THE STUDY OF MAJOR MILK PROTEINS GENETIC POLYMORPHISMS IN THE MAIN CATTLE, BUFFALO, SHEEP AND GOAT BREEDS FROM ROMANIA WITH THE AIM OF USING THEM AS GENETIC MARKERS IN BREEDING AND TRACEABILITY
**CONTENTS**

**INTRODUCTION** ................................................................. 31

**PART I: BIBLIOGRAPHIC STUDY**

**CHAPTER I**
COMPOSITION OF CATTLE, BUFFALO, SHEEP, GOAT
MILK AND MILK SYNTHESIS MECHANISM IN
LACTATING MAMMARY GLAND.................................................. 32

**CHAPTER II**
MOLECULAR TECHNIQUES USED FOR MILK
PROTEINS POLYMORPHISMS STUDY............................................. 33

**CHAPTER III**
CURRENT STAGE OF THE RESEARCH CONCERNING
THE STUDY OF MILK PROTEINS POLYMORPHISMS
IN CATTLE, BUFFALO, SHEEP AND GOAT.................................... 34

**CHAPTER IV**
THE IMPORTANCE OF MAJOR MILK PROTEINS
POLYMORPHISMS STUDY BY THEIR POSSIBLE
USE IN ANIMAL BREEDING AND DAIRY
PRODUCTS TRACEABILITY.......................................................... 37

**PART II: PERSONAL RESEARCH**

**CHAPTER V**
THE RESEARCH AIM AND OBJECTIVES........................................ 39

**CHAPTER VI**
CHROMOSOME MAPPING OF THE GENES CODING
FOR MAJOR MILK PROTEINS BY FLUORESCENCE
IN SITU HYBRIDISATION (FISH)................................................... 40
CHAPTER VII
THE STUDY BY PCR-RFLP OF THE TWO GENES POLYMORPHISM CODING FOR K-CASEIN AND β-LACTOGLOBULIN IN CATTLE BELONGING TO ROMANIAN SIMMENTAL BREED ..........................................................41

CHAPTER VIII
COMPARATIVE STUDY BY IEF AND PCR-RFLP OF MAJOR MILK PROTEINS GENETIC POLYMORPHISMS IN CATTLE BELONGING TO ROMANIAN SIMMENTAL BREED ........................................................................................................42

CHAPTER IX
THE CHARACTERIZATION OF MAJOR MILK PROTEINS POLYMORPHISMS IN SOME ROMANIAN CATTLE, BUFFALO, SHEEP AND GOAT BREEDS BY IEF TECHNIQUE .................................................................42

CHAPTER X
MOLECULAR CHARACTERIZATION OF THE NEW ALLELES IDENTIFIED IN αS1-CASEIN AND β-CASEIN LOCI IN ROMANIAN GREY STEPPE CATTLE, MODAVIAN VARIETY AND ROMANIAN BUFFALO ..................................................................................................................47

CHAPTER XI
IDENTIFICATION OF DECLARED AUTHENTICITY/ORIGIN OF MILK AND DAIRY PRODUCTS USING THE MAJOR MILK PROTEINS POLYMORPHISMS AS GENETIC MARKERS ..................................................................................................................48

CHAPTER XII
GENERAL CONCLUSIONS AND RECOMMENDATIONS ..................................................................................................................49

SELECTIVE BIBLIOGRAPHY ........................................................................................................................................50
INTRODUCTION

Lactogenesis can be divided in two phases: **Phase I** - begins in the last stage of gestation and is characterized by cellular, structural and functional differentiation processes of the secretor epithelium; **Phase I** - begins immediately after birth and involves the ending of cellular differentiation, which coincides with milk secretion and its synthesis in significant quantities. These two phases are essential having as final effect the beginning of milk production (the start of lactation).

In ruminants’ milk there are 6 major milk proteins types, which are coded by 6 non-allelic genes specifically expressed in mammary epithelial cells during lactation: \( \alpha \text{S1-casein} \) (\( \alpha \text{S1-CN} \)), \( \beta \)-casein (\( \beta \)-CN), \( \alpha \text{S2-casein} \) (\( \alpha \text{S2-CN} \)), K-casein (K-CN), \( \beta \)-lactoglobulin (\( \beta \)-LG) and \( \alpha \)-lactoalbumin (\( \alpha \)-LA).

The mutations occurring in these genes structure during the years, lead to the appearance of many alleles in these loci. These variations, named generically polymorphisms, are caused by genes restructuration (substitution of a nucleotide with another, deletions, insertions etc), which can have as effects: modification of genes expression levels, inactivation of genes expression and modification in aminoacid sequence of the proteins codified by these genes.

Some of these genetic variants from the 6 loci have a positive effect on milk casein content, coagulation time, curd firmness and cheese making efficiency, others having a negative effect on these parameters; some \( \beta \)-CN genetic variants (A\(_1\), B, C) were associated with some illnesses in humans as Diabetes Mellitus type 1, Ischemic Heart disease, Sudden Infant Death Syndrome, Schizophrenia and Autism; some of \( \alpha \text{S1-CN} \) genetic variants are incriminated for causing allergies in children, following milk consumption.

Using of the information provided by the 6 major milk proteins genetic variants, cannot be done without a characterization of Romanian farm species/breeds.
The characterization accomplished in my Ph.D. Thesis, provides extremely valuable information about the alleles frequencies in the six milk proteins loci in local Romanian species/breeds. This can open new possibilities of using these polymorphisms as genetic markers in animal breeding, in manipulation of milk components by genetic selection in order to create alternative milk formulas in order to overcome some illnesses, or as genetic markers in authenticity/origin identification of dairy products.

PART I: BIBLIOGRAPHIC STUDY

CHAPTER I
COMPOSITION OF CATTLE, BUFFALO, SHEEP, GOAT MILK AND MILK SYNTHESIS MECHANISM IN LACTATING MAMMARY GLAND

The scientific research established that milk has a complex composition, mainly concerning the casein fraction, whey proteins, enzymes and antimicrobial components. The milk represents a highly valuable food source containing more than 100 components in suspension or emulsion needed for a normal development of humans and animals (Banu et al., 1998).

The milk proteins are divided in two main groups according to their behaviour in acid pH = 4,6:

1. **The soluble fraction** in acid pH (whey proteins), is composed of \( \beta \)-lactoglobulin (\( \beta \)-LG), \( \alpha \)-lactoalbumin (\( \alpha \)-LA) and other minor proteins. In the cheese making process by adding chimosin, this fraction remains in whey.

2. **The insoluble fraction in acid pH (whole casein)** is composed of \( \alpha \)S1-casein (\( \alpha \)S1-CN), \( \alpha \)S2-casein (\( \alpha \)S2-CN), \( \beta \)-casein (\( \beta \)-CN) and K-casein (K-CN). The caseins are found in fresh milk in a high number of solid particles in suspension, bounded by \( \text{Ca}_9(\text{PO}_4)_6 \) molecules. These particles are named micelles.
The protein content of cattle milk is in average 3.3%, in buffalo milk 4.1%, in sheep milk 5.3% and in goat milk 3.7% (Iurcă et al., 1998). From the total milk proteins the casein fraction represents 80% and whey proteins 20%.

The proportion of the casein fraction has a significant influence on milk manufacturing properties, quantity and quality of obtained cheese, texture and flavour of different cheeses and other dairy products (Buchberger et al., 2000).

The 6 major milk proteins are coded by 6 non-allelic genes, which are specifically expressed in mammary gland during lactation.

The Mendelian segregation analysis of the 4 non-allelic genes coding for the 4 casein types in ruminants, revealed that they are located on the chromosome pair 6 in the following order: αS1-CN, β-CN, αS2-CN, K-CN (Threadgill et al., 1990; Hayes et al., 1993a). The gene coding for β-LG was located on the chromosome pair 11 (Hayes et al., 1993b) and that coding for α-LA was located on the chromosome pair 5 (Soulier et al., 1989; Vilotte et al., 1987).

The endocrine system, mainly the pituitary gland, plays a central role in all aspects involving growth and development of mammary gland (mamogenesis), starting of lactation and milk synthesis (lactogenesis), maintaining of milk synthesis (galactopoiesis) and then its activity ending, by two major hormones: growth hormone or somatotroph (GH or STH) and prolactin (PRL). The hormonal signals induce the transcriptional activation of the genes coding for major milk proteins and specific protein synthesis.

CHAPTER II
MOLECULAR TECHNIQUES USED FOR MILK PROTEINS POLYMORPHISMS STUDY

The studies accomplished from the early ’50 using paper electrophoresis (PE) evidenced the milk proteins polymorphism, β-LG being the first protein in which a polymorphism was identified by Aschaffenburg and Drewry (1955).
Due to the development of more advanced techniques for protein analysis as: SGE (starch gel electrophoresis), AE (agarose gel electrophoresis), PAGE (polyacrylamide gel electrophoresis), IEF (isoelectric focusing), HPLC (high performance liquid chromatography), Maldi TOF-MS (mass spectrometry), the knowledge concerning the polymorphisms of major milk proteins in farm species was enriched substantially.

The appearance of the DNA based techniques: PCR (polymerase chain reaction), PCR-RFLP (restriction fragments length polymorphism), SSCP (single strand conformation polymorphism), RT-PCR (real time polymerase chain reaction) and DNA sequencing techniques, allowed the identification of many polymorphisms in the genes structure, having repercussions on gene expression and aminoacid composition of milk proteins.

CHAPTER III
CURRENT STAGE OF THE RESEARCH CONCERNING THE STUDY OF MILK PROTEINS POLYMORPHISMS IN CATTLE, BUFFALO, SHEEP AND GOAT

The mutations which appeared over the years in the structure of these genes coding for the six major milk proteins, lead to the appearance of many genetic variants in these loci.

These variations, named polymorphisms, indicate the fact that each milk protein is found in several forms codified by autosomal codominant genes named alleles, meaning that both alleles are expressed in heterozygous individuals.

**Alpha S1-CN** is a composed of 199 aminoacids, having a molecular weight of 23,614 kDa. **In cattle in αS1-CN locus** 10 genetic variants were identified so far: A, B, C, D, Eyak, Ebali, F, G, H, (Farrell et al., 2004) and I^RV^ (Bâlteanu et al., 2007a; Bâlteanu et al., 2008a, b). The I^RV^ allele was renamed I^SM^ because it was not evidenced in other Romanian breeds, being specific to Romanian Grey Steppe cattle, Moldavian variety (SM
Studiul polimorfismelor genetice ale proteinelor majore din lapte la principalele rase de taurine, bubaline, ovine și caprine din România în scopul utilizării lor ca markeri genetici în ameliorare și trasabilitate

= Moldavian variety. **In buffalo**, 3 genetic variants were identified so far: A, B (Chianese et al., 2009) and B\(^{RV}\) (Bâlteanu et al., 2007c; Bâlteanu et al., 2008c). Following the characterization in \(\alpha S1\)-CN locus of Romanian Buffalo populations it was established that this genetic variant is specific to this breed, being renamed according to the region in which it was identified **\(\alpha S1\)-CN B\(^{BT}\)** (BT = Transylvanian Buffalo). **In sheep**, 9 genetic variants were identified so far: A, B, C, D, E, F, G, H, I (Pirisi et al., 1999; Giambra et al., 2010). **In goat**, 17 genetic variants were identified so far: A, B\(_1\), B\(_2\), B\(_3\), B\(_4\), C, D, E, F, G, H, I, L, M, N, 01 and 02 (Ramunno et al., 2004).

**Beta-CN** is a protein composed of 209 aminoacids, having a molecular weight of 23,983 kDa. **In cattle** in \(\beta\)-CN locus 14 genetic variants were identified so far: A\(_1\), A\(_2\), A3, B, C, D, E, A\(_3\)m, B\(_2\), A\(_4\), H, F, A\(_5\), G (Farrell et al., 2004). **In buffalo**, 3 genetic variants were identified so far: A, B (Ferranti et al., 1998) and C\(^{RV}\) (Bâlteanu et al., 2007c; Bâlteanu et al., 2008c). Following the characterization of \(\beta\)-CN locus of Romanian Buffalo populations it was established that this genetic variant is specific to this breed, being renamed according to the region in which it was identified **\(\beta\)-CN C\(^{BT}\)** (BT = Transylvanian Buffalo). **In sheep** there is no clear evidence of polymorphism at the protein level (Chianese et al., 1995). **In goat** 8 genetic variants were identified so far: A, A\(_1\), B, C, D, E, 0 and 0' (Cosenza et al., 2005a; Caroli et al., 2006).

**Alpha S2-CN** is a protein composed of 207 aminoacids, having a molecular weight of 25,150 kDa. **In cattle** in **\(\alpha S2\)-CN** locus 4 genetic variants were identified so far: A, B, C, D (Farrell et al., 2004). **In buffalo**, 3 genetic variants were identified so far: A, B, C (Ferranti et al., 1998). **In sheep**, 2 genetic variants were identified so far: A, B (Rossi et al., 1984; Mauriello et al., 1990). **In goat**, 9 genetic variants were identified so far: A, B, C, E, F, G, D, 0 and 0' (Erhardt et al., 2002).

**Kappa-CN** is a protein composed of 169 aminoacids, having a molecular weight of 19,007 kDa. **In cattle** in **K-CN** locus 11 genetic variants were identified so far: A, B, C, B\(_2\), E, F, G, Az, H, I, J (Farrell et al., 2004). **In buffalo**, 2 genetic variants were identified so far: A, B (Mitra et al., 1998). **In sheep**, 1 genetic variant was identified so far (Chianese et al., 1996; Ceriotti et al., 2004b). **In goat**, 16 polymorphic sites were
identified, corresponding to 13 protein variants and 3 silent mutations, involving 15 polymorphic sites just in exon 4: A, B, B', B", C, C', F, G, H, I, J, L, D, E, K, M (Jann et al., 2004b).

**Alpha-LA** is a protein composed of 123 aminoacids, having a molecular weight of 14,175 kDa. In cattle in α-LA locus, 3 genetic variants were identified so far: A, B, C (Farrell et al., 2004). In buffalo, 2 genetic variants were identified so far: A, B (Chianese et al., 2004). In sheep, 2 genetic variants were identified so far: A, B (Dall’Olio et al., 1989). In goat, 2 genetic variants were identified so far: A_1 and A_2 (Cosenza et al., 2005b).

**Beta-LG** is a protein composed of 162 aminoacids, having a molecular weight of 18,277 kDa. In cattle in β-LG locus 13 genetic variants were identified so far: A, B, C, D, Dr, Dyak, E, F, G, W, H, I, J (Farrell et al., 2004). In buffalo, 2 genetic variants were identified so far: A, B (Chianese et al., 2004). In sheep, 3 genetic variants were identified so far: A, B (Kolde et al., 1983; Schlee et al., 1993) and C (Erhardt et al., 1989). In goat 2 genetic variants were identified so far: A and B (Yahyaoui et al., 2000).

**CHAPTER IV**

**THE IMPORTANCE OF MAJOR MILK PROTEINS POLYMORPHISMS STUDY BY THEIR POSSIBLE USE IN ANIMAL BREEDING AND DAIRY PRODUCTS TRACEABILITY**

**The influence of major milk proteins genetic variants on milk quantity**

In cattle the influence of major milk proteins genetic variants on milk quantity was evaluated in several studies. The BB genotypes form αS1-CN locus, A_2A_2 from β-CN locus, AA from β-LG locus and AA (+15) genotype from α-LA locus, were associated with a higher milk quantity (Bovenhuis et al., 1992; Bleck et al., 1993b; Ikonen et al., 1999, Ng-Kwai-Hang et al., 2006).
In sheep the studies concerning the influence of β-LG genotypes on milk quantity are contradictory. Bolla et al. (1989) and Caroli et al. (1995) observed in Sarda sheep that BB genotype is associated with a higher milk quantity. In another study made on a Sicilian sheep breed, Di Stasio et al. (1992) observed that AB genotypes had the highest milk production. In Valle del Belice breed, Giaccone et al. (2000) observed that AA genotype is associated with a higher milk quantity.

In goat Kumar et al. (2006) observed that individuals with AA genotypes at the β-LG locus had a higher milk production when compared to other genotypes.

The influence of major milk protein genetic variants on milk quality, milk manufacturing properties and cheese making efficiency

In cattle the CC genotype from αS1-CN locus, the BB genotype from K-CN locus and the BB genotype from β-LG locus were associated with a higher casein and whole protein content in milk (Jakob et al., 1994; Fitzgerald et al., 1997; Lunden et al. 1997). The BB genotype from β-LG locus was correlated with a higher fat content in milk (Wedholm et al., 2006; Heck et al., 2009). The milk from CC genotypes in αS1-CN, the BB genotypes from β-CN locus and BB genotypes from K-CN has a shorter coagulation time (Delacroix - Buchet et al., 1994; Fitzgerald et al., 1997; Kubarsepp et al., 2005). The BB genotype from K-CN locus has a positive influence on cheese yield, in comparison with AA genotypes, the observed differences being situated between 2,7-15%, depending on the cheese type produced (Van den Berg et al., 1992; Fitzgerald et al., 1997; Walsh et al., 1998).

The effects of αS1-CN polymorphism on goat milk quality were studied in French breeds (Mahe et al., 1993; Delacroix et al. 1996, Manfredi et al., 2000). The results obtained indicated that A allele in comparison with E and F has a significant effect (positive) on protein and fat content and milk manufacturing properties. In cheese making experiments the following differences were obtained concerning the cheese quantity: +7,4% between AA and EE genotypes, +6,9% between EE and FF and +14,8% between
AA and FF genotypes. The cheese obtained from AA genotypes had a weak specific goat flavour (because of a weaker lipolysis), in comparison with FF genotypes which had a more pronounced goat flavour (because of a higher lipolysis). Similar results were obtained in Spanish breeds (Sanchez et al., 1998), Italian breeds (Meggiolaro et al., 2000), Norwegian breeds (Vegarud et al., 1999) and SUA (Clark et al., 2000).

In sheep the CC genotype from αS1-CN locus has a positive influence on milk composition and cheese yield. In cheese making experiments the following differences were obtained concerning the cheese yield, +3.5% and +8.6% respectively, between the CC genotype, in comparison with CD and DD genotypes (Pirisi et al., 1999). The BB genotype from β-LG locus was associated with a higher fat content in milk and lower in lactose, having a favorable effect on cheese yield (Celuk et al., 2006).

The use of major milk protein polymorphism as genetic markers in authenticity identification of milk and dairy products

At international level a high number of methods were proposed for the identification of possible adulteration caused by undeclared inter-specific milk mixtures. In most cases dairy products authenticity identification is based on major milk protein polymorphism analysis: a) Polyacrylamide gel electrophoresis in native conditions (Kaminarides et al., 2002); b) Polyacrylamide gel electrophoresis in denaturant conditions (Veloso şi colab., 2002); c) Isoelectric focusing electrophoresis (Addeo et al., 1990; Anonymous, 2001; Bâlteanu, 2005; Bâlteanu et al., 2007b,c,d; Vlaic et al., 2008; Bâlteanu et al., 2008c); d) Capilarity electrophoresis (Lee et al., 2001); e) Immunochemical methods - ELISA (Hurley et al., 2004); f) Reverse phase high performance liquid chromatography (Veloso et al., 2002); g) Mass spectrometry (Siciliano et al., 2000); h) DNA based techniques (Bottero et al., 2002).
PART II: PERSONAL RESEARCH

CHAPTER V
THE RESEARCH AIM AND OBJECTIVES

The study of major milk protein polymorphisms in local species / breeds had as a goal the evaluation of the possibility of using this information in several directions as:

- improving milk quality (especially of casein content), its manufacturing properties and cheese making efficiency;
- obtaining of a hypoallergenic milk by selection of individuals with no αS1-CN in milk, which is the main milk allergen;
- obtaining of a healthier milk which is containing β-CN genetic variants which don’t have a negative effect on human health;
- authenticity / origin identification of milk and dairy products.

The using of information provided by these genetic markers cannot be done without the knowledge of genetic polymorphisms occurring in the six loci in cattle, buffalo, sheep and goat local breeds. This characterization can provide extremely valuable information about the allele frequencies in the four casein loci and the two whey protein loci in local breeds.

In the context of actual research knowledge in this research field, the proposed project has as major objectives research directions with importance on international level, but little or not known at the national level:

- Setting up of a FISH protocol (fluorescence in situ hybridisation) using BAC probes (bacterial artificial chromosome), in order to map on metaphase chromosomes and interphase nuclei the genes coding for major milk proteins;
- The characterization of genetic polymorphisms in the 6 loci coding for the 6 major milk proteins in some cattle breeds: Romanian Simmental, Romanian Black and White, Brown, Red Holstein, Transylvanian Pinzgau, Romanian Grey Steppe cattle;
- The characterization of genetic polymorphisms in the 6 loci coding for the 6 major milk proteins in Romanian Buffalo;
- The characterization of genetic polymorphisms in the 6 loci coding for the 6 major milk proteins in Carpathian goat;
- The characterization of genetic polymorphisms in the 6 loci coding for the 6 major milk proteins in some local ovine breeds: Turcana, Carabasa, Tigaie, Rusty Tigaie, Cluj Merinos and Botosani Karakul;
- The molecular characterization of new discovered allele (renamed $\alpha$S1-casein $I^{SM}$) in Romanian Grey Steppe Cattle, Moldavian variety, in order to study the possibility of using it as biodiversity genetic marker;
- The molecular characterization of new discovered alleles in Romanian Buffalo breed renamed $\alpha$S1-casein $B^{BT}$ and $\beta$-casein $C^{BT}$;
- The study of the possibility to use milk proteins polymorphisms from native farm species/breeds milk, to identify authenticity and origin of milk, cheeses and other dairy products, considering also the 2 new casein allele identified in Romanian Buffalo.

**CHAPTER VI**

**CHROMOSOME MAPPING OF THE GENES CODING FOR MAJOR MILK PROTEINS BY FLUORESCENCE IN SITU HYBRIDISATION (FISH)**

In these experiments I used two detection systems of the hybridised probe: a system with a single antibody labelled with a red fluorophore (TRITC), antibody recognising specifically the digoxigenin, which was used to label one of the BAC probes (the probe containing WAP gene), and a double system, in which a primary antibody is specifically recognising the biotine (used to label the second probe containing the casein cluster), which is further recognised by a secondary antibody labelled with a green fluorophore (FITC).
The hybridising with mentioned probes was done on metaphase chromosomes, obtained from rabbit fibroblasts cell cultures and mouse mammary tumour cells HC11. The hybridisation was also done on bi-dimensional and three-dimensional interphase nuclei with a preserved structure, prepared from HC11 cells. In all cases were obtained specific hybridisation signals, which allowed a correct mapping on the chromosomes and specific chromosomes territories of studied genes.

CHAPTER VII
THE STUDY BY PCR-RFLP OF THE TWO GENES POLYMORPHISM CODING FOR K-CASEIN AND β-LACTOGLOBULIN IN CATTLE BELongING TO ROMANIAN SIMMENTAL BREED

The study of K-CN şi β-LG genes polymorphism had a goal testing some PCR-RFLP protocols, which two allow the identification of the two common A and B allele from these loci. In the case of K-CN a 350 bp fragment was amplified and digested with Hinf I, three restriction profiles being observed corresponding to AA, AB and BB, from this locus. In the case of β-LG, a 262bp fragment was amplified and digested with Hae III, three restriction profiles being observed corresponding to AA, AB and BB, from this locus. The calculation of genotypes frequencies in K-CN locus, revealed a higher frequency of AB genotype, in comparison with the other two. The calculation of genes frequencies this locus revealed a higher frequency of A allele, in comparison with B allele. The calculation of genotypes frequencies in β-LG locus revealed a higher frequency of AB genotype, in comparison with the other two. The calculation of genes frequencies this locus revealed a higher frequency of B allele, in comparison with A allele.

The obtained results are indicating that the two DNA tests can be used in identification of the common A and B allele from the two loci, allowing a precocious genotyping of cattle for the two genetic markers.
CHAPTER VIII
COMPARATIVE STUDY BY IEF AND PCR-RFLP OF MAJOR MILK PROTEINS GENETIC POLYMORPHISMS IN CATTLE BELONGING TO ROMANIAN SIMMENTAL BREED

The study was done on 14 individuals belonging to Romanian Simmental breed, which were comparatively genotyped in K-CN and β-LG loci by PCR-RFLP and IEF. This experiment had as a goal the evaluation of genotyping efficiency at the DNA level (by PCR-RFLP) and at the expression level directly from milk (by IEF), in the loci coding for major milk proteins.

Following the comparative analysis were identified in K-CN and β-LG loci the same genotypes by PCR-RFLP and IEF. Moreover by IEF were identified in the same gel the genotypes in the other 4 loci coding for αS1-CN, αS2-CN, β-CN and α-LA, allowing a correct identification of all allele, even of those which are very rare.

The obtained results are evidencing the fact that IEF technique allows to establish more precisely the genotypes in the 6 loci, even of the rare allele, difficult and costly identifiable by PCR.

CHAPTER IX
THE CHARACTERIZATION OF MAJOR MILK PROTEINS POLYMORPHISMS IN SOME ROMANIAN CATTLE, BUFFALO, SHEEP AND GOAT BREEDS BY IEF TECHNIQUE

In cattle were genotyped by IEF in the 6 loci coding for major milk proteins a number of 693 individuals belonging to 7 breeds, as follows: Romanian Simmental (236 individuals), Romanian Black and White (230 individuals), Red Holstein (13 individuals), Brown (134 individuals), Transylvanian Pinzgau, black variety (26...
individuals) and red variety (30 individuals) and Romanian Grey Steppe cattle, Moldavian variety (24 individuals).

In αS1-CN locus two common genetic variants were identified, namely B and C. The frequency of B allele was 0,9 in improved breeds in direction of higher milk production, the highest frequency being observed in Red Holstein breed. The lowest frequency was observed in Black Pintzgau. The frequency of C variant, associated in many cattle breeds with a higher milk processing quality, was the biggest in Black Pintzgau (0,308), this allele being absent in Red Holstein breed. In αS1-CN casein locus a new allele, renamed I\textsuperscript{SM}, was identified in Romanian Grey Steppe cattle, Moldavian variety, with an isoelectric point between those of B and C genetic variants. The frequency of this new allele, identified by IEF, was 0,041 in analysed populations.

In β-CN locus five genetic variants were identified: A\textsubscript{1}, A\textsubscript{2}, A\textsubscript{3}, B, C. The frequency of ancestral allele A\textsubscript{2} was the highest in almost all breeds (over 0,5) in comparison with the others allele. The exception was observed in Red Pintzgau, in which the frequency of A\textsubscript{1} variant was higher (0,483), in comparison with the other alleles. The A\textsubscript{3} variant was detected only in one individual belonging to Romanian Simmental breed. The B variant, associated in many cattle breeds with a higher milk processing quality, has an extremely reduced frequency in all breeds (between 0,039-0,117). The C variant was detected with a very low frequency in Romanian Simmental, Brown and Pinzgau, black variety, breeds. In Grey Cattle the presence of C variant was detected only in individuals from Tazlau şi Tupilati, this being a proof of infusion with Brown breeds. It was not detected in individuals, considered pure breed, from SCDCB Dancu.

In αS2-CN locus only one genetic variant was identified, namely A variant, with the frequency 1.

In K-CN locus three genetic variants, named A, B and C, were identified. The highest frequency of A variant was observed in Black and White breed (0,835) and of B variant was observed in Brown (0,619). In Romanian Simmental breed the frequency of A allele is very high (0,679). The C variant was detected with a low frequency in Romanian Simmental and Brown breeds.
In α-LA locus only one genetic variant was identified, namely B, with the frequency 1.

In β-LG locus three genetic variants were identified: A, B and C. The frequency of A variant was the highest in Romanian Simmental and Grey Steppe cattle (over 0.5). The frequency of B allele is high in all breeds, in Red Holstein breed being 0.769.

In buffalo were genotyped by IEF in the 6 loci coding for major milk proteins a number of 139 individuals belonging to Romanian Buffalo. In αS1-CN and β-CN loci were identified with a 0.86 frequency two common allele: A in αS1-CN and B in β-CN locus. Two new genetic variants, named αS1-CN B^SM and β-CN C^SM, were identified with a frequency of 0.14. They were observed in linkage on chromosome 6. In the other four loci just one allele was identified with the frequency 1.

In sheep were genotyped by IEF in the 6 loci coding for major milk proteins a number of 282 individuals belonging to six breeds, as follows: Turcana (44 individuals), Carabasa (40 individuals), Tigaie (45 individuals), Rusty Tigaie (28 individuals), Cluj Merinos (35 individuals), Botosani Karakul, black variety (15 individuals), dark grey variety (15 individuals), brown variety (15 individuals), light grey variety (15 individuals), pink variety (15 individuals), white variety (15 individuals).

In all six ovine breeds genotyped, the milk proteins polymorphisms was very reduced, the only polymorphic locus being β-LG, in which two genetic variants were identified A and B. The A variant had the highest frequency (over 0.5) in Turcana, Carabasa, Cluj Merinos, and Karakul respectively, having the highest frequency in Karakul, white variety. The B allele had a higher frequency in comparison with A allele in Tigaie and Rusty Tigaie breeds.

In αS1-CN, β-CN, αS2-CN, K-CN and α-LA just one allele was identified in each locus with the frequency 1. The presence of C allele in αS1-CN locus has a positive meaning, because this allele was associated in many studies with higher milk processing parameters and higher cheese yield.

In goat were genotyped by IEF in the 6 loci coding for major milk proteins a number of 283 individuals belonging to Carpathian goat.
In $\alpha S1$-CN four allele categories, associated with four expression levels were identified. The results indicate a 0.569 frequency of high expression allele in $\alpha S1$-CN (A, B, C) and a frequency of 0.431 in the case of medium, low expression and null alleles. If we consider that in Romania there are 700,000 goats and the calculated frequency of E, F and 0 allele is 43.1%, we can deduce that a number of 301,700 of Carpathian goat are producing a milk with a low protein content, which are significantly affecting milk quality and cheese making efficiency. These results can explain the heterogeneity of Carpathian goat concerning some milk quality parameters. In $\alpha S1$-CN locus, two new possible alleles were identified, named $\alpha S1$-CN 0 and $\alpha S1$-CN X. The first was identified with a 0.021 frequency and is characterised by the absence of $\alpha S1$-CN in milk. It was observed always in linkage with an unknown profile in $\beta$-CN locus, proving the linkage of the two alleles on the same chromosome. The second genetic variant, $\alpha S1$-CN X, was observed just in one individual in a heterozygous condition with B allele. Based on the electrophoresis profile, it seems to present in the sequence a major restructuration, probably caused by an exon skipping phenomenon as that found in F allele. Based on the electrophoresis bands colour intensity, was concluded that this variant belongs to high expression alleles category.

In $\beta$-CN locus three genetic variants were identified. The common A and C variants were identified with a 0.979 frequency. They were not differentiated by IEF due to the same isoelectric point. In some individuals was identified, with a frequency of 0.021, an unknown genetic variant named $\beta$-CN X, always in linkage with $\alpha S1$-CN 0.

In $\alpha S2$-CN locus three genetic variants were identified. The variant C of $\alpha S2$-CN had the highest frequency (0.553), in comparison A variant (0.431), the E allele having a low frequency (0.016).

In K-CN locus two genetic variants were identified. The A variant was predominant (0.912), the B variant having a low frequency (0.088). The presence of B variant in Carpathian goat has a positive significance, as long as this allele was associated in different studies with a high milk processing quality.

In $\beta$-LG şi $\alpha$-LA, just one allele was identified in each locus with the frequency 1.
MOLECULAR CHARACTERIZATION OF THE NEW ALLELES IDENTIFIED IN αS1-CASEIN AND β-CASEIN LOCI IN ROMANIAN GREY STEPPE CATTLE, MODAVIAN VARIETY AND ROMANIAN BUFFALO

In Romanian Grey Steppe cattle

A new genetic variant in cattle αS1-CN locus, renamed I\textsuperscript{SM}, was identified in Romanian Grey Steppe cattle, Moldavian variety, being complete characterized using a combined methodology of protein, cADN and DNA sequencing.

By Maldi Tof-MS analysis, a substitution in the position 192 of mature protein was identified. In this region I\textsuperscript{SM} variant is containing Gly as C variant and not glutamic acid, present in B variant. The other substitutions which are making the difference between I\textsuperscript{SM} variant and B or C were not clearly identified by Maldi Tof-MS analysis.

Comparative analysis of chromatograms obtained following sequencing (with the forward and reverse primer, respectively) of the 2 samples carriers of α\textsubscript{s1}-CN I\textsuperscript{SM} (B\textsuperscript{SM}, C\textsuperscript{SM}) and 2 reference samples (BB, CC), revealed the mutations characterizing this new genetic variant: substitution of an adenine from C and B allele (exon 11, nucleotide 297, codon 99 gaA coding for Glu) with a thymine in the I\textsuperscript{SM} allele (gaT coding for Asp); substitution of an adenine from the B allele (exon 17, nucleotide 620, codon 207 gaA coding for Glu) with guanine in C and I\textsuperscript{SM} alleles (gGa coding for Gly). This last substitution was confirmed at the protein level by Maldi Tof-MS analysis.

There are 2 additional substitutions observed in the B, C and I\textsuperscript{SM} alleles (sequenced from Romanian Grey Steppe cattle, Moldavian variety), that differ from some sequences published in GenBank, with no effect on aminoacid sequence of protein.

Based of the substitution of an adenine from position 21 of exon 11 from C allele (adenine found in B allele too), with a thymine in I\textsuperscript{SM} allele, a PCR-RFLP test was developed based on the amplification and restriction with BseGI enzyme of a 422 base
pairs (bp) fragment from αS1-CN gene, including entire exon 11. Because the presence of this new αS1-CN allele was noticed only in pure breed individuals, its origin in this breed is unquestionable. This theory is also sustained by the high frequency of this allele in the genotyped Grey Steppe cattle population: 0,128, frequency relatively high for a new genetic variant. This suggests that I<sup>SM</sup> variant had an even higher frequency at the time when this breed and its progenitors were the only ones existing in Romania.

Based on similarities and differences observed at the molecular level, between αS1-CN B, C and I<sup>SM</sup> allele, was concluded that this new allele was issued directly from the ancestral C allele, bringing the first molecular clue about phylogenetic relationships among the Romanian Grey Steppe cattle and <i>Bos genus</i> representative breeds from Africa and Asia.

**In Romanian Buffalo**

The cADN sequencing of αS1-CN A (common allele) and B<sup>SM</sup> alleles from Romanian Buffalo, revealed two point mutations characterizing B<sup>SM</sup> allele, different from the other known allele in Mediterranean and Indian Buffalo. The deletion of entire exon 6 is a very interesting type of restructuration, shortening the sequence of mature protein with 8 aminoacids, in comparison with the normal protein which has 199 aminoacids. Three further mutations were identified, which are similar or different from those found in common allele.

The cADN sequencing of β-CN B (common allele) and C<sup>Bt</sup> alleles, leaded to identification of four further mutations differentiating the two variants and affecting the predictable protein sequence. Furthermore two additional silent substitutions were identified with no effect on protein sequence, as long the resulting codons are coding for the same aminoacids.

As long as these new identified allele are specific to Romanian Buffalo they could be easily used as genetic markers for Romanian Buffalo cheeses authenticity/origin identification.
CHAPTER XI
IDENTIFICATION OF DECLARED AUTHENTICITY/ORIGIN OF MILK AND DAIRY PRODUCTS USING THE MAJOR MILK PROTEINS POLYMORPHISMS AS GENETIC MARKERS

If in other EU countries, products adulteration cases are quite rare, in Romania undeclared cow milk (produced at a lower price and in higher quantities), is often added into buffalo, goat or ewe milk, normally fated to produce „authentic” dairy products, especially in the case of industrial processed milk.

The official European methodology for cheeses authenticity identification is based on the IEF analysis of gamma casein fractions, resulting from the hydrolysis of β-CN by plasmin, having the disadvantage that it cannot be differentiated goat and sheep milk.

Up-to-date in Romania there is no identification methodology of undeclared inter-specific milk addition in dairy products.

The researches described in this Chapter had as a main goal the study of possibility to use milk proteins polymorphisms, found in whole six loci from native farm species studied, as genetic markers to identify authenticity and origin of dairy products. Also, the study of possibility of using the two alleles identified in Romanian Buffalo β-CN C^{SM} and αS1-CN B^{SM}, as genetic markers for origin/authenticity identification of Romanian Buffalo cheeses, was another goal of this research.

The authenticity analysis concerning the possible identification of undeclared inter-specific milk mixtures was done on 12 cheese types form Romanian market, as follows: 2 types of cattle cheeses, 3 types of buffalo cheeses, 3 types of sheep cheeses and 4 types of goat cheeses.

In the case of cattle cheeses it was found that these were produced only form cattle milk. This is not surprising because the using of other milk types in cattle cheese production in less probable. In the case of buffalo cheeses, it was observed that the majority of them were produced exclusively from cattle milk. In the case of sheep
cheeses it was observed that two of them were produced exclusively from sheep milk, and one had a 50% cattle milk. In the case of goat cheeses it was observed that three of them were exclusively produced from goat milk, the forth one being falsified 100% with cow milk.

The comparison between two buffalo cheeses types, a Mozzarella type produced form Italian buffalo milk from Campania region, Italy and a salty cheese produced from Romanian Buffalo from Transylvania, clearly revealed that they can be differentiated by the presence in salty cheese of the αS1-CN B\textsuperscript{SM} and β-CN C\textsuperscript{SM} genetic variants, which are absent in Italian Buffalo. Therefore the possibility of using these genetic variants (which were not observed so far in other European or Asian buffalo breeds), as genetic markers for authenticity origin and traceability identification of Romanian buffalo cheeses is possible immediately.

CHAPTER XII
GENERAL CONCLUSIONS AND RECOMMENDATIONS

1. The cytogenetic experiments allowed a correct mapping on metaphase chromosomes and specific chromosome territories of the four caseins and WAP gene. As a consequence I recommend these protocols for mapping of genes on specific chromosomes territories, the study of linkage, crossing over and of the karyotype, in order to identify deletions, duplications, inversions, translocations, which are not visible by classical cytogenetic methods.

2. Following the comparative analysis were identified in K-CN and β-LG loci the same genotypes by PCR-RFLP and IEF. Moreover by IEF were identified in the same gel the genotypes in other 4 loci coding for αS1-CN, αS2-CN, β-CN and α-LA, allowing a correct identification of all alleles, even of those which are very rare.

3. Were genotyped by IEF in the 6 loci coding for major milk proteins a number of 693 individuals belonging to 7 cattle breeds, 139 individuals belonging to a local buffalo breed, 282 individuals belonging to six local sheep breeds and 283 individuals belonging
to a local goat breed. Following genotyping alleles and genotypes frequencies, in the 6 loci coding for major milk proteins, were calculated.

4. As a consequence I recommend the PCR-RFLP technique for precocious genotyping of cattle for the two markers and the IEF technique for genotyping in the 6 milk protein loci, to a more precisely genotyping and for the evaluation of alleles expression levels.

5. There were identified and molecular characterized 3 new genetic variants, one in Romanian Grey Steppe cattle in αS1-CN locus and two in αS1-CN and β-CN loci in Romanian Buffalo. In Carpathian goat breed were identified 3 new possible genetic variants, two in αS1-CN locus and one in β-CN, which are in progress of characterization.

6. Using a combined analysis for detection of casein fraction and whey proteins polymorphisms found in Romanian farm species, were rapidly identified the species in the bulk milk samples and analysed cheeses. It can be concluded that the genetic variants of major milk proteins found in Romanian farm species (the majority of them being found in all world breeds), can be successfully used as genetic markers in identification of authenticity/origin dairy products.

**SELECTIVE BIBLIOGRAPHY**


Studiul polimorfismelor genetice ale proteinelor majore din lapte la principalele rase de taurine, bubaline, ovine și caprine din România în scopul utilizării lor ca markeri genetici în ameliorare și trasabilitate

Studiul polimorfismelor genetice ale proteinelor majore din lapte la principalele rase de taurine, bubaline, ovine și caprine din România în scopul utilizării lor ca markeri genetici în ameliorare și trasabilitate