THE STUDY OF GENETIC POLYMORPHISMS WITHIN THE MAJOR MILK PROTEINS OF DOMESTIC PIGS
(SUMMARY OF Ph.D. THESIS)

Scientific coordinators:
Prof. AUGUSTIN VLAIC, Ph.D.
Prof. ROBERT RENAVILLE, Ph.D.

CLUJ-NAPOCA
2011
CONTENTS

RESEARCH AIM, OBJECTIVES AND IMPORTANCE .......................................................... 3

CHAPTER I. CURRENT STATE OF RESEARCH REGARDING THE MAJOR PORCINE MILK PROTEINS ............................................................................................... 6

1.1. STUDIES CONCERNING THE CASEIN FRACTION ............................................ 7

1.1.1. The study of porcine αS1-casein polymorphisms .................................... 8

1.1.2. The study of porcine αS2-casein polymorphisms .................................... 8

1.1.3. The study of porcine β-casein polymorphisms ......................................... 9

1.1.4. The study of porcine κ-casein polymorphisms........................................ 10

1.2. THE STUDY OF PORCINE WHEY PROTEINS ............................................. 11

1.2.1 The study of porcine α-lactalbumin polymorphisms ................................. 11

1.2.2. The study of porcine β-lactoglobulin polymorphisms............................ 12

1.2.3. Studies regarding porcine WAP polymorphisms..................................... 13

CHAPTER II. PORCINE MILK PROTEIN POLYMORPHISM STUDY, AT PROTEIN LEVEL ................................................................. 14

2.1. MATERIALS AND METHODS ......................................................................... 14

2.2. RESULTS AND DISCUSSIONS ....................................................................... 14

2.3. CONCLUSIONS......................................................................................... 17

CHAPTER III. PORCINE β-CASEIN c.647A>G POLYMORPHISM.............................. 18

3.1. MATERIALS AND METHODS .................................................................... 18

3.2. RESULTS AND DISCUSSIONS .................................................................... 19

3.3. CONCLUSIONS....................................................................................... 23

CHAPTER IV. PORCINE β-CASEIN c.294T>G POLYMORPHISM.............................. 24

4.1. MATERIALS AND METHODS .................................................................... 24

4.2. RESULTS AND DISCUSSIONS .................................................................... 24

4.3. CONCLUSIONS....................................................................................... 26

CHAPTER V. PORCINE β-CASEIN c.175G>A POLYMORPHISM.............................. 27

5.1. MATERIALS AND METHODS .................................................................... 27

5.2. RESULTS AND DISCUSSIONS .................................................................... 27

5.3. CONCLUSIONS....................................................................................... 30

SELECTIVE REFERENCES................................................................................. 31
RESEARCH AIM, OBJECTIVES AND IMPORTANCE

Biochemical polymorphism of proteins is a widely spread phenomenon in all animal species, meaning that in individuals belonging to the same species a certain protein may be present in more than one molecular form.

Research aim

The research aim of the present thesis was to identify and characterize at a molecular level the main polymorphisms that occur in the case of the 7 major proteins found in the milk of domestic pigs. This goal is of utmost importance and scientific relevance, having in mind that similar studies are scarce at an international level and that in Romania and Belgium this subject has never been approached.

Research objectives and methodological activities

The objectives and the methodological activities of the research, in accordance to those established in the research project, were:

1. The identification of the main polymorphisms that occur in the case of the major porcine milk proteins

This objective aimed mainly to identify the major porcine milk protein polymorphisms and also the individuals carrying different variants. In order to achieve this goal, the following actions were undertaken:

- Identification of swine farms in order to find animals bred in pure breed (knowing the fact that in industrial pig farming only hybrids are used);
- Collecting samples of milk/colostrum from lactating sows belonging to local (unimproved) and imported (highly improved) porcine breeds;
- Identification (using IEF) of the genetic polymorphisms and of the individuals carrying polymorphic variants;
- Determination of the genetic structure in the analyzed populations.

2. The characterization of the identified polymorphisms, at DNA level

Within this objective not only the one polymorphism identified by IEF, but three \( \beta \)-casein polymorphisms (the other two discovered while sequencing) have been fully characterized at a molecular level.
To accomplish this objective the following activities have been carried out:

- RNA extraction from samples (from animals) identified to be carriers of polymorphic variants;
- Primer design and cDNA synthesis for samples identified as polymorphic;
- Sequencing the identified variants;
- Sequence alignments; theoretical restriction maps;
- Development of genotyping protocols (PCR-RFLP);
- Genotyping (using the developed tests) – in order to establish gene and genotype frequency.

3. Testing the influence of different protein variants upon the growth dynamic of litters during suckling

This objective aimed to determine the influences of different major milk protein variants upon the growth dynamic of the litters during the suckling period.

To accomplish this objective the following activities have been undertaken:

- Identifying sows that present polymorphisms at one or more milk proteins loci;
- Confirming polymorphisms (with the aid of IEF technique or PCR tests);
- Choosing sows that will constitute experimental lots (having the highest level of genetic similarity, meeting same conditions during lactation, having approximately the same age etc.);
- Keeping track of the growth dynamic of the litters belonging to experimental lots (from farrowing to weaning);
- Centralizing and statistically interpreting the obtained data.

4. Analysis of the relations between the sow’s genotype at the β-casein locus (CSN2) and the reproductive performances of the sows

Numerous studies, reported the presence of loci involved in prolificacy on the 8th chromosome pair in pigs (SSC 8), loci involved in the quantitative traits determinism (quantitative trait loci – QTL). According to our knowledge, there is no molecular marker that can be used in marker assisted selection in pigs, on the SSC 8, for prolificacy. For this purpose we used the “candidate genes” approach, from the premises that the CSN2 locus is also situated on SSC 8, by studying the relations between the sows’ genotypes at the β-casein locus and their reproduction performances; aiming to highlight the existence
of some possible linkage groups, which could later allow the identification of QTLs involved in prolificacy, based on β-casein genotypes.

Activities carried out:

- Genotyping of 150 sows, regarding the CSN2 locus (blood sampling, DNA extraction, amplification, migration, enzyme restriction, migration, genotype summarizing);
- Summarizing and processing the recorded performances regarding the analyzed traits and the statistical interpretation of the results.

The importance of the research resides in the fact that the results obtained regarding the study of genetic polymorphisms of the major porcine milk proteins, will substantially contribute to the enrichment of the global scientific knowledge in the field. Research conducted during the last three years have shown that among the seven major porcine milk proteins, β-casein presents the highest degree of polymorphism, within and between porcine breeds, both in Romania and in Belgium. Among the numerous results presented in the thesis I would like to point out the discovery of a new porcine β-casein allele (GenBank Accession Number GU827390) and the complete characterization of an alternative splicing phenomenon that leads in certain sows to the synthesis of a much shorter β-casein isoform. Three original genotyping protocols have been developed. Also, a porcine β-lactoglobulin polymorphism was identified – the characterization of the two variants is in progress.

The study of the associations between the genetic variants of the porcine β-casein and production and reproduction traits are elements of absolute novelty, similar studies have never been carried out in this species.

Having in mind that genes encoding major milk proteins are used worldwide in marker assisted selection in other species (αS1-casein in goats – associated with milk protein content and flavour; κ-casein and β-lactoglobulin in cattle – associated with milk quantity and quality, bovine β-casein – associated with human health etc.) it is probable that highly polymorphic genes such as porcine β-casein will be the next selection markers in pig breeding.
CHAPTER I
CURRENT STATE OF RESEARCH REGARDING MAJOR PORCINE MILK PROTEINS

Milk proteins are divided in two major groups: the casein fraction and the whey proteins. The first group, characterised by a limited mobility at 4.6 pH (they precipitate), comprises the four caseins: $\alpha_{\text{s1}}$, $\alpha_{\text{s2}}$, $\beta$- and $\kappa$-casein. Whey proteins remain soluble in an acid pH; they are represented, in most species, by $\alpha$-lactalbumin and $\beta$-lactoglobulin. In porcine milk a third major whey protein was first identified in 1987 (Finkelstein and Hurley, cited by Gallagher et al., 1996), called at the time novel whey protein. Relatively recently (1998), this third major porcine whey protein was characterized, namely WAP (whey acidic protein), a protein also encountered in mouse, rat, rabbit and camel milk (Simpson et al., 1998).

The first milk protein polymorphism was highlighted in bovine milk, 1955, by Aschaffenburg and Drewry, when 2 polymorphic variants of the $\beta$-lactoglobulin were described and named A and B, respectively. Since, the genetic polymorphism of milk proteins was intensely studied in dairy species.

In Romania, among the first studies concerning the genetic polymorphism of milk proteins were conducted by the Animal Genetics Department of USAMV Cluj-Napoca, studies focusing on bovine $\kappa$-casein (Vlaic et al., 2001) and since, the subject of milk protein polymorphisms had an outstanding continuity within the department (Vlaic et al., 2008, 2010; Bâlteanu et al., 2007; Bâlteanu, 2010; Pop et al., 2009; Crîtîuteu et al., 2009, 2011).

Bâlteanu (2010), using IEF and several PCR-RFLP genotyping protocols, studied this aspect in buffalo, cattle, sheep and goat breeds. Interesting results concerning $\alpha_{\text{s1}}$-casein polymorphism correlations with cheese making efficiency in Carpathian goat breed were also obtained at Cluj-Napoca, under the supervision of Prof. dr. Vlaic Augustin (Pop et al., 2009).

According to the information found in the field literature, the polymorphism of porcine milk proteins was never investigated in Romania (present study excepted).
In Belgium, interesting results leading to a patent, concerning bovine κ-casein genotype and milk production parameters were obtained by Renaville and Gengler (2003).

Data found in the field literature reveals that porcine milk protein polymorphisms were never investigated in Belgium (present study excepted).

Worldwide, the genetic polymorphism of milk proteins was intensely studied in dairy species, but having in mind that these species are not the focal point of the present thesis they will not be discussed (for a comprehensive review see Bâlteanu, 2010).

The first porcine milk protein polymorphisms were detected in the 1960s (Glasnak, 1966, 1968; Gerrits and Kraeling, 1967, Gerrits et al., 1969). Concerning major porcine milk proteins, the most notable results were obtained by the research team lead by Jason Lee Alexander, United States of America, which, in the ‘90s, were the first ones to determine the sequences of all known genes (at the time) encoding the major porcine milk proteins (α_{S1}-CN, α_{S2}-CN, β-CN, α-LA, β-LG).

Countless other studies, presented bellow, were focused on this thematic, but most of them were only conducted at a protein level.

1.1. STUDIES CONCERNING THE CASEIN FRACTION

The casein fraction of porcine milk consists of 4 phosphoproteins synthesized by the epithelium of the lactating mammary gland: α_{S1}-casein, α_{S2}-casein, β-casein and κ-casein. These are the most important porcine milk proteins, representing more than 60% of the total milk proteins (Aimutis et al., 1981). Regardless of the species, caseins are characterized by a limited solubility between the pH limits of 4 to 5 (they precipitate).

Caseins (the generic name of the 4 proteins belonging to this fraction) are present in milk under the shape of micelles. Casein micelles are colloidal aggregates formed of caseins and calcium phosphate. They have been identified under this shape in the milk of all investigated species. Having a transport protein role, the micelles transform milk into a slightly viscous fluid, allowing thus the transport of high quantities of calcium and
phosphorus. If the caseins were absent within the mammary gland, these minerals would precipitate (Horne, 2002).

In domestic pigs, these proteins are encoded by codominant autosomal genes situated on the 8th chromosomal pair (Archibald et al., 1994).

1.1.1. The study of porcine αS1-casein polymorphisms

Porcine αS1-casein, as a protein, without the signal peptide (15 AA), has 191 AA, a molecular mass of 22576.03 Da and an isoelectric point (pI) of 5.88 (Șuteu et al., 2009).

At protein level, Erhardt et al. (1989) used ion-exchange chromatography to isolate a protein which he called porcine αsl-casein; however, the migrating profile of this protein in Urea-PAGE suggests that it is unlikely to be αS1-casein since it has a slower electrophoretic mobility than porcine β-casein (Gallagher et al., 1996).

No polymorphic variants of this protein have been so far recorded using protein-based techniques.

At DNA level the first ones to determine the full cDNA sequence of porcine αS1-casein were Alexander and Beattie (1992a). They reported that it contained 191 amino acids and had a molecular weight of 24273 Da, in the non-phosphorylated form.

Alexander and Beattie (1992a) show that, concerning porcine αS1-casein, a length polymorphism is present (caused by altered RNA splicing). Clones lacking (at translational level) regions Gly93-Gln112 and Glu107-Gln112 are described.

Archibald et al. (1994) also report the existence of a polymorphism within this gene, using RFLP analyses. Porcine αS1-casein cDNA clones restricted with TaqI produced two restriction products: ~ 8.7 kb/ ~ 7.7 kb.

Locus identifier: CSN1S1.

1.1.2. The study of porcine αS2-casein polymorphisms

Porcine αS2-casein, as a protein, without the leader peptide, has 220 AA, a molecular mass of 25866 Da and an isoelectric point of 5.78 (Șuteu et al., 2009). The preprotei n has a length of 235 AA (signal peptide included – 15 AA).
At protein level, it was first isolated and characterized by Cerning-Beroard, in 1984, with the use of affinity chromatography, showing that this would contain 200-201 AA (6 to 7 less than bovine α_s2-casein). Moreover, this study revealed that porcine α_s2-casein has a lower content of Thr, Leu and Arg and a higher ones of Ser, Glx, Ile and His than the bovine one. The molecular weight of the phosphate-free polypeptide described by Cerning-Beroard was ~23000 Da; its molecular weight, if 15 or 16 phosphate groups were attached, would be around 24500 Da.

At DNA level the first ones to determine the full sequence of porcine α_s2-casein cDNA were Alexander et al. (1992). They reported a cDNA length of 1093 bp and calculated based on the amino acid composition (220 AA) a molecular weight, in the non-phosphorylated form, of 24273 Da. Several discrepancies can be seen between the protein sequence published in the article and the one uploaded in the NCBI database for the positions 126, 128-139, 141-142 and 145.

In the case of porcine α_s2-casein, Archibald et al., (1994) as a result of using several restriction enzymes, report a BamHI polymorphism. Following the electrophoresis of the restricted products two migrating profiles could be observed of ~15.6 kb and ~12.1 kb respectively.

Locus identifier: CSN1S2.

1.1.3. The study of porcine β-casein polymorphisms

Without the leader peptide (15 AA), porcine β-casein contains 217 AA, has a molecular weight of 24377.32 Da and an isoelectric point of 5.77 (Şuteu et al., 2009).

At a protein level, porcine β-casein was first isolated by Woychik and Wondolowski, in 1969, but it was not characterized (Gallagher et al., 1996).

The full sequence of this protein was obtained based on the cDNA sequence (Alexander and Beattie, 1992b).

Porcine β-casein contains higher levels of Ala, Leu and Lys than bovine β-casein. The porcine variant does not contain Trp and Cys. The protein described by Alexander and Beattie (1992b) has 6 possible phosphorylation sites and a molecular weight of 24397 Da, in the non-phosphorylated form.
The first porcine β-casein polymorphism was described by Glasnak (1966, 1968), who identified β₁-casein\textsuperscript{B}, β₁-casein\textsuperscript{C} and β₂-casein. In 1969, Gerrits identifies a polymorphism in what he then called casein\textsubscript{3}, and names the two variants A and B respectively. Later it was proven that the A and B variants described by Gerrits corresponded to the β₁-casein\textsuperscript{B} and β₁-casein\textsuperscript{C}, identified by Glasnak (Gallagher \textit{et al.}, 1996; Şuteu \textit{et al.}, 2009).

In 1987 and 1989, according to a study conducted by a German research team, lead by G. Erhardt, a new polymorphism of porcine β-casein is presented and the polymorphic variants are named A and B (Gallagher \textit{et al.}, 1996). In 1992, Ng-Kwai-Hang and Groschclaude (\textit{cited by Gallagher \textit{et al.}, 1996}) conclude that Erhardt’s A and B variants are probably the same two variants described above.

At DNA level, the sequence of this gene was also established by Alexander and Beattie (1992b). They reported a cDNA length of 1100 bp, excluding the polyA tail. Based on the amino acid composition (217 AA), the molecular weight of this protein under the non-phosphorylated form, was estimated at 24397 Da.

Locus identifier: CSN2.

Concerning the β-casein locus, Archibald \textit{et al.} (1994) identify following PCR-RFLP analyses, a Sac\textsubscript{I} type polymorphism, after restriction resulting ~16 kb and ~14.4 kb fragments, respectively.

1.1.4. The study of porcine κ-casein polymorphisms

Without the leader peptide (21 AA), porcine κ-casein, has a length of 167 AA, a molecular weight of 18877.46 Da and an isoelectric point of 8.68 (Şuteu \textit{et al.}, 2009).

At protein level, it was first isolated by Woychik and Wondolowski, in 1969 (Gallagher \textit{et al.}, 1996). Cerning-Beroard (1984) isolated and characterized porcine κ-casein, stating that it contains a similar number of amino acid as the bovine κ-casein, although the porcine protein contains 177 AA, while the bovine one contains 169 (Mercier \textit{et al.}, 1976). Porcine κ-casein had higher levels of Asx, Gly, Val and Phe, but lower levels of Glx and Ile, when compared to the bovine one.
Another variant of this protein is the one derived from porcine κ-casein cDNA (Levine et al., 1992). It contains 10 AA less than the one described by Cerning-Beroard (1984).

**At DNA level**, the first ones to determine the full porcine κ-casein cDNA sequence were Levine et al. (1992). They reported a cDNA length of 851 bp. For the mature protein, based on the amino acid composition (167AA), they estimated the molecular weight for this protein in non-phosphorylated form at 18689 Da.

Locus identifier: *CSN3*.

In the case of porcine κ-casein, Archibald et al. (1994) state that an *EcoRI* type polymorphism might exist, but that the obtained fragments had a low intensity.

1.2. THE STUDY OF PORCINE WHEY PROTEINS

Whey proteins, also called “serum proteins” are represented in all dairy species by two major proteins: α-lactalbumin and β-lactoglobulin. In porcine milk a third major whey protein was first identified, namely WAP (whey acidic protein), a protein also encountered in mouse, rat, rabbit and camel milk (Simpson et al., 1998; Sylvie Rival et al., 2001; Şuteu et al., 2009).

1.2.1. The study of porcine α-lactalbumin polymorphisms

The mature porcine α-lactalbumin has a length of 122 AA, a molecular weight of 14155.14 Da and an isoelectric point of 4.68 (Şuteu et al., 2009). Including the leader peptide (19 AA), the preprotein has 141 AA.

**At a protein level**, it was first isolated by Karlsson, in 1966, but it was never characterized (Gallagher et al., 1996).

Schmidt and Ebner (1971) are the ones to isolate and characterize porcine α-lactalbumin stating that it contains 118 AA, and that its molecular weight is 14218 Da. By using starch gel electrophoresis, Schmidt and Ebner (1971) highlight two variants of
this protein, which they name (according to their migration profile) “fast” and “slow”, respectively.

Bell et al. (1981) underline the presence of two genetic variants of this protein and name them A and B, respectively.

**At DNA level**, the porcine α-lactalbumin cDNA sequence was determined by Das Gupta et al. (1992). They reported an 83.1% similarity between the porcine and the bovine variants at a nucleotide sequence level and a 75.2% similarity for the amino acid sequences.

Bleck et al. (1995) identified within this gene a single nucleotide polymorphism and uploaded the two sequences in the GenBank: L31944 and L31945.

Locus identifier: *LALBA*.

**1.2.2. The study of the porcine β-lactoglobulin polymorphisms**

**At a protein level**, mature porcine β-lactoglobulin, has 160 AA, a molecular weight of 18877.46 Da and an isoelectric point of 4.60 (Şuteu et al., 2009). Including the signal peptide (18 AA), the preprotein has 178 AA.

According to Gallagher et al. (1996) this protein was first isolated by Karlsson, in 1966.

Its first polymorphism was identified by Kalan et al. (1967, 1968), two variants being described (β-LG A and β-LG B). In 1969, Kraeling and Gerrits, using vertical electrophoresis in polyacrylamide gel, identified a polymorphism in the whey fraction of porcine milk, naming the identified variants: Whey\(^1\)\(^A\), Whey\(^1\)\(^B\), respectively. Afterwards (Kalan et al., 1971) states that Whey\(^1\)\(^A\), Whey\(^1\)\(^B\) are the exact same proteins as β-LG A and β-LG B. Kemmer also indicated in 1969 two variants of porcine β-lactoglobulin, but they were not described (Gallagher et al., 1996).

Bell et al. (1970) report the isolation of A and B β-lactoglobulin, but these are not characterized either. Later, Bell et al. (1981) identify a third polymorphic variant of porcine β-lactoglobulin, which they name C. The authors state that this differentiates itself from β-LG A through the substitution of Glu with Asp in position 9 and the substitution of His with Gln in position 68.
By using isoelectric focalization (IEF), Conti et al. (1989) identify 3 β-lactoglobulin variants, with 4.4; 4.65 and 4.9 as isoelectric points, and they name them β-LG 1, β-LG 2 and β-LG 3, respectively.

At DNA level, this gene was sequenced by Alexander and Beattie (1992c). They reported that the full length of porcine β-lactoglobulin cDNA was 768 bp, and this codifies a mature protein comprising 160 AA, with a molecular weight of 17811 Da.

For one out of the five sequenced clones, Alexander and Beattie (1992c) noticed an extra codon. This codon (AAA) placed in the 318-325 region, encodes lysine and is considered by authors as the proof of a genetic polymorphism at this locus.

Locus identifier: LGB.

1.2.3. Studies regarding porcine WAP polymorphisms

At protein level, porcine WAP, without the leader peptide (19 AA), has 113 AA, a molecular weight of 22985.86 Da and an isoelectric point of 5.04 (Şuteu et al., 2009).

This protein, known as the main whey protein in mice, rats, rabbits and camels was first isolated and described in porcine milk, by Simpson et al. (1998). The molecular weight of porcine WAP in SDS PAGE proved to be considerably higher than 11.7 kDa, estimated based on the AA composition, indicating that the protein is glycosylated (Simpson et al., 1998).

At DNA level, WAP gene was cloned and described by Sylvie Rival et al. (2001). It has a length of 2087 bp and comprises 4 exons. The gene was mapped, using the FISH technique, around the telomeric region of the 18th chromosome pair (Sylvie Rival et al., 2001).
CHAPTER II
PORCINE MILK PROTEIN POLYMORPHISM STUDY, AT PROTEIN LEVEL

2.1. MATERIALS AND METHODS

Twenty-five porcine milk and colostrum samples have been collected from different porcine breeds and analyzed using IEF, as described by Ţuteu et al. (2009).

2.2. RESULTS AND DISCUSSIONS

Because this technique has never been used before to study milk protein polymorphisms in domestic pigs, the interpretation is slightly. Thus, the isoelectric points (pI) of all 7 proteins were calculated (Tab. 1). The isoelectric points were calculated based on the amino acid composition using the software found on www.uniprot.org webpage. It must be mentioned that these isoelectric points belong to the proteins in a non-phosphorylated state.

Table 1.
Characterization of the major porcine milk proteins
(signal peptide is not included)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Protein length (AA)</th>
<th>Isoelectric point (pI)</th>
<th>Molecular mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{s1}$-casein</td>
<td>191</td>
<td>5,88</td>
<td>22576,03</td>
</tr>
<tr>
<td>$\alpha_{s2}$-casein</td>
<td>220</td>
<td>5,78</td>
<td>25866,00</td>
</tr>
<tr>
<td>$\beta$-casein</td>
<td>217</td>
<td>5,77</td>
<td>24377,32</td>
</tr>
<tr>
<td>$\kappa$-casein</td>
<td>167</td>
<td>8,68</td>
<td>18877,46</td>
</tr>
<tr>
<td>$\alpha$-lactalbumin</td>
<td>122</td>
<td>4,68</td>
<td>14155,14</td>
</tr>
<tr>
<td>$\beta$-lactoglobulin</td>
<td>160</td>
<td>4,60</td>
<td>17863,51</td>
</tr>
<tr>
<td>WAP</td>
<td>113</td>
<td>5,04</td>
<td>11985,86</td>
</tr>
</tbody>
</table>
Based on these isoelectric points, the following protein succession was determined in the IEF gels: β-LG; α-LA; WAP; β-CN; α_{S2}-CN; α_{S1}-CN; κ-CN (Şuteu et al., 2009). Using the current migrating conditions (pH = 4.2-7), the κ-casein variants cannot be visualized, having a pI = 8.68 (Tab. 1).

The analysis of the electrophoretic profiles of the colostrum samples revealed the existence of a polymorphism in individuals belonging to the Large White breed. Given the gel region where the different electrophoretic profiles were observed and taking into account the isoelectric points of the porcine milk proteins we concluded that this was a β-casein polymorphism (Şuteu et al., 2009). The protein variants that presented different electrophoretic profiles, belonging to β-casein, were named A and B. The genetic marker being codominant, the two variants generated 3 genotypes: AA, AB and BB.

Other porcine β-casein polymorphisms have been reported (Glasnak et al., 1966; Gerrits et al., 1969), but given the fact that the techniques used for detection were different, we cannot compare the two variants (A and B) with the ones previously described, reason why, the characterization of this polymorphism at a DNA level was required.

Figure 1 presents the migration profiles of the milk samples belonging to the Mangalita and Bazna individuals, while Large White colostrum samples were used as reference samples. Analyzing this gel, it can be observed that the above mentioned β-CN polymorphism occurs also in the two local breeds.

Also, in Figure 1, a second porcine milk protein polymorphism can be observed, identified using this technique. It is a β-LG polymorphism, protein which in all investigated individuals (excepting 2 individuals belonging to the Mangalita breed) focused in a single point, resulting a single band. The variant present (in homozygous or heterozygous condition) in all investigated individuals, due to its high frequency was named A. The variant which appeared only in two individuals was named B. The characterization at a molecular level of these 2 protein variants (by cDNA sequencing of the genes encoding them) and the comparison with the known sequences will allow the assignment of a final nomenclature. Until the full characterisation of these two β-lactoglobulin variants, it is impossible to compare it with those existing in the field literature (reviewed by Gallagher et al., 1996).
Figure 1. Comparative IEF profiles of some porcine milk and colostrum samples

Lanes 1 and 2 – Bazna breed. Lanes 3 and 4 – Mangalita breed.
Lanes 5, 6 and 7 – Large White breed (reference samples). Original.

Regarding $\alpha_{S1}$-CN, $\alpha_{S2}$-CN, $\alpha$-LA and WAP, in the case of the analyzed individuals, there are monomorph. The identified genotypes, at CNS2 and BLG loci, are presented in Tables 2 and 3.

Table 2.

<table>
<thead>
<tr>
<th>BREED</th>
<th>Number of samples</th>
<th>Observed $\beta$-casein genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AA</td>
</tr>
<tr>
<td>Large White</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Pietrain</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Duroc</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Mangalita</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Bazna</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>
By using the IEF technique, the only goal was to identify the proteins that present different variants (polymorphisms) and of course the carrier individuals. Due to this reason and to the small number of individuals used in this research, in this chapter the gene and genotype frequency was not calculated, the genetic structure of these breeds, on these loci, is to be determined after the genotyping at DNA level.

2.3. CONCLUSIONS

1. Using the IEF technique, a porcine β-casein polymorphism was identified, 2 variants having been observed, named β-CN A and β-CN B, respectively. Pietrain and Duroc breeds exhibit only β-CN A while Large White, Mangalita and Bazna breeds exhibit both variants.

2. In the case of β-LG, a new variant was detected, temporarily named β-LG B. Further studies will lead to a complete characterization and to a final nomenclature of this variant. For the time being, this variant could be restricted to the Mangalita breed.

3. Regarding αS1-CN, αS2-CN, α-LA and WAP, in the case of the analyzed individuals, they are monomorph.
CHAPTER III
PORCINE β-CASEIN c.647A>G POLYMORPHISM

3.1. MATERIALS AND METHODS

**Milk sample collection**
RNA extraction has been carried out, for the individuals carrying different variants of porcine β-casein, from the same milk samples used for IEF (previous chapter).

**Blood sample collection**
A total number of 31 blood samples were collected from individuals belonging to Bazna (12 individuals) and Mangalita (19 individuals) breeds, from the S.C.D.A. Turda farm, in order to extract DNA.

For the Large White, Pietrain and Landrace breeds (Belgium), blood samples have been kindly provided by Professor Nadine Buys (U.K. Leuven) and the DNA was extracted by PROGENUS SA.

**RNA extraction from milk somatic cells**
In order to sequence the porcine β-casein cDNA, RNA was extracted using the PureZOL reagent from individuals identified as carriers of different variants of this protein (AA, AB and BB).

**Obtaining and sequencing the porcine β-casein cDNA**
To obtain the cDNA, a reverse transcription of the extracted RNA samples was carried out, using RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas), following the instructions provided by the producer.

In order to obtain the porcine β-casein cDNA, 2 µl of whole cDNA were amplified using the GoTaq Flexi DNA Polymerase (Promega) amplification kit and the primer pair BCZP-F and BCZP-R. The primers were designed using the Fermentas REviewer software (http://www.fermentas.com) using as reference the sequence of Alexander and Beattie (1992), GenBank X54974, the primers having the following sequences:

- BCZP-F 5’-3’: GCA AGA GCG AAG GAA GAA CT
- BCZP-R 5’-3’: CAA AAG TGA GGA GAG GGC AT
Development of a new PCR-RFLP protocol to highlight the c.647A>G polymorphism of porcine β-casein

The development of a new genotyping protocol involved a multitude of steps such as: sequence alignments, construction of theoretical restriction maps, primer design (described in detail within the thesis).

3.2. RESULTS AND DISCUSSIONS

The porcine β-casein cDNA sequencing of the three reference samples (AA, AB, BB) confirmed the identified polymorphism at this locus (Şuteu et al., 2009).

In Figure 2 a fragment of the sequencing chromatograms can be observed, where the single nucleotide polymorphism (SNP) that caused the occurrence of the 2 protein variants can be observed. This substitution took place in the 647 position of the cDNA, but on the chromatograms it appears in the 578 position due to the fact that the forward primer attached after the coding region of the signal peptide.

In the case of the BB genotype in the 647 position of the cDNA an adenine is present, which in the case of the AA genotype was substituted by a guanine (c.647A>G). In the case of the heterozygous individual, 2 superimposed peaks can be observed, one coloured in green and one coloured black, denoting the fact that in this position both guanine (specific to A variant) and adenine (specific to B variant) are present (Fig. 2).

The c.647A>G substitution modifies the amino acid from the position 201 of the mature protein, because the CAA codon encodes glutamine (Q) while the CGA codon encodes arginine (R).

The sequencing results are in accordance with the results obtained following the IEF migration. In other words, the arginine amino acid (R) from the 201 position gives the mature protein an isoelectric point of pl = 5.99 – a fact pointed out at IEF level by a more cathodic migration profile (A variant). On the other side, if glutamine (Q) is present in the 201 position of the mature protein, the isoelectric point of the protein becomes pl = 5.77 – conferring the protein (B variant) a more anodic migration profile, compared to
the A variant. Of course, the IEF migration profile of the porcine β-casein in the case of heterozygous individuals is characterized by the occurrence of 2 bands.

Figure 2. Sequencing chromatograms of the porcine β-casein cDNA from 3 individuals carrying different genetic variants of these gene. The mutational event is marked with a square (original).

To facilitate discussions regarding the 2 variants, they will be named depending on the amino acid present in the 201 position – that differentiates them (201R and 201Q – Tab. 4).

Table 4.

Final nomenclature of the porcine β-casein variants as a result of c.647A>G SNP

<table>
<thead>
<tr>
<th>Final naming of the β-casein genotypes</th>
<th>Variant naming at IEF migration</th>
<th>Nucleotide in the 647 position of the cDNA</th>
<th>Amino acid in the 201 position of the mature protein</th>
<th>Isoelectric point of the mature protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>201R</td>
<td>AA</td>
<td>GG</td>
<td>Arginine (R)</td>
<td>5.99</td>
</tr>
<tr>
<td>201Q</td>
<td>BB</td>
<td>AA</td>
<td>Glutamine (Q)</td>
<td>5.77</td>
</tr>
<tr>
<td>201QR</td>
<td>AB</td>
<td>AG</td>
<td>Glutamine/Arginine Q/R</td>
<td>5.77/5.99</td>
</tr>
</tbody>
</table>
Regarding the SNP from the position 647 of the cDNA (the A/G substitution), analyzing the known sequences, it can be observed that in the case of their majority in this position, adenine was identified by sequencing (GenBank NM_214434, X54974, EU025876, EU213063), while only two sequences deposited in GenBank reported the presence of guanine (EU242520, GU827390). Since sequences similar to those obtained in this research after sequencing (regarding the 647 position of the porcine β-casein cDNA) were already present in the GenBank, we chose not to deposit them in this database.

By carrying out sequence alignments, it was observed that at DNA level, the nucleotide in the 647 position of the cDNA is part of the 6th exon. This is the largest exon of the β-casein gene in domestic pigs, having a length of 519 bp.

Amplification of the DNA samples was carried out using the primers E6BP-F and E6BP-R. These are complementary to regions inside the introns that flank the 6th exon, the fragment resulting from amplification having a length of 1216 bp (Fig. 3).

![Electrophoretic profiles of three porcine DNA samples, amplified using the E6BP-F and E6BP-R primers. Lane 1 – GeneRuller 1kb DNA Ladder (Fermentas); lane 2 – 201QR individual; lane 3 – 201Q individual; lane 4 – 201R individual; lane 5: negative control (original).](image)

Figure 3. Electrophoretic profiles of three porcine DNA samples, amplified using the E6BP-F and E6BP-R primers. Lane 1 – GeneRuller 1kb DNA Ladder (Fermentas); lane 2 – 201QR individual; lane 3 – 201Q individual; lane 4 – 201R individual; lane 5: negative control (original).

The restriction of the amplification products was carried out using the *StyI* enzyme (Fermentas, FastDigest). In the case of the 201Q allele, the presence of an adenine in the
473 position of the 6\textsuperscript{th} exon (647 position of the cDNA) creates an additional restriction site for the above mentioned enzyme. According to the theoretical restriction map of the amplified fragment, after restriction, the following fragments are obtained:

- \textit{201Q}: 679bp + 434bp + 85bp + 18pb;
- \textit{201R}: 1113bp + 85bp + 18pb;
- \textit{201QR}: 1113bp + 679bp + 434bp + 85bp + 18bp

\textbf{Figure 4.} The migration profiles belonging to the 3 porcine \(\beta\)-casein genotypes, following the restriction of the amplification products with \textit{StyI} enzyme (Fermentas FastDigest). Lane 1 - unrestricted product; lane 2 – AmpliSize Molecular Ruler 50-2000 bp Ladder (BioRad); lane 3 – \textit{201QR}; lane 4 – \textit{201R}; lane 5 – \textit{201Q} (original).

The electrophoretic migration of the DNA samples amplified using the E6BP-F and E6BP-R primers and restricted using the \textit{StyI} enzyme (Fig. 4), confirms the A/G substitution identified by sequencing the porcine \(\beta\)-casein cDNA (Fig. 2), also confirming the trustworthiness of the theoretical restriction map. So, all the above mentioned facts constitute a protocol for genotyping, adequate to point out the Q201R polymorphism of the porcine \(\beta\)-casein gene.

The genotyping carried out up to date, using the above mentioned protocol, led to the following gene and genotype frequencies:
Table 5.
The genetic structure of the investigated populations, at the β-casein locus, regarding the Q201R polymorphism (c.647A>G)

<table>
<thead>
<tr>
<th>BREED</th>
<th>N</th>
<th>Number of individuals for each genotype</th>
<th>Genotype frequencies</th>
<th>Gene frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>201Q</td>
<td>201QR</td>
<td>201R</td>
</tr>
<tr>
<td>Landrace</td>
<td>41</td>
<td>0</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>Pietrain</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Large White</td>
<td>36</td>
<td>2</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Landrace (CRA–W)</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>Mangalita</td>
<td>13</td>
<td>0</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Bazna</td>
<td>11</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

Gene and genotype frequencies (Tab. 5) show that in all investigated breeds 201R variant is predominant. The highest frequency of 201Q variant is encountered in the case of the Bazna breed (0.445).

3.3. CONCLUSIONS

1. The BCZP F and BCZP R primers, whose design was realized within the research team, are adequate for amplifying and sequencing the porcine β-casein cDNA.

2. By sequencing the porcine β-casein cDNA, the polymorphism identified using the IEF technique, was characterized. A c.647A>G substitution in the 647th position of the cDNA leads to the introduction of a glutamine in the 201st position of the mature proteins (if a guanine exists) or the introduction of a arginine (when a adenine exists). The amino acid in this position modifies the pI of the mature protein, phenomenon that constituted the basis for occurrence of different IEF migration profiles. This polymorphism was named “Q201R”, in order to facilitate discussions.

3. The E6BP-F and E5BP-R primers successfully amplify the 6th exon of the CSN2 gene, where the c.647A>G SNP is found at a DNA level. Because the adenine 647th position of the cDNA creates a supplemental restriction site for the StyI enzyme, a genotyping protocol could be established based on this enzyme.
CHAPTER IV
PORCINE β-CASEIN c.294T>G POLYMORPHISM

4.1. MATERIALS AND METHODS

The working steps and the reagents are the exact same ones as described in the previous chapter.

Within the gene, this newly identified point mutation (c.294T>G) is also located in the 6th exon, which allowed the use of the same primer set (E6PB-F and E6PB-R).

According to the theoretical restriction maps, the guanine from the position 294 of the cDNA, in the case of G allele, creates a restriction site for the HpaII enzyme (Fermentas FastDigest); this site has never been identified in none of the known sequences until now (Şuteu et al., 2010, 2011a).

4.2. RESULTS AND DISCUSSIONS

During the experiments carried out in order to characterize the Q201R polymorphism of the porcine β-casein (previous chapter), as a result of the sequence alignments, a new allele was identified at this locus.

Since this was an unknown allele at that time, its sequence was uploaded in the GenBank database, receiving the identification number GU827390. Analyzing the sequencing chromatograms (Fig. 5), it can be seen that they are identical, the only difference between them being the framed nucleotide. At a cDNA level, this point mutation took place in the 294 position, where all porcine β-casein sequences existent in the GenBank database feature a thymine (Tab. 6). The guanine present in the 294 position, that differentiates the new allele from the rest of the sequences present in GenBank, determined us to name it the “G allele”.
Figure 5. The sequencing chromatogram belonging to the porcine β-casein G allele (upper part) compared to a normal allele (lower part). The mutational event is marked with a square (original).

The T/G substitution in the 294 position of the G allele cDNA is actually a “silent mutation”, as both the CCT and the CCG codons, encode the proline amino acid (P). Nevertheless, taking into account all the encountered mutations in the case of this allele (Tab. 6), the protein obtained by theoretical translation of the nucleotide sequence is a distinct one, no absolute similarities with any other known porcine β-casein sequence being recorded, after sequence alignment.

Table 6.

Sequence alignment of all known β-casein sequences

<table>
<thead>
<tr>
<th>Sequence (GenBank Acc. No.)</th>
<th>Position within the cDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>GU827390</td>
<td>T A G G T G C G G</td>
</tr>
<tr>
<td>NM 214434</td>
<td>T A T G T T T G A</td>
</tr>
<tr>
<td>X54974</td>
<td>T A T G T T T G A</td>
</tr>
<tr>
<td>EU24520</td>
<td>A G T A T G C G G</td>
</tr>
<tr>
<td>EU025876</td>
<td>T A T G T T T G A</td>
</tr>
<tr>
<td>EU213063</td>
<td>T A T G C T T A A</td>
</tr>
</tbody>
</table>
Based on the fact that the guanine from the mentioned position creates a restriction site for the \textit{Hpa}II endonuclease, an adequate protocol for distinguishing porcine β-casein G allele has been developed (detailed by Şuteu \textit{et al.}, 2010, 2011a). Genotyping carried out up to date, indicate that G allele is not strictly limited to the Mangalita breed, were it was discovered (Tab. 7).

Table 7.

The genetic structure of the investigated populations, at the β-casein locus, regarding the c.249T>G polymorphism

<table>
<thead>
<tr>
<th>BREED</th>
<th>N</th>
<th>Number of individuals for each genotype</th>
<th>Genotype frequency</th>
<th>Gene frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TT          TG       GG</td>
<td>P_{TT}  H_{TG} Q_{GG}</td>
<td>p_T   q_G</td>
</tr>
<tr>
<td>Landrace</td>
<td>40</td>
<td>27          11       2</td>
<td>0.675  0.275 0.05</td>
<td>0.8125 0.1875</td>
</tr>
<tr>
<td>Pietrain</td>
<td>38</td>
<td>11          19       8</td>
<td>0.29   0.5 0.21</td>
<td>0.54   0.46</td>
</tr>
<tr>
<td>Large White</td>
<td>32</td>
<td>20          12       0</td>
<td>0.625  0.375 0</td>
<td>0.8125 0.1875</td>
</tr>
<tr>
<td>Landrace (CRA–W)</td>
<td>26</td>
<td>17          9        0</td>
<td>0.65   0.35 0</td>
<td>0.825  0.175</td>
</tr>
<tr>
<td>Mangalita</td>
<td>15</td>
<td>5           10       0</td>
<td>0.33   0.67 0</td>
<td>0.67   0.33</td>
</tr>
</tbody>
</table>

4.3. CONCLUSIONS

1. The experiments that aimed to characterize the Q201R polymorphism of the swine β-casein, especially sequencing this gene’s cDNA, allowed the identification of a new allele at this locus, characterized by the presence of a guanine in the 294\textsuperscript{th} position of the cDNA. This new allele was named allele “G” and its sequence was uploaded into GenBank database, receiving the identification number GU827390.

2. The E6BP-F and E5BP-R primers successfully amplify the 6\textsuperscript{th} exon of the CSN2 gene, where the above described SNP (c.294T>G) is found at a DNA level.

3. The guanine from the 294\textsuperscript{th} position of the porcine β-casein cDNA (position where all the other known sequences contain thymine) creates a restriction site for the \textit{Hpa}II enzyme. This fact allowed the development of a PCR-RFLP type test that can differentiate the G allele.
CHAPTER V
PORCINE $\beta$-CASEIN c.175G>A POLYMORPHISM

5.1. MATERIALS AND METHODS

Having in mind that most of the materials and methods used for the characterization of this polymorphism have already been presented in the previous chapters, only the ones that differ will be described in detail.

After agarose gel electrophoresis of $\beta$-casein cDNA amplification products, besides the expected fragment (with a length of 693 bp), another shorter fragment could be visualized (174 bp). Initially, the fragment was considered to be a non-specific amplification product, but due to the fact that after modifying the amplification conditions (increase in the primers annealing temperature, reduction of the primer concentration) the product persisted, we decided to sequence it. In order to do this, the fragment was cut out of the gel, purified and amplified again with the same primers BCZP F and BCZP R. The purification of this fragment was carried out using the Zymoclean Gel DNA Recovery Kit.

5.2. RESULTS AND DISCUSSIONS

Total RNA samples extracted from the somatic cells recovered from milk were reverse transcribed, subsequently being amplified using the primer pair BCZP F and R, porcine $\beta$-casein specific primers. After the sample migration in agarose gel, besides the expected fragment, of 693 bp, a second amplification product, much shorter (174 bp), could be observed.

The 174 bp fragment (Fig 6, lane 3) was gel purified, reamplified and sequenced using the same primers (BCZP F and R). Sequencing the 174 bp fragment revealed that this was a variant of the porcine $\beta$-casein, lacking the 6$^{th}$ exon (519 nt). The sequence of
this new transcriptional variant of the porcine β-casein was uploaded into the GenBank database, receiving the identification number HM114445.

Figure 6. The migration profile of the porcine β-casein cDNA resulted after sample amplification using the BCZPF and R primers. L: 50 bp DNA Ladder (Fermentas). Lane 1: cDNA of a sow exhibiting the alternative splicing phenomenon. Lane 2: porcine β-casein cDNA correctly spliced. Lane 3: the 174 bp fragment cut out and purified from the gel and reamplified. NC: negative control (original).

In order to find the causes that led to the occurrence of this phenomenon of alternative mRNA splicing, a primer pair was created, complementary to portions of the introns that flank the 6th exon (E6BP F and R). These primers were used to amplify this DNA region, in sows where the alternative splicing phenomenon manifested (Fig. 6, lane 1), and in sows whose RNA was correctly spliced (Fig. 6, lane 2). The sequencing of this DNA fragment indicates that a point mutation at the level of the intro-exon junction is the cause for the alternative splicing phenomenon (Fig. 7).

The substitution identified by sequencing (c.175G>A) affects the first nucleotide of the 6th exon. In the sows were the first nucleotide was A, the mRNA was alternatively spliced, 2 transcripts being produced. When guanine is present, the mRNA is properly spliced.

The c.175G>A mutation causes an amino acid substitution (p.Asp59Asn) in the normal length protein (217 AA).
In all the analyzed individuals, the 693 bp fragment was present, in a larger quantity than the 174 bp fragment, indicating that the normal length protein (217 AA) is found in larger quantities even in the milk from the sows having the AA genotype (alternatively spliced), compared to the 44 AA isoform.

The c.175G>A substitution (Fig. 7) can be highlighted using PCR-RFLP, due to the fact that when guanine is present in this position, it creates a supplementary restriction site for the PfoI enzyme.

The β-casein variant deduced based on theoretical translation of the HM114445 sequence has a length of just 44 AA compared to the 217 AA, when the RNA is properly spliced. The alternative splicing phenomenon and the mutation causing it (c.175G>A) were confirmed by sequencing in 6 unrelated individuals, belonging to 3 different breeds.

This substitution causes the alternative splicing phenomenon due to the fact that in the case of the large introns (in this case 1206 nt) and large exons (in this case 519 nt) the
intron/exon architecture plays a crucial role in splicing sites recognition by the spliceosomal machine (Hertel et al., 2008).

As the biological role of this new variant of swine β-casein is not yet known and taking into account that similar events documented in other species had a major impact on the milk quality, is possible that the β-casein (CSN2) may be the next marker used in porcine MAS.

A preliminary association study performed on 26 Landrace sows didn’t reveal any influences upon the growth dynamics of the litters during nursing.

5.3. CONCLUSIONS

1. By sequencing, the substitution of the first nucleotide from the 6th exon of porcine β-casein was discovered (c.175G>A). In the sows presenting A in this position, the mRNA of the β-casein is alternatively spliced, two variants of the same protein being present in the milk. The first variant is the normal one, having a length of 217 AA, while the second is more shorter, having only 44 AA, due to the fact that the c.175G>A substitution modifies the intron-exon architecture; in the case of A presence, the 6th exon (519 nt) is removed during the precursor mRNA maturing (alternative splicing).

2. The mature RNA molecule coding the short variant of the β-casein (44 AA) was reverse transcribed and sequenced. The sequence was uploaded in GenBank receiving the HM114445 identifier.

3. Based on the theoretical transcription of the nucleotide sequences, it has been shown that the c.175G>A mutation, in the case of the normal variant of the β-casein, determines a substitution of the amino- cid situated in the 59th position of the protein, aspartane being replaced by asparagine.

4. Based on the fact that G from the 175th position of the cDNA creates an additional restriction site for the PfoI endonuclease, a new PCR-RFLP test was developed to allow the differentiation at allelic level.
SELECTIVE REFERENCES


