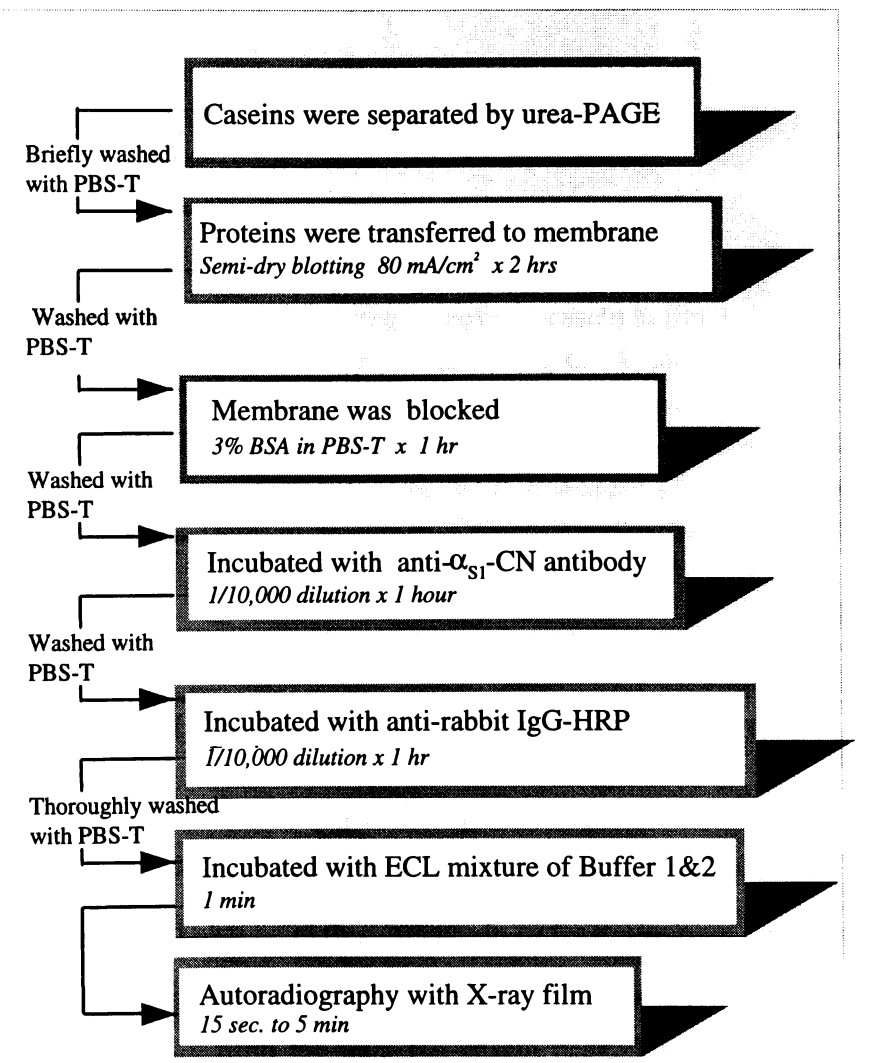



**Figure 4.4.2.1 Flow diagram for ECL immunodetection.**

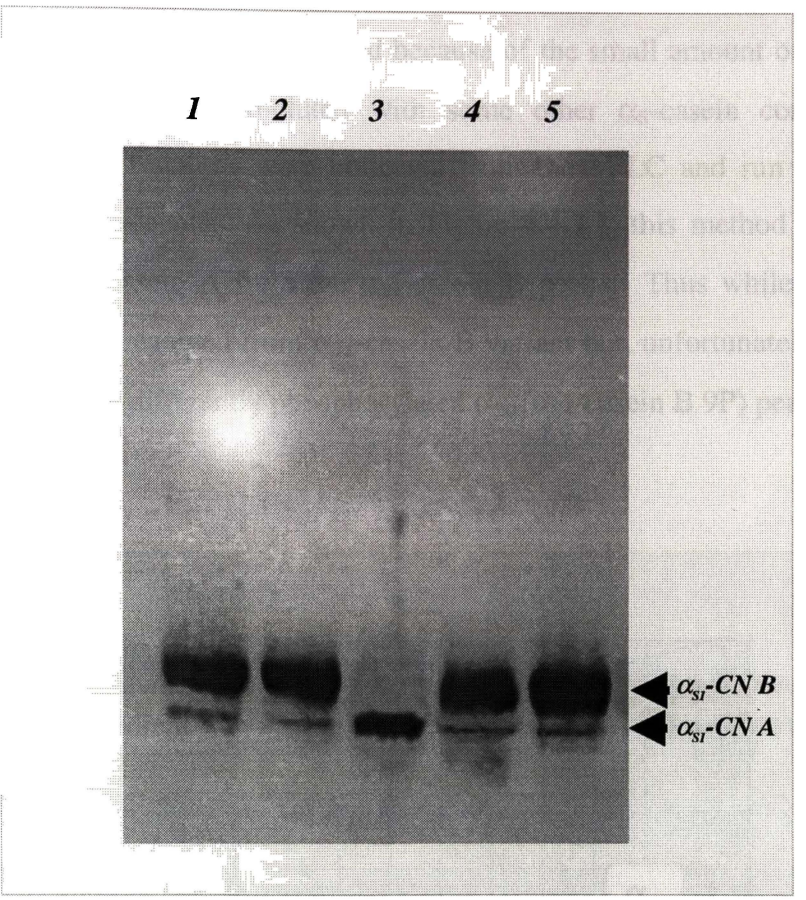



**Figure 4.4.2.2 Western blotting analysis using an anti-$\alpha_{s1}$-casein antibody.** *Lane 1, lane 2 and lane 5: caseins from individual milk samples; lane 3: $\alpha_{s1}$-CN A standard; lane 4: $\alpha_{s1}$-CN B standard. The casein samples were electrophoresed on discontinuous urea-PAGE (16 x 20 cm, 1.5 mm thick) in a BioRad Protean™ II Slab Cell. The electrophoresis was at 20 mA for 16 hours. Proteins were then transferred from the gel to Hybond™-C super supported pure nitrocellulose membrane and detected by an anti-$\alpha_{s1}$-CN polyclonal antibody.*

## 4.4.3 Isolation of $\alpha_{S1}$-CN variant A protein from normal milk

Examination of the standard caseins separation profiles such as Andrews *et al.* (see Figure 4.1.1.1) reveals poor resolution of $\alpha_{S1}$ and $\alpha_{S0}$. One would assume that the $\alpha_{S1}$-casein A variant peak would also be in this region. In order to improve on the results of Andrews *et al.*(1985), a higher concentration (8M) of urea was used to try to separate the minor $\alpha_{S1}$-casein A peak from the bulk $\alpha_{S1}$-casein peak. The sample solvent and eluting buffers both contained urea and a more gradual salt gradient was employed for protein elution. The result is shown in Figure 4.4.3.1. The $\alpha_{S0}$-casein zone was better

resolved from the $\alpha_{S1}$-casein peak but still overlapped partially. Understandably, the $\alpha_{S1}$-casein A protein fraction was not detected because of the small amount of protein, added to which it might have co-eluted with some other $\alpha_S$-casein component. Nevertheless, the protein fractions were collected from the FPLC and run on urea-PAGE to check the components. As shown in Figure 4.4.3.2, this method failed to completely separate $\alpha_{S1}$-casein A from the $\alpha_{S1}$-casein B peaks. Thus while the $\alpha_{S1}$-casein A has been largely resolved from $\alpha_{S1}$-casein B variant but, unfortunately it was not well resolved from the differently phosphorylated $\alpha_{S0}$ ($\alpha_{S1}$-casein B 9P) peak.

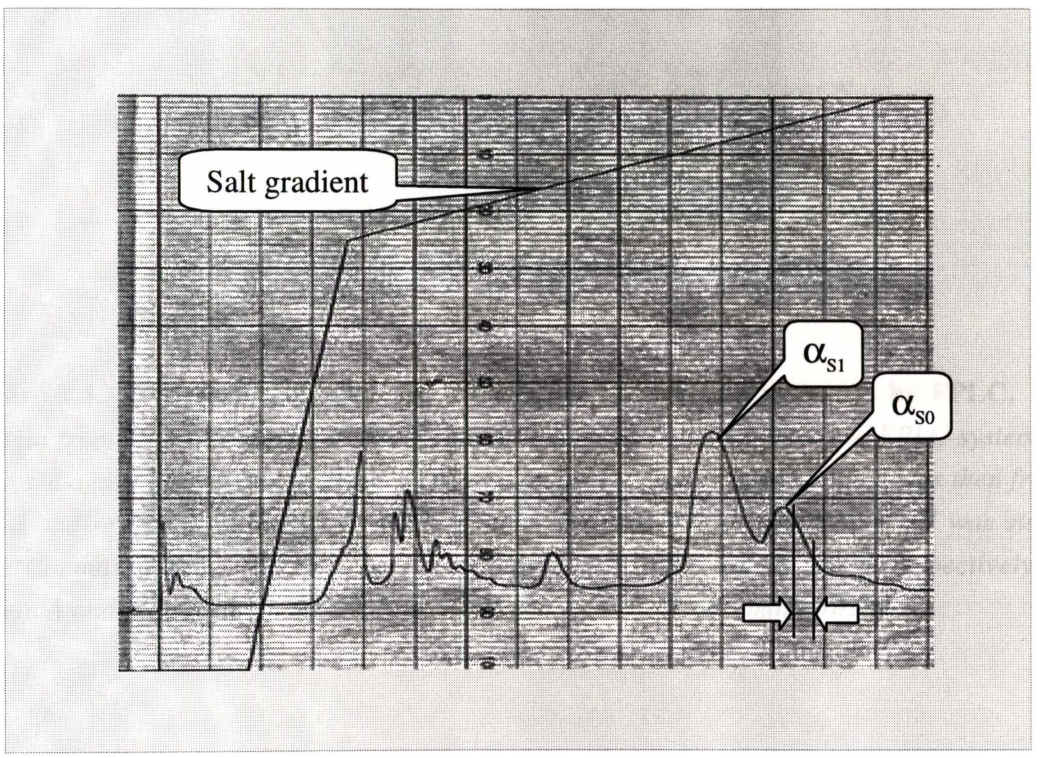



**Figure 4.4.3.1 Bovine caseins separated by FPLC.** *A Mono Q HR 5/5 column was fitted onto a Phamacia liquid chromatography controller LCC-500 system. The casein from normal cows was prepared by acid precipitation and re-dissolved in FPLC buffer I (Tris.Cl 20 mM pH 9, urea 8 M, ME 10 mM) at a concentration about 10 mg/ml. A 500-µl casein sample was loaded onto the column. The flow rate of the mobile phase was 1 ml/min. After a 4 min wash with salt free buffer I, buffer II (buffer I containing 0.35 M NaCl) was increased gradually to constitute 75% of the mobile phase at the end of the next 5 min. A more shallow salt gradient was then applied in the next 20 min increasing the buffer II portion in the mobile phase from 75% to 100%. The protein fractions were collected every half-minute. The white arrows indicate the fraction which was further analysed by urea-PAGE (Figure 4.4.3.2).*

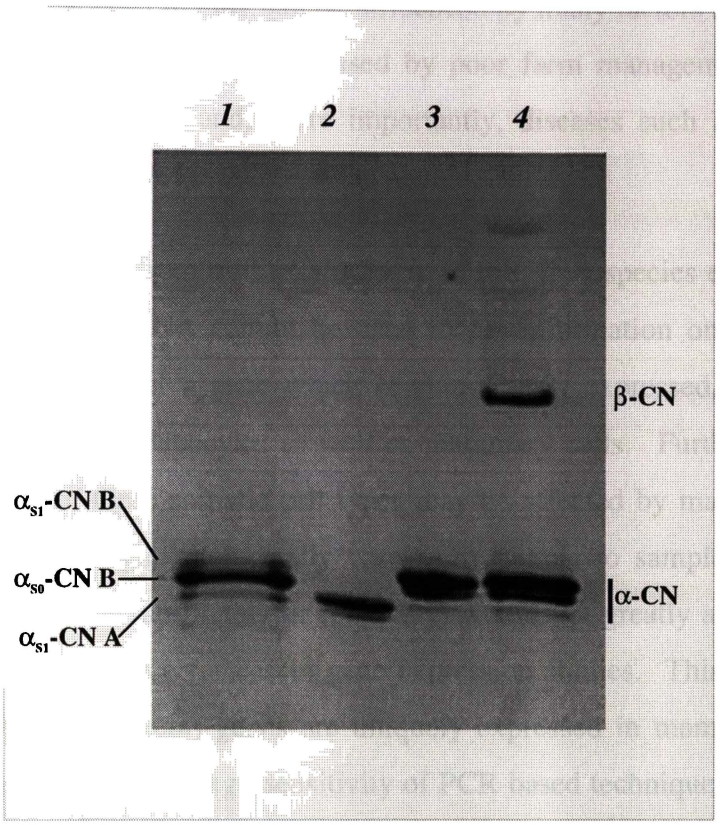



**Figure 4.4.3.2 Urea-PAGE analysis of the $\alpha_{S0}$-casein fraction prepared by FPLC.** *The protein from the trailing part of the $\alpha_{S0}$-casein peak was collected from the FPLC system and dialysed against MQ water to remove the urea and salts. The protein solution was then freeze-dried and re-dissolved in protein loading buffer for urea-PAGE. The PAGE gel was stained with Coomassie Blue. Lanes 1, proteins isolated by FPLC; lane 2 and 3, respectively, $\alpha_{S1}$-casein A and B standards; lane 4, The casein sample before chromatography.*

## 4.5 Discussion

### 4.5.1 The validity of sampling RNAs from milk

Cells from milk are a surprisingly good source of RNA for gene expression studies and useful data can result from relatively simple PT-PCR protocols. It is well known that milk contains low levels (i.e. 5 to 20 x $10^4$ cells/ml) of somatic cells, mainly consisting of lymphocytes, neutrophils (or polymorphonuclear PMN) leucocytes and epithelial cells in the approximate ratio 1:1.5:14 (Kitchen 1981). Obviously, the descriptions of cellular types and the cell counts in milk are far from conclusive. This is partly due to the morphological similarity between many epithelial cells and macrophages and

sometimes the PMN leucocytes. Moreover, the somatic cell count levels and the relative amounts of the cells in milk can be influenced by many factors such as stage of lactation, number of lactations, stress caused by poor farm management, nutritional problems, climatic conditions and, more importantly, diseases such as mastitis and ephemeral fever (Kitchen 1981).

Clearly, the RNA isolated from milk is a mixture of the RNA species expressed by all the cell types. Thus, the RNA cannot be used to get information on the mammary specific expression levels of a gene which is ubiquitously expressed, or even those expressed at low level in leucocytes as well as mammary cells. Furthermore, as the relative proportions of these somatic cell types may be affected by many factors (e.g. mastitis), such RNAs might artefactually vary from sample to sample. Fortunately, such contaminating RNA (absolutely or relatively) would not greatly affect the use of these cells as an RNA source for casein gene expression studies. This is because a) casein (and other milk protein) genes are uniquely expressed in mammary epithelial cells (Tucker 1988) and b) the high sensitivity of PCR based technique only need very small amounts of RNA for analysis. Under most circumstance, the secretory cells exfoliated in milk are sufficient to be used for casein mRNA studies.

However, an important question is, given that we can recover secretory cells or the remnants thereof, how different are these from those in lactating tissue? In other words, how legitimate is it to extrapolate from gene expression in exfoliated cells to that in the mammary gland itself? There is no simple answer to this question but several issues are clearly relevant. First, are secretory cells lost from the organ as a result of some kind of cyto-pathological event such as apoptosis, necrosis or, alternatively, are otherwise healthy cells occasionally exfoliated as some kind of benign phenomenon? Second and related to this, are the cells lost randomly and representatively from throughout the secretory organ? These two questions are somewhat difficult to answer, but at least some of the cells from the bovine mammary gland are viable and can be cultured and many appear intact cytologically (Buehring 1990).

An added uncertainty is whether or not the mammary epithelial cells are predominantly from ductal or alveolar regions. Indeed, histological examination of sections of lactating bovine mammary tissue, reveals occasional epithelial cells detached from

alveoli and "floating" in the vacoules of lactating mammary glands. These cells do not appear grossly abnormal and the hallmarks of involution are not apparent in the surrounding tissue (Molenaar 1995).

The third issue is the stability of mRNA species in exfoliated cells. Experiments with primary sheep mammary cell cultures suggest that, in dispersed cells, the levels of milk gene specific mRNAs persist for at least 8 and perhaps as long as 20 hours (Wheeler *et al.* 1993).

In summary, while it is unlikely that exfoliated cells have identical mRNA levels to those in the mammary gland itself, it is reasonable to assume that they do reflect the general levels of expression. It is also likely that the exfoliated cells have been "sampled" from throughout the mammary gland, something which is difficult to achieve with a biopsy of somewhat less than 1 g of tissue from one region of the gland.

In studies not reported here, we have also carried out experiments aimed at detecting the minor $\alpha_{S1}$-casein exon skipped mRNA in snap frozen mammary tissue from slaughtered lactating cows (supplied by Dr. H. Davey, AgResearch, Ruakura) and also from the cDNA contained in a Stratagene cDNA library generated from a lactating cow. In both cases, the levels of the minor $\alpha_{S1}$-casein exon 4 skipped RT-PCR product were similar to those obtained from "milk" RNA. We have also extended the use of this "milk" RNA to α-lactalbumin, κ-casein and β-lactoglobulin gene expression studies (data not shown). In addition, similar procedures have been successfully used to examine salivary gland gene expression in RNA from saliva samples.

## 4.5.2 Methods of quantitative analysis

### 4.5.2.1 Quantitative RT-PCR

Several techniques, such as a Northern blot, the RNase protection assay (RPA), the reverse transcriptase polymerase chain reaction (RT-PCR) and *in situ* hybridisation are commonly used to measure changes in gene expression. Northern blotting and RNase protection assay are probably the most accurate and precise. Unfortunately, Northern blotting requires significant amounts of undegraded RNA. RPA methods also require reasonable levels of RNA and polymorphic differences between individuals can cause

problems. The amount of RNA obtained from milk is reasonably small, somewhat less than 1 μg per 50ml and somewhat degraded, so the only method that could really be used in these experiments was quantitative RT-PCR.

Many different techniques based on PCR amplification have been utilised to evaluate specific mRNA levels (Wang *et al.* 1989; Hayward-Lester *et al.* 1995; Kohsaka *et al.* 1993; Tsai and Wiltbank 1996). The quantitative competitive RT-PCR (QC-RT-PCR) method has several advantages (Foley *et al.* 1993; Reischl and Kochanowski 1995). It relies on the use of an internal standard that "mimics" or closely imitates the target RNA species with respect to primer binding and other variables affecting PCR amplification. Ideally, the experimental and standard target sequences amplify with the same efficiency but can be distinguished from each other following agarose gel electrophoresis.

The measurement of $\alpha_{S1}$-casein exon 4 skipping, is somewhat akin to the QC-RT-PCR. The sequences of the two RT-PCR products to be measured, namely the 255 bp normal and 216 bp skipped fragments, are identical except for a 39 bp (exon 4) internal deletion. Clearly, the two fragments are able to be amplified by the same primers and one would assume that the amplification efficiencies are similar, but this is not completely certain. The major source of variability usually not controlled for in competitive RT-PCR assays is RNA purity and integrity. Fortunately, this problem does not arise in this study because the two fragments are produced from the same RNA sample, the same cDNA synthesis reaction and the same PCR amplification reaction. So, overall one would assume that conditions are about as optimal as they could be for a QC-RT-PCR reaction and that the ratio of these two RT-PCR products reflects the original ratios of the mRNAs.

The major problem in these experiments is the plateau effect which could be affecting the relative amplification of the two products if too many PCR cycles are used. One would ideally like to use somewhat less than 20 cycles but this would give insufficient product for quantitation on simple agarose gel electrophoresis with ethidium bromide staining.

### 4.5.2.2 GeneScan analysis

GeneScan software sizes and quantitates DNA fragments automatically, allowing faster and more accurate analysis than traditional methods such as radiolabelling. In this study, it provided a simple means for quantitating results from the RT-PCR reactions.

Even so, the GeneScan system had some obvious shortcomings. Although the peak areas could be accurately quantitated, different ratios were obtained in some cases, depending on the number of PCR cycles employed. As stated earlier, this might reflect a plateau effect, but could also arise from inappropriate background subtraction especially in the case of small peaks. The accuracy of the ratios of exon 4 skipping and normal product was also limited. The predominant difficulty arose from the big difference of the absolute amounts of the normal and exon-skipped fragments. Possibly, real time systems such as Taqman (Perkin Elmer) which measure the cumulative product produced at every cycle would have some advantages, but it would not be applicable unless a probe spanning the skipped region could be used.

As mentioned before, there would be some advantages in using end-fluorescently labelled primers but more cycles would have to be used as these are less sensitive than methods incorporating labelled dUTPs (as used here) which give a higher signal intensities from the PCR products.

As [F]-dUTPs method incorporating fluoresent dye-labeled dUTPs during the primer-extension step, all the amplification products contained multiple fluorophores (unlike the one fragment, one fluorophore of the 5'-end-labelling method). Although this produces a greater signal, the number of fluorophores might differ from one another from fragment to fragment and as the fluorophores affect DNA mobility to some degree, the fragment peaks tends to be wider and peak area can be overestimated (ABI Prim GeneScan Analysis 2.1 User's Manual).

Perhaps more importantly, the exon 4 skipped fragment is 39 bp shorter than the 255 bp normal product. The number of T in 216 bp exon 4 skipped fragment is ~21% (14/66) less than that in 255 bp full-length product. Thus, it is reasonable to predict that a given number of 255 bp fragments will be measured 21% more than the same number of 216 bp exon 4 skipped fragments and this does not count the affects by the wider or even

split peak. Therefore, the ratio of exon 4 skipped product should be actually higher (at least 20% more) than 1-4% estimated directly from the electropherograms.

### 4.5.3 "Leaky" exon skipping of the $\alpha_{s1}$-casein gene

Exon skipping, is the most common pattern of aberrant splicing in mammals including humans; it is observed in the expression patterns of numerous genes including those coding for major milk proteins. The known mutations that cause exon skipping involve either exonic or intronic sequences, mostly, as one would predict, in splice-site consensus regions. However, about 20% of the mutations responsible for the aberrant exon skipping do not occur in, and do not appear to directly affect DNA regions within or near the splice junctions (Nakai and Sakamoto 1994; Krawczak *et al.* 1992).

In the major milk protein families, it is well known that mutations within the splice site consensus sequences of bovine $\alpha_{S1}$-, $\alpha_{S2}$- and even (human) β-casein genes cause related exons to be out-skipped completely during the expression of these gene (Table 4.5.3.1). Interestingly, it is also reported that partial exon skipping may occur, i.e. in the transcripts of $\alpha_S$ casein gene families in a number of species. Partial exon skipping has been found in goat, ovine. pig and human $\alpha_{S1}$-casein pre-mRNA transcripts, as well as in sheep $\alpha_{S2}$-casein pre-mRNA transcripts (Alexander and Beattie 1992; Brignon *et al.* 1990; Boisnard *et al.* 1991; Ferranti *et al.* 1997; Ferranti *et al.* 1995; Johnsen *et al.* 1995). Generally, in most of these cases, the exon skipped mRNAs account for a considerable proportion, as much as 50%, of the total mRNA product of the genes. In addition, the protein products of part of these alternative spliced mRNA forms have also been identified in caprine and ovine milk (Ferranti *et al.* 1995 & 1997).

The *cis*-elements responsible for the partial exon skipping of these genes are yet to be examined because (excluding the case of the goat $\alpha_{S1}$-casein gene) the genomic organisation of these genes are not available. The goat $\alpha_{S1}$-casein gene expresses multiple forms of $\alpha_{S1}$-casein peptide chains through exon skipping events involving numbers of exons. A single nucleotide deletion in exon 9 and two insertions, of 11 and 3 bases in length, in the down-stream intron, were identified as mutations potentially related to the exons skipping events in variant F of this gene (Leroux *et al.* 1992).

However, much more analysis is required to elucidate the complicated splicing process for this or other genetic variants of this gene.

For bovine, the $\alpha_{S1}$-casein proteins and gene are the most thoroughly studied. However, the leaky exon skipping described here has not been documented before, except indirectly in the McKnight *et als* (1989) paper which described the isolation of bovine $\alpha_{S1}$-casein A variant cDNA from a homozygous B cow! The reason for not discovering this phenomenon earlier in normal animals is simply that the bovine exon 4 skipped mRNA and mature protein product occur in small amounts and would not be easy to definitively identify.

**Table 4.5.3.1 Complete exon skipping in major milk protein genes**

| *species* | Bovine | Bovine | Human |
|---|---|---|---|
| *Gene* | $\alpha_{s1}$-CN A | $\alpha_{s2}$-CN | β-CN |
| *Exon skipping* | Exon 4 | Exon 8 | exon 3 |
| *Mutations possibly related* | Point mutations within 5' splice site | G→T transversion at the last nt of exon 8 | ? |
| *Truncated protein* | variant A | variant D | β-CN |
| *Ref.* | Mohr *et al.* 1994; this thesis | Bouniol *et al.* 1993 | Martin and Leroux 1992 |

The result of this and previous work suggested that partial or leaky exon skipping is likely to be widespread in the genes coding for the $Ca^{2+}$-sensitive caseins. What are the mechanisms that cause such splicing phenomenon to occur in these genes? It seems there is not a simple answer to this question. It is well known these genes have a somewhat close evolution tie, as demonstrated by the alignment analysis of the amino acid sequences, e.g. the identity of ovine $\alpha_{S1}$-CN C is 97.8% to caprine $\alpha_{S1}$-CN B and 89% to bovine $\alpha_{S1}$-CN C. It is reasonable to assume that the homology of the genes themselves is also extensive. Some sequence elements shared by these genes, within or away from the splice sites, may play a role in the occurrence of the partial exon skipping. Nevertheless, to characterise these *cis*-elements more detailed knowledge about the genomic organisation and expression regulation of these genes, as well as the splice mechanisms is needed.

*CHAPTER V*

# SCANNING THE $\alpha_{S1}$-CASEIN GENE FOR OTHER EXON SKIPPING EVENTS AND CRYPTIC MUTATIONS

*Additional anomalous RT-PCR products in the exon 4 region of the $\alpha_{s1}$-CN mRNA*
*$\alpha_{s1}$-Casein exon 17 skipping*
*Screening of the full length coding region*
*Discussion*
*Exon skipping events in the region of exon 4*
*The putative protein from exon 17 skipped mRNA*

## 5.1 Introduction

Traditional methods of mRNA analysis such as Northern analysis do not readily permit the detection of minor mRNA products, especially if they are of similar size (± 10%) of the main product. These problems can be overcome by using RT-PCR, especially if specific regions of the mRNA molecule are subject to analyses. This chapter describes efforts to:

- detect exon skipping in the $\alpha_{S1}$-casein gene in addition to that already described in chapter IV. Examination of the splice region sequences (section 3.4.1) would lead one to predict that leaky exon skipping might occur in other exons.
- predict additional $\alpha_{S1}$-casein protein products that might arise from such aberrant exon skipping.
- scan the other genetic variants of $\alpha_{S1}$-casein that can not be detected by routine protein assays.

Reasonable predictions can be made about the properties of such aberrant proteins based on the predicted size of the protein, on the regions of the casein sequence which are expected to be altered or deleted and the net charge of the remaining amino acid residues etc. In actual fact, even though very accurate predictions can be made about the amino acid sequences of aberrant products, the purification and analyses of such products from the complex protein mixture which constitutes milk is not simple. Nevertheless, such minor components may have important consequences when milk is being processed for specialised uses, e.g. when manufacturing nutraceutical type products in which it may be important only to include certain variants of milk proteins and to avoid peptides with specific amino acid peptide sequences.

## 5.2 Additional Anomalous RT-PCR Products in the Exon 4 Region of the $\alpha_{s_1}$-CN mRNA (Exon 4&5 Skipping)

The GeneScan results indicated that one of the examined animals (Figure 4.3.2 E) consistently produced a minor peak at position 192 nt, in addition to that at 216 nt which is expected for an exon 4 skipped product. Thus, the size of the peak (that is

255-192 = 39+24) suggesting that a 24 nt exon, which commonly occurs in the $\alpha_{S1}$-casein gene, may also be skipped in addition to exon 4 of 39 nt.

### 5.2.1 Sequence analysis of the 192 bp RT-PCR product

In order to sequence the 192 bp RT-PCR product, it was necessary to separate this minor band from the bulk of the 255 bp product using a 3% agarose gel. As the 192 bp band cannot be seen directly on the gel, the gel slice around the 192 band area was excised and the DNA in this gel slice was isolated using a Wizard PCR Preps Kit. The purified DNA was then used as template for the next round of PCR amplification. The product of second round PCR was electrophoresed on a 3% agarose gel. The resulting 192 bp band was stabbed with a pipette tip and again used as a template for another amplification reaction under the same PCR conditions. This re-amplified 192 bp PCR product (Figure 5.2.1.1) was purified by using Wizard PCR Preps Kit before being sequenced on an ABI 377 sequencer.

The sequence obtained from this purified product (Figure 5.2.1.2) does indeed suggest that exon 4&5 skipping has occurred, but the actual sequence in the region of what should be the third to seventh bases of sequence in exon 6 is not clear, suggesting that the 192 bp band actually contains a mixture of products, possibly arising from a heterozygotic animal.

If there were actually two different alleles in the animal, with a mutated region in one of the alleles, one would expect the sequence of the 255 bp full length product to also reflect this heterozygosity. When the 255 bp product was sequenced (Figure 5.2.1.3), no such abnormality was seen, suggesting that the changes seen in the 192 bp fragment are possibly a result of aberrant splicing! Alternatively, it is possible that one allele of the $\alpha_{S1}$-casein gene is not giving rise to any full length (255 bp) product at all.

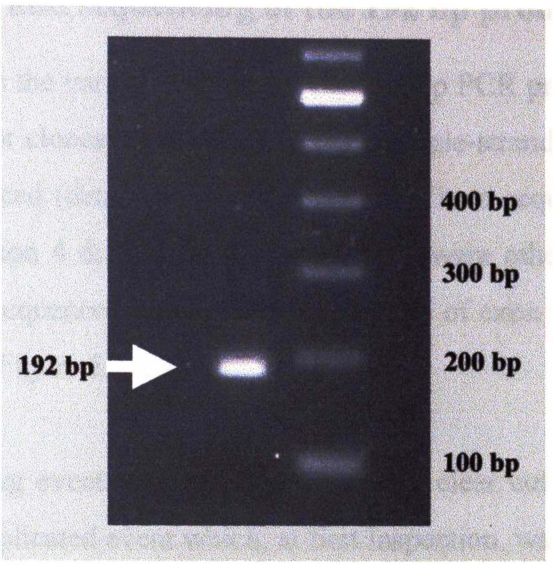



**Figure 5.2.1.1 The isolated 192 bp product.** *The re-amplified product was electrophoresed on a 3% agarose gel to check the amount and specificity of 192 bp production.*

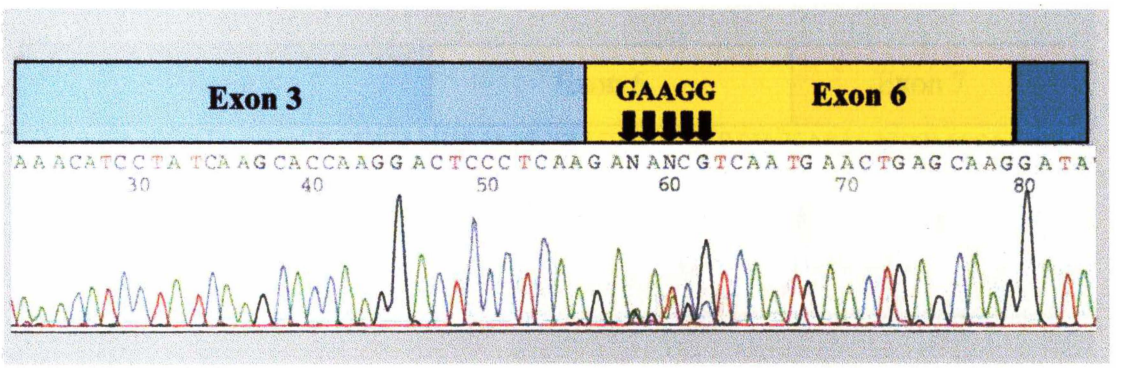



**Figure 5.2.1.2 Sequencing result for the 192 bp product.** *Black arrows indicated those nucleotides in exon 6 that are changed or not clear. The letters above the arrows show the nucleotides of normal exon 6.*

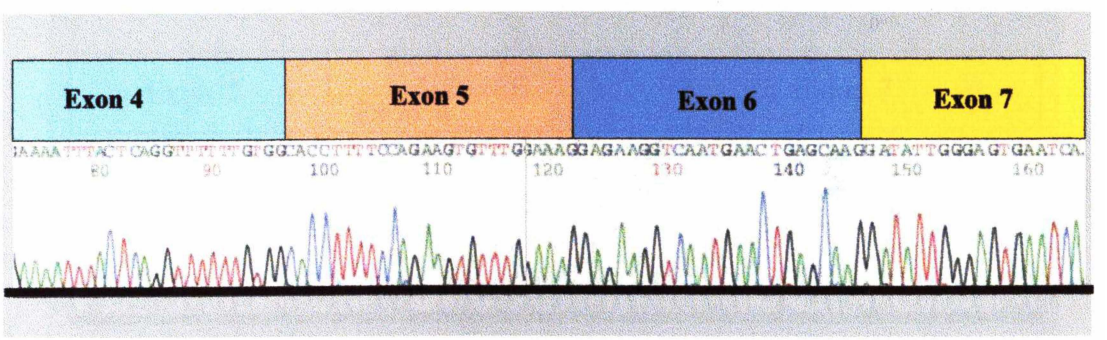



**Figure 5.2.1.3 The sequence of the 255 bp product from the cow with the putative heterozygotic exon 6.**

## 5.2.2 The cloning and sequencing of the 192 bp product

To distinguish between the various sequences, the 192 bp PCR product was cloned into pBluescript II, and six clones were converted into single-stranded phage DNA using VCSM 13 and sequenced (detailed in section 2.3.6). Three sequences were identical, all having complete exon 4 & 5 internal deletions but were otherwise normal (Figure 5.2.2.1). Three other sequences had a complete deletion of exon 5 and partial deletions of exon 4 and exon 6 (Figure 5.2.2.2).

Thus, the exon skipping events seemed to consist of a clear cut exon 4 & 5 skipping event plus a more complicated event which, at first inspection, was not easy to explain.

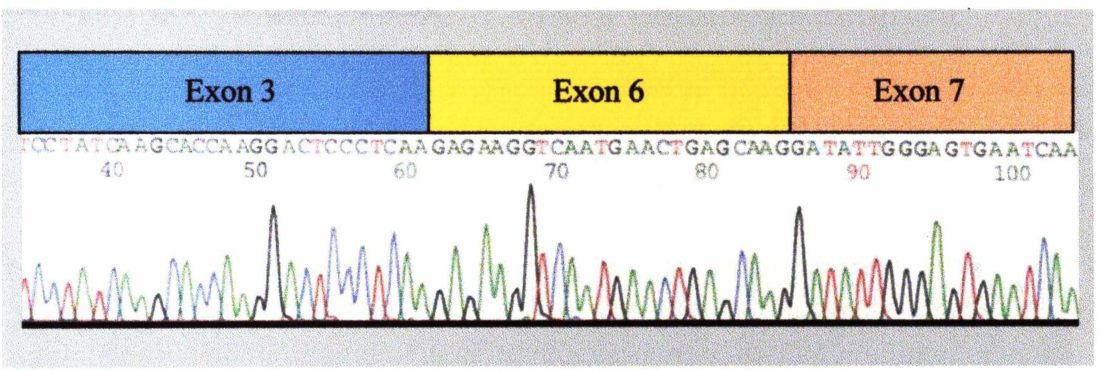



**Figure 5.2.2.1 The sequence of cloned 192 bp product with clean exon 4 and 5 skipping.**

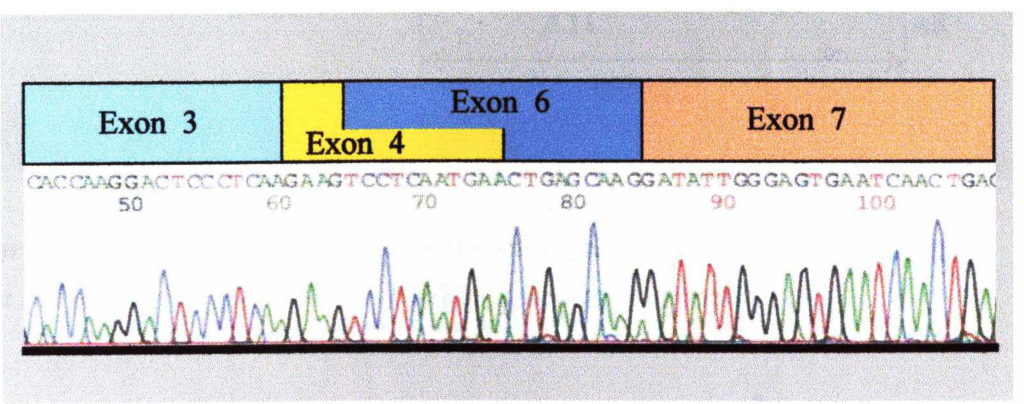



**Figure 5.2.2.2 The sequence of cloned 192 bp product with exon 5 skipping and exon 4 and 6 partial deletion.**

## 5.2.3 DNA sequencing of flanking regions of the genomic DNA

In order to check for any peculiarities in the genomic DNA sequences which might account for the anomalous mRNA sequences, the intron 3 to intron 6 genomic region from this cow was PCR'd and sequenced using primers A1/A11 and A8/A10 (Figure 5.2.3.1 and Figure 5.2.3.2).

No sequence mutations were observed at any of the splice sites within this region. The sequence of the A1/A11 product is identical to the published sequence (Appendix III). However, a one base (A) deletion was seen in intron 5. This deletion is located 97 bp upstream of exon 6 and it is not obvious how it could play a role in the complicated exon skipping events. As this single base deletion occurs on only one allele, a normal sequence and a n-1 sequence is seen after the point of deletion on the sequencing electropherogram (Figure 5.2.3.3).

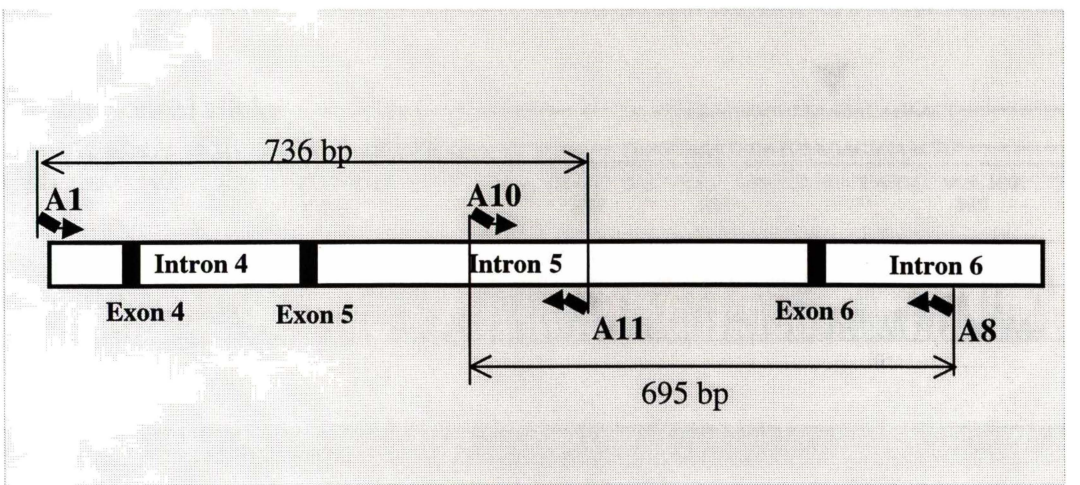



**Figure 5.2.3.1 A schematic diagram of the amplification of exon 4 to exon 6 region.**
*A1 is 55 bp upstream of exon 4, A8 is 167 bp downstream of exon 6.*

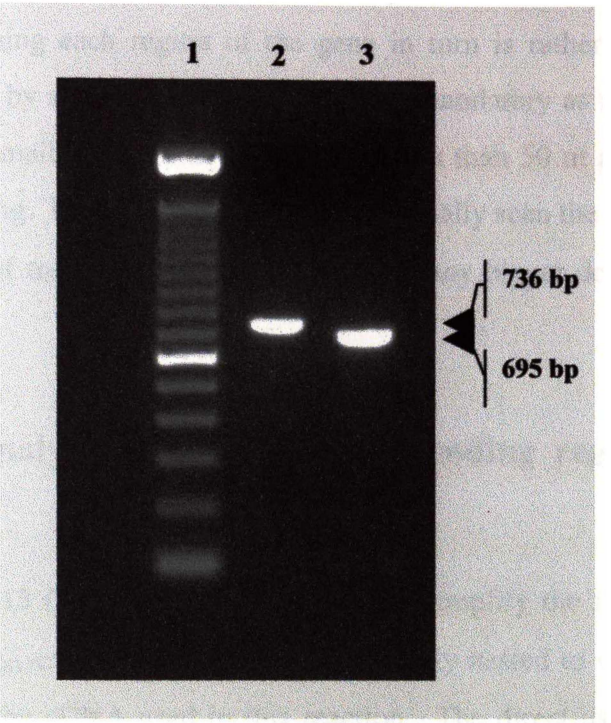



**Figure 5.2.3.2 The PCR products amplified for sequencing.** *Lane 1, 100 bp DNA molecule makers; Lane 2, PCR amplified by A1/A11 primer pair; Lane 3, PCR product of A8/A10 primer pair.*

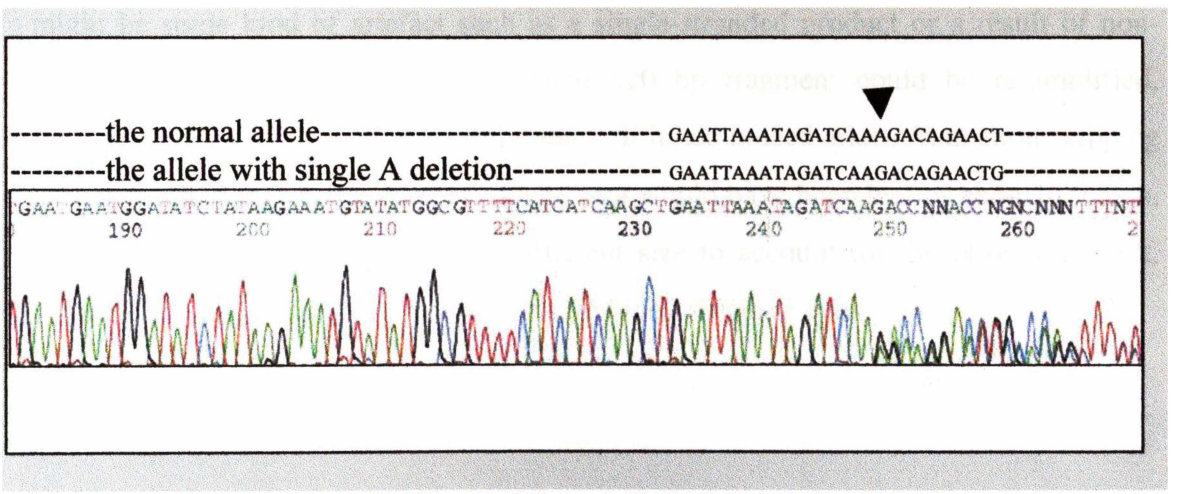



**Figure 5.2.3.3 This sequence of the PCR fragment amplified using A8/A10 primers.**

## 5.3 $\alpha_{S1}$-Casein Exon 17 Skipping

Having shown that a variety of 'leaky" exon skipping events appear to occur in the exon 4 region of the casein gene, albeit at relatively low levels, the next question was whether or not "leaky" exon skipping events occurred elsewhere. The analysis of splice

site sequences (Table 3.4.1) would lead one to predict that this is a reasonable expectation, but scanning each region of the gene in turn is rather laborious. Such scanning, short region by short region is more-or-less mandatory as most of the exons of this gene are very small, 14 of total 19 exons are less than 50 nt and some of them are as short as 24 nt long. It was decided, instead, to initially scan the full length coding region of the mRNA of $\alpha_{S1}$-casein to see if there were any bigger deletions within the $\alpha_{S1}$-casein mRNAs

## 5.3.1 RT-PCR analysis of the full length coding region of $\alpha_{s1}$-CN mRNA

The primer pair A7/A13 (Figure 5.3.1.1) was used to amplify the full length protein coding region of the $\alpha_{S1}$-casein mRNA (A13 is internally nested to the reverse primer A12 used to produce the cDNA used in this reaction. The direct use of A12 did not give a satisfactory RT-PCR product). As showed in Figure 5.3.1.2, in all cows tested, the expected full length product of 676 bp was obtained, but an additional minor band of approximately 520 bp was also visible. It was initially thought that this extra band might be some kind of artefact such as a single-stranded product or a result of non-specific priming. However, the isolated 520 bp fragment could be re-amplified, proving it was not a single-stranded product. If it had indeed arisen from exon skipping in the $\alpha_{S1}$-casein gene, the most obvious event would have been loss of exon 17 which, at 155 nt long is the only exon of sufficient size to account for the observed result. Alternatively, it is possible a multi exon combined skipping event.

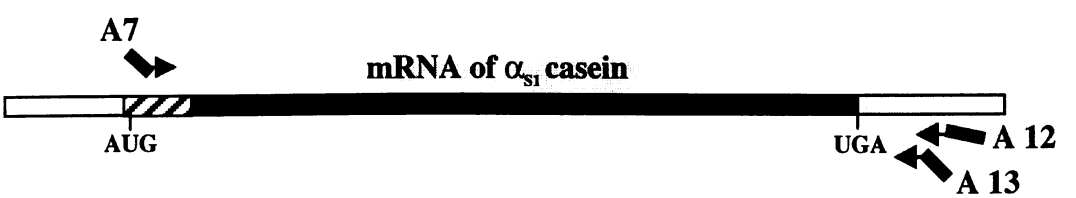



**Figure 5.3.1.1 The primers used to amplify the full length $\alpha_{s1}$-casein coding region.** *The open box is non coding exons; hatched box is signal peptide coding region; the black box indicate the region coding mature $\alpha_{s1}$-casein protein.*

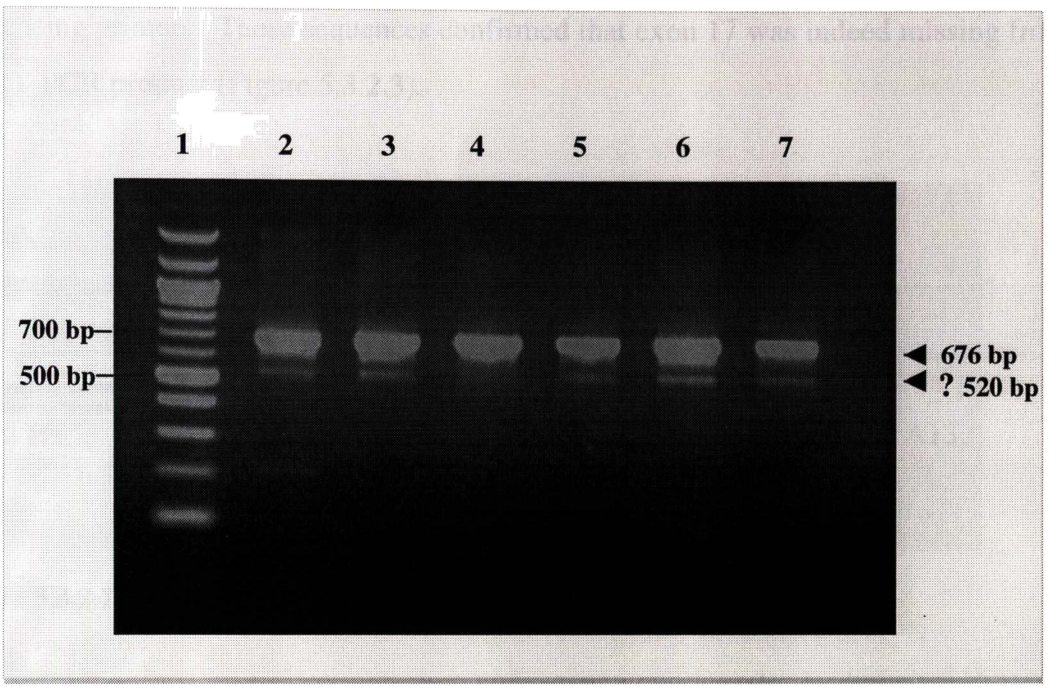



**Figure 5.3.1.2 RT-PCR using primers A13/A7.** *Lane 1, 100 base pair DNA molecule maker; lane 2-land 6 individual cows.*

## 5.3.2 Characterisation of the minor band

### 5.3.2.1 *Pvu* II digestion.

A *Pvu* II restriction enzyme digestion experiment was carried out to examine the origin of the minor product. The full length 676 bp A7/A13 product has two *Pvu* II recognition sites and will give three fragments sized at 370 bp, 165 bp and 141 bp after digestion. As one of the *Pvu* II sites is located within exon 17, the product with the deletion of exon 17 will only generate two fragments, of sizes 370 bp and 151 bp (Figure 5.3.2.1).

The result is shown in Figure 5.3.2.2. The three bands from the full length product are exactly as expected. The digestion pattern of the minor (±520 bp) product indeed suggests that exon 17 is missing.

### 5.3.2.2 Sequencing results

The truncated product was isolated and sequenced from each end, using A7 and A13 as sequencing primers. These sequences confirmed that exon 17 was indeed missing from the RT-PCR product (Figure 5.3.2.3).

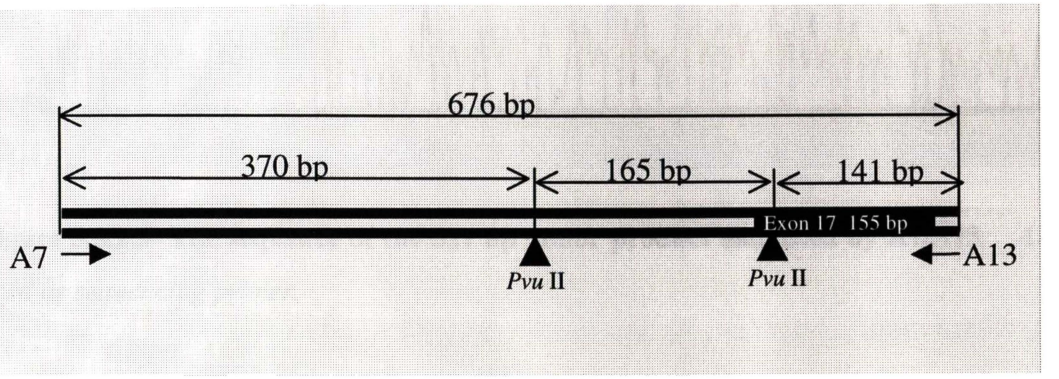



**Figure 5.3.2.1 Restriction map for *Pvu* II digestion.**

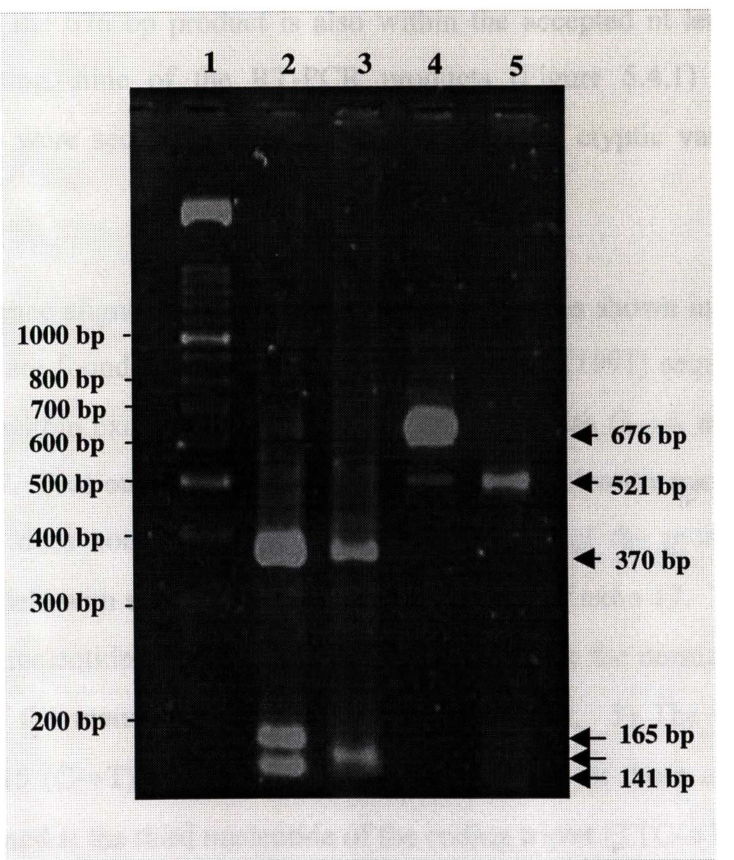



**Figure 5.3.2.2 The result of *Pvu* II digestion of full length and truncated $\alpha_{s1}$-casein cDNA.** *Lane 1, 100 bp ladder; lane 2, Pvu II digested 676 bp product; lane 3, Pvu II digested isolated 521 bp minor product. Lane 4 and lane 5 are the undigested controls of 676 bp and 521 bp respectively.*

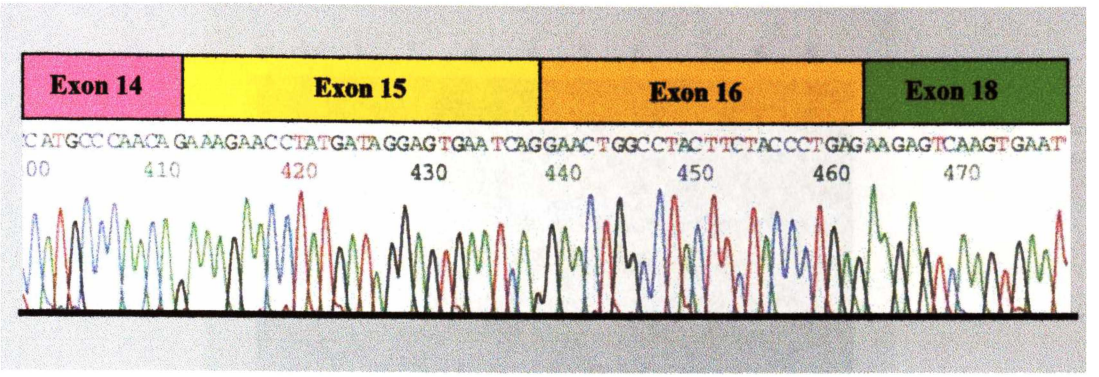



**Figure 5.3.2.3 The sequence of the 521 bp minor product amplified by A7/A13.** *A7 was used as sequencing primer.*

## 5.4 Screening of the Full Length Coding Region

As the A7/A13 RT-PCR product covers the full region which codes for the mature $\alpha_{S1}$-casein protein and the 676 bp product is also within the accepted nt length range of automatic sequencing, nine of the RT-PCR products (Figure 5.4.1) from normal individual animals were sequenced to examine the potential cryptic variants of $\alpha_{S1}$-casein.

The multiple sequence alignment result of these nine animals is shown in Figure 5.4.2. Three nucleotides are found to differ from that of Koczan's (1991) sequence. All of these variants are within exon 17. These are: 1) an allele with G→A mutation at +8 position of exon 17. Five out of the nine animals have this nucleotide transition. It is a silent mutation so does not change the amino acid residue of the protein. 2) Two animals have an allele with a point mutation (C→G) at +22 of exon 17. This mutation occurs at the first nucleotide of the coding triplet and therefore the normal Leu residue at position 155 of the mature $\alpha_{S1}$-casein will change to Val. 3) The third mutated nucleotide is at +15 (C→T) of exon 17 and is found in one of these animals. This mutation is positioned at the third nucleotide of the coding triplet (TTC→TTG) and will not alter the amino acid residue of the protein.

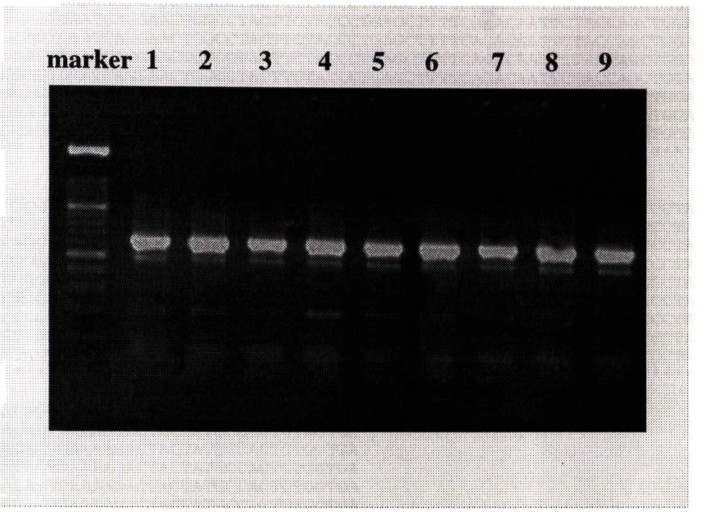



**Figure 5.4.1 The RT-PCR product for $\alpha_{S1}$-casein cryptic variant scanning.** *Lane 1 to lane 9 are the A1/A11 amplification products of the nine normal cows.*

```
          Exon 2|                exon 3              |                exon 4              |
1A7       AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
20A7      AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
5A7       AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
10A7      AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
14A7      AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
7A7       AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
2A7       AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
12A7      AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
11A7      AGGCCTAAACATCCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCACC
          ********************************************************************************

exon 5          |          exon 6          |          exon 7          |    exon 8
1A7       TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
20A7      TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
5A7       TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
10A7      TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
14A7      TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
7A7       TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
2A7       TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
12A7      TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
11A7      TTTTCCAGAAGTGTTTGGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATTGGGAGTGAATCAACTGAGGATCAAGCCA
          ********************************************************************************

                     |    exon 9                |                exon 10   |
1A7       TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
20A7      TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
5A7       TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
10A7      TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
14A7      TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
7A7       TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
2A7       TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
12A7      TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
11A7      TGGAAGATATTAAGCAAATGGAAGCTGAAAGCATTTCGTCAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCAC
          ********************************************************************************

                          exon 11                          |          exon 12
1A7       ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
20A7      ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
5A7       ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
10A7      ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
14A7      ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
7A7       ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
2A7       ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
12A7      ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
11A7      ATTCAAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAAAGTACC
          ********************************************************************************

          |          exon 13         |                exon 14              |
1A7       CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
20A7      CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
5A7       CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
10A7      CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
14A7      CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
7A7       CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
2A7       CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
12A7      CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
11A7      CCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAAGAGGGAATCCATGCCCAACAGAAAGAAC
          ********************************************************************************
```

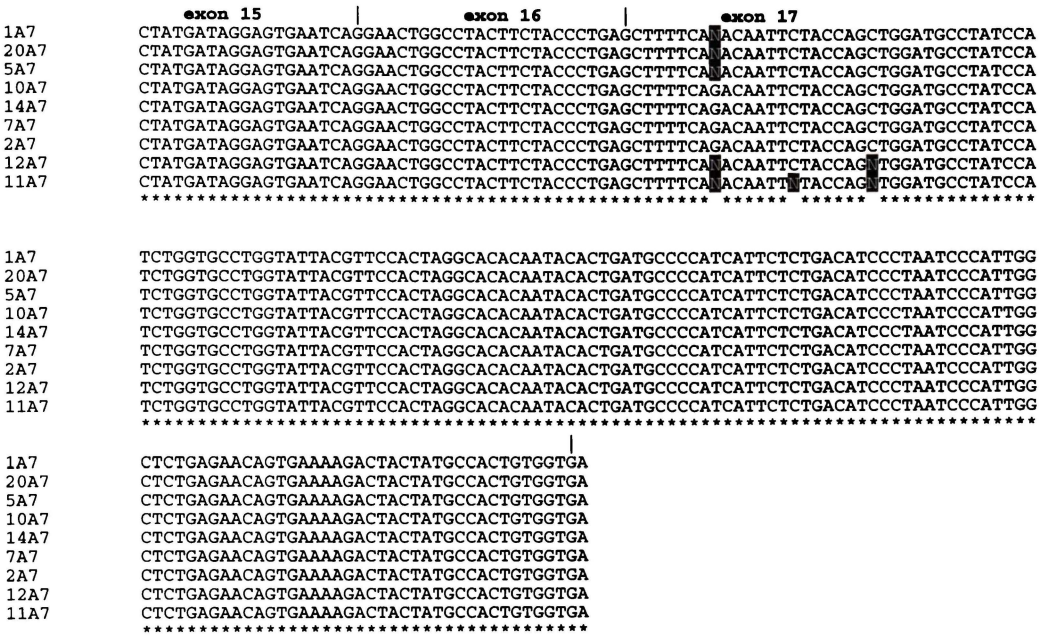



**Figure 5.4.2 The alignment of $\alpha_{S1}$-casein gene coding sequences from nine normal cows.**

## 5.5 Discussion

### 5.5.1 Exon skipping events in the region of exon 4

The investigations described here reveal a number of aberrant RT-PCR products. If these accurately reflect the species of mRNA being produced *in vivo*, it is clear that a number of quite aberrant mRNA molecules are being produced by an exon skipping phenomenon. Initially, it was felt that what was observed was possibly an artefact of the reverse transcriptase or the PCR reaction and the minor bands were just the result of some mispriming or recombination during the course of the extensive series of PCR reactions required to detect and amplify up some of these minor products. However, with one exception, which will be discussed below, this is unlikely to be the case for the following reasons. First, the sequences of all the products were authentic $\alpha_{S1}$-casein sequences and were not of homologous genes e.g. $\alpha_{S2}$-casein, or portions thereof. Second, the regions deleted corresponded exactly to exons of the $\alpha_{S1}$-casein gene, and it is highly improbable, if not impossible for recombinational events between products to give rise to exactly exon deleted products as some artefact of the PCR reaction! Third, other products which were accurately sized by GeneScan analyses, although not

actually purified and sequenced, also were highly suggestive of exon skipped products because their size was reduced below that of the full length product by an exact 24 or 39 nts, or a combination of the two. Taken together, these observations suggest that a variety of exon skipping events are occurring at a low but significant level.

The one exception, in one animal, which may result from an artefact in the PCR reaction is the unusual product which appeared to involve exon 4 & 5 skipping, but on cloning and sequencing seemed to have 7 bases of exon 4 replacing 7 bases of exon 6. Initially it was felt that this might arise from a cryptic splice site, but inspection of the intron sequences of the $\alpha_{S1}$-casein gene in this specific cow (which agreed with the $\alpha_{S1}$-casein in Genbank, 55 bp upstream of exon 4 to 167 bp downstream of exon 6) did not reveal any such sites. Thus, although one cannot be absolutely certain, it would appear that this particular RT-PCR product could represent some kind of recombination or crossing over in which a partially synthesised strand misprimes onto that which is used as a template.

## 5.5.2 The putative protein translated from exon 17 skipped mRNA

It was found that about 5-10% percent of $\alpha_{S1}$-casein mRNA was missing the last coding exon 17, which contains 155 nucleotides. When translated, this represents the last 51 amino acids at the carboxy terminal of the mature protein.

The key questions which arise from this discovery are whether or not the truncated mRNA gives rise to a protein product and if so, what is its exact amino acid sequence and how much is produced? The simplest assumption to make is that translation of the truncated mRNA will continue along the sequence until the first termination codon is encountered. Carrying out this exercise in silico, one finds that termination would occur exactly at the end of the next exon (exon 18) with the TGA termination signal consisting of the final TG in exon 18 and the first A in the final exon (exon 19) of the mRNA. This is interesting on several counts. First, the additional amino acid sequence in the putative protein is coded for by a complete exon. Second, the termination codon sequence has exactly the same origins (the TG in the last coding exon plus the first A in next exon) as in the normal mRNA and third, the actual 44 base sequence of exon 18 is

very highly conserved across the $\alpha_{S1}$-casein mRNA sequences of water buffalo, goats and sheep, even though this exon is supposedly not translated into protein in any of these mammals (It is worth noting that reasonable homology also exists with pig, and with the rat and mouse $\alpha_{S1}$-caseins, and in all cases the same termination codon is found. However, this exon does not appear to be conserved in the rabbit, human and camel $\alpha_{S1}$-caseins , nor in guinea pig casein B). The net result is that exon 18 rather than exon 17 would be the last coding region of the putative protein. Moreover, this leaves the final 385 base exon 19 intact and presumably able to carry out all of the processing functions of a 3'-UTR (untranslated region) just as it does in the normal mRNA. It is thus reasonable to assume that this represents a case of alternative exon usage (i.e. of exon 18 rather than exon 17) and that the protein product probably does exist. This is in complete contrast to deleterious point mutations or a deletion in a gene which completely disrupts the normal coding sequence, leading to a random termination signal which may no longer be supported by an authentic 3' UTR region with normal regulatory and poly A signals.

***Predicted protein sequence***

What is the nature of the putative protein? This is easy to predict at the primary amino acid sequence level. The amino acids encoded by exon 17 would be missing (these are amino acid residue 164-214 in the normal protein in Figure 5.5.2) and replaced by those encoded by exon 18 in the truncated protein (from amino acid residue 164 to 177 in putative protein in Figure 5.5.2). The final processed protein would now be 37 residues shorter, i.e. 162 rather than the 199 in the normal $\alpha_{S1}$-casein protein.

One would predict that this truncated protein would be phosphorylated in the normal way but, given the rather anomalous mobility of some caseins on PAGE, it is hard to predict just how easily it would be to resolve by electrophoresis. Clearly, selective cleavage and mass spectrometric analysis would be the method of choice for detecting peptides unique to this putative protein.

Finally, how much of this protein is likely to be produced in milk? The best prediction is about five percent of total $\alpha_{S1}$-casein, reflecting the relative amount of the two

mRNA transcripts in the mammary gland. However, unlike the A variant of $\alpha_{S1}$-casein, for which there is much data showing that the exon-4 skipped mRNA is translated with similar efficiencies to the normal mRNA, one must point out that we have no *a priori* evidence for the truncated mRNA product being translatable.

```
                10        20        30        40        50        60
                |         |         |         |         |         |
Normal          MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVALFPEVFGKEKVNELSKDIG
Putative        MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVALFPEVFGKEKVNELSKDIG
                ************************************************************

                70        80        90        100       110       120
                |         |         |         |         |         |
Normal          SESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK
Putative        SESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK
                ************************************************************

                130       140       150       160       170       180
                |         |         |         |         |         |
Normal          VPQLEIVPNSAEERLHSMKEGIDAQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWY
Putative        VPQLEIVPNSAEERLHSMKEGIDAQQKEPMIGVNQELAYFYPE----------------
                *****************************************

                190       200       210
                |         |
Normal          YVPLGTQYTDAPSFSDIPNPIGSENSEKTTISLW----------
Putative        ----------------------------------KSQVNSEGLQSYGL
                                                  |         |
                                                  164       177
```

**Figure 5.5.2 Sequences alignment of the normal $\alpha_{S1}$-casein and the putative protein generated by exon 17 skipping.**

# *CHAPTER VI*

# CONCLUSIONS

*RNA sampling from milk permits large scale gene expression studies*
*The $\alpha_{S1}$-casein A variant*
*Synthesis of small amounts of A variant $\alpha_{S1}$-CN protein in normal animals*
*Other mRNA transcripts arising via alternative splicing from the bovine $\alpha_{S1}$-CN gene*
*Do exon skipping events occur in all $\alpha_{S1}$-CN genes?*
*A new approach for detecting milk protein genetic variants*
*Future work*

This investigation has centred on a particular class of milk protein variants that arise primarily from RNA processing aberrations rather than mutations in the coding sequences of the genes *per se*. Originally, it was thought that, in dairy cattle, the A variant of $\alpha_{S1}$-casein might represent an isolated example of this phenomenon. However, the current investigations have shown a much more complex pattern of RNA processing abnormalities occurring during the processing of the $\alpha_{S1}$-casein gene. In this final chapter, the basis for "leaky" exon skipping is discussed, comparisons are made with data obtained by other workers on multiple mRNA transcripts in mammary specific genes, and the phenomenon is placed in a biological context.

## 6.1 RNA Sampling From Milk Permits Large Scale Gene Expression Studies

### 6.1.1 Practical and ethical advantages

A major practical consideration in many gene expression studies in large animals and humans is the limited availability of tissue samples from internal organs. Thus, while for cell culture and small animal work, it is usually straightforward both to design experiments and to collect tissue samples under controlled conditions, in large animals such as cows it is becoming prohibitively expensive to set up large scale experiments in which samples are collected post-slaughter; in addition, the individual variability (both due to genetic and environmental factors) of farm animals makes analysis much more complicated than with inbred animals such as laboratory mice. Moreover, while it is possible to obtain samples of many tissues by biopsy or during surgical procedures, such samples may not be representative of the organ as a whole and, in any case, it may prove impossible to repeatedly sample an organ over a period of time.

In the case of human studies, the ethical constraints are obviously even more severe, although paradoxically, obtaining large numbers of pathology and surgical specimens from patients is not a problem. It is the sampling of tissues from healthy individuals that is virtually impossible.

Undoubtedly, the major scientific advantage of the non-invasive sampling technique examined in this thesis is the ability to sample very large numbers of animals or, alternatively, to repeatedly sample an individual (animal) under a variety of conditions.

In animal experiments, the statistical power resulting from the use of both large sample sizes and the elimination of individual variation, more than outweighs the uncertainties over the authenticity of the mRNA recovered from exfoliated cells. Indeed, in this investigation of exon skipping in the $\alpha_{S1}$-casein gene, the sloughed off cells provide an ideal screening tool which can then be followed up with more comprehensive DNA sequencing, gene expression and protein studies in those animals which show abnormal levels of exon skipping.

The method has potential in some medical applications, particularly the investigation of pathological conditions in the breast and salivary gland and the urinary tract. It also is directly applicable to monitoring transgene expression in the increasing number of large animals transgenics designed to use the mammary gland as a bioreactor for producing "pharmaceuticals"

Finally, it is believed that non-invasive sampling circumvents the animal ethics and welfare problems that inevitably arise when obtaining tissue samples for RNA extraction. Whilst, it might still be necessary to biopsy and to sacrifice some animals for definitive and comprehensive studies, the methods should enable large scale studies of gene expression studies in secretory organs with minimal impact on animals.

### 6.1.2 Basic research advantages

For many RNA analyses experiments, it does not matter greatly whether or not the RNA is recovered from cultured cells (which may or may not be transfected with the gene of interest), short term tissue culture or tissue itself. This is because the aim is often just to characterise the major (unique?) mRNA species. However, when investigating phenomena such as alternative splicing or exon skipping, quite different results are obtained in a variety of *in vitro* systems from what pertains *in vivo* (Lewin 1997; Dirksen *et al.* 1995; Schischmanott *et al.* 1997). Thus, it is imperative that RNA is analysed directly from the tissue under investigation. This requirement places severe restrictions on investigations of large animals and humans and, apart from pathological and surgical samples, the only convenient source of cellular RNA is blood. For this reason, the vast majority of reports on aberrant RNA processing in humans are connected with disease conditions, often involving mutations in the splice consensus

sequence of genes. Reports concerning RNA processing variants in normal human tissues are somewhat rare. The situation is similar with large animals.

Thus, the ability to recover epithelial cells from bovine milk is an enormously powerful tool when investigating gene expression in the mammary gland.

## 6.2 The $\alpha_{S1}$-Casein A Variant

Because of its unique chemistry and processing properties, the A variant of $\alpha_{S1}$-casein has aroused interest for a number of years. The primary structure of bovine $\alpha_{S1}$-casein A was documented in the early 1970s (Grosclaude *et al.* 1972) when it was pointed out that due to the internal deletion of the 13 amino acid residues, the A variant had lost a major chymosin cleavage sites and had a less hydrophobic nature than its other counterparts in the $\alpha_{S1}$-casein family. These changes dramatically alter the physico-chemical properties of the $\alpha_{S1}$-casein A protein and may give it useful processing properties, although pilot processing experiments still have not confirmed whether or not these properties will be of economic importance (M. Boland, pers. comm.).

### 6.2.1 What range of mutations could give rise to the A variant?

As stated earlier, the initial search for mutations in the $\alpha_{S1}$-casein variant A gene was in a German Red herd and, somewhat surprisingly, only one mutation could be found and this was quite distant from the splice site (Mohr *et al.* 1994). *A priori*, one would not predict that the T→A single base substitution at +6 in the intron 4 splice consensus sequence would be sufficient to cause exon 4 skipping. If the explanation proffered by Mohr *et al.* is actually correct, namely the vulnerability of a particularly weak splice acceptor, it might well mean that this mutation is not sufficient to render the exon 4 skipping 100% effective. Indeed, one might predict that, if Mohr *et al.* were to analyse the $\alpha_{S1}$-casein mRNA and protein from a homozygous A variant animal, significant levels of normal gene expression might be found.

The situation is quite different with the A variant described in the present work. The deleted A at position +4 of the 5' splice sequence of intron 4 that was found in New Zealand DRI $\alpha_{S1}$-casein A cow completely destroys the consensual nature of the donor

sequence. Thus, one would expect the exon 4 skipping in these animals to be absolute, or nearly so. Although, analyses were not carried out at the mRNA level to confirm this prediction, they were at the protein level, conforming that purified DRI A variant $\alpha_{S1}$-casein protein from homozygous animals did not contain a detectable normal $\alpha_{S1}$-casein protein component (Figure 4.4.1.1).

One would assume that at least some of the other $\alpha_{S1}$-casein A variants which have been reported around the world have arisen from yet different mutations from the two described above, and that these could "target" any one of the six or so bases in the splice consensus sequence. These might result in varying degrees of "leakiness" in exon skipping. Indeed, it is possible that the exon 4 processing events are particularly sensitive to a range of mutations (in acceptor as well as donor sites – and possibly in other gene regions). Thus, it is possible that if one was to scan a large number of animals, varying degrees of exon 4 skipping may be occasionally discovered and these may reflect various mutations.

### 6.2.2 PCR tests

From what has been described above, both specifically for the German Red and DRI mutations, and also in more general terms for other A variant animals yet to be characterised, it is clear that one simple PCR test will not be diagnostic for genotyping all A variant genes. Obviously, the first step in characterising any new A variant is to sequence the $\alpha_{S1}$-casein gene in the region of exon 4. Developing a test based on mutations found in this region may not be straightforward because of the lack of restriction enzyme recognition sites within the 5' consensus sequence. However, in most cases an ACRS based method similar to that designed to genotype the DRI A variant should be applicable. These methods do have some limitations, namely, a propensity to not always amplify both alleles efficiently in a one step PCR reaction from genomic DNA which can result in interpretation problems. This can be overcome by using semi-nested ACRS reactions (Section 3.3) and, for even more definitive analyses, two ACRS reactions, designed to recognise (cut) one or the other of the two alleles (normal and mutated). After three seasons of using the ACRS method to scan DRI animals, it can be stated that this method is simple in practice and is capable of mass screening large numbers of animals providing an effective means of genotyping

both newborn calves and bulls. Indeed, it compares more than favourably with some of the relatively complicated methods used to genotype β-casein alleles (Damiani *et al.* 1992).

## 6.3 Synthesis of Small Amounts of A Variant $\alpha_{S1}$-CN Protein in Normal Animals

The fact that an even a minor perturbation of the consensual donor sequence in intron 4 can cause exon skipping in German Red A variant animals, led to the proposal by Mohr *et al.* that this splice acceptor signal 3 in the normal $\alpha_{S1}$-casein gene was very weak. Thus, it is not entirely surprising that "leaky" exon 4 skipping occurs in normal animals (Chapter 4). Indeed, early work by McKnight *et al.* (1989) had provided direct evidence that exon 4 skipped cDNA species existed in a library constructed from the mammary gland of normal (B/B variant) animals. McKnight *et al.* scarcely commented on this paradoxical result, and they made no effort to quantitate the levels of exon 4 skipped cDNA, nor did they check whether or not A variant protein was present in the milk. In actual fact, it would have been relatively simple for them to check this possibility, as even a few percent of A variant $\alpha_{S1}$-casein protein would have been readily detectable on the PAGE systems that they used.

In the study presented here, the direct analysis of mRNA from milk samples enables the scanning of $\alpha_{S1}$-casein expression in a relatively large number of animals. After analysing RNA from more than 20 individual dairy cows (mixture of B and C genotypes) and the pooled samples from a whole herd of cows, it was concluded that all produce small amounts (2-5 percent) of aberrant exon 4 skipped $\alpha_{S1}$-casein mRNA. A similar result was obtained when RNA from lactating mammary gland tissue, and cDNA from a Stratagene cDNA mammary gland library were analysed.

Although the work presented here was only semi-quantitative, it would seem that no really large individual fluctuations in the percent "leaky" exon 4 skipping are occurring. That is, no samples reflected either very low (less than 1%) or very high (greater than 10%) exon skipping. Unfortunately, the RT-PCR methods employed here (either direct agarose gel or GeneScan assays) suffered from a number of large quantitation errors which made more precise estimates of the percent exon 4 skipping

rather difficult. Even more sophisticated methods involving the TaqMan or Lightcycler techniques which permit product quantitation at the end of every PCR cycle might be rather difficult to implement in this particular type of analysis in which two products, which differ only in size, are being simultaneously detected. Another possible, and more general, problem which could complicate the analysis would arise if the RNA recovered from the milk did not authenticly represent the RNA being expressed in the mammary gland, though there is no evidence to suggest that this is a major problem. Fortunately, in the case "leaky" $\alpha_{S1}$-casein exon 4 skipping, we do not have to concern ourselves overly with these problems because simpler analyses can be done directly on milk samples, once the relevant minor A variant band has been identified.

The availability of high purity A variant $\alpha_{S1}$-casein samples and also an affinity purified polyclonal antibody to $\alpha_{S1}$-casein (both courtesy of the N.Z.DRI) made the analysis of protein samples by PAGE extremely definitive. Indeed, even without confirmation by Western blotting, the minor protein band running at exactly the same position as the A variant control, was strongly suggestive of "leaky" exon 4 skipping in normal B and C variant animals. Over 40 samples (both of purified caseins and whole proteins from fresh milk) and also purified $\alpha_{S1}$-casein (prepared as part of this investigation and also from N.Z.DRI) were analysed, and the minor band running at the position of the A variant confirmed by immunoblotting to be an $\alpha_{S1}$-casein species. As with the mRNA analysis, the minor protein band constituted a few percent of total $\alpha_{S1}$-casein protein.

Again, as with the mRNA analysis, no marked variation in the amount of the minor band, as a percent total $\alpha_{S1}$-casein was seen from individual to individual.

The discovery of the co-existence of this minor variant with the major $\alpha_{S1}$-casein (B or C) protein points to a hitherto unsuspected complexity in the milk casein proteins. Thus, besides the different phosphorylated, glycosylated and degradation products, one must consider the possibility that some minor and ill-defined casein products, actually arise from alternative splicing. The extent of this phenomenon, *in toto*, is unknown although additional investigations, about to be discussed, suggest that the A variant example is unlikely to be unique.

## 6.4 Other mRNA Transcripts Arising via Alternative Splicing from the Bovine $\alpha_{S1}$-Casein Gene

Although, one could probably have predicted the existence of "leaky" exon 4 skipping, *a priori*, based on the work of Mohr *et al.* (1994) and McKnight *et al.* (1989), the additional observations made in the present investigation concerning alternative splicing of the $\alpha_{S1}$-casein gene were quite unexpected. Of particular interest is the leaky exon 17 skipping which was detected in all normal cows examined. Possibly of less universal importance is the combined exon 4 and 5 skipping which was identified in an individual animal. As noted earlier, the lengths of the exons of $\alpha_{S1}$-casein are all multiples of three, which means that removal of exons does not disrupt the reading frame of the codons. Although this would lead one to predict that variant $\alpha_{S1}$-casein products could, in theory, be produced from a whole range of exon skipped mRNAs, actually demonstrating the existence of such protein products is not straightforward. The ratio of these skipped mRNA to that normal full length transcript is higher than that of exon 4 skipping (around 5-10 percent).

In case of exon 17 skipping, exon 18 is very likely able to act, instead of exon 17, as the last coding exon of this gene (as discussed detailed in section 5.5.2). Although this latter exon is normally a non-coding, examination of the sequence suggests that is reasonable to assume that such an aberrant mRNA product could actually give rise to an $\alpha_{S1}$-casein protein product. One would expect this to constitute five to ten percent of total $\alpha_{S1}$-casein, if one assumes that, as for exon 4 skipping, protein levels reflects the relative amount of the two mRNA transcripts in the mammary gland.

For the exon 4 - 5 skipped product, one might also predict a truncated protein product and, indeed, GeneScan analyses suggested that, in some animals there were traces of other (24 nt) exon skipped mRNA products which, conceivably, might also give rise to variant $\alpha_{S1}$-casein protein products. While it is easy to predict the primary structure of such putative proteins, as emphasised earlier, it is much more difficult to experimentally detect them. Possibly, the availability of "standards" for the various predicted protein variants would help, as it indeed does in the case of the A variant, but these would be extremely difficult to synthesise, and the most attractive means of analysis is probably mass spectrometry of digest fragments (see below).

## 6.5 Do Exon Skipping Events Occur in All $\alpha_{S1}$-Casein Genes?

Both regions of the $\alpha_{S1}$-casein mRNAs examined in detail in the present investigations, namely exon 4 and exon 17, show evidence of "leaky" exon skipping. In addition, as alluded to above, preliminary GeneScan results, (data not shown) also suggests additional aberrant skipping events in the exon 4 and 5 regions of the gene. One thus wonders what the full extent of "leaky" exon skipping will be when other regions of the gene (mRNAs) are examined in detail! Furthermore, although individually rather small, the cumulative fraction of such aberrant mRNA species and, possibly, the protein products could be substantial.

Although, exon skipping events of the type described in this thesis have not been previously reported for the bovine $\alpha_{S1}$-casein gene, similar phenomena have been observed in a number of other species. Thus, partial exon skipping has been found in goat, ovine, pig and human $\alpha_{S1}$-casein pre-mRNA transcripts, as well as in sheep $\alpha_{S2}$-casein pre-mRNA transcripts (see Appendix IV). Generally, in most of these cases, the exon skipped mRNA species account for a considerable proportion, as much as 50% of the total transcripts, which is much higher than that occurring in bovine. In addition, protein products corresponding to alternatively spliced mRNA forms have been identified in caprine and ovine milks (Ferranti *et al.* 1995 & 1997).

This leads one to ask what mechanisms could cause such frequent aberrant splicing in these genes? It seems there is not a simple answer to this question. One possible clue is the common evolutionary origin of these proteins, as demonstrated by the degree of alignment of the amino acid sequences. It is reasonably to assume that the homology of the genes is also extensive and that some sequence elements shared by these genes, within or away from splice sites, may play a role in the "leaky" exon skipping. Unfortunately, the *cis*-elements responsible for the exon skipping in these genes have not been identified except in the case of the goat $\alpha_{S1}$-casein gene (Appendix IV). The goat $\alpha_{S1}$-casein gene expresses multiple forms of $\alpha_{S1}$-casein peptide chains through exon skipping events involving numbers of exons. A single nucleotide deletion in exon 9 and two insertions, of 11 and 3 bases in length, in the down-stream intron, were identified as mutations potentially related to the exon skipping events in variant F of the goat $\alpha_{S1}$-casein gene (Leroux *et al.* 1992). This situation, with a particular variant

undergoing a high level of exon skipping, apparently because of extensively mutated gene sequences, is quite different from the bovine examples discovered in the present study in which, the exon 4 and exon 17 skipping events appear to occur in all "normal" animals.

Possibly, the combined exon 4 and 5 skipping observed in the current investigation in one individual cow does have more similarities with the goat variant F example, as a point deletion was found in this particular animal. However, it is difficult to see how exon skipping would be disrupted, given the site of the deletion, well within intron 5, some 500 and 100 bp away from, respectively, the upstream and downstream splice boundaries. Nevertheless, one cannot dismiss the importance of such mutations until the *cis*-elements, genomic organisation and regulation of expression of the $\alpha_{S1}$-casein gene are better understood.

The unusual exon skipping involved partial exon 4 and partial exon 6 is particularly difficult to understand. Inspection of the sequences of relevant regions of the $\alpha_{S1}$-casein gene in this specific cow did not reveal any cryptic splice site or suggest any new splice site which could, by an RNA processing mechanism, create the observed product. Although one cannot rule out some biological process leading to this aberrant RNA, the repeated sequence in exon 4 and exon 6 (Figure 5.2.2.2) does suggest a possible PCR artefact as the result of some kind of recombination or crossing over in which a partially synthesised strand is used as a template.

## 6.6 A New Approach for Detecting Milk Protein Genetic Variants (Full Length cDNA Scans)

Interspecies comparisons have confirmed the high rate of evolution of milk proteins, particularly caseins. It is easy to accept that relatively fast evolution and/or low "functional constraints" on milk protein genes and their protein products have led to the accumulation of a relative large number of genetic polymorphisms in these genes. Indeed, this argument is supported by the occurrence of a relatively low level of polymorphism in the one milk protein with an overt functional role, α-lactalbumin which is an essential component of the lactose synthetase complex and for which there

is only one common "variant" has been observed in milk from Western cattle. (Eigel *et al.* 1984)

Electrophoresis and analytic chromatography are the two most commonly used methods used to analyse milk gene genetic polymorphisms. Indeed, so far, most of the milk protein polymorphisms have been identified by such protein analyses. However, as the efficacy of polymorphism identification based on such protein analyses relies on relatively large changes in the molecular weight or/and electric charge of the protein, it is reasonable to assume that there are some protein variants that cannot be detected by such traditional methods because of the more subtle changes in physico-chemical properties.

Perhaps the only physico-chemical method that is readily adaptable to detecting such subtle differences is a mass spectrometric approach. Electrospray ionisation mass spectrometry (ESI-MS) is a powerful analytic technique used for forming multiple charged gas-phase molecule ions from neutral, intact molecules and subsequently determining their molecular masses, with extremely high accuracy (Mann and Wilm 1995; Burr *et al.* 1997; Chin and Ng-Kwai-Hang 1997). However, this sophisticated tool is not an idea method for large scale protein polymorphism screening. Firstly, ESI-MS does not examine amino acid residue composition of a given protein or peptide sample but its mass. In addition, even though it is very sensitive, the degree of accurate in mass determination is as high as 0.01-0.05%, ESI-MS can not distinguish all the potential amino acid residue substitution whereas intact milk protein molecule is analysed directly. A more accurate result can be achieved by analysing enzyme digested and HPLC fractionated peptides of these proteins, but this would be very laborious and time-consuming.

A more attractive approach, based on the success in the present investigation of obtaining mammary specific RNAs from milk, is to use RT-PCR to scan the full length coding mRNA. This mRNA→ protein "reverse" method would be able to:

1) Identify "cryptic mutations" in which net electric charge was unaltered – such mutations are difficult to detect using routine protein techniques;

2) Allow detection of silent mutation(s) in which the amino acid residues of the protein does not change but the mutation may link to other relevant genetic events, e.g. mutations in non-coding regions of the gene individual or QTLs.

3) Allow multiple mRNA species to be detected – these could be of the same size e.g. normal full-length transcripts from different alleles, or represent aberrant transcripts such as those documented in the present study. As discussed above, proteins resulting from such aberrant transcripts would be difficult to detect with routine analytical screens at the protein level.

In the past, such large scale scanning of nucleic coding sequences has been impractical for two reasons. First, if genomic DNA from individual animals is used, it is incredibly time consuming and wasteful to scan different regions of the gene by setting up the large number of PCR reactions necessary to ensure each exon, often only tens of bases long, are analysed! Second, although RNA has been recognised as the desirable starting material for screening the coding sequences of genes from individual animals, it has not been practical to biopsy mammary tissue from large numbers of cows.

In this study, such full length scanning has been carried on nine individual animals. Three nucleotides were found to be different from the published $\alpha_{S1}$-casein mRNA sequence (Koczan *et al.* 1991). All of these variants were within exon 17. Two of the changes were silent mutations confined to the third position within coding triplets and did not result in amino acid substitutions. The third exchange (C→G) at +22 of exon 17 occurred in two of the nine animals sequenced. This base substitution is at the first position of the triplet therefore the normal Leu (CTG) residue at position 155 of mature $\alpha_{S1}$-casein will be predicted to change to Val (GTG).

Searching for new genetic polymorphisms was not a major aim of this project, but clearly it would be a straightforward, though somewhat laborious exercise, to scan a large number of animals in this way. In terms of cost, time and analytical power, this method is far preferable to protein based methods although, clearly, amino acid sequencing would have to be carried out to confirm variant proteins were actually produced from animals which yielded novel mRNAs.

## 6.7 Future Work

### 6.7.1 Exon skipping events in $\alpha_{S1}$-casein and other milk protein genes

As mentioned earlier (Sect 3.4.1), there are other putative weak consensus signals in the $\alpha_{S1}$-casein gene. By setting up appropriate RT-PCR reactions each of these regions could readily be scanned for leaky exon skipping. Indeed, in some cases it would be possible to scan several of the shorter exons at once, especially if the GeneScan method was used (the advantage of GeneScan over standard agarose gel electrophoresis is the very high accuracy with which the size of a minor product can be established). The method would be readily adaptable to the other casein genes and, given the relative simplicity of the technique, even the whey proteins could be analysed. Indeed, the power of the mRNA analyses is that it is simpler to do the experiments than to theorise over which exons in a gene may or may not be subject to skipping!

The major weakness in this type of investigation is ascertaining whether or not the corresponding aberrant protein product is actually made! Thus, the investigations described in this thesis were made much easier because the exon skipped product, $\alpha_{S1}$-casein variant A, was already a well characterised protein. One would predict that virtually all exon skipped casein products produced in small amounts are going to be very difficult to detect by standard PAGE. However, in the case of whey proteins, the analyses would presumably be much simpler.

### 6.7.2 Qualitative and quantitative analysis of putative protein products produced from exon skipped mRNAs

As aforementioned, the putative minor protein products are hard to detect by traditional methods. Advanced protein analysis methods might need to be applied singly, or in combination, to identify these putative proteins. Methods such as high performance liquid chromatography or capillary electrophoresis on line with electrospray ionization mass spectrometry; selective cleavage and FAB/MS (Fast-atom-bombardment mass spectometry analysis) etc. would be useful for detecting peptides unique to these putative protein products. In addition, as the isoelectric point (p*I*) of the truncated proteins are, in some cases*, likely to be different from that of the normal protein,

* *Theoretical calculations indicate that $\alpha_{S1}$-CN variant A would shift (increase) the pI by +0.03 units and the putative exon 17 deleted protein would shift the pI by +0.21 units from that of the common $\alpha_{S1}$-casein B and C variants.*

careful analysis by high resolution polyacrylamide gel isoelectric focusing (PAGIF) and/or the immunobilized pH gradients (IPGs) techniques would possibly provide a relative simple way for recognising the putative protein in fractionated casein samples.

### 6.7.3 The dynamics of exon skipping

A question of both basic and applied interest is whether or not the leakiness of exon skipping varies both from animal to animal and throughout the lactation cycle. The preliminary data presented here on exon 4 and exon 17 skipping does not reveal any marked variation but, for a variety of technical reasons, the data must be considered semi-quantitative. Thus, although care was taken to ensure that the RT-PCR reaction cycles were kept to a minimal (16 to 20 cycles in the case of GeneScan reactions), one would really need to employ a fluorescence quantitation system such as the Lightcycler (Wittwer *et al.* 1997) or Taqman (Heid *et al.* 1996; Wang and Brown 1999) systems to ensure estimations based on linear reactions for both the minor and major products. These systems would also have the advantage of allowing the simultaneous fluorescence detection of both the major and minor products in the one reaction.

The small amount of minor product also gives a sensitivity problem, and the setting of the baseline on GeneScan runs was a critical parameter which could result in large fluctuations in the apparent amounts of exon skipped product. At the protein level, the estimation of the amounts of $\alpha_{S1}$-casein variant A in milk also presented difficulties. The Western blots, although very specific, were difficult to quantitate although it is quite possible that densitometric scans of stained protein gels would be satisfactory. Thus, in case of exon 4 $\alpha_{S1}$-casein skipping, small format alkaline urea-PAGE gels could well prove satisfactory, so it would be a rather simple exercise to analyse large number of animal for any fluctuations in the amount of this exon 4 skipped product.

Thus, the fact that no significant variations in the leakiness of exon skipping were seen, may simply reflect the limitations of the methods applied to date. The screening of large numbers of animals using more quantitative methods might well reveal significant variations from individual to individual and, possibly, an occasional animal in which the level of exon skipping is either markedly depressed or elevated. Additionally,

systematic investigations of animals throughout their lactation cycle, are needed to established whether or not this is also an important parameter.

The one individual in which a unique pattern of exon skipping was observed, was that involving both exons 4&5; cDNA and genomic sequencing did not reveal any obvious mutations which would be responsible for this anomalous RNA processing and it was not possible to analyse the milk (the animal was not lactating in the season following the discovery of the exon skipped mRNA species). Given the small number of animals analysed to date, it is quite possible that other animals with extreme exon skipped "phenotypes" will be discovered.

### 6.7.4 The mechanism of leaky exon skipping

Several scenarios come to mind when seeking an explanation for leaky exon skipping in the mammary gland. The first scenario is that the exon skipping occurs because of defective operation of the *cis*- and *trans*-elements responsible for the normal regulation of the RNA processing machinery in the mammary gland. It is possible, that some cells (a few percent) are highly defective and these are producing almost exclusively exon skipped protein, while the other 95% are normal. Although unlikely, this possibility could be investigated using *in situ* hybridisation of lactating mammary gland sections with oligonucleotide probes specific for the A variant mRNA (see section 4.2.2). A more likely explanation for defective processing is the very high rate of (casein) mRNA synthesis in the lactating mammary gland; possibly this "saturates" the RNA processing machinery in some way. Thus, RNA from very early involuting or late pregnant mammary gland might be predicted to demonstrate less exon skipping, but this analysis could be complicated by the fact that, even in pregnancy, cells in the mammary gland tend to be turned either "full on" or "full off" rather than uniformly exhibiting a gradual rise in milk protein mRNA levels (Molenaar *et al.* 1992). A much more instructive set of analyses may involve looking for evidence of exon 4 $\alpha_{S1}$-casein mRNA skipping under conditions in which the gene is expressed at much lower levels than *in vivo* in the mammary gland. Obvious systems would be mammary tissue explants, cell cultures and even transfection assays involving bovine $\alpha_{S1}$-casein partial cDNA-partial genome DNA constructs introduced into HC11 or comma D mouse cell lines. Perhaps even more instructive, would be the analysis of $\alpha_{S1}$-casein mRNAs produced by so called illegitimate or ectopic gene expression in non-mammary tissues of cows. Such experiments would be technically difficult as, in lymphocytes for example, there is only

likely to be an average of 0.1 to 1 $\alpha_{S1}$-casein mRNA transcripts per cell so it would be necessary to used nested primers in order to detect a small fraction of exon 4 skipped mRNA.

In summary then, the discovery of an leaky exon skipping mechanism in a tissue which is so amenable to investigation as the bovine mammary gland, offers a unique resource for elucidating the basic parameters of a phenomenon which is ubiquitous and of fundamental importance to a whole range of developmental, regulatory and pathological conditions in mammals. It is, however, a phenomenon which is, paradoxically, extremely difficult to study under *in vivo* conditions. One can confidently predict that screening large numbers of animals under a variety of experimental conditions, will lead to the discovery of factors which influence the levels of exon skipping, whether they be physiological, cellular, endocrinological, mutational or pathological. Additionally, these studies will yield information on the detailed composition of milk which may be of nutritional and economic importance.

# *Appendices*

## I. Average concentrations of some constituents in the milk of a number of species.

| *Species* | *Fat(g/l)* | *Casein (g/l)* | *Milk serum protein (g/l)* | *Lactose (mM)* | *Calcium (mM)* |
|---|---|---|---|---|---|
| Cow | 37 | 28 | 6 | 133 | 30 |
| Goat | 45 | 25 | 4 | 114 | 22 |
| Sheep | 74 | 46 | 9 | 133 | 58 |
| Pig | 68 | 28 | 20 | 153 | 104 |
| Horse | 19 | 13 | 12 | 172 | 17 |
| Human | 38 | 4 | 6 | 192 | 7 |
| Rat | 103 | 64 | 20 | 90 | 80 |
| Guinea-pig | 39 | 66 | 15 | 83 | 41 |
| Rabbit | 183 | 104 | 32 | 60 | 214 |

(*Source: Davies, Holt and Christie 1983*)

## II. Organization of milk protein genes

| *Exon/intron number* | *Exon/ intron size (bp)* | | | | | |
|---|---|---|---|---|---|---|
| | $\alpha_{s1}$-*CN* | $\alpha_{s2}$-*CN* | *β-CN* | *κ-CN* | *α-LA* | *β-LG* |
| 1 | 53/1373 | 44/1785 | 44/1935 | 65/2540 | 160/ | 135/ |
| 2 | 63/1970 | 63/1809 | 63/724 | 57/5923 | 159/ | 140/ |
| 3 | 33/893 | 27/789 | 27/112 | 33/2011 | 76/ | 74/ |
| 4 | 39/440 | 21/88 | 27/1895 | 574/1848 | 330 | 111/ |
| 5 | 24/596 | 42/389 | 24/932 | 171/ | | 105/ |
| 6 | 24/602 | 27/792 | 42/1320 | | | 42/ |
| 7 | 24/692 | 27/440 | 498/601 | | | 183/ |
| 8 | 24/1500 | 27/1042 | 42/730 | | | |
| 9 | 33/768 | 24/96 | 312/ | | | |
| 10 | 24/90 | 45/260 | | | | |
| 11 | 54/1531 | 123/529 | | | | |
| 12 | 42/647 | 27/1386 | | | | |
| 13 | 13/1207 | 27/1695 | | | | |
| 14 | 42/896 | 24/91 | | | | |
| 15 | 27/946 | 45/1104 | | | | |
| 16 | 24/795 | 120/796 | | | | |
| 17 | 155/590 | 45/4386 | | | | |
| 18 | 44/1789 | 266/ | | | | |
| 19 | 385/ | | | | | |
| **exon/intron** | 1138/16370 | 1024/17437 | 1079/7409 | 900/12322 | ***** | ***** |
| **Total (bp)** | 17508 | 18483 | 8498 | 13000 | 2023 | 4723 |

## III The full length sequence of B.taurus gene for alpha-s1-casein.

(Genbank Accession X59856, Authors: Koczan,D; 28-MAY-1991)

*mRNA*: join (2132..2184, 3558..3620, 5591..5623, 6517..6555, 6996..7019, 7661..7684, 8287..8310, 9003..9026, 10527..10559, 11328..11351, 11442..11495, 13027..13068, 13716..13739, 14947..14988, 15885..15911, 16858..16881, 17677..17831, 18422..18465, 19255..19639)

```
   1 gatcttccca acccagatgg gcatgaaaaa ggagagaaat aaaaaggact taacagaagc
  61 agaagaaatt aagaagaggt ggcaagaata gtattacaga agaactgtat ttaaagatct
 121 taatgaccca gatagccaca gttgtgtagt ctctcatcta cagcttaaaa ctcaatgttc
 181 aaaaaactaa gttcatagca tactgcccca tcacttcatg gaaaatagtg ggggaggggg
 241 agaaggtgga agtagtgtca gattttattt tcttggactc aaaatcactg cagacagtga
 301 ttatagccat gaaattaaat gacgcttact ccttgaaaag aaagttctga aaaacctaga
 361 caacatatta aaaagcagtg acatcacttt actgataagt gtcttnatag tcaaagctat
 421 ggtttttcca gtagccatgt acagatgtga gaattggact atgaagaagg atgagtgtca
 481 aaggactgat gttttcaaat tgtggtggat acactccttt gcatgcgtgc taagtcattt
 541 cagtcatgtc caactctttg caacccagtg gactgtcgtc tgccaggttc ctctgtccat
 601 gggattctcc aggcaagagc aacggagtgg gttgtcattt cctccaccag gggatcttcc
 661 caatccagat attgaacctg catctctaat gtttcctgca ttggcaggca ggttctttac
 721 cactagtgcc acctggaaag tccggattac actcctggga aagacaaaag tagagtatta
 781 caatgcagca aggatttttg ttctcagctc cttgaataaa ttatagtgaa tagaaaacat
 841 tagtatcttg ttgaaattga tgtgaaacag atagtaagga agataatatc taaagaaaac
 901 ttcaatatgg gaaattatag tcttttctat cttcaaagtg gacagcctga acagttttga
 961 aatttctttt aatacaaaat aatgttcctg tcatacaact gtgaatacac tgaaaatatc
1021 actatagatt ttttaaagta tataatatga ttccttttctt ataaacaatg agttgcaatc
1081 aacaagtttt taaagctctc acttgtatag atttattttt agcacataat atttttctac
1141 aatgtacaat gccagttaat tctaggagta caattaagaa ttggagagat aggaattttt
1201 ttcttttact tgtttacttt aaaagatgga aaatcagagt tatggtttat ttttcgcaat
1261 atttaaaaat tataattctt gaataactat taattttaat taaataatct gtaatgagaa
1321 tcctcctacc aatgtaggag acgtgagttt gactcccggg tagggaagat accctgcaga
1381 aggaaatggc aacccactcc aatattatta cttgggaaat cccatggaca gaggagactg
1441 gcaggctgca gtccatgggg gtcacaaaga actggacacg acttagaaac taaacaacaa
1501 caatttatac cagaatgaat gaactagtta ccacaactag tacacccaaa atgaacaaaa
1561 aatagcttgg tggtataatt aaaatgccac caaaatttat acaataatta tattttcttt
1621 ttgcaggaaa aagattagac cacatataat gtaacttatt tcacaaggta aataattata
1681 ataaataata tggattaact gagttttaaa aggtgaaata aataatgaat tcttctcatg
1741 gtcttgtatg ttaataaaaa ttgaaaaatt ttgaagaccc cattttgtcc caagaatttc
1801 atttacaggt attgaatttt tcaaaggtta caaaggaaat tttattgata taataaatgc
1861 atgttctcat aataaccata aatctagggt tttgttgggg ttttttttttg tttgttaatt
1921 tagaacaatg ccattccatt tcctgtataa tgagtcactt ctttgttgta aactctcctt
1981 agaatttctt gggagaggaa ctgaacagaa cattgatttc ctatgtgaga gaattcttag
2041 aatttaaata aacctgttgg ttaaactgaa accacaaaat tagcatttta ctaatcagta
2101 ggtttaaata gcttggaagc aaaagtctgc catcaccttg atcatcaacc cagcttgctg
2161 cttcttccca gtcttgggtt caaggtatta tgtatacata taacaaaatt tctatgattt
2221 tcctctgtct catctttcat tcttcactaa tacgcagttg taacttttct atgtgattgc
2281 aagtattggt actttcctat gatatactgt tagcttaaaa atatatttgc aaatgttgat
2341 actatctatc tcagagctat aggtgaaaaa ttaaatactt ttataaagac caaattgatc
2401 atttttaaac gaaattctta tatactgaaa atgtagatac ataacttcag tatagattta
2461 tggtaaaata atttgaatca tttttgtcaa attctgtaaa aagttgtcat acagaataat
2521 ttataatatt tttgttttca tagaaataac atttctggta gaatatttca aggccatttt
2581 tattttgtgt aattaggtta ataaaattaa ttttataagg aaatgtcaat gacagacaat
2641 tagatataaa tgactacttt tataaagatg attaaatttg gatatttgta aggatacaaa
2701 tatatgaaaa cagtagactc atttggggcc tataaatatg tctttttttaa caaatgcagg
2761 tagattctac agtttgtaaa ctgaagcagc ctatataaaa taatctgtca ttagtttgct
2821 gactaaggta taaacaaatt tcatgtataa tctaattttt cttatgtatc tgaaacgcat
2881 ttttccagca catataaatg tatgtatttt tgggtcttgc aatttaatgg aactctagga
2941 gtcaaatgtg atatgtttga cttatgattc tgtttaatca tcttcattca gtcatgtcat
3001 gatatatcaa cccagcaaaa ttaagtaata gctagatcct tttaaaaatt taatgaagtt
3061 aatagtttct acataatgca caatgttttt catgaagact ctgaaagagc aggctaaaga
3121 taaagacatt ttaaaaaatt acagatatta aatgtaattt accagtgggt tttagtttat
3181 caattttaac aaatccaatg atctaagagg aaatttcttt ttaatttttt ttgtagtatt
3241 ttaaaatttg taatatttaa atattgatgc ttctctattc ctctgacaaa ccctactatt
3301 actttcagga tcaaatgctt tactttaaag tgtagtaaga tattggtatt tcttatctta
3361 tataagcact aagcaaaata atttgaatgg taaatattta tattgaagag caaaattaaa
3421 aactaaatga ataaaaatat tactttcaag tgcaacaact tttatcataa tatactcctt
3481 tgttggaaaa attaagaatt ttttttttcat gaatcaaatt ttattataag acctaactat
3541 tttatttttct tacatagatc ttgacaacca tgaaacttct catccttacc tgtcttgtgg
3601 ctgttgctct tgccaggcct gtgagtacag tagagaattt agaagattct agattcttgt
3661 ttaaagtcat ctcaaatgca atttgatgca agtctcatca agtgcaagat attgtagtca
3721 taaagaattt gatggtctct aattagctat taaagtgttg atattaaagc tatggtaatc
```

3781 cttcacattt tgatcattat tatttagttt attcaaagac ttaactctaa ataaaatttc
3841 taacttgaaa tataaacacc tcacaattaa aaattttaaa aaaaaagaaa taaggtaatt
3901 aacaatacaa gtaaagagca tcaaagaagg taactagtct ctttgcctgg gtccattata
3961 ggcttaacat atttgtaaac atatatatat ccaatattgt attaccaaat atactgctgc
4021 tgctgctaag tcacttcagt cgtgtccgac tctgtgcgac accattgatg gcagcccacc
4081 agctctgcca tccctgggat tctccaggca agaacactgg agtgggttgc catttccttc
4141 tccaatgcat gaaagtgaaa agtgaaagtg aagtctctca gtcgtgtccg actcctagca
4201 accccatgga ctgccaggct cctccatcca tgggattttc caggcaagag tactggagtg
4261 ggttgccatt gccttctcca accaaatata ctattaaacc ctatattcca gtgtatccac
4321 ttttggaacc taaatcaaac cctcatttga gatgctcatg ccaaccaata tttcccaagg
4381 tacagaaagt tgggctcatt cagctgattc aaagatctaa taatttggtc cttgtagaaa
4441 gaaaaactag ataatgtaaa gtaacttaag tttcttctca aaaaacaaat tcagttatta
4501 atgtgaaaca aaaagttatt ctgttccttt gtgacctctg ctgctaagtc gcttcagtcg
4561 tgtccgactc tgtgcgaccc catagacagc agcccacaag gctcccccat ccctgggatt
4621 ctccaagcaa gaacactgga gtgggttgcc atttccttct ccaatgcatg aaagtgagaa
4681 gggaaagtga agtcgctcag tcatgtctga ctcttagcga cctcatggac tgcagcccac
4741 caggctcctc tgtccatggg attttccagg caagagtact ggagtggggt gccattgcct
4801 tctccgttgt gacctctaga taatgataaa taaataaata ggaatcaact gacaagaaag
4861 tgattcaaat aagataatag ttttggatat ttggacactc aaactatcaa atatagatga
4921 aaaagtttct gaaatgctga gatattctat tgttaaactc ttaaactctt attttctaaa
4981 ttgtaaataa tgattgaagg atcactaata atccagcttc ttaaccaata gagttctgtc
5041 tgtgctaaac cctaagcatc aaaagatgga aatatctgac agtaagttac aaaaaaagga
5101 tccaagttct tcagaaatgt gtcattggag tagtccatat cttttccttt tatcagtgaa
5161 acagatatag atccccccagc aaacagattc tttaatcctt ttccaaagaa aacatcattt
5221 tttaatgcta acatttaaca aacataaatc ttgttcccac agttaaaatg cagattgagt
5281 taaaatttta tataatttaa tttatgataa naaataaaat ccagacaaac agtatttcag
5341 attatttttt gtctttttat atacttttct ccaccatatt ctaaaaacag aagataattt
5401 acttttcttg atttttgtca ataatttttt tttccctccc aggggaactt ggtgtcaaat
5461 ttagctgtta aaatacaaac ttcttaaata gcactattaa atgtatagta ttacatgtgc
5521 cttgtgatat tattattatt gtattttgag tgcttttggt tttacaattc ttgcattttt
5581 tttttaacag aaacatccta tcaagcacca aggactccct caagtaagtg ttctattcta
5641 tgttccaaga actcactgta aattgtgtaa cttaagtgat gataaattgc taatatatat
5701 attgtagtct cattccttcc ttctctagta aacagccagt ttcacattcg ctgaggtgta
5761 atatcttcaa ctattgagct gaatattgat ctgttctcac aaacctttt agagaagagg
5821 gcatttggac ttacatattt atgattaata aaaattttat tatgcaagag cagtagttaa
5881 acagaatatg tatatgtggt ctattttact gtattattga ttctttctat ctcttctcct
5941 tgcactcaaa catatgattt acaacttgat ctcaatttac acactgagta ttaaattaga
6001 tactctatca cacactttta caagagtcat taatcttcca ccctatccac tactcagttc
6061 atttttaat gtttacagaa tgagccagac agcagagaga taaggttctt tctttacccc
6121 aactaatttc tctttcagtt agctatgata aatactgaat atcttcttca tttcaatttt
6181 catttgtata ccttcatttc aattgtatcc agatactcag atagttgatt accctctcac
6241 ttctctgttt tcattacatt cctgtttaat tcctcctttg acatttgata agtatgaact
6301 gactaaacct ataggcctgg taaagaagga accaaatatg gtgtctcttt cagattttaa
6361 aataaaatca taaaactgaa actaattatc tcttatttta cctaagaaat attttggttt
6421 acctaaataa atggagaatt tgtgttcaaa tggaaaaaca ttctcctttt ctgactgtgt
6481 ttttcacttg tacaattcac aatttaattc ctacaggaag tcctcaatga aaatttactc
6541 aggtttttg tggcagtaag tattatctac ttcttcttca atgacaaatg tatttttctg
6601 gaaaaatcaa ctcaatttct atttacaaac atttttcac ttggtcaaat attttcattt
6661 actcacccaa gcaaccacaa agaatattga aattatatat gcaaatatta aaaagctgct
6721 ttaaatatta atctgtacct atacatttga ctctctgtaa aacacttacg ctgttttgtg
6781 cataatctat gtaaacttta caatttccat tctcattatg atggaatgtt ttatctctac
6841 aataccctgg atgtgttgtg tcttctgtaa tttcatgact gaattgtcaa atagaacaca
6901 aatagaaatt ttatttcaag acaagtattt aaaagatttg ataggcaacc caatttagcc
6961 tgaatgattt ttaatataat ctttttccct tgtagccttt tccagaagtg tttggaaagg
7021 taagaaattc tgaacagaat atactgcaga attaacaaag cattttattc tatgttattt
7081 atggtgttat gccaattctt tattgccttt ttgtgaaaaa atcacttaga ttattattag
7141 attattgaac cagacagaga gatgaagata actcagaact atgacttaaa caagatgaaa
7201 gataataata gtcaaattta ctggaacaaa aatttttaaa tatttttaag cacacagtaa
7261 tgaactgaca gcttaaataa agaaaactta gcaaggagat aatacaagaa ataatattgc
7321 aaaaaatatt agtgcattca caaaaagcat atttatttt taatttgcag cagcaaaatg
7381 taagatacat ttctttttt ttttttcat ggaatgaaaa aaaatttat tatattcatt
7441 ttctccattt tatgttttaa ggttaacatt ttatctttac tatcttgcat tatcaaatga
7501 caactcagaa atgcaagctt aaaagaggaa ttggtaaagt ggagaaagct gtgcagttct
7561 gtctttgatc tatttaattc agcttgatga tgaaaacgcc atatacattt cttatagata
7621 tccattcatt catgtttctc tcttcttttc tgtccttaag gagaaggtca atgaactgag
7681 caaggtaagg aacataaatg atatttaaat tattttaaag ttatctcaaa atcatatttt
7741 gcaacctaca attgattggt tttgacaaag agaatttttg agggaaaaat tttcgtatga
7801 aaatacgaaa attaaggtgg tatgtcaaaa aatgcttgtg cagtatgaaa aacattttag
7861 aaaaaaaaat ttcatgtaac atttatactt ttcttgaact ttcaaaatga ctgatgagat
7921 ctactgagaa ttctgaatat tgatttcatt gatttaactt tgcaaccatg aggagacaag
7981 gcattagtat aaatgaatga atggataaaa gaatgaatta gtgcaattat ggactgacag
8041 actattttca ctgatcgcca aaacagaaaa aatgttttat tgagtactac atgaagcaat
8101 atatctgctc caagaatgtt ttacatagtc acaaatgatt tttttacttt ctttaaacac
8161 aagtttttga tcaaggttgt ttttcaaagt tttaatagct ttgttattaa aacaacatac
8221 aggtatttaa aaatttacac cttcttaatt attcttcata cctgactaag taaatttttct
8281 tggcaggata ttgggagtga atcaactgag gtaagattct ttattctaaa actattaaat
8341 ataatataag ggaaataaag aagtaaaaaa tatcctaatc aaatatcctt taagtatcct
8401 aatcaaaatg aatgaacaat tctctaagga caaaactaaa acagatatct ctaattcaaa

```
 8461 gaaaaaaaga tgcaacatgt atgttgacca aaattggctg gataactaaa ccaacatagt
 8521 attttgagct taaattctat ggagcatctc actgctctct atactgctat ggtcatggaa
 8581 actggacaca acatactaca tggagaaaat tcattttgct taaatattta tacatcattg
 8641 aatcaatgct ttgtgtactt ttatttatca cataatcttt aatccaatgc tctggtttgt
 8701 catccaaaag ccaagttaaa aaaaaaagaa gttaggctag ttaaaggaaa atgtaggata
 8761 tttacaatga ttcatcaaaa catagaaaaa tctacagtac tttttattga actctaaaat
 8821 gcagattaaa gagaaaaatt cttcagttag cctggtaggt aggctttttt taaattttcc
 8881 ttaccacttc actattgcca cccatttcta tttcctttgg catccatttt atttggtaat
 8941 tatcatttta tattggaatg tttgtcatga aaaataaaat taattctctt ttctttccta
 9001 aggatcaagc catggaagat attaaggtaa gatctttatt ttaataaact ctacacttat
 9061 atatcataaa taggatatgc tctatgcttt aagaaagcta tccactgcaa ttgtgtgatt
 9121 gaactcctaa agcagtgcta tcatacccaa gaatgtacaa tgttgtgcca gataaagtta
 9181 aagtaaggag aggaaatttg acatttacat gatcaggaga aactttgctt ctttattaaa
 9241 atcaagagta aaaagcaaaa tgtgcatgag tgtttaaaaa tgaaacaacc gactcaaaaa
 9301 attaaaacct tgtttctgtc atcccaggta agagttcact ctcttgtgtg ttgagggtgg
 9361 tctttgggca gaaagaatct attgcagcaa ttatgtagga tcagtgggct atgctgttgt
 9421 tcagtcgctc agtcatgcct gactctgtga ccccattgac gagcatgcca ggcttacctg
 9481 tccttcacca tctcccagag cttgctcaaa ctcacgtcca ttcagtcagt gatgctatcc
 9541 aaccatctca tcctctgaca tccccgtctc ctcctgccct caatctttcc caggaccagg
 9601 gtcttttcta aagagtgggt tctttgcatc aggtggccaa agtattggaa cttcagcttt
 9661 agcatcagtt ctcccaatga atattcagga ttgacttcct ctagtgattc ctccactctt
 9721 ctccagtagc atgttgggca cctactgacc tggggagttc atctttcagt gtcatatctt
 9781 ttaacctttt catactgttc atggggttcc caaggcaaga atgctgaagt ggttttccat
 9841 tccattctcc aatgggctat gctacctaat ctcaaaacat gggaaaatat tcattagccc
 9901 gcaatacttt ttactcaact ctcccaagag atgactattt tgctgcacat tagaatcaat
 9961 atcttagtat gtaacttggg aacagaaatc atatccatat gtaaatatga gcatctgtga
10021 cttaggaaaa aaattaaatg tttcacagag tagaactttt aaaagctcat gtaattttaa
10081 gttccattct gaaattgatt gttctatttt aatatttcca atttatttac atcaaaattg
10141 ccccaagtgt atatgggaat ataaactcaa cccagggttt caggtagccc agctcaatag
10201 gattcttcat tccaggtttc atttttatca caaaaattca tatgcatttg cattctactc
10261 aagaaagtat ctgtatttgt tttaataaaa aggttttaag ttacaaagta agttattaca
10321 ccccccctct caacatattt taaataaaat tgacaatcca aaaaaaagta aattttattg
10381 gccctgaatc tttttatacc caattgtttt cactaaaaac tagttagcaa cccagtatga
10441 aagtgtgttt caaactgtgt ttcataatag tttccttcta attctaaaag tctcagaggc
10501 agtaacaatg attttctttc ttttagcaaa tggaagctga aagcatttcg tcaagtgagg
10561 tataccattt ttatgttaat tcaatatccc aattataaaa tgtttatgaa agtttgttga
10621 accataaagt ttacatgtcc cttaaggttt cattgcataa ggcactatgt atgcagctct
10681 atcctaattt taacatacaa ggctgtcaac cccaaaatgt aagaaaacaa ttatttgtca
10741 gagacaaatt atagatttct ttctgagcaa ataatcctat ctgtatatct gtggtgtcca
10801 gacattattt ttctaatttt tttttccttc cagattttgg tagaactaaa tcttgtcact
10861 gaaaattacn ataatggcta tgttctaaat tgtgaaattt aaacatgata tatcctttgt
10921 ggaactcaat caattctatg ttaaagagac actggaaccc tttgctattt gttaacctag
10981 aggagaaaaa gagaaggcat acaaggtgat gaagtcttaa actttgggag cagaacttca
11041 tccaatttac aaaataaacc cagttccatt atgcggagtt tctgtgaagt cagttaaaag
11101 tgaatacttc ataatccaac attttgaata acttccctgg tgaaagtagt gagaagagaa
11161 aatataacat gtactttgac tgttcatctt atctaaatta cccaggaatt tgtggctaaa
11221 atcagtttac cacagataag ctatgatgtg tctggttaat tagcattttt atcttgaatg
11281 taaattaatg tcataaaact aacaatacat gttttttatt ttttaaggaa attgttccca
11341 atagtgttga ggtgagatat atctactaaa tttaaaatat attaacatca tccaggatat
11401 cttaaaattt aattaaactt tttatttttt gaatttttta gcagaagcac attcaaaagg
11461 aagatgtgcc ctctgagcgt tacctgggtt atctggtaaa attttattaa aagttaatca
11521 aagaccaatg tatcagggaa tgagcaagaa tgttgtattg ataaattatc tctccttttc
11581 aatatctgct aaacttaaag taagcagtct aacagattct agcagtatac tgatcccttc
11641 tgaaaataaa actgacaact tttttaatcc cagatattta attcatactc tcatttgaca
11701 aattgtgtat tacagtttac tttcagatct gattagcaca ttattgaaat ggcacgctat
11761 agatttgacc agacttgagt ttgactctaa gctctacgtt ttaccagcat gtggtattag
11821 acagtttcct tatctgctgt aagtctataa cctaagttcc aaagtgacaa aataataata
11881 aatatttaaa gcatataaag caattaatac aaataagaca tagtgcaaag taaagtgaag
11941 tcgctcagtc atgtccgact ctttgccact ccatggacta tanctaccag gctcctctgt
12001 tcatgggatt ttccaggcaa catttccttc tgcagaggat tttcccaacc caggaatcga
12061 acccaggtct ctcgcactgt agacagacgg tttaccgtct gagccatttt gaattgcgaa
12121 gtaagccctc ggtaaatgtg gctagtatca gaaacatcat ttttcccttc atccaatatt
12181 ctgatttaat ttgcaaaaat cagaaaacaa atggggcttc ccaggtggtg ctagtggtaa
12241 ggaacccctc taccaatgca ggagactcag gcttgatcac tgggttgggg aaaatcccct
12301 ggaggagggc atggcaaccc actctagtat tcttaccttg agaattccat gaacaaagga
12361 gcctggaggg ctatatccat aggatcacag agagttgggt acaactgagg tgactaatca
12421 ctgcacagaa agcaaatctg ttttcaattt tcttagcaat tctgatataa tattcactgt
12481 aatcaaaact gcctcagttc agttcagttc agtcactcag tcatgtccaa ctctttgtga
12541 ccccctggac cgcaacacac catgactccc tgtccatgac caacacccga agcttgctca
12601 agctcatgtc cattcagtga tgccatccaa ccatctcatc ctctgtcctc ttctcctctc
12661 ctgccttcaa tctttcccag aatcagatct tttccaatgc atcagttctt caaatcaggt
12721 ggccaaagta ttggagtttc agcaaaagtc tttccagtga atattcagga ctgatttctt
12781 ttaggattga ctgaatactt atttatatct gactgcatta taatatattt taagttaaaa
12841 ataatgtaac aaagtagttt tcctaatatt aaaaaaaaaa aagaagaaga attaggctaa
12901 tccaaattct ctgtggtaga aacaacacta aagtaagatt attagtttct cattatttac
12961 tcctgggaaa gagatactat gatagatggt atccacgaaa ttgacaatat tttttctttt
13021 gaataggaac agcttctcag actgaaaaaa tacaaagtac cccagctggt aatattttat
13081 tataataata caaaattaag tctacagaat taaaataatt aaatgaaatt tactttgact
```

13141 aaattctaca tcaaatcatg ctagagcctc ccaaatgatt cgttatactc tggaatttat
13201 tctgtattta ttttcccaaa ttcagttctg agtggtgatt ccgtatatat gcctccttgt
13261 gaaagcacag atcattgcca aataaacatt ccatttaaga aaacagtgta gactctgtag
13321 aaggaatcag ggatcccata cattcaaatt gcttttgtca aaatttttcta agaaagaatc
13381 ggcaagttga ttgatctcat tgtctctaga cccgtttcct ttatagttac tttcatttcc
13441 tgacattgtt ttgaatattg aatgagcaat ctcctattca gttatccctg aggctagctt
13501 gtggcaagtc caataaagtg ttcataaagt ctctccttcc cctaatccct aagctctata
13561 aaatttgcta tgttcatgag acctttgaca atattatgaa gattttgttt tgtttaatta
13621 cagataagct atgatgtgtc tggttaatta gcatttttat tttgaatgta aattaatgtc
13681 ataaaactaa taatacatgt tttttatttt ttaaggaaat tgttcccaat agtgctgagg
13741 tgagatatat ttactaaatt taaaatatat taaaatatat taaatgcact ataagatgtg
13801 catttgaact atgaagtcat aaactatctg ctgatagtac ggatagatca aaaaagtatc
13861 cattagcaca atccagctta cactattatc aatccttttg agtgtcatag tttttcactt
13921 gggaggccat aaatttttct ttttagttat accaaataag cttggttaat taggacttta
13981 gggaaaagga cttgtcttaa tttagcaaaa ttacattttc acattcttat tctgttacaa
14041 ttgtaggaag gaggaagtca ataatgaatt taatttggac ctcagtgtca ctcttgcttc
14101 agtttccaag ttgcattctg ttgccttgga cttagaagcc aaataaattt agatatatca
14161 gtaccccaat caaacttagc attactaatc taagaaatta catgaataat atattttagc
14221 accttagtta tgatatgatt aggccataac ctccataagt gttagaatga tatggtcctg
14281 ctcttagtgt tcattgttca tagtgttcac tgctcatagt attcattatt atttatttca
14341 tagatggtcc ctagcattta gcgtagaccc ccagtgaatg aaagtgccta gcaaatgtct
14401 gtgattcaca catgagtgaa atcaatgtac taaaataaag tgatctctct ttctgaggaa
14461 gctattcatg accatataag attattaaat ctagaatatc ttagtatttt cttcaagtgt
14521 tgtttttata gtgtgtgttc ctctttcttt tttaacagaa ataaatataa tccaagggag
14581 aagtttttcc tactacccat atttgataat gttttattta aaatttataa aagataaaat
14641 gtaaacaagt gattaatagt catgaacaga aaatatgtag tcctagggaa ttggtggatt
14701 attttccccc caaatatgaa acttatagat tctttatcct tccaaattat agtaagtata
14761 acattgagtt gaatattttt ttctaaagtt agaaaatgaa agctatattg ctaattttaa
14821 caatttggc agagaatacg tttatactaa tcttcacaga tgatgacata ttcttggtaa
14881 ttgtaaaatg aatgtgttat atcagaagat atctaagtaa cttagaacac atttccttct
14941 atttaggaac gacttcacag tatgaaagag ggaatccatg cccaacaggt aatattttgc
15001 ttaatgaatt acatactgat aatatgttgc aaagtttaaa tacgtttgct tttaaatagc
15061 ttcacaattt tgaagagact tctttcttct caggtggtta ttttttactc ccaactttct
15121 gaaacaagtt gtagtagtca atctgagaaa ggagaactat ttgtcattct tctttggaat
15181 gggagctgct ctattctgtc cgttctcccc tgacagcagc acccttttgta attctaaaca
15241 ctatgaagtg ttcatttcca tggttatgtt aatttctctt tttaaggtat tcctaaattt
15301 tatatcttta tgtccatgat tttattctta aattcaattt atatttagga aattcaaact
15361 cataattatg ctctttcaag caagccattt ggagatttat taatactaat ttaaaccttt
15421 agaagaaact gtctggagag atattaaatg tttctgtcaa aaaacctttc tgatgttctg
15481 ttttcccttc catctgtaca gtgtatttat ctgtaatata taatagaaac tgattcttcc
15541 tctgccctac atttcattga ttcaatagtg ttcctggggc ttctgatttc atttttaagg
15601 tgacaaccag atactgtgaa atatgtgctg ctattacaat tctgattcct atcaaaagct
15661 atgccaattt caatcactac tgtcatttgg tgaatcttat gaggattgtg tatgttttgt
15721 atgtcttaga agaaatagta tctagataat gcaattggca aaattattag tgtgaattcc
15781 cagatctcta acatgaaaag catttcaaaa agtttgcttc ttacattttt tggttttatt
15841 cagccttaaa agatcacccc tactcttttt ttttttcttt ccagaaagaa cctatgatag
15901 gagtgaatca ggtaagtgtg tgtctgtctg tgtgtatttt aatactgccc caaactatct
15961 attggtaacc actgtttttt agaagttatt ttactagttg catcaagaag tagtgctctg
16021 ctagtatttc ttagttgttg actggtgttg gactctaaca ttgctggttc cagcattgct
16081 atgaaagatc agatctaagt tgccaaagga aatgaaagaa caattctatg gccatccctg
16141 agtgaagaga gcttcaggtc aacttttaag ttagaacaac ctcaatcaac tattcatatt
16201 gttcaaattt gacttttttat taatctacag tactgtctgg gcttcttgag ttttcagtct
16261 tagatctaat cttatgacct tgagcaaact gtgttcttct ctctaagccc tgctctttca
16321 tttgtgaaat taatatgaat ctatcaattc attgtcttcc aaaactcaca tatcttgagt
16381 tttgcagtgt ttttttgtgt gtcttcactt ttaaaggtgg taactgcatt cagttcagtt
16441 tctttggtcc attgatgatt gagcttacat atagtatgct aatttaaaaa ataaaccata
16501 caaagaaaac catggggttt tataaatatg aatctataag ggattaaaca agattcaaaa
16561 cacagtagtc atcaataaat tttatgataa ttattcctta atatgaatca tatttcgctc
16621 tcaatttctt ctaaaattct ctcttcactc atgggccaaa gtaagagaaa agaaaagaca
16681 attatgtact tcactgaatt ctaaagagta cactatcctc acctttccat gtaatttttа
16741 aaaaagaggg taatcattct aatcttaatg tgaacatttg gtagtaatac tttagtctat
16801 tatggcattt aactgggttg gaatcacaaa actatttttc cctctctctc ttttaaggaa
16861 ctggcctact tctaccctga ggtgaattta ttttattttt tatattaaca ctaataagag
16921 aaaatctcag atatcatatt tattataatc attaggatag gcaaccacag caaaacttga
16981 aatgctttct ttggtttttc taattccatc aaagaacatg tattttctac ctgaatatat
17041 aaagaattac attaatacac agaaacaaat aaaaggccat gatagattaa cagaggtacc
17101 tattccaaaa tttcagaaaa gaaataatgt ttttgcacca aaaccttata gttatgaagg
17161 ggcactgggc taggaattag agaattcttt tctctttatt cttggcattt ctgaaatccc
17221 tgtattattc tgggacgagg actctccctt tctttgagct cattgtcttc atggaaaact
17281 aaactataga acagaaaggc aaaaatgacc ttctctagtc caaaaaccca tgagttaca
17341 cattatataa taattaaatt cacactacca ctttcctctg aactttgata gctggacctg
17401 tgttagtaat gcaggttctt agagttcaat actgactatt gatcttaaag aagtgattta
17461 ttcattaccc ctcaggtacc ccaagaaaag ccctttccaa aatttcctta ctgtgactta
17521 ccatagggaa gagtttgatt tctctgtttt cctcacagag taaacatctc ttgtgatgcg
17581 aatagccatg tctgaaatga aggcaatgat tcattttcag agattcaaaa ctgatttctc
17641 atacactgtt gctttttcaa tggtctttct ctctagcttt tcagacaatt ctaccagctg
17701 gatgcctatc catctggtgc ctggtattac gttccactag gcacacaata cactgatgcc
17761 ccatcattct ctgacatccc taatcccatt ggctctgaga acagtgaaaa gactactatg

17821 ccactgtggt ggtaagttca tttaaatgac tgcatattgt tgccttatca aaggaaataa
17881 aagaaaacat aatataaaaa tagatttaga ataagcatga cacataaatg cttagtgtcc
17941 tatactacaa ttttctgaaa tggaaaattg atgataactt tctggtatat ggctaatgtt
18001 aatccattac tcaggaacat gtggagcagt gctatctatt cgataagtga taatcattct
18061 gatgaaaata ggaaaatttt ctctccaaag taaaaattca actttatcct ccttgcactt
18121 ttgctaatct ttaaatgcct ttcttttgat tatatccatg atatacatta gaatgcattg
18181 tgggggataa actgcagatt ttggcattcc taaagtccca acttgaaatc ctgatctttt
18241 ttttttttt ttttgctta cttgaaatag tataatgatg cttgttttta taaccttgaa
18301 ggtgattaaa tataataatc cattaagcat actgctggga aaattagtgc tcatttttg
18361 attcagcaaa attttattac tgaatacctt actcacattt taccaatttt tgctccttca
18421 gaagagtcaa gtgaattctg agggactcca cagttatggt ctttggtaag ttggaaactg
18481 cttgtctaat cattgatcct cttttcatat gagagctcag tacaaaagta caacgtgtag
18541 actataaagt tgttttgctg gtcctctagt ctagctatat ttaaacacat tacacttaga
18601 taatatcaat aattaaattg gcttcaacat tttttgtatt atagtaatat caaatttaag
18661 tagcaataaa actagctaat tttaattaat ataatttatt atattgaaat ttttatacaa
18721 gtataggagt gtgggtgtta ctaattctgg tgaccccaac agtggaatcc tattctttat
18781 gatcactgaa atagaaatac ttaatacgtc acatagttaa atcaagtgtt tgtctattaa
18841 gaagaaaaca aatttaaatg tccatagatg tgttttgagt acttcaattt actatatagg
18901 agcttgtagt tgtgtgtgta tgtgtatgtt ccctttctag tggtggaaac ttcccttata
18961 acagatcaaa agatggaaaa taaaagtctg acgtacttga agtttttact ttgaattatt
19021 ttgccatatt ttctgttact gcaaaagaaa gtgcaaccta gtcataaatt gcctgtaaaa
19081 gttaaaaatc aggagcagtg ggtatgtgtt agtaacagga gacacatata actttgtgaa
19141 cttcatctta agatgaagag aaaatgacta ttaaagtgta tcttatcata acagtacctt
19201 ctcccttcaa aacatgcagc ataactaacc acatatttct tttttggttt acagatggtt
19261 ctgaaaattc catgctctac atgtcttttc atctatcatg tcaaaccatt ctatccaaag
19321 gcttcaactg ctgttttaga atagggcaat ctcaaattga aggcactctt tcttcttgag
19381 ttctctactg tattttagat agtgtaacat ccttaagtga aattgtccta acagcttgtt
19441 acctaaattc cagtagtatc atgctggtat aaaggccact gagtcaaagg gaattaaagt
19501 cttcattaaa tttctgtatg gaaaatgttt taaaagcctt tgaatcactt ctcctgtaag
19561 tgccatcata tcaaataatt gtgtgcatta actgagattt tgtctttctt cttttcaata
19621 aattgcattt taaggcacta ttcctatttt ttgtcattat tccattggaa ggaatttaca
19681 caaccttgtg agtttgtgtg tatataacat tttgttttca ctaaattttt atgacatttt
19741 caaccacatt ttaatgaaaa aattcaaatg ttcacttcta gctgatcctg gtagattata
19801 aactgagtct aagatctttc atttgaagtc aactgtttat agaatatttt ccatgtgaac
19861 atggacgtgg ctgcagagag aacagcagtg tagtagttgt gggtgcactg gcaatttgaa
19921 caacctggtt atactgttta atttggcagg acttcttgga gctcaatcca tttgattata
19981 cttgagttta taactaggta gattacaaag taaagattgt aaaggttaac acaggacaca
20041 cagtaacaga gcattcatag tgagctgttc agtgtcctcc aaaactctaa tgagtaatta
20101 cttctcaaaa tgacaacagc tgcttactct gatatcttta aagccccagg tgaaatttat
20161 tttccgggct ttattctagg attcccaacc aacccaatct gtcatgactc tttaagaacc
20221 ttgtatgacc caggtcacaa aacacaggta ataaggtctc ttgtacaaca gtttagacca
20281 tctatatcag ttgttctgtt tactgattat gtagatatgt atacataaca ttgaaacatg
20341 taatttttgc cttgaaaatt ttttttaaa aatccactta gtattaggtc ttctgcctcc
20401 tctatgttaa gtacaaggta caacaacttg ttattcaatc tagagaaggt ttgtcaacat
20461 actcagaata ctggacctgg cagaagttaa cggagggtgc agattcatac atcttcataa
20521 aatgtatttc ccagcttctg tcatcatgtc agttcccaaa gaatcaggaa aacacaattt
20581 catgtgttgt gaatcctata gataatgcca tccatttaat gaatgtcttg taaacagtag
20641 catggtaaaa gtcacgtttt ctgtagatta aactgtggta caaacaggcc aatatcactg
20701 atgaacatag atgcaaaaat ccttaacaaa attctagcaa tcagaatcca acaacacatt
20761 aaaaagatca tacaccatga ccaagtgggc tttatcccag ggatgcaagg attcttcaat
20821 atctgcaaat caatcaatgt aatacaccac attaacaaat tgaaaaataa aaaccatatg
20881 attatctcaa tagatgcaga gaaagccttt gacaaaattc aacacccatt tatgataaaa
20941 actctccaga aagcaggaat agaaggaata tacctcaaca taataaaagc gatatatgac
21001 aaacccacat caaacattat cctcaatggt gaaaaattga aancattccc tctaaagtca
21061 ggaacaagac aagggtgccc actttcacca ttactattca acatagtctt ggaagttttg
21121 gccacagcaa tcagagcaga aaaagaaata aaaggaatcc aaattccaaa agaagaagta
21181 aaactctcac tatttgcaga tggcatgatc ctctacatag aaaaccctaa aaactccact
21241 agaaaattac tagaactaat caatgactat agtaaagttg caggatataa aatcaacaca
21301 cagaaatccc ttgcattcct atacactaat aatgagaaaa cagaaagaga aattaaggaa
21361 acaattccat tcaccattgc aacagaaaca ataaaatact taggaatata tctacctaaa
21421 gaaactaaag acctatacat agaaaactat aaaacactgg tgaagaatca agaggacact
21481 aatagatgag aatataccat gttcatggat tggaagaatc aatatagtga aaatgagtat
21541 actacccaaa gcaattatag attcaatgca atccctatca agctaccaac agtattcttc
21601 atagagctag aacaaataat ttcacaattt gtgtggaaat acaaaaaacc tcgaatagcc
21661 aaagcgatct tgagaaagaa gaatggaact ggaggaatca acctacctga cttcaggctc
21721 tactacaaag ccacagtcat caagacagta tggtactggc acaaagacag aaatatagat
21781 caatggaaca aaatagaaag cccagagata aacccatgca catatggaca ccttatcttt
21841 gacaaaggag gcaagaatat acaatggatt aaagacaatc tctttaacaa gtggtgctgg
21901 gaaatctggt caaccacttg taaaagaatg aaactagagc actttctaac accatacaca
21961 aaaataaact caaaatggat taaagatcta aacgtaagac cagaaaatat aaaactccta
22021 ggggagaaca taggcaaaac actctctgac atacatcaca gcaggatcc

## IV. Exon skipping reported in major milk protein genes

| *Ref.* | *Species* | *Gene* | *Exon Skipping* | *Related Mutations* | *Truncated Protein* |
|---|---|---|---|---|---|
| Mohr et al 1994 | Bovine | $\alpha_{S1}$-CN A | Exon 4 (completely) | Point mutations within splice site | Yes (variant A) |
| Ferranti et al 1997 | Caprine | $\alpha_{S1}$-CN A, B, C | Exon 13; Exon 16 (partially) | Non mutation detected | Yes (7-16% of full-length) |
| Leroux et al. 1992 | Caprine | $\alpha_{S1}$-CN F<br>$\alpha_{S1}$-CN D | Exon 9,10, 11, and >7 other minor transcrits<br>Exon 9 | A single nt deletion in exon 9 and two insertion in downstream intron | Yes (variant F and D, low lever); likely yes for the minor transcripts |
| Alexander & Beattie 1992 | Porcine | $\alpha_{S1}$-CN | 18 nt and 60 nt exons (Partially) | Unknown | Unknown |
| Ferranti et al 1995 | ovine | $\alpha_{S1}$-CN A, C, D. | Exon 16 (partially) | Non mutation detected | Yes |
| Johnsen et al. 1995 | human | $\alpha_{S1}$-CN | Exon 8 (partially) | Unknown | Unknown |
| Bouniol et al. 1993 | Bovine | $\alpha_{S2}$-CN | Exon 8 (completely) | G→T transversion at the last nt of exon 8 | Yes (variant D) |
| Boisnard et al. 1991 | Ovine | $\alpha_{S2}$-CN | A 27 nt exon or 44 nt untranslated exon (partially) | Unknown | Yes (30-40%) |

## V. Single base-pair substitutions in 5' splice sites of human genes

| *Gene* | *Exons* | *Mutation* | *Disease* | *Comment* |
|---|---|---|---|---|
| *Mutations of the invariant G (+1) at 5' ss* | | | | |
| APOA2 | 4 | AA(g→a)taagt | ApoA2 deficiency | |
| APOA2 | 4 | AA(g→a)taagg | ApoA2 deficiency | |
| AR | 8 | TG(g→t)taagg | Androgen insensit | Cryptic, 123 nt upstream. (AGgtgtg?) |
| ARSA | 8 | AG(g→a)tagga | Leukodystrophy | |
| C3 | 41 | GG(g→a)taagg | Complement C3 def | Cryptic, 61 nt upstream. (??gtgagt) |
| CETP | 16 | TC(g→a)taagt | Lipid transfer def. | |
| CFTR | 27 | AA(g→t)tatga | Cystic fibrosis | |
| CFTR | 27 | AG(g→t)taata | Cystic fibrosis | |
| COL1A2 | 49 | TG(g→a)tatgc | Ehlers Danlos VII | 100% skipping |
| COL3A1 | >45 | AG(g→a)tgagt | Ehlers Danlos IV | 100% skipping |
| COL3A1 | >45 | GC(g→a)taagt | Aortic aneurism | 1% skipping; 38% cryptic, 24 nt dnstm |
| COL3A1 | >45 | AG(g→a)taaac | Ehlers Danlos IV | 70% skipping; 21%cryptic (GGtataa) |
| COL3A1 | >45 | AT(g→a)tgagt | Ehlers Danlos IV | 100% cryptic, 30 nt downstream |
| F9 | 8 | AG(g→t)tcata | Haemophilia | |
| | | AG(g→t)tactt | | |
| | | AG(g→a)tcata | | |
| F11 | 15 | AG(g→a)tgagt | Factor XI deficiency | |
| GNAS | 13 | AG(g→c)tttgt | Osteodystrophy | |
| HBB | 3 | AG(g→a)ttggt | Thalassaemia β | No wt or skipped RNA, some cryptic |

| | | | | |
|---|---|---|---|---|
| | | GG(g→a)tgagt | Thalassaemia β | Crypic, 47 nt dnstm, some skipping |
| HEXA | 14 | TG(g→c)taagg | Tay-Sachs disease | 50% skipping, 50% cryptic |
| PAH | 13 | CC(g→atgagt) | Phenylketonuria | |
| | | CA(g→a)taagt | Phenylketonuria | 100% skipping |
| PBGD | 15 | CG(g→a)tgagt | Porphyria, acute | |
| PROC | 9 | AG(g→a)tggga | Protein C deficiency | |
| RB | 27 | AC(g→a)taagc | Retinoblasma | 100% skipping |
| | | AG(g→t)tattg | Retinoblasma | |
| UROD | >7 | AG(g→c)tgagt | Porphyria, cutaneous | 100% skipping |
| *Mutations of the invariant t (+2) at 5' ss* | | | | |
| COL1A2 | 49 | TGg(t→c)atgc | Ehlers Danlos VIIB | 100% skipping |
| F9 | 8 | TGg(t→g)aagc | Haemophilia | |
| HBB | 3 | AGg(t→g)tggt | Thalassaemia β | |
| | | AGg(t→c)tggt | Thalassaemia β | |
| OTC | 10 | Agg(t→c)atgc | OTC deficiency | |
| RB | 27 | Agg(t→c)tagt | Retinoblastoma | |
| SPTB | ? | CGg(t→a)gagc | Elliptocytosis | Some skipping |
| *Other mutations at 5' ss* | | | | |
| AT3 | 7 | A(G→A)gtgagt | AT3 deficiemcy | |
| COL1A2 | 49 | T(G→A)gtatgc | Ehlers Danlos VII | aa change, 3% wt RNA, 7% skipping |
| F8 | 26 | T(G→A)gtaagc | Haemophilia A | |
| HEXA | 14 | T(G→A)gtaacc | Tay-Sachs disease | No aa change, 3% et RNA, 7% skipping |

| | | | | |
|---|---|---|---|---|
| PBGD | 15 | T(G→A)gtaggg | Porphyria, acute | 100% skipping |
| HBB | 3 | A(G→C)gttggt | Thalassaemia β | aa change, few wt RNA, 4 crptic forms |
| LCAMB | ? | ATgt(g→c)agt | LCAM deficiency | 3% wt RNA, 97% skipping |
| OTC | 10 | AGgt(a→g)tgc | OTC deficiency | 100% skipping |
| SPTB | ? | GGgt(g→t)agt | Elliptocytosis | Skipping (rate not determined) |
| CAT | 13 | TGgtag(g→a)t | Acatalasaemia | 100% skipping |
| COL1A1 | 51 | CTgtaa(g→a)t | Osteogenesis imp. II | 100% skipping, temp.-dep. |
| COL3A1 | >45 | ATgtga(g→t)t | Ehlers Danlos IV | 100% skipping, temp.-dep. |
| F8 | 263 | CAgtga(g→a)t | Haemophilia A | |
| HBB | 3 | AGgttg(g→c)t | Thalassaemia β | 50% wt RNA, 3 cryptic forms |
| | 3 | AGgttg(g→a)t | Thalassaemia β | |
| | 3 | AGgttg(g→t)t | Thalassaemia β | 15% wt RNA, 85% cryptic forms |
| HPRT | 9 | ATgtaa(g→a)t | Lesch-Nyhan syndrom | |
| F9 | 8 | AGgttt(g→a)t | Haemophilia B | |
| PS | 15 | CGgtaa(g→a)c | Protein S deficiency | |
| PROC | 9 | CGgtga(g→a)t | Protein C deficiency | |
| | 9 | CGgtga(g→c)t | Protein C deficiency | |
| | 9 | CGgtga(g→t)t | Protein C deficiency | |
| F9 | 8 | CAgtgag(t→c) | Harmophilia B | |
| HBB | 3 | CAgtgag(t→c) | Thalassaemia β | Almost normal wt RNA amount |
| F9 | 8 | AGgtaagt(a→g) | Harmophilia B | |

Data adapted from the review by Krawczak et al 1992

# *List of publications / presentations*

**(1996-1999)**

## SCIENTIFIC ABSTRACTS

**Alternative expression forms of the bovine $\alpha_{S1}$-casein gene**

*Tao Xie and Richard J. Wilkins*
*Poster presented at the Queenstown International Molecular Biology Meeting 1998.8*

**Expression of the bovine $\alpha_{S1}$-casein gene – when the cow skips over the moon**

*Tao Xie and Richard J. Wilkins*
*Poster presented at the Queenstown International Molecular Biology Meeting 1997.8*

## PUBLICATIONS

**Designing DNA tests for the rarer milk protein polymorphism**

*Richard J. Wilkins and Tao Xie*
*IDF Bulletin, Milk Polymorphism II, Brussels 1998*

**Two distinct gene mutations - one milk protein polymorphism - the example of the $\alpha_{S1}$-casein A variant**

*Richard J. Wilkins and Tao Xie*
*IDF Bulletin, Milk Polymorphism II, Brussels 1998*

## PROVISIONAL PATENT

**Improvements in and relating to RNA testing in dynamic systems**

*Wilkins R.J., Xie T. and Cursons R.T.M.*
*NZ Patent Application No. 328492*

## TECHNOLOGY TRANSFER

**Occurrence of variant of the $\alpha_{S1}$-casein protein in milk**

*Richard J. Wilkins and Tao Xie*
*Industrial report for Dairy Research Institute 7th December 1998*

**Genotyping of dairy cattle–analysis of dairy research institute animals for the $\alpha_{s1}$-casein A variant (DRI) allele**

*Richard J. Wilkins and Tao Xie*
*Industrial report for Dairy Research Institute 20th March 1998*

**Development of a PCR test for the alpha-S1-casein variant A (DRI) and the genotyping of the calves**

*Richard J. Wilkins and Tao Xie*
*Report to the Dairy Research Institute. 1997*

**Development of a PCR test for the beta-lactoglobulin C variant and the analysis of New Zealand sires.**

*Richard J. Wilkins and Tao Xie*
*Report to the Livestock Improvement Corporation 1996*

**κ-Casein genotyping of cows suspected of erroneously being protein typed as A variants**

*Richard J. Wilkins and Tao Xie*
*Report prepared for the Livestock Improvement Corporation Ltd. April 1996*

# *References*


ABI Prism GeneScan Analysis 2.1, User's Manual. (1996). Perkin-Elmer Corporation, Foster City, California, U.S.A.

Aebi M, Horning H, and Weissmann C (1987). 5' cleavage site in eukaryotic pre-mRNA splicing in determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* **50**: 237-246.

Aebi M, Horning H, Padgett R, Reiser J and Weissmann C (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* **47**: 555-565.

Aggler J, Park CS and Bissell MJ (1988). Regulation of milk protein and basement membrane gene expression: The influence of the extracelluar matrix. *J. Diry Sci.* **71**: 2830-2842.

Aleandri R, Buttazzoni LG, Schneider JC, Caroli A and Davoli R (1990). The effects of milk protein polymorphisms on milk components and cheese-producing ability. *J. Dairy Sci.* **73**: 241-255.

Alexander LJ and Beattie CW (1992). The sequence of porcine $\alpha_{s1}$-casein cDNA: evidence for protein variants generated by altered RNA splicing. *Anim. Genet.* **23**: 283-288.

Alexander LJ, Beattie CW, Hayes G, Pearse MJ, Steward AF and Mackinlay AG (1992). Isolation and characterisation of the bovine beta-lactoglobulin gene. *Genbank* submission number X14710.

Alexander LJ, Steward AF, MacKinlay AG, Kapelinskaya TV, Tkach TM and Gorodetsky SI (1988). Isolation and characterisation of the bovine kappa-casein gene. *Eur. J. Biochem.* **178**: 395-401.

Amills M, Francino O, Jansa M and Sanchez A (1997). Isolation of genomic DNA from milk samples by using chelex resin. *J. Dairy Res.* **64**: 231-238.

Andersson Y, Lindquist S, Bergstrom S and Hernell O (1997). Three variants of parathyroid hormone-related protein messenger RNA are expressed in human mammary gland. *Pediatr. Res.* **41**: 380-383.

Andersson-Eklund L and Rendel J (1993). Linkage between amylase-1 locus and major gene for milk fat content in cattle. *Anim. Genet.* **24**: 101-103.

Andrews AT, Taylor MD and Owen AJ (1985). Rapid analysis of bovine milk proteins by fast protein liquid chromatography. *J. Chromatogr.* **348**: 177-185.

Anema SG and Creamer LK (1993). Effect of the A and B variants of both $\alpha_{s1}$- and κ-casein on bovine casein micelle solvation and κ-casein content. *J. Dairy Res.* **60**: 505-516.

Annan WD and Manson W (1969). A fractionation of the αs-casein complex of bovine milk. *J. Dairy Res.* **36**: 259-268.

Atamas SP and White B (1997). Nonradioactive method for quantitation of PCR products without hybridisation with a specific probe. *BioTech.* **22**: 20-22.

Ausubel FM, Brent R, Kinston RE, Moor DD, Seidman JG, Smith JA and Struhl K (1987). Current protocols in molecular biology. Eds. John Wiley and Sons. New York.

Baker BS (1989). Sex in flies: the splice of life. *Nature* **340**: 521-524.

Basch JJ, Douglas FW Jr, Procino LG, Holeinger VH and Farrell HM Jr (1985). Quantitation of caseins and whey proteins of processed milks and whey protein concentrates, application of gel electrophoresis, and comparison with Harland Ashworth procedure. *J. Dairy Sci.* **68**: 23-31.

Bass BL and Weintraub H (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**: 1089-1098.

Bech AM and Kristiansen KR (1990). Milk protein polymorphism in Danish cattle and influence of genetic variants on milk yield. *J. Dairy Res.* **57**: 53-62.

Berget SM (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**: 2411-2414.

Bjourson AJ and Cooper JE (1992). Band-stab PCR: a simple technique for the purification of individual PCR products. *Nucleic Acids Res.* **20**: 4675.

Boisnard M, Hue D, Bouniol C, Mercier JC and Gays P (1991). Multiple mRNA species code for two non-allelic forms of ovine $\alpha_{S2}$-casein. *Eur. J. Biochem.* **201**: 633-641.

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

Bonsing J, Ring JM, Stewart AF and Mackinlay AG (1988). Complete nucleotide sequence of the bovine β-casein gene. *Aust. J. Biol. Sci.* **41**: 527-37.

Boulanger A, Grosclaude F and Mahe MF (1984). Polymorphism of caprine (Capra hircus) $\alpha_{S1}$- and $\alpha_{S2}$-casein. *Genet. Sel. Evol.* **16**: 157-175.

Bouniol C, Printz C and Mercier JC (1993). Bovine $\alpha_{S2}$-casein D is generated by exon VIII skipping. *Gene* **126**: 289-293.

Bovenhuis H, van Arendonk JAM and Korver S (1992). Associations between milk protein polymorphisms and milk protein traits. *J. Dairy Sci.* **75**: 2549-2559.

Breathnach R, Benoist C, O'Hare K, Cannon F and Chambon P (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA* **75**: 4853-4857.

Brignon G, Mahe MF, Ribadeau-Dumas MF, Mercier JC and Grosclaude F (1990). Two of the three genetic variants of goat $\alpha_{S1}$-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 JC and Pelissier JP (1977). Complete amino acid sequence of bovine $\alpha_{S2}$-casein. *FEBS lett.* **76**: 274-279.

Buehring GC (1990). Culture of mammary epithelial cells from bovine milk. *J. Dairy Sci.* **73**: 956-963

Burr RG, Moore CH and Hill JP (1997). ESI-MS phenotyping of bovine β-lactoglobulin genetic variants in a New Zealand dairy cattle population. In *Milk Protein Polymorphism.* pp340 International Diary Federation Brussels.

Chabot B (1996). Directing alternative splicing: cast and scenarios. *TIG* **12**: 472-478.

Chaplin LC (1986). Hydrophobic interaction fast protein liquid chromatography of milk proteins. *J. Chromatogr.* **363**: 329-335.

Charron M, Shaper J and Shaper N (1998). The increased level of β1,4-galactosyl-transferase required for lactose biosynthesis is achieved in past by translational control. *Proc. Natl. Acad. Sci. USA* **95**: 14805-14810.

Chin D and Ng-Kwai-Hang KF (1997). Application of mass spectrometry for the identification of genetic variants of milk proteins. In *Milk Protein Polymorphism.* pp332 International Diary Federation Brussels.

Chomczynski P (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *BioTech.* **15**: 532-536.

Chu S, Murray CB, Liu MM and Zeitlin PL (1996). A short CIC-2 mRNA transcript is produced by exon skipping. *Nucleic. Acids Res.* **24**: 3453-3457

Chung CT, Niemela SL and Miller RH (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* **86**: 2172-2175

Church GM and Gilbert W (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**: 1991-1995.

Clark AC (1996). Genetic modification of milk proteins. *Am. J. Clin. Nutr.* **63**: 633s-638s.

Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16**: 9677-9688.

Coker CJ, Creamer LK, Burr RG and Hill JP (1997). The action of chymosin or plasmin on $\alpha_{s1}$-casein A, B and C. In *Milk Protein Polymorphism* Int. Dairy Federation Brussels.

Cooper DN and Krawczak M (1990). The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum. Genet.* **85**: 55-74.

Cooper TA and Mattox W (1997). The regulation of splice site selection and its role in human disease. *Am. J. Hum. Genet.* **61**: 259-266.

Cottrez F, Auriault C, Capron A and Groux H (1994). Quantitative PCR: validation of the use of a multispecific internal control. *Neucleic Acids Res.* **22**: 2712-2713.

Coulter L.R, Landree MA and Cooper TA (1997). Identification of a new class of exonic splicing enhancers by in vivo selection. *Mol. Cell. Biol.* **17**: 2143-2150.

Creamer LK (1974). Preparation of $\alpha_{s1}$-casein A. *J. Dairy Sci.* **57**: 341-344.

Creamer LK (1993). Effect of the A and B variants of both $\alpha_{s1}$ and κ-casein on bovine casein micelle solvation and κ-casein content. *J. Dairy Res.* **60**: 505-516.

Creamer LK and MacGibbon AKH (1996). Some recent advances in the basic chemistry of milk proteins and lipids. *Int. Dairy J.* **6**: 539-568.

Creamer LK and Richardson T (1984). Anomalous behaviour of bovine alpha s1- and beta-caseins on gel electrophoresis in sodium dodecyl sulfate buffers. *Arch. Biochem. Biophys.* **234**: 476-486.

Creamer LK, Xoerb HF, Olson NF and Richardson T (1982). Surface hydrophobicity of alpha s1-I, alpha s1-casein A and B and its implication in cheese structure. *J. Dairy Sci.* **65**: 902-906.

Croves ML (1975). Methods of gel electrophoresis of milk proteins. In: *American Dairy Science Association, Champaign, IL.* Edited by Swaisgood H. p.26.

Dalrymple MA and Garner I (1998). Genetically modified livestock for the production of human protein in milk. *Biotech. Genet. Eng. Rev.* **15**: 33-49.

Damiani G, Pilla F, Leone P and Caccio S (1992). Direct sequencing and bidirectional allele specific polymerase chain reaction of the bovine β-casein B variant. *Anim. Genet.* **23**: 561-566.

David VA and Deutch AH (1992). Detection of bovine $\alpha_{s1}$-casein genomic variants using the allele-specific polymerase chain reaction. *Anim. Genet.* **23**: 425-429.

Davies DT, Holt C and Christie WW (1983). The composition of milk. in *Biochemistry of lactation*, pp. 1-11, ed. Mepham T.B. Elsevier, Amsterdam & New York

Davidow LS (1992). Selecting PCR designed mismatch primers to create diagnostic restriction sites. *Comp. Appl. Biosci.* **8**: 193-194.

Davies JD (1974). Human colostrum cells: their relation to periductal mononuclear inflammation. *J. Pathiol.* **112**: 153-160.

Delouis C and Richard P (1993). Lactation. In *Reprodution in Mammals and Man.* Paris, Ellipes. English revised, ed. 503-530.

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

Dietz HC and Kendzior RJ Jr (1994). Maintenance of an open reading frame as an additional level of scrutiny during splice site selection. *Nature Genet.* **8**: 183-188.

Dirksen WP, Sun Q and Rottman FM (1995). Muptiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. *J. Biol. Chem.* 270: 5346-5352.

Dobbeling U, Boni R, Haffner A, Dummer R and Burg (1997). Method for simultaneous RNA and DNA isolation from biopsy material, culture cells, plants and bacteria. *BioTech.* **22**: 88-90.

Eigel WN, Butler JE, Ernstrom CA, Farrell HM, Jr, Harwalkar VR, Jenness R and Whitney RM (1984). Nomenclature of proteins of cow's Milk: Fifth Revision. *J. Dairy Sci.* **67**: 1599-1631.

Erhardt G (1993). A new $\alpha_{s1}$-casein allele in bovine milk and its occurrence in different breeds. *Anim. Genet.* **24**: 65-66.

Falaki M, Prandi A, Corradini C, Sneyers M, Gengler N, Massart S, Fazzini U, Burny A, Portelle D and Renaville R (1997). Relationships of growth hormone gene and milk protein polymorphisms to milk production traits in Simmental cattle. *J. Dairy Res.* **64**: 47-56.

Farr VC, Stelwagen K, Cate LR, Molenaar AJ, McFadden TB and Davis SR (1996). An improved method for the routine biopsy of bovine mammary tissue. *J. Dairy Sci.* **79**: 543-549.

Farrell HM and Thompson MP (1971). Biological significance of milk protein polymorphism. *J. Dairy Sci.* **54**: 1219-1228.

Feagin JE, Jasmer DP and Stuart K (1987). Developmentally regulated addition of nucleotides within apocytochrome β transcripts in Trypanosoma brucei. *Cell* **49**: 337-345.

Ferranti P, Addeo F, Malorni A, Chianese L, Leroux C and Martin P (1997). Differential splicing of pre-messenger RNA produces multiple forms of mature caprine $\alpha_{S1}$-casein. *Eur. J. Biochem.* **249**: 1-7.

Ferranti P, Malorni A, Nitti G, Laezza P, Pizzano R and Chianese L (1995). Primary structure of ovine $\alpha_{S1}$-caseins: localization of phosphorylation sites and characterization of genetic variants A, C and D. *J. Dairy Res.* **62**: 281-296.

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

Fleet JC (1995). A new role for lactoferrin: DNA binding and transcription activation. *Nutr. Rev.* **53**: 226-227.

Foley KP, Leonard MW and Engel JD (1993). Quantitation of RNA using the polymerase chain reaction. *Trends Genet.* **9**: 380-385.

Fries R (1993). Mapping the bovine genome: methodological aspects and strategy. *Anim. Genet.* **24**: 111-116.

Fuerst C and Solkner J (1994). Additive and nonadditive genetic variances for milk yield, fertility, and lifetime performance traits of dairy cattle. *J. Dairy Sci.* **77**: 1114-1125.

Funk DA (1993). Optimal genetic improvement for the high producing herd. *J. Dairy Sci.* **76**: 3278-3286.

Gallagher DS, Schelling CP, Groenen MMA and Womack JE (1994). Confirmation that the casein gene cluster resides on cattle chromosome 6. *Mammal. Genome* **5**: 524

Giometti CS, Williams K and Tollaksen SL (1997). A two-dimensional electrophoresis database of human breast epithelial cell proteins. *Electrophoresis* **18**: 573-581.

Girardet JM and Linden G (1996). PP3 component of bovine milk: a phosphorylated whey glycoprotein. *J. Dairy Res.* **63**: 333-350.

Grabowski PJ, Nim FUH, Kuo HC and Burch R (1991). Combinatorial splicing of exon pairs by two-site binding of U1 small nuclear ribonucleoprotein article. *Mol. Cell. Biol.* **11**: 5919-5928.

Graml R, Weiss G, Buchberger J and Pirchner P (1989). Different rates of synthesis of whey protein and casein by alleles of the β-lactoglobulin and $\alpha_{S1}$-casein locus in cattle. *Genet. Sel. Evol.* **21**: 547-554.

Green A, Roopra A and Vaudin M (1990). Direct single stranded sequencing from agarose of polymerase chain reaction products. *Nucleic Acids Res.* **18**: 6163-6164.

Green MR and Pastewka JV (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.

Groenen MA, Dijkhof RJ, van der Poel JJ, van Diggelen R and Verstege E (1992). Multiple octamer binding sites in the promoter region of the bovine $\alpha_{S2}$-casein gene. *Nucleic Acids Res.* **20**: 4311-4318.

Groenen MAM and van der Poel JJ (1994). Regulation of expression of milk protein genes: a review. *Livestock Product. Sci.* **38**: 61-78.

Groenen MAM, Dijkhof RJM, Verstege AJM and van der Poel JJ (1993). The complete sequence of the gene encoding bovine $\alpha_{S2}$-casein. *Gene* **123**: 187-193.

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

Grosclaude F (1988). Le polymorphism genetique des principales lactoproteines bovines. Relations avec la quantite, la composition et les aptitudes fromageres gu lait. *INRA Prod. Anim.* **259**: 1569-1571.

Grosclaude F, Joudrier P and Mahe MF (1978). Polymorphisme de la casein $\alpha_{S2}$ bovine: etroite liaison du locus $\alpha_{S2}$-Cn avec les loci α-Cn, β-Cn et κ-Cn; mise en evidence d'une deletion dabs le variant $\alpha_{S2}$-Cn D. *Ann. Genet. Sel. Anim.* **10**: 313-327.

Grosclaude F, Mahe M, Brignon G, Di Stasio L and Jeunet R (1987). A Mendelian polymorphism underlying quantitative variation of goat $\alpha_{S1}$-casein. *Genet. Sel. Evol.* **19**: 399-412.

Grosclaude F, Mahe MF, Mercier JC and Ribadeau-Dumas B (1972). The localization of a deletion of 13 amino acids differentiating variant A from variant B and C of bovine $\alpha_{S1}$-casein in the N-terminal part. *FEBS lett.* **11**: 109-113.

Grosclaude F, Mahe MF, Mercier JC and Ribadeau-Dumas B (1972). Characterisation des variants genetiques des caseins $\alpha_{S1}$ et β bovines. *Eur. J. Biochem.* **26**: 328-337.

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

Grundemann D and Schomig E (1996). Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. *BioTech.* **21**: 898-903.

Guthrie C (1991). Messager RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* **253**: 157-163.

Guthrie C and Patterson B (1988). Spliceosomal snRNAs. *Annu. Rev. Genet.* **22**: 387-419.

Guyette WA, Matusik RJ and Rosen JM (1979). Prolaction-mediated transcriptional and post-transcriptional control of casein gene expression. *Cell* **17**: 1013-1023

Hartmann PE and Prosser CG (1984). Physiological basis of longitudinal changes in human milk yield and composition. *Fed. Proc.* **43**: 2448-2453.

Hayes JF, Ng-Kwai-Hang KF and Moxley JE (1983). Heritability of milk casein and genetic and phenotypic correlations with production traits. *J. Dairy Sci.* **67**: 841-846.

Heid CA, Stevens J, Livak KJ and Williams PM (1996). Real time quantitative PCR. *Genome Res.* **6**: 986-994.

Helfman DM and Ricci WM (1989). Banch point selection in alternative splicing of tropomyosin pre-mRNAs. *Nucleic Acids Res.* **17**: 5633-5650.

Hengen PN (1995). Quantitative PCR an accurate measure of mRNA? *TIBS* **20**: 476-477.

Hengen PN (1996). Carriers for precipitating nucleic acids. *TIBS* **21**: 224-225.

Hengen PN (1996). Eliminating ghost bands from plasmid preps. *TIBS* **21**: 441-442.

Hengen PN (1996). Preparing ultra-competent *Escherichia coli*. *TIBS* **21**: 75-76.

Her C and Weinshilboum R (1996). Long PCR: selective suppression by restriction endonuclease digestion. *Biotech.* **21**: 764-766.

Hill J.P (1993). The relationship between β-lactoglobulin phenotypes and milk composition in New Zealand dairy cattle. *J. Dairy Sci.* **76**: 281-286.

Hoagland PD, Thompson MP and Kalan EB (1971). Amino acid composition of $\alpha_{S3}$-, $\alpha_{S4}$- and $\alpha_{S5}$-caseins. *J. Dairy Sci.* **54**: 1103-1110.

Hodges D and Berstein SI (1994). Genetic and biochemical analysis of alternative RNA splicing. *Adv. in Genet.* **31**: 207-281.

Hoj S, Fredholm M, Larsen NJ and Nielsen VH (1993). Growth hormone gene polymorphism associated with selection for milk fat production in lines of cattle. *Anim. Genet.* **24**: 91-96.

Holt C (1992). Structure and stability of bovine casein micelles. *Adv. Protein Chem.* **43**: 63-151.

Holt C and Sawyer L (1988). Primary and predicted secondary structures of the caseins in relation to their biologic functions. *Protein Eng.* **2**: 251-259.

Hopper KE and Mckenzie HA (1973). Minor components of bovine α-lactalbumin A and B. *Biochem. et Biophysica. Acta.* **295**: 352-363.

Horton BS (1995). Commercial utilization of minor milk components in the health and food industries. *J. Dairy Sci.* **78**: 2584-2589.

Houdebine LM (1995). The production of pharmaceutical proteins from the milk of transgenic animals. *Reprod. Nutr. Dev.* **35**: 609-617.

Huang CH, Reid ME and Blumenfeld OO (1993). Exon skipping caused by DNA recombination that introduces a defective donor splice site into the human glycophorin A gene. *Biol. Chem.* **268**: 4945-4952.

Huang M, Arnheim N and Goodman MF (1992). Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.* **20**: 4567:4573.

Huang SH (1994). Inverse polymerase chain reaction - an efficient approach to cloning cDNA ends. *Mol. Biotech.* **2**: 15-22.

Ichihara Y and Kurosawa Y (1993). Construction of new T vectors for direct cloning of PCR products. *Gene* **130**: 153-154.

Ido E and Hayami M (1997). Construction of T-tailed vectors derived from a pUC plasmid: a rapid system for direct cloning of unmodified PCR products. *Biosci. Biotechnol. Biochem.* **61**:1766-1767.

Ikonen T, Ruottinen O, Erhardt G and Ojala M (1996). Allele frequencies of the major milk proteins in the Finnish Ayrshire and detection of a new κ-casein variant. *Anim. Genet.* **27**: 179-181.

Imafidon GI, Farkye NY and Spanier AM (1997). Isolation, purification, and alteration of some functional groups of major milk proteins: a review. *Critic. Rev. in Food Sci. and Nutri.* **37**: 663-689.

Javier Lopez A (1998). Aternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* **32**: 279-305.

Jayagopala Reddy NR, Wilkie BN and Mallard BA (1996). Construction of an internal control to quantitate multiple procine cytokine mRNAs by RT-PCR. *BioTech.* **21**: 868-875.

Jenness R (1979). The composition of human milk. *Semin Perinatol.* **3**: 225-239.

Johnsen LB, Rasmussen LK, Petersen TE and Bergland L (1995). Characterization of three types of human $\alpha_{S1}$-casein mRNA transcripts. *Biochem. J.* **309**: 237-242.

Josephson RV (1972). Electric focusing of bovine milk caseins. *J. Dairy Sci.* **55**: 1535-1542.

Karatzas CN and Turner JD (1997). Toward altering milk composition by genetic manipulation: current status and challenges. *J. Dairy Sci.* **80**: 2225-2232.

Kitchen BJ (1981). Review of the progress of dairy science: bovine mastitis: milk compositional changes and related diagnostic tests. *J. Dairy Res.* **48**: 167-188.

Kleffe J, Hermann K, Vahrson W, Witting B and Brendel V (1996). Logitlinear models for the prediction of splice sites in plant pre-mRNA sequences. *Nucleic Acids Res.* **24**: 4709-4718.

Koczan D, Hobom G and Seyfert HM (1991). Genomic organization of the bovine alpha-S1 casein gene. *Nucleic Acids Res.* **19**: 5591-5596.

Koczan D, Hobom G and Seyfert HM (1993). Characterization of the bovine $\alpha_{S1}$-casein gene C-allele, based on a Mae III polymorphism. *Anim. Genet.* **24**: 74.

Kohsaka H, Taniguchi A, Richman DD and Carson DA (1993). Microtiter format gene quantification by covalent capture of competitive PCR products: application to HIV-1 detection. *Nucleic Acids Res.* **21**: 3469-3472.

Krainer AR and Maniatis T (1988). RNA splicing. In: *Transcription and Splicing*. Edited by Hames B.D. and Glover D.M. IRL Press Oxford, pp131-220.

Krawczak M and Cooper DN (1991). Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum. Genet.* **86**: 425-441.

Krawczak M, Reiss J and Cooper DN (1992). The mutational spectrum of single base-pair substitutions in mRNA splicing junctions of the human genes: causes and consequences. *Hum. Genet.* **90**: 41-54.

Kudo M, Kitamura-Abe S, Shimbo M and Ida Y (1992). Analysis of context of 5'-splice site sequences in mammalian pre-mRNA by sub-class method. *Comput. Appl. Biosci.* **8**: 367-376.

Kuivaniemi H, Kontusaari S, Tromp G, Zhao M, Sabol C and Prockop DJ (1990). Identical G+1 to A mutations in three different introns of the type III procollagen gene (COL3A1) produce different patterns of RNA splicing in three variants of Ehlers-Danlos syndrome IV. *J. Biol. Chem.* **265**: 12067-12074.

Kumosinski TF, Brown E.M and Farrell HM Jr (1991). Three-dimensional molecular modelling of bovine caseins: $\alpha_{s1}$-casein. *J. Dairy Sci.* **74**: 2889-2895.

Kuo HC, Nasim FU and Grabowski PJ (1991). Control of alternative splicing by the different binding of the U1 small nuclear ribonucleoprotein particle. *Science* **251**: 1045-1050.

Kuzmanoff K.M, Andersen J and Beattie CW (1991). Isolation and characterization of Monoclonal antibodies monospecific for bovine α-casein and β-casein. *J. Dairy Sci.* **74**: 803-810.

Kwok S, Kellogg DE, Mckinney N, Spasic D, Goda L, Levenson C and Sninsky JJ (1990). Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* **18**: 999-1005.

Laemili UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 682-685.

Laird JE, Jack L, Hall L, Boulton A, Parker D and Graig RK (1988). Structure and expression of the quinea-pig α-lactalbumin gene. *Biochem. J.* **254**: 85-94.

Lamond AI, Konarska MM and Sharp PA (1987). A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev.* **1**: 532-543.

Langford CJ, Klinz FJ, Donath C and Gallwitz D (1984). Point mutations identify the conserved intron-contained TACTAAC box as an essential splicing sequence in yeast. *Cell* **36**: 645-653.

Larsen B and Thymann M (1966a). Studies on milk protein polymorphismes in Danish cattle and the interaction on the controlling genes. *Acta Vet. Scand.* **7**: 189-205.

Larsen B and Thymann M (1966b). Studies on milk protein polymorphism and the interaction of the controlling genes. In *Polymorphisms Biocheniques des Animaux. Proc. Xth Eur. Conf. Anim. Blood Grps Biochem. Polym.*, Paris, pp. 421-425.

Lawyer FC, Stoffel S, Saiki RK, Myambo K, Drummond R and Gelfand DH (1989). Isolation, characterisation, and expression in *Escherichia coli* of the DNA polymerase gene from Thermus aquaticus. *J. Biol. Chem.* **264**: 6427-6436.

Lear W, McDonnell M, Kashyap S and Boer PH (1995). Random primer p(dN)6-digoxigenin labelling for quantitation of mRNA by Q-RT-PCR and ELISA. *BioTech.* **18**: 78-83.

Leroux C, Mazure N and Martin P (1992). Mutations away from splice site recognition sequences might *cis*-modulate alternative splicing of goat $\alpha_{s1}$-casein transcripts. *J. Biol. Chem.* **267**: 6147-6157.

Leveziel H, Metenier L, Guerin G, Cullen P, Provot C, Bertaud M and Merier JC (1991). Restriction fragment length polymorphism of ovine casein genes. close linkage between the $\alpha_{s1}$-, β-, and κ-CN loci. *Anim. Genet.* **22**: 1-10.

Lewis SA (1996). Removal of Coomassie blue precipitates from polyacrylamide gels. *BioTech.* **21**: 820.

Libert F, Lefort A, Okimoto R, Womack J and Georgest M (1993). Construction of a bovine genomic library of large yeast artificial chromosome clones. *Genet.* **18**: 270-276.

Libri D, Piseri L and Fiszman MY (1991). Tissue-specific splicing in vivo of the β-tropomyosin gene: dependence on an RNA secondary structure. *Science* **252**: 1842-1845.

Lien S and Rogne S (1993). Bovine casein haplotypes: number, frequencies and applicability as genetic markers. *Anim. Genet.* **24**: 373-376.

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, Gomez-raya L, Steine T, Fimland E and Rogne S (1995). Associations between casein haplotypes and milk yield traits. *J. Dairy Sci.* **78**: 2047-2056.

Lin CY, McAllister AJ, Ng-Kwai-Hang KF, Hayes JF, Batra TR, Lee AJ, Roy GL, Vesely JA, Wauthy JM and Winter KA (1989). Relationships of milk protein types to lifetime performance. *J. Dairy Sci.* **72**: 3085-3090.

Lindersson M, Lunden A and Andersson L (1995). Genotyping bovine milk proteins using allele discrimination by primer length and automated DNA sizing technology. *Anim. Genet.* **26**: 67-72.

Lindquist S, Hansson L, Hernell O, Lonnerdal B, Normark J, Stromqvist M and Bergstrom S (1994). Isolation of mRNA and genomic DNA from epithelial cells in human milk and amplification by PCR. *BioTech.* **17**: 692-696.

Lipkin E, Shalom A, Khatib H, Soller M and Friedmann A (1993). Milk as a source of deoxribonucleic acid and as a substrate for the polymerase chain reaction. *J. Dairy Sci.* **76**: 2025-2032.

Liu X, Robinson GW, 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.

Manley JL and Tacke R (1996). SR proteins and Splicing control. *Genes Dev.* **10**: 1569-1579.

Mann M and Wilm M (1995). Electrospray mass spectrometry for protein characterisation. *TIBS* **20**: 219-224.

Marchuk D, Drumm M, Saulino A and Collins FS (1990). Construction of T-vectors: a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19**: 1154.

Martin P and Leroux C (1992). Exon-skipping is responsible for 9 amino acid residue deletion occurring near the N-terminal of human β-casein. *Biochem. Biophys. Res. Commun.* **183**: 750-757.

Martin P and Leroux C (1994). Characterisation of a further goat $\alpha_{s1}$-casein variant generated by exon-skipping. *Proceeding of the XXIV International Conference on Animal Genetics.* P. 88 ISAG, Prague.

Martin P, Brignon G, Furet JP and Leroux C (1996). The gene encoding $\alpha_{s1}$-casein is expressed in human mammary epithelial cells during lactation. *Lait* **76**: 523-535.

Mathy NL, Lee RP and Walker J (1996). Removal of RT-PCR inhibitors from RNA extracts of tissues. *BioTech.* **21**: 770-774.

Matsuo M, Masumura T, Nishio H, Nakajima T, Kitoh Y, Takumi T, Koga J and Saito H (1991). Exon skipping during splicing of dytrophin mRNA precursor due to an intra-exon deletion in the dystrophin gene of the Duchenne muscular dystrophy kobe. *J. Clin. Invest.* **87**: 2127-2131.

Matsushita T, Tanomoto M, Yamamoto K, Sugiura I, Hamaguchi M, Takamatsu J and Saito H (1989). Nucleotide sequence analysis of hemophilia B with inhibitor phenotype. *Blood* **74** [Suppl]: 251a.

McGann TC, Donnelly WJ, Kearney RD and Buchheim W (1980). Composition and size distribution of bovine casein micelles. *Biochim. Biophys. Acta* **630**: 261-270.

McKenzie HA and Hopper KE (1973). Minor components of bovine α-lactalbumin A and B. *Biochem. Biophys. Acta* **295**: 352-363.

McKnight RA, Jimenez-Flores R, Kang Y, Creamer LK and Richardson T (1989). Cloning and sequencing of a complementary deoxyribonucleic acid coding for a bovine $\alpha_{s1}$-casein A from mammary tissue of a homozygous B variant cow. *J. Dairy Sci.* **72**: 2464-2473.

Mclean DM, Bruce Graham ER and Ponzoni PW (1984). Effects of milk protein genetic variants on milk yield and composition. *J. Dairy Res.* **51**: 531-546.

Medrano JF and Sharrow L (1989). Milk protein typing of bovine mammary gland tissue used to generate a complementary deoxyribonucleic acid library. *J. Dairy Sci.* **72**: 3190-3196.

Menon RS, Chang YF, Jeffers KF and Ham RG (1992). Exon-skipping in human β-casein. *Genomics* **12**: 13-17.

Mepham TB (1992). Biosythesis of milk protein. In *Advanced dairy chemistry volume I: proteins,* edited by Fox P.F., Elsevier Science Publishers LTD, London.

Mepham TB (1987). Physiology of lactation. Open University Press, Milton Keynes & Phildelphia.

Mercier JC, Grosclaude F and Ribadeau-Dumas B (1971). Structure primaire de la caseine $\alpha_{S1}$-bovine Sequence complete. *Eur. J. Biochem.* **23**: 41-45.

Mercier JC, Grosclaude F and Ribadeau-Dumas B (1973). Structure primaire de la caseine kB bovine. *Eur. J. Biochem.* **35**: 222-235.

Mohr U, Koczan D, Linder D, Hobom G and Erhardt G (1994). A single point mutation results in a allele-specific exon skipping in the bovine $\alpha_{S1}$-casein mRNA. *Gene* **143**: 187-192.

Molenaar AJ (1995). Milk protein gene expression in the ruminant mammary gland. *D.Phil Thesis*, University of Waikato, New Zealand.

Molenaar AJ, Davis SR and Wilkins RJ (1992). Expression of alpha-lactalbumin, alpha-S1-casein, and lactoferrin genes is heterogeneous in sheep and cattle mammary tissue. *J. Histochem. Cytochem.* **40**: 611-618.

Mount SM (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**: 459-472.

Murphy LD, Herzog CE, Rudick JB, Fojo AT and Bates SE (1990). Use of the polymerase chain reaction in the quantitation of mdr-1 gene expression. *Biochem.* **29**: 10351-10356.

Nakai K and Sakamoto H (1994). Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* **141**: 171-177.

Nakata K, Kanehisa M and DeLisi C (1985). Prediction of splice junctions in mRNA sequences. *Nucleic Acids Res.* **13**: 5327-5340.

Neilan BA, Leigh DA, Rapley E and McDonald BL (1994). Microsatellite genome screening: rapid non-denaturing, non-isotopic dinucleotide repeat analysis. *BioTech.* **17**: 708-712.

Nelson KK and Green MR (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev.* **3**: 1562-1571.

Newman A (1997) RNA splicing: out of the loop. *Curr. Biol.* **7**: R418-R420.

Ng-Kwai-Hang KF (1997). A review of the relationship between milk protein polymorphism and milk composition/milk production. In *Milk Protein Polymorphism.* Int. Diary Federation Brussels.

Ng-Kwai-Hang KF and Grosclaude F (1992). Genetic polymorphism of milk Proteins. *In Advanced dairy chemistry volume I: proteins,* edited by Fox P.F., Elsevier Science Publishers LTD, London.

Ng-Kwai-Hang KF, Hayes JF, Moxley JE and Monardes HG (1984). Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. *J. Dairy Sci.* **67**: 835-840.

Ng-Kwai-Hang KF and Kroeker EM (1984). Rapid separation and quantification of major caseins and whey proteins of bovine milk by polyacrylamide gel electrophoresis. *J. Dairy Sci.* **67**: 3052-3056.

Nilsen TW (1994). RNA-RNA interactions in the spliceosome: unraveling the ties that bind. *Cell* **78**: 1-4.

Nishimura A, Morita M, Nishimura Y and Sugino Y (1990). A rapid and highly efficient method for preparation of competent *Escherichia coli* cells. *Nucleic Acids Res.* **18**: 6169.

Ohno K and Suzuki K (1988). Multiple abnormal β-hexosaminidase μ chain mRNAs in a compound-heterozygous Ashkenazi Jewish patient with Tay-Sachs disease. *J. Biol. Chem.* **263**: 18563-18567.

Okayama H, Curiel DT, Brantly ML, Holmes MD and Crystal RG (1989). Rapid, nonradioactive detection of mutations in the human genome by allele-specific amplification. *J. Lab. Clin. Med.* **114**: 105-113.

Oshima Y, and Gotah Y (1987). Signals for the selection of a splice site in pre-mRNA. computer analysis of splice junction sequences and like sequences. *J. Mol. Biol.* **195**: 247-259.

Padgett RA, Grabowki PJ, Konarska MM, Seiler S and Sharp PA (1986). Splicing of messager RNA precursors. *Annu Rev. Biochem.* **55**: 1119-1150.

Park O and Mayo KE (1991). Transient expression of progesterone receptor messenger RNA in ovarian granulosa cells after the preovlatory luteinizing hormone surge. *Mol. Endocrinol.* **5**: 967-978.

Passey R, Glenn W and Mackinlay A (1996). Exon skipping in the ovine $\alpha_{S1}$-casein gene. *Comparat. Biochem.* **114**: 389-394.

Penotti FE (1991). Human pre-mRNA splicing signals. *J. Theor Biol.* **150**: 385-420.

Peterson RF (1963). High resolution of milk proteins obtained by gel electrophoresis. *J Dairy Sci.* **46**: 1136.

Pfeffer U, Fecarotta E, Arena G, Forlani A and Vidali G (1996). Alternative splicing of the estrogen receptor primary transcript normally occurs in estrogen receptor positive tissues and cell lines. *J. Steroid Biochem Mol. Biol.* **56**: 99-105.

Pierre S, Jolivet G, Devinoy E, Theron MC, Malienou-N'Gassa R, Puissant C and Houdebine LM (1992). A distal region enhances the prolactin induced promoter activity of the rabbit alpha S1-casein gene. *Mol. Cell Endocrinol.* **87**: 147-156.

Pope B and Kent HM (1996). High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Res.* **24**: 536-537.

Popov LS (1996). Some aspects of structure and expression of milk protein genes (a review). *Mol. Biol.* **30**: 742-753.

Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ and Scott J (1987). A novel form of tissue-specific RNA procession produces apolipoprotein-B48 in intestine. *Cell* **50**: 831-840.

Prinzenberg EM, Anglade P, Ribadeau-Dumas B and Erhardt G (1998). Biochemical characterisation of bovine $\alpha_{s1}$-casein F and genotyping with sequence-specific primers. *J. Dairy Res.* **65**: 223-231.

Prinzenberg EM, Hiendleder S, Ikonen T and Erhardt G (1996). Molecular genetic characterisation of new bovine kappa-casein alleles CSN3F and CSN3G and genotyping by PCR-RFLP. *Anim. Genet.* **27**: 347-349.

Reed R (1989). The organisation of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**: 2113-2123.

Reed R and Maniatis T (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell* **41**: 95-105.

Reed R and Maniatis T (1986). A role for exon sequences and splice site proximity in splice site selection. *Cell* **46**: 681-690.

Reischl U and Kochanowski B (1995). Quantitative PCR: a survey of the present technology. *Mol. Biotechnol.* 3: 55-71.

Reynolds JA and Tanford C (1970). Binding of dodecyl sulfate to proteins at high binding ratios. Possible implications for the state of proteins in biological membranes. *Proc. Natl. Acad. Sci. USA* **66**: 1002.

Ribadeau-Dumas B, Brignon G, Grosclaude F and Mercier JC (1972). Primary structure of bovine beta casein complete sequence. *Eur. J. Biochem.* **25**: 505-514.

Richardson T, Oh S, Jimenez-Flores R, Kumosinski T, Brown EM and Farrel HM (1992). Molecular modelling and genetic engineering of milk proteins. In *Advanced dairy chemistry volume I: proteins,* edited by Fox P.F., Elsevier Science Publishers LTD, London.

Rijnkels M, Kooiman PM, de Boer HA and Pieper FR (1997). Organisation of the bovine casein gene locus. *Mammal. Genome* **8**: 148-152.

Robberson BL, Cote GJ and Berget SM (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell Biol.* **10**: 84-94.

Rogozin IB and Milanesi L (1997). Analysis of donor splice sites in different eukaryotic organisms. *J. Mol. Evol.* **45**: 50-59.

Rolland MP, Bitri L and Besancon P (1993). Polyclonal antibodies with predetermined specificity against bovine $\alpha_{s1}$-casein: application to the detection of bovine milk in ovine milk and cheese. *J. Dairy Res.* **60**: 413-420.

Ron M, Yoffe O, Ezra E, Medrano JF and Weller JI (1994). Determination of effects of milk protein genotype on production traits of Israeli Holsteins. *J. Dairy Sci.* **77**: 1106-1113.

Rosen JM, Supowit SC, Gupta P, Yu-Lee LY and Hobbs AA (ed.) (1981). Regulation of casein gene expression in hormone-dependent mammary cancer. In *Hormones and Breast Cancer.* Cold Spring Harbour Laboratory pp 397-424

Rust S, Funke H and Assmann G (1993). Mutagenically separated PCR (MS-PCR): a highly specific one step procedure for easy mutation detection. *Nucleic Acids Res.* **21**: 3623-3629.

Ryan KR and Cooper TA (1996). Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. *Mol. Cell Biol.* **16**: 4014-4023.

Sadler AM, Kiddy CA, McCann RE and Mattingly WA (1968). Acid production and curd toughness in milk of different $\alpha_{S1}$-casein types. *J. Dairy Sci.* **551**: 28-38.

Sambrook J, Fritsch EF and Maniatis T (1989). Molecular cloning. a laboratory manual. second edition. Eds. Cold Spring Harbour Laboratory Press, New York.

Sameshima Y, Akiyama T, Mori N, Mizoguchi H, Toyoshima K, Sugimura T, Terada M and Yokota J (1990). Point mutation of the $P^{53}$ gene resulting in splicing inhibition in small cell lung carcinoma. *Biochem. Biophys. Res. Commun.* **173**: 697-703.

Sawyer L and Holt C (1993). The secondary structure of milk proteins and their biological function. *J. Dairy Sci.* **76**: 3068-3078.

Schagger H and Jagow GV (1987). Tricine-sodium dodecyl sulfate polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analyt. Biochem.* **166**: 368-379.

Schild TA and Geldermenn H (1996). Variants within the 5'-flanking regions of bovine milk-protein-encoding genes. III. genes encoding the Ca-sensitive caseins $\alpha_{S1}$, $\alpha_{S2}$ and β. *Theor. Appl. Genet.* **93**: 887-893.

Schlieben S, Erhardt G and Senft G (1991). Genotyping of bovine κ-casein (κ-$CN^A$, κ-$CN^B$, κ-$CN^C$, κ-$CN^E$) following DNA sequence amplification and direct sequencing of κ-$CN^F$ PCR product. Anim. Genet. **22**: 333-342.

Schischmanott PO, Yaswen P, Parra MK, Lee G, Chasis JA, Mohandas N and Conboy JG (1997). Cell shape-dependent regulation of protein 4.1 alternative pre-mRNA splicing in mammary epithelial cells. *J. Biol. Chem.* 272: 10254-10259.

Schmidhauser C, Casperson GF, Myers CA, Sanzo KT, Bolten S and Bisell MJ (1992). A novel transcriptional enhancer in involved in the prolactin- and extracellular matrix- dependent regulation of β-casein gene expression. *Mol. Biol. of the Cell* **3**: 699-709.

Schmidt DG (1979). Properties of artificial casein micelles. *J. Dairy Res.* **46**: 351-355.

Schmitt-Ney M, Doppler W, Ball RK and Groner B (1991). β-casein gene promoter activity is regulated by the hormone mediated relief of transcriptional repression and a mammary gland specific nuclear factor. *Mol. Cell. Biol.* **11**: 3745-3755.

Shapiro MB and Senapathy P (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**: 7155-7174

Sharma RC and Schimke RT (1996). Preparation of electro-competent *E. coli* using salt-free growth medium. *BioTech.* **20**: 42-44.

Siebert PD and Huang BC (1997). Identification of an alternative form of human lactoferrin mRNA that is expressed differentially in normal tissues and tumour derived cell lines. *Proc. Natl. Acad. Sci. USA* **94**: 2198-2203.

Slattery CW (1976). Review: casein micelle structure; an examination of models. *J. Dairy Sci.* **59**: 1547-1556.

Smith CW and Nadal-Ginard B (1989). Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell* **56**: 749-758.

Smith CW, Chu TT and Nadal-Ginard B (1993). Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell. Biol.* **13**: 4939-4952.

Solovyev VV, Salamov AA and Lawrence CB (1994). Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* **22**: 5156-5163.

Somasekhar MB and Mertz JE (1985). Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. *Nucleic Acids Res.* **13**: 5591-5609.

Sommer SS (1992). PCR amplification of specific allele. *Science* **255**: 514.

Sommer SS, Groszbach AR and Bottema CD (1992). PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *BioTech.* **12**: 82-87.

Soulier S, Mercier JC Vilotte JL, Anderson J, Clark J and Provot C (1989). The bovine and ovine genomes contain multiple sequences homologous to α-lactalbumin-encoding gene. *Gene* **83**: 331-338.

Stewart AF, Willis IM and Mackinlay AG (1984). Nucleotide sequences of bovine alpha S1- and kappa-casein cDNAs. *Nucleic Acids Res.* **12**: 3895-3907.

Stinnakre MG, Vilotte JL, Soulier S and Mercier JC (1994). Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. *Proc. Natl. Acad. Sci. USA* **91**: 6544-6548.

Strange ED, Malin EL, Van Hekken DL and Basch JJ (1992). Chromatographic and electrophoretic methods used for analysis of milk proteins. *J. Chromatogr.* **624**: 81-102

Studencki AB and Wallace RB (1984). Allele-specific hybridisation using oligonucleotide probes of very high specific activity: discrimination of the human β-A- and β-S-globin genes. *DNA* **3**: 7-15.

Sugiyama K, Hitomi Y, Adachi H and Esumi H (1994). Cell type specific patterns of mRNA splicing in hepatoma cells transfected with the mutated albumin minigene of Nagase albuminemic rats. *Cancer Lett.* **83**: 221-227.

Swaisgood HE (1993). Review and update of casein chemistry. *J. Dairy Sci.* **76**: 3054-3061.

Talerico M and Berget SM (1990). Effect of 5' splice site mutations on splicing of the preceding intron. *Mol. Cell. Biol.* **10**: 6299-6305.

Tang X, Nakata Y, Li HO, Zhang M, Gao H, Fujita A, Sakatsume O, Ohta T and Yokoyama K (1994). The optimisation of preparations of competent cells for transformation of *E.coli*. *Nucleic Acids Res.* **22**: 2857-2858.

Thompson MP (1966). DEAE-cellulose-urea chromatography of casein in presence of β-mercaptoethanol. *J. Dairy Sci.* **49**: 792-795.

Thompson MP and Gorden WG (1967). Differential solubility of $\alpha_{s1}$-casein A in calcium chloride solution at 1 and 33 °C. *J. Dairy Sci.* **50**: 941.

Thompson MP and Kiddy CA (1964). Genetic polymorphism in caseins of cow's milk. III. isolation and properties of $\alpha_{s1}$-caseins A, B, and C. *J. Dairy Sci.* **47**: 626-632.

Thompson MP, Gorden WG, Boswell RT and Farrel HM Jr (1969). Solubility solvation, and stabilisation of $\alpha_{s1}$- and β-caseins. *J. Dairy Sci.* **52**: 1166-1173.

Thompson MP, Kiddy CA, Pepper L and Zittle CA (1962). Variations in the αs-casein fraction of individual cow's milk. *Nature* **195**: 1001-1002.

Threadgill DW and Womack JE (1990). Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Res.* **18**: 6935-6942.

Towbin H, Staehelin HT and Gordon J (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA.* **76**: 4350-4354.

Treisman R, Orkin SH and Maniatis T (1983). Specific transcription and RNA splicing defects in five cloned β–thalassaemia genes. *Nature (London)* **302**: 591-596.

Troutt AB, McHeyzer-Williams MG, Pulendran B and Nossal GJ (1992). Ligation-anchored PCR: a simple amplification technique with single-sided specificity. *Proc. Natl. Acad. Sci. USA.* **89**: 9823-9825.

Tsai SJ and Wiltbank MC (1996). Quantification of mRNA using competitive RT-PCR with standard curve methodology. *BioTech.* **21**: 862-866.

Tsang TC, Harris DT, Akporiaye SF, Schluter SF, Bowden GT and Hersh EM (1996). Simple method for adapting DNA fragments and PCR products to all of the commonly used restriction sites. *BioTech.* **20**: 51-52.

Tucker HA (1988). Lactation and its hormonal control. In *"The physiology of reproduction"* Knobil E and Neil J *et al.* Ed., Raven Press, Ltd., New York.

Udy GB, Tower RP, Snell RG, Wilkins RJ, Park SH, Ram PA, Waxman DJ and Davey HW (1997). Requirement of STAT5b for sexual dimorphism of body growth rates and liver gene expression. *Proc. Natl. Acad. Sci. USA* **94**: 7239-7244.

Ugozzoli L and Wallace RB (1991). Allele-specific polymerase chain reaction. *Methods* **2**: 42-48.

Van Hekken DL and Thompson MP (1992). Application of PhastSystem to the resolution of bovine milk proteins on urea-polyacrylamide gel electrophoresis. *J. Dairy Sci.* **75**: 1204-1210.

Velmala R, Vilkki J, Elo K, and Maki-Tanila A (1995). Casein haplotypes and their association with milk production traits in the Finnish Ayrshire cattle. *Anim. Genet.* **26**: 419-425.

Vilotte JL, Soulier S, Printz C and Mercier JC (1991). Sequence of the goat α-lactalbumin-encoding gene: comparison with the bovine gene and evidence of related sequences in the goat genome. *Gene* **98**: 271-276.

Wang AM, Doyle MV and Mark DF (1989). Quantification of mRNA by the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **86**: 9717-9721.

Wang T and Brown MJ (1999). mRNA quantification by real time TaqMan polymerase chain reaction: validation and comparison with RNase protection. *Anal. Biochem.* **269**: 198-201.

Wang Z and Rossman TG (1994). Isolation of DNA fragments from agarose gel by centrifugation. *Nucleic Acids Res.* **22**: 2862-2863.

Wassarman DA and Steitz JA (1991). Alive with dead proteins. *Nature* (London) **349**: 463-464.

Watson JW, Gordon KE, Robertson M and Clark AJ (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.

Wheeler TT, Callaghan MR, Davis SR and Wilkins RJ (1993). An appraisal of the utility of primary cell culture from sheep udders to investigate the control of mammary function. *Proc. N.Z. Soc. Anim. Product.* **53**: 151-154.

Whelan KF and Taylor DE (1996). Use of a DNA sequencing gel apparatus for analysis of polypeptides. *BioTech.*. **21**: 805-808.

Whitney RM, Brunner JR, Ebner KE, Farrel HM Jr, Josephson RV, Morr CV and Swaisgood HE (1976). Nomenclature of the proteins of cow's milk: fourth revision. *J. Dairy Sci.* **59**: 795-815.

Winkelman AM and Wickham BW (1997). Associations between milk protein genetic variants and production traits in New Zealand dairy cattle. In *Milk Protein Polymorphism.* Int. Dairy Federation. Brussels.

Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA and Balis UJ (1997). The LightCycler: a microvolume multi-sample fluorometer with rapid temperature control. *BioTech.* **22**: 176-81.

Wong DW, Camirand WM and Pavlath AE (1996). Structures and functionalities of milk proteins. *Critic. Rev. in Food Sci. Nutri.* **36**: 807-844.

Wu J and Manley JL (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev.* **3**: 1553-1561.

Wu JY and Maniatis T (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**: 1061-1070.

Yoshimura M and Oka T (1989). Isolation and structural analysis of the mouse β-casein gene. *Gene* **78**: 267-275.

Yu-Lee LY, Richter-Mann L, Couch CH, Steward AF, MacKinlay AG and Rosen J (1986). Evolution of the casein multigene family: conserved sequences in the 5' flanking and exon regions. *Nucleic Acids Res.* **14**: 1883-1902.

Zhang MQ and Marr TG (1995). Correlations and constraints among different splicing sequence features in human genes. In: *Notes of gene-finding and gene structure prediction workshop.* Penn Tower Hotel, Philadephia