Elucidation of African elephant beta casein phosphorylation state and casein micelle structure

Moses Madende

Submitted in accordance with the requirements for the degree

Philosophiae Doctor

In the Faculty of

Natural and Agricultural Sciences

Department of Microbial, Biochemical and Food Biotechnology

University of the Free State

Bloemfontein

South Africa

Supervisor: Prof. G. Osthoff

Co-supervisor: Dr. G. Kemp

January 2017

ACKNOWLEDGEMENTS

Foremost, I would like to thank the Almighty and ever faithful God. For it is through His endless mercies and grace that carried me through to completion of this journey. May all glory, praise and honour be directed to Him.

I would also like to express my innermost gratitude to my supervisor Prof. Garry Osthoff. You have been an amazing mentor, thank you for your patience and continuous support of my PhD study and also for encouraging my research while giving me an opportunity to grow as a research scientist. Your positive impact on my research has been invaluable. For this and more, I am indebted to you.

To my co-supervisor Dr. Gabre Kemp, thank you for your insightful comments and the hard questions that prompted me to approach my research from various perspectives. I am also grateful for your expert and sincere guidance as well as your resourcefulness in improving the quality my research.

Special and heartfelt thanks are reserved for my dear family. Thank you for believing in me, for your inspiration and continuous spiritual and mental support. It was only through your support structure that I was strengthened during trying times. This milestone is dedicated to you. To Dr. M.L. Malale, thank you for your patience and willingness to get the best out of me.

I also place on record my gratitude to the National Research Foundation (NRF) for funding this project.

Finally, my sincere thanks go to my friends and fellow colleagues for the stimulating discussions and the late nights we spent together. My deepest appreciation for all the wonderful times we shared. I also thank everyone else who in one way or the other contributed to the successful completion of this Ph.D. In particular, I am grateful to Prof Koos Albertyn for providing laboratory resources for the experimental work.

"I have fought the good fight. I have completed the race. I have kept the Faith".

2 Timothy 4 verse 7

CONTENTS

	Page
Title page	
Acknowledgements	1
Contents	3
Motivation	8
Chapter 1: Literature review	
1.1. Introduction	10
1.2. Evolution	12
1.2.1. Evolution of mammals and lactation	12
1.2.2. Evolution of casein genes	17
1.3. Milk biosynthesis and secretion	18
1.4. Milk proteins	20
1.4.1. Caseins	21
1.4.1.1. α_{s1}-casein	22
1.4.1.2. α_{s2}-casein	25
1.4.3.3. β-casein	28
1.4.3.4. κ-casein	31
1.5. Casein micelle	34
1.6. African elephant milk	41
1.7. Analysis of mammary gland products	42
1.7.1. Proteomics of milk	43

1.8. Protein structure prediction 45

1.9. Comparative genomics 48

1.10. Discussion and conclusions 49

1.11. Aims of the study 51

1.12. References 52

Chapter 2: Interspecies comparison of casein micelles by high resolution field emission scanning electron microscope

2.1. Introduction 60

2.2. Materials and methods 62

2.2.1. Sample preparation 62

2.2.2. SEM preparation 63

2.2.3. SEM imaging 63

2.3. Results 64

2.3.1. Cow milk casein micelles 64

2.3.2. Sheep milk casein micelles 66

2.3.3. Human milk casein micelles 67

2.3.4. Horse milk casein micelles 68

2.3.5. African elephant milk casein micelles 69

2.3.6. Frozen milk casein micelles 70

2.4. Discussion 72

2.5. Conclusions 77

2.6. References 79

Chapter 3: Elucidation of African elephant beta casein phosphorylation state

3.1. Introduction 82

3.2. Materials and methods 86

3.2.1. Sample preparation 86

3.2.2. RP-HPLC fractionation 87

3.2.3. Electrophoresis separation 87

3.2.4. Enzymatic dephosphorylation 88

3.2.5. LC MS/MS (Orbitrap) analysis of 2D gel spots 88

3.2.6. In-liquid digestion and phospho-enrichment 89

3.2.7. LC MS/MS (Triple TOF) analysis of phospho-peptides 91

3.2.8. Mass spectrometry data analysis 91

3.3. Results 93

3.3.1. Determination of β-casein sequence 93

3.3.2. Determination of β-casein phosphorylation 94

3.4. Discussion 101

3.5. Conclusions 104

3.6. References 106

Chapter 4: Structure modeling of casein proteins

4.1. Introduction 109

4.2. Materials and methods 111

4.2.1. Homology modeling 111

4.3. Results 112

4.3.1. Alpha caseins 112

4.3.2. Beta caseins 117

4.3.3. Kappa caseins 120

4.4. Discussion 123

4.5. Conclusions 130

4.6. References 132

Chapter 5: Comparative genomics of casein genes

5.1. Introduction 135

5.2. Materials and methods 137

5.2.1 Comparative genomics 137

5.3. Results 137

5.3.1. α_{s1}-casein 137

5.3.2. α_{s2}-casein 141

5.3.3. β-casein 145

5.3.4. κ-casein 149

5.4. Discussion 154

5.5. Conclusions 157

5.6. References 159

Chapter 6: General discussion and conclusions 161

6.1. Future research 169

6.2. References 170

Summary 173

Opsomming 175

MOTIVATION

Milk has been the subject of scientific research for over 150 years and as a result, it is perhaps the best characterized, in chemical terms, of our common foods (Fox and McSweeney, 1998). Secreted by female mammalian species, milk is solely intended to meet all the nutritional requirements of the neonate (McSweeney and Fox, 2013). In addition to energy provision, milk constituents provide a plethora of physiological functions, mostly served by proteins. However, it was apparent from the aforementioned studies that milk and milk fractions are characterized by a wide array of proteins whose concentration spans across several orders of magnitude. Milk proteins possess many functional properties that have attracted great interest from the dairy industry. Milk caseins present an interesting group of milk proteins mainly because of their involvement in the formation of casein micelles, these are amorphous complexes of the individual caseins with large amounts of colloidal calcium and phosphate (Phadungath, 2005).

Interestingly, the exact structure of the casein micelle is still not fully understood (Holt et al., 2013). This can largely be attributed to analytical instrument limitations that result due to the relatively larger size of the casein micelle. As a result, casein micelle structure is currently described by several models. However, most of these models are based on data obtained from dairy mammals, mostly bovine milk caseins. Bovine milk possess all the types of caseins which naturally exist in specific proportions (Ginger and Grigor, 1999).

The same cannot be said about some of the mammalian species, for example, human and African elephant milk, which are naturally devoid of one or two of the caseins (Madende et al., 2015). Such mammalian species may possess casein micelles that are structurally different from their bovine milk counterparts. Moreover, the individual casein proteins may possess different properties that enable them to fulfil their biological functions in milk, regardless of the absence of the other caseins.

The aforementioned provides the scope of our current research where African elephant milk caseins are investigated in order to improve the understanding of the structure of casein micelles. Apart from being a non-dairy ancient mammalian species product, African elephant milk lacks both alpha caseins and contains high levels of β-casein compared to κ-casein (Madende et al., 2015). These unique properties make elephant milk caseins ideal for the casein micelle structure studies.

CHAPTER 1

LITERATURE REVIEW

1.1. INTRODUCTION

Mammals can be described as warm-blooded vertebrate animals that possess mammary glands which are utilized in the production of milk (Lemay et al., 2009). Milk secretion is a common feature among all mammals ranging from large to small as well as arctic and tropical mammalian species (Oftedal, 2012). Milk serves as a highly digestible, concentrated and nutritionally balanced food for the neonate. Secreted milk is extremely varied in composition, this is mainly due to the unique nutritional and physiological requirements of each species (Lefèvre et al., 2010).

Protein composition of mammalian milk also varies considerably even among the same species, for example as observed in cow milk (Bijl et al., 2014). Genetic polymorphism and post-translational modification (PTM) are the main factors attributed to these variations in major milk proteins. Caseins in milk undergo several PTMs, these include phosphorylation at serine and occasionally threonine residues by casein kinases as well as glycosylation at threonine residues (Swaisgood, 1993; Ginger and Grigor, 1999; Phadungath, 2005).

From a nutritional view point, caseins are a source of amino acids but they also provide phosphate binding sites which subsequently enable the binding of minerals such as calcium (Ginger and Grigor, 1999).

Bovine milk contains mainly four gene products of caseins which are termed α_{s1}-, α_{s2}-, β- and κ-caseins (Farrell et al., 2006). The presence and proportions of caseins tend to vary among mammalian species as some mammals are devoid of one or two of the caseins in their milk whereas others contain multiple copies of specific caseins. To highlight the above, human milk contains high levels of β-casein and very low levels of α_{s1}-caseins, whereas α_{s2}-casein is completely absent (D'Alessandro et al., 2010). Horse milk on the other hand has very low levels of κ-casein with a higher β-casein content (Iametti et al., 2001) and two different α_{s2}-caseins (Martin et al., 2013). Sheep milk has the highest milk protein content amongst the dairy ruminants. The milk protein content is predominated by β-casein whereas the κ- and α-caseins content is approximately equivalent (Martin et al., 2013). Caseins in milk exist as amorphous aggregates of the individual caseins with colloidal calcium and phosphates, these aggregates are known as casein micelles (Rollema, 1992; Horne, 1998; De Kruif, 1999).

Although all mammalian milk contain casein micelles, the exact structure of the casein micelle is not clear, as a result, a number of models have been proposed for its structure and are all based on bovine caseins studies (Swaisgood, 1993; Farrell et al., 2006). It is worth mentioning that bovine caseins that have been used as standards for casein micelles studies share characteristics with caseins of non-dairy

mammalian species (Phadungath, 2005; Lemay et al., 2009). In the present study, we investigate the structure of casein micelles in African elephant milk. Furthermore, the study aims to identify and characterize β-casein phosphoforms that are possibly present in African elephant milk. This work was done to possibly give an insight into the structural aspects of non-dairy casein micelles and to further elucidate the role of β-casein in casein micelle formation. Although African elephant caseins and casein micelles were central to the current study, a comparative study of casein micelles structure of cow milk and those of horse, human, sheep and African elephant milk was also done.

1.2. Evolution

1.2.1. Evolution of mammals and lactation

Lactation can be described as the profuse secretion of milk via the mammary gland to feed the neonate (Lefèvre et al., 2010). The term *mammalia* was first coined by the prominent taxonomist Carolus Linnaeus in 1758 (Oftedal, 2012). This group of animals included both terrestrial and aquatic animals that possess mammary glands as a defining morphological feature. Apart from mammalian species, no other organism produces abundant glandular secretions to nourish its offspring, which is comparable to milk in complexity, magnitude and duration of secretion. The evolution of mammals can be traced back 310 million years ago (MYA) from the synapsids era although they only appeared approximately 166 to 210 MYA towards the end of the Triassic period (Lemay et al., 2009; Lefèvre et al., 2010). The class *Mammalia* is divided into two subclasses namely, *Theria* and *Prototheria* (monotremes) (Fox and McSweeney, 1998).

This earliest split in mammalian phylogeny occurred approximately 166 MYA. *Prototheria* are ancient mammals that lay eggs, extant species include the echidna and platypus. Unlike the *Prototheria*, the *Theria* give birth to live young. Another split then occurred within the *Theria* subgroup which established the infraclasses, *Metatheria* (marsupials) and *Eutheria* (placentals) lineages about 140 MYA (Lefèvre et al., 2010). *Metatheria* include kangaroos and opossums, whereas the *Eutheria* include mammals such as humans, rats, bovines etc (Lemay et al., 2009).The latter group compose approximately 95 % of all mammals and their young are much more mature at birth compared to the *Prototheria* (Fox and McSweeney, 1998). Figure 1.1 illustrates mammalian lineages with approximate divergence times.

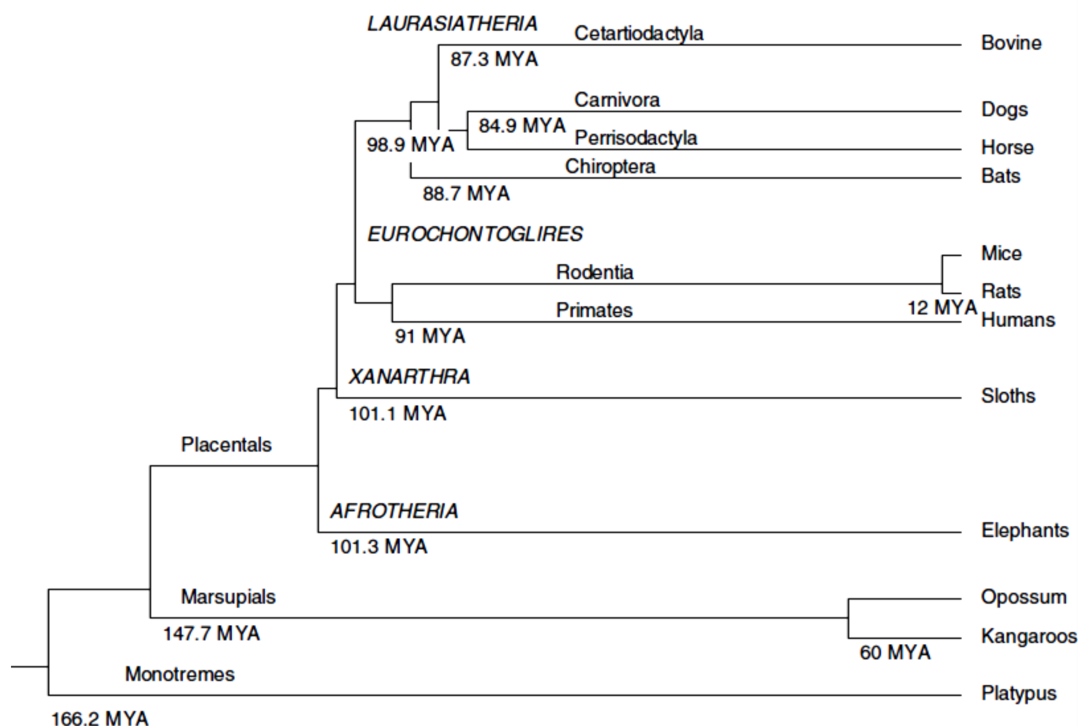

Figure 1.1. Splitting topology and divergence of representative monotremes, marsupials and placental (eutherian) mammals. The time of origin of each major branch is represented in million years ago (MYA). Source: (Lemay et al., 2009).

Lactation seems to have been established before the divergence of extant mammalian lineages (Oftedal, 2002). Evolutionary studies suggest that ancestral mammary glands secretions had antimicrobial properties. However, in an effort to provide maternal care for the neonate that is more effective, efficient and adaptable, lactation developed even further over the course of evolution. Furthermore, the evolution of a placenta in placental species led to a much more developed embryo and an ability to lactate (Oftedal, 2012). Following the development of exceedingly nutritious milks, evolution created diversity in milk with regards to composition, length of lactation, quantity of milk produced, amount of time between nursing and the extent to which lactation contribute to the nutrition of the offspring (Capuco and Akers, 2009; Lefèvre et al., 2010).

The primary milk constituents evolved prior to the appearance of mammals. Furthermore, some of the milk constituents may have origins as ancient as the split of synapsids from sauropsids (Oftedal, 2012). The nutrient composition of milk differs widely across mammalian species depending on several factors including stage of lactation and the specific demands of the offspring. For example, milk of seals may contain up to 60 % fat during early lactation whereas in wallabie's milk the fat content at the same stage of lactation is remote (Green et al., 1980; Lang et al., 2005; Brennan et al., 2007). In essence, the particulars of lactation have evolved in such a way that the various reproductive and environmental demands of different species are met (Capuco and Akers, 2009). Apart from nutrient supply, milk also provides immunological agents and promotes endocrine maturation in the offspring (Goldman, 2012). Thus, milk makes provision for short-term and lasting requirements of the neonate, which can be very species-specific.

The milk of egg laying monotremes is very different from that of most other mammals (Capuco and Akers, 2009). The hatchlings are largely altricial and completely milk dependent for nutrition. The mammary glands are arranged into two areas of the abdomen, since the glands do not terminate into teats, the secreted milk is licked off by the young from the glandular surface, the areolae. Monotremes exhibit a much longer period of lactation where the young develops extensively compared to a much shorter gestation period.

Like monotremes, marsupials rely on milk as the sole source of nutrition. Although marsupials give birth to live offsprings (oviparous), the neonate is still very altricial. The lactation period of marsupials is much longer and the milk composition changes extensively to meet the developing nutritional requirements of the neonate (Oftedal, 2012). Unlike monotremes and marsupials, eutherians have a longer gestation period. Milk composition of eutherian mammals is mostly complex during the entire lactation period. The evolution of lactation and the dynamics of the different milks produced are depicted in Figure 1.2.

The development and function of the mammary gland is under systematic and local control. Lactogenic hormones such as insulin, cortisol and prolactin are responsible for the induction of milk protein gene expression. The differentiation of secretory cells and the start of milk synthesis and secretion are regulated to synchronize with parturition (Capuco and Akers, 2009). The secretion of copious milk is largely determined by the decline in progesterone levels although this is not the case in marsupials. Milk also contains milk-borne factors that are involved in mammary

gland function. These factors, such as a protein known as feedback inhibitor of lactation, play an important role in regulating mammary epithelial function and survival, mostly during involution. Alpha lactalbumin has also been shown to regulate mammary function by causing apoptosis of the gland after milk stasis (Svensson et al., 2000).

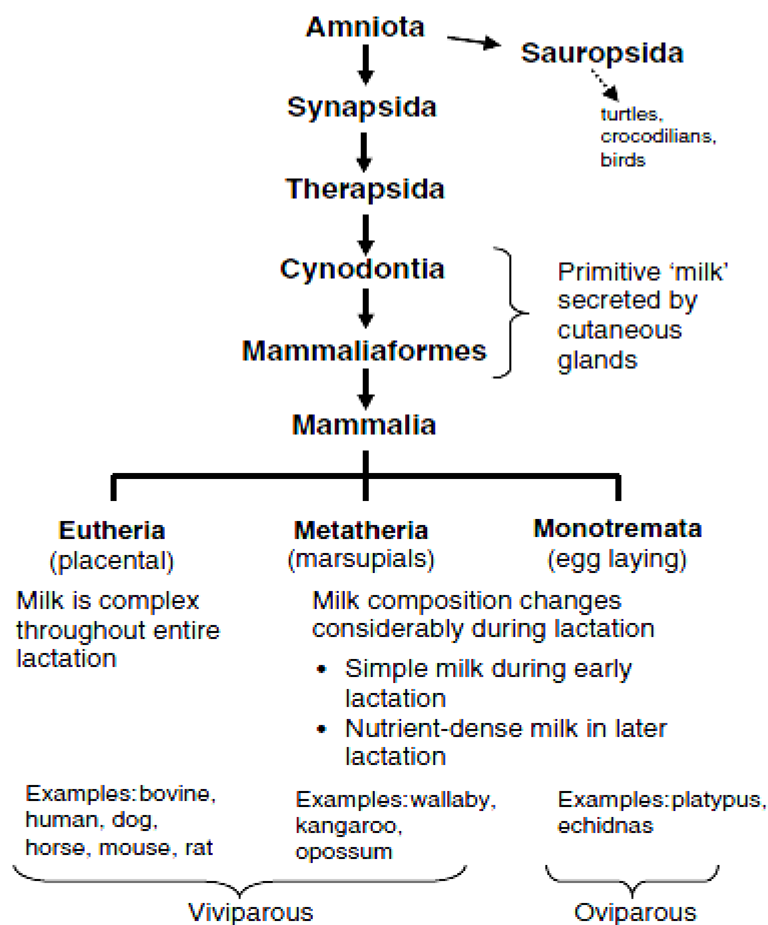

Figure 1.2. Evolution of mammalian lactation and the dynamics of milk secreted. Source: (Capuco and Akers, 2009).

1.2.2. Evolution of casein genes

Caseins are synthesized in the mammary gland during lactation and are among the cardinal proteins that evolved in the lineage leading to mammals (Kawasaki et al., 2011). In milk, caseins exist as part of large complexes with colloidal calcium phosphate known as casein micelles. Caseins are divided into two groups, the calcium sensitive caseins (α_{s1}-, α_{s2}- and β-casein) and calcium insensitive caseins (κ-casein)(Ginger and Grigor, 1999). Casein sequences show high rates of substitutions which further complicates the elucidation of casein evolution. Caseins have evolved from a gene family of secreted calcium-binding phosphoproteins (SCPP), specifically the odontogenic ameloblast-associated gene (ODAM) that arose by gene duplication (Kawasaki, 2009). Although casein sequences are diverse, the organization and orientation of casein genes is highly conserved (Rijnkels, 2002).

It has been shown that the calcium sensitive casein genes evolved from a putative common ancestor referred to as *CSN1/2* (Kawasaki et al., 2011). Six and four exons that comprise the *CSN1/2* are found in both *SCPPPQ1* and *ODAM* genes respectively. With regards to calcium insensitive caseins, five of the exons in the follicular dendritic cell secreted peptide (*FDCSP*) gene are also found in the calcium sensitive gene. Furthermore, the phylogenetic distribution of the *FDCSP* and *SCPPPQ1* suggest that they both evolved from the *ODAM* gene (Kawasaki et al., 2011). Considering the above, it is likely that calcium sensitive casein genes directly originated from *SCPPPQ1* gene, whereas calcium insensitive casein genes originated from *FDCSP* gene via two different evolution pathways. The expression of

SCPPPQ1, *FDCSP* and *ODAM* has been detected in dental tissues, therefore suggesting that caseins evolved as calcium binding proteins.

1.3. Milk biosynthesis and secretion

Milk constituents are either directly synthesized and secreted from the mammary epithelial cells into the alveolar lumen or alternatively transported across the epithelial barrier from other sources (Larson, 1979; Anderson et al., 2007; Lönnerdal, 2007; Shennan, 2008). Figure 1.3 depicts the five major milk biosynthesis and secretion pathways in the secretory epithelial cells into the alveoli lumen where milk components collect.

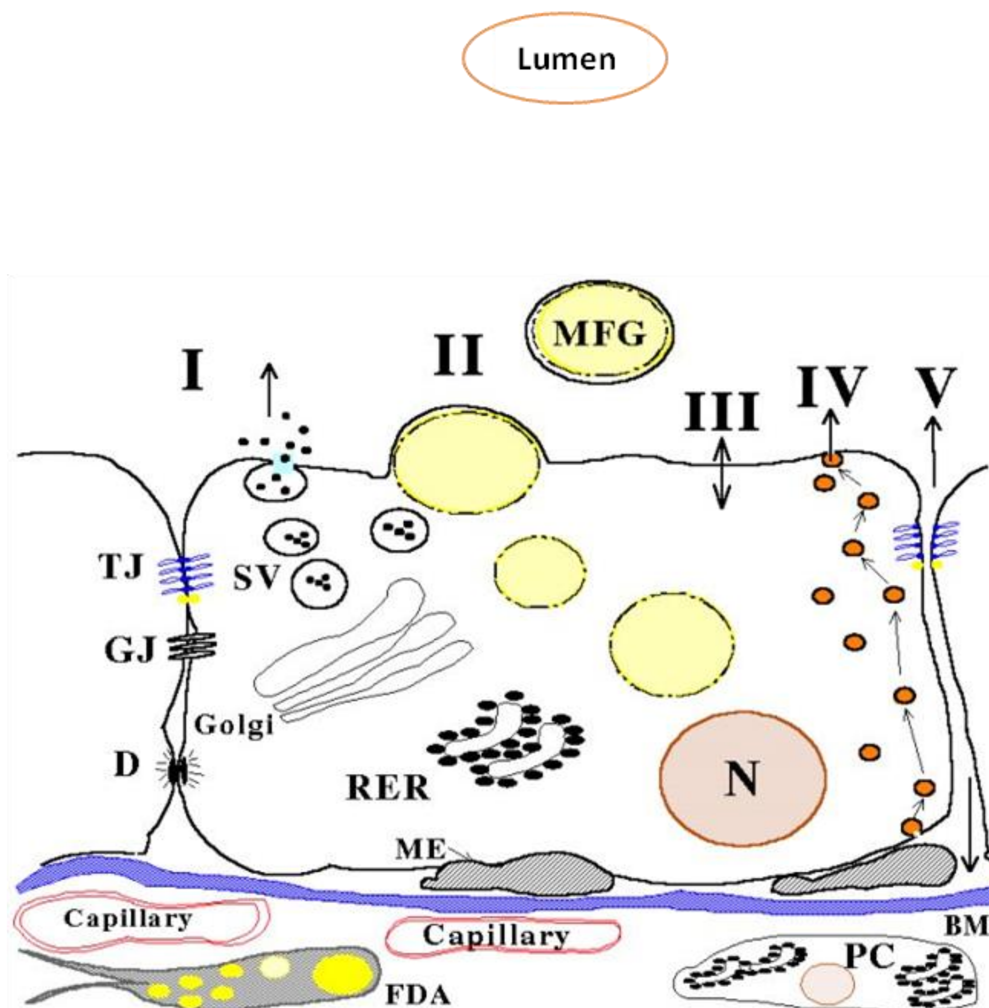

Figure 1.3. Cellular mechanisms for biosynthesis and secretion of milk components in the lactating mammary tissue (alveolar cell) as described in the text (Neville et al., 2001). The different biosynthesis and secretion stages are represented by roman numerals I-V.

Amino acids, which are essentially the building blocks for proteins, are extracted from the blood by the mammary gland with several sodium-dependent or sodium independent systems facilitating the actual transport and targeting a specific group of amino acids (Burgoyne and Duncan, 1998). Inside the secretory epithelium the basic protein synthesis pathway occurs which is the same as in other tissues. The biosynthesis of milk proteins is initiated by hormones that induce specific gene expression.

There is evidence that milk expression is also under epigenetic regulation, it was recently shown that DNA methylation at specific sites on the α_{s1}-casein promoter was able to down regulate the expression of α_{s1}-casein during mammary gland involution (Vanselow et al., 2006). Since milk proteins are secretory proteins, they have to be exported into the milk pool in the alveolar lumen. Milk protein transport and secretion is by exocytosis (I) (Figure 1.3). Also transported via this pathway is lactose, some minerals and water (Shennan and Peaker, 2000). All the major milk proteins are synthesized with *N*-terminal signal peptides which target the respective mRNAs to the ER for translocation of the nascent peptides across the ER membrane (Neville et al., 2001).

The proteolytic removal of the signal peptide and PTMs of proteins then occurs in the ER lumen. Folding of the protein into an appropriate 3D structure, to become a

functional protein, as well as association with for example carbohydrates, ions or phosphates occur during the transportation to and inside the Golgi apparatus (Burgoyne and Duncan, 1998). Here proteins, together with lactose, are encapsulated in a secretory vesicle that buds off from the Golgi apparatus. Consequently the secretory vesicles reach and fuse with the apical membrane, and release their contents of proteins, lactose, ions, and water into the milk pool of the alveolar lumen.

Lipids are incorporated into milk by budding off as lipid droplets from the cell apex and are consequently secreted into milk with a membrane, derived from intracellular sources and the cell surface pathway, as the milk fat globule membrane (II) (Heid and Keenan, 2005). Membrane bound transporters enable the transport of the rest of the minerals, some small molecules and water across the basal/lateral and apical membranes (III). Milk constituents that are not derived from milk-secreting cells, including immunoglobulin, serum albumin and peptide hormones, are conveyed across the mammary epithelium by transcytosis (IV) (Mather and Keenan, 1998). The paracellular route enables the equilibration of constituents between cells during times when the epithelial tight junctions are permeable (V).

1.4. Milk proteins

In bovine milk, caseins represent about 80 % of the total milk proteins and compose of four gene products known as α_{s1}-, α_{s2}-, β- and κ-caseins (Anahory et al., 2002). They are synthesized in the mammary gland and have very little secondary structure due to their relatively high proline content (Farrell et al., 2004). Whey proteins

account for up to 20 % of total milk proteins, of which α-lactalbumin (α-LA) and β-lactoglobulin (β-LG) are the main whey proteins (Fox and Mcsweeney, 1998). Because of their relative abundance, whey and casein proteins have been widely studied through mass spectrometry (MS) and electrophoresis. Peptone, that is low molecular weight peptides and proteins in the milk fat globule membrane (MFGM), compose the rest of the milk proteins in mammals. Low molecular weight peptides in the whey fraction are also known as miscellaneous minor proteins, they include transferrin, lactoferrin, lactollin, ceruplasmin, glycoprotein-A, kinogen, M-1 glycoprotein epidermal growth factor, glycolactin, angiogenin among others (Farrell et al., 2004).

1.4.1. Caseins

Caseins in bovine milk can be described as a portion of milk proteins that can be iso-electrically precipitated from raw milk by acidification to pH 4.6 (Eigel et al., 1984). These proteins have evolved from a group of proteins known as secreted calcium (phosphate)-binding phosphoproteins (SCPP) (Holt et al., 2013). Caseins are secreted into the mammary gland in response to lactogenic hormones and as a result, they are exclusively limited to lactation and milk (Lefèvre et al., 2010; Fox and Brodkorb, 2008). Moreover, caseins are subdivided into mainly two groups, calcium sensitive (α_{s1}-, α_{s2}- and β-caseins) and non-calcium sensitive caseins (κ-casein). The former group precipitates in the presence of higher calcium concentration whereas the latter group remains soluble (Ginger and Grigor, 1999). Caseins exert properties and sequences that are different from each other. This is a consequence of several factors such as varying levels of post-translational modifications (PTMs), mutational

changes in casein genes, proteolysis by indigenous milk proteases or oxidation of cysteine to disulphide bonds (Swaisgood, 1993). Due to their large size, caseins cannot be crystallized and therefore renders their secondary structure determination by X-ray crystallography virtually impossible (Swaisgood, 1993). However, methods such as small-angle neutron scattering (SANS), small-angle X-ray scattering (SASX) and molecular modeling have been used successfully to predict the secondary structures of caseins (Swaisgood, 1993; Qi, 2007). Caseins are not globular in structure, thus they lack strategically placed cysteine residues that stabilize the structure of globular proteins (Swaisgood, 1993). Additionally, casein sequences show that they are amphipathic in nature, a feature that is responsible for their unique functional properties. Due to their varied application in the dairy industry, caseins have become a popular target for study and therefore are amongst the most studied food proteins (Ginger and Grigor, 1999).

1.4.1.1. α_{s1}-casein

Amongst calcium sensitive caseins, α_{s1}-casein characteristically shows greater solubility in the presence of calcium. Bovine α_{s1}-casein is the more abundant protein fraction in bovine milk constituting approximately 40 % of total casein (Ginger and Grigor, 1999; Farrell et al., 2004). The primary structure of bovine α_{s1}-casein is composed of 199 amino acid residues and does not have any cystein residues in its sequence. Bovine α_{s1}-casein is highly phosphorylated and exist in two forms, the major form with 8 phosphate groups/molecule bound and the minor form with 9 phosphate groups/mol bound (Ginger and Grigor, 1999). All α_{s1}-casein contain a region known as multiple phosphorylation site, phosphorylation occurs at threonine

(Thr) or serine (Ser) residues (Ginger and Grigor, 1999). This family of caseins has the highest net negative charge in neutral pH buffer with only monovalent cations present (Farrell et al., 2004). The multiple sequence alignment of α_{s1}-caseins is shown in Figure 1.4. These sequences were determined directly by amino acid sequencing and confirmed by cDNA sequencing, or inferred from DNA sequencing of other eutherian species α_{s1}-casein genes.

The reference form of bovine α_{s1}-casein contains three hydrophobic regions, residues 1-44, 90-113 and 132-199. The amino acids in these regions are highly conserved between species (Martin et al., 2013). Residues 41-80 consists of eight glutamates, seven seryl-phosphates and three aspartates thus making it very polar. Bovine α_{s1}-casein is the major constituent of bovine caseins and contains a very acidic region between residues 38 and 78 that is responsible for calcium binding (Farrell et al., 2004). Circular dichroism (CD) or Raman (FTIR) spectral analysis indicates the presence of approximately 14 % α-helix, 40 % β-sheet and 24 % turn-like structures (Michael Byler et al., 1988). In addition, plasmin, which hydrolyzes bonds adjacent to lysine or arginine, cleaves this protein most rapidly at several sites, the major cleavage site being between residues 23 and 24. These regions are accessible to enzyme attack and must be sufficiently exposed to solvent to allow enzyme-substrate complexes. α_{s1}-casein also possess a highly conserved 15 amino acid signal peptide sequence, albeit the rest of the sequences varies among mammalian species (Ginger and Grigor, 1999). Three of the seven variants of bovine α_{s1}-casein have been studied in more detail. The rare A variant has a 13 amino acid residue deletion whereas the B Variant and C variant differ in amino acid substitutions.

```
Cow           MKLLILTCLVAVALA RPKHPIKHQGLPQEVL--------NENLLRFFVAPFPE-------
Sheep         MKLLILTCLVAVALA RPKHPIKHQGLSPEVL--------NENLLRFVVAPFPE-------
Goat          MKLLILTCLVAVALA RPKHPINHQGLSPEVP--------NENLLRFVVAPFPE-------
Bison         MKLLILTCLVAVALA RPKHPIKHQGLPQEVL--------NENLLRFFVAPFPE-------
W.Bufallo     MKLLILTCLVAVALA RPKQPIKHQGLPQGVL--------NENLLRFFVAPFPE-------
Human         MRLLILTCLVAVALA RPKLPLRYPERLQNPSESS-EPIPLESR-------------EEYM
Horse         MKLLILTCLVAVALA RPKLPHRQPEIIQNEQDSR-EKVLKERKFPSF--------ALEYI
Donkey        MKLLILTCLVAVALA RPKLPHRHPEIIQNEQDSR-EKVLKERKFPSF--------ALEYI
Pig           MKLLIFICLAAVALA RPKPPLRHQEHLQNEPDSR-EELFKERKFLRF--------PEVPL
Rabbit        MKLLILTCLVATALA RHKFHLGHLKLTQEQPESSEQEILKERKLLRFVQTVPLELREEYV
Camel         MKLLILTCLVAVALA RPKYPLRYPEVFQNEPDSI-EEVLNKRKILDL--------AVVSP
Llama         MKLLILTCLVAVALA RPKYPLRYPEVFQNEPDSI-QEVLNKRKILEL--------AVVSP
              *:***: **.*.*** * *                      :

Cow            --------VFGKEKVNELSKDIGSESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQ
Sheep          --------VFRKENINELSKDIGSESIEDQAMEDAKQMKAGSSSSSEEIVPNSAEQKYIQ
Goat           --------VFRKENINELSKDIGSESTEDQAMEDAKQMKAGSSSSSEEIVPNSAE-KYIQ
Bison          --------VFGKEKVNELSKDIGSESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQ
W.Bufallo      --------MFGK--------DVGSESTEDQAMEDIKQMEAESISSSEEIVPISVEQKHIQ
Human          NGMNRQRNILRE-KQTDEIKDTRNESTQNCVVAEPEKMESSISSSSEEM-----------
Horse          NE------LNRQ-RE--LLKEKQKDEHKEYLIEDPEQQESSSTSSSEEVVPINTEQKRIP
Donkey         NE------LNRQ-RE--LLKEKQKDEHKEYLIEDPEQQESSSTSSSEEVVPINTEQKRIP
Pig            LS------QFRQ-EIINELN-------RNHGMEGHEQ-RGSSSSSSEEVVGNSAEQKHVQ
Rabbit         NELNRQRELLRE-KENEEIKGTRNEVTEEHVLADRET-EASISSSSEEIVPSSTKQKYVP
Camel          -I------QFRQ-ENIDELKDTRNEPTEDHIMEDTER-KESGSSSSEEVVSSTTEQKDIL
Llama          -I------QFRQ-ENIDELKDTRNEPTEDHIMEDTER-TVSGSSSSEEVVSSTTEQKDIL
                          :               .:  :   :       *****:

Cow            -KEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNS------------AEERLHSMKEGIHA
Sheep          -KEDVPSERYLGYLEQLLRLKKYNVPQLEIVPKS------------AEEQLHSMKEGNPA
Goat           -KEDVPSERYLGYLEQLLRLKKYNVPQLETVPNS------------AEEQLHSMKEGNPA
Bison          -KEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNS------------AEERLHSMKEGIHA
W.Bufallo      -KEDVPSERYLGYLEQLLRLKKYNVPQLEIVPNL------------AEEQLHSMKEGIHA
Human          -----SLSKC---AEQFCRLNEYNQLQLQAAHAQEQIRRMNENSH---------------
Horse          -REDMLYQHT---LEQLRRLSKYNQLQLQAIHAQEQLIRMK-----------------EN
Donkey         -REDMLYQHT---LEALRRLSKYNQLQLQAIYAQEQLLRMK-----------------EN
Pig            KEEDVPSQSY---LGHLQGLNKYKLRQLEAIHDQEL---------------HRTNEDKHT
Rabbit         -REDLAYQPY---VQQ-------------------QL--------------LRMKERYQI
Camel          -KEDMPSQRY---LEELHRLNKYKLLQLEAIRDQKLIPRVKLSSHPYLEQLYRINEDNHP
Llama          -KEDMPSQRI---LEELHRLNKYKLLQLEAIRDQKLIPRVKLSSHPYLEQLYRINEDNHP
                      .

Cow            QQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSE
Sheep          HQKQPMIAVN--------QLFRQFYQLDAYPSGAWYYLPLGTQYTDAPSFSDIPNPIGSE
Goat           HQKQPMIAVNQELAYFYPQLFRQFYQLDAYPSGAWYYLPLGTQYTDAPSFSDIPNPIGSE
Bison          QQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSE
W.Bufallo      QQKEPMIGVNQELAYFYPQLFRQFYQLDAYPSGAWYYVPLGTQYPDAPSFSDIPNPIGSE
Human          ----------------VQVPFQQLNQLAAYPYAVWYYP-QIMQYVPFPPFSDISNPTAHE
Horse          SQRKPMRVVNQEQAYFYLEPFQPSYQLDVYPYAAWFHPAQIMQHVAYSPFHDTAKLIASE
Donkey         SQRKPMRVVNQEQAYFYLEPFQPSYQLDVYPYAAWFHPAQIMQHVAYSPFHDTAKLIASE
Pig            QQGEPMKGVNQEQAYFYFEPLHQFYQLDAYPYATWYYPPQ---YIAHPLFTNIPQPTAPE
Rabbit         QEREPMRVVNQELAQLYLQPFEQPYQLDAYLPAPWYYTPEVMQYVLSPLFYDLVTPSAFE
Camel          QLGEPVKVVT--------QPFPQFFQLGASPYVAWYYPPQVMQYIAHPSSYDTPEGIASE
Llama          QLGEPVKVVTQEQAYFHLEPFQQFFQLGASPYVAWYYPPQVMQYIAHPSSHDTPEGIASE
                                   :    ** .     *::      :       :     . *

Cow            NSEKTT-MPLW---
Sheep          NSGKIT-MPLW---
Goat           NSGKTT-MPLW---
Bison          NSGKTT-MPLW---
W.Bufallo      NSGKTT-MPLW---
```

```
Human           NYEKNNVMLQW---
Horse           NSEKTDIIPEW---
Donkey          NSEKTDIIPEW---
Pig             KGGKTEIMPQW---
Rabbit          SAEKTDVIPEWLKN
Camel           DGGKTDVMPQW---
Llama           DGGKTDVMPQWW--
                .  *    :  *
```

Figure 1.4. A multiple sequence alignment of 12 α_{s1}-casein protein sequences. The 15 amino acid long signal peptide sequence is indicated in bold and italics. An asterisk (*) indicates positions which have a single, fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties. A period (.) indicates conservation between groups of weakly similar properties.

1.4.1.2. α_{s2}-casein

Bovine α_{s2}-caseins, the most highly and variably phosphorylated of the calcium sensitive caseins, consists of 207 amino acids (Eigel et al., 1984; Martin et al., 2013). In addition, this group of caseins is also the least hydrophobic of all bovine caseins (Farrell et al., 2004). It occurs in milk in several forms with phosphorylation ranging from 10-13 phosphate groups (Eigel et al., 1984). Human milk appears to be devoid of α_{s2}-casein, thus sequence comparison is limited to a few eutherian species. Figure 1.5 shows a multiple sequence alignment of α_{s2}-caseins. The genes encoding α_{s2}- and β-caseins are more closely related to each other than genes encoding for α_{s1}-caseins, as shown by amino acid multiple sequence comparison (Ginger and Grigor, 1999). The majority of α_{s2}-casein peptides have an internal disulphide bond between cysteine residues 36 and 40 forming a small loop in the structure. Additionally, a small proportion of this protein exists as disulfide bonded dimers in bovine. Moreover, α_{s2}-casein dimers can either be antiparallel or parallel. Parallel dimers involve disulphide bonds forming between cysteine 36 and 40 in one protein and cysteine 36 and 40 in the other (Farrell et al., 2004). The opposite applies for

antiparallel dimers and dimer formation does not influence the interaction of α_{s2}-casein with other caseins.

Hydrolysis by plasmin of α_{s2}-caseins occurs at several sites; primarily in the afore-noted C-terminal regions so that, at neutral pH, these positively charged residues are primarily at the surface and could actively participate in the binding of inorganic phosphate (Farrell et al., 2004). There are several other α_{s2}-like caseins that have been identified from other species such as rat (γ -casein), mouse(γ- and ε-casein), guinea pig (casein A) and rabbit (α_{s2a}- and α_{s2b}-caseins) (Ginger and Grigor, 1999). Like α_{s1}-casein, α_{s2}-casein possesses a classical 15 amino acid residues long signal peptide that is highly conserved. Four genetic variants of bovine α_{s2}-casein are recognized, these are termed variants A-D (Eigel et al., 1984). CD and FTIR spectral analysis indicates that there is an increased level of α-helix with 30-40 % in addition to approximately 20 % turn-like structures and 20 % β-sheet in α_{s2}-casein (Michael Byler et al., 1988).

```
Cow          MKFFIFTCLLAVALA KNTMEHVSSSEESI-ISQETYKQEKNMAINPSKENLCSTFCKEVV
Sheep        MKFFIFTCLLAVALA KHKMEHVSSSEEPINISQEIYKQEKNMAIHPRKEKLCTTSCEEVV
Goat         MKFFIFTCLLAVALA KHKMEHVSSSEEPINIFQEIYKQEKNMAIHPRKEKLCTTSCEEVV
W.Bufallo    MKFFIFTCLLAVALA KHTMEHVSSSEESI-ISQETYKQEKNMAIHPSKENLCSTFCKEVI
Horse        MKFFIFTCLLAVALA KHNMEHRSSSEDSVNISQEKFKQEKYVVIPTSKESICSTSCEEAT
Donkey       MKFFIFTCLLAVALA KHNMEHRSSSEDSVNISQEKFKQEKYVVIPTSKESICSTSCEEAT
Pig          MKFFIFTCLLAVAFA KHEMEHVSSSEESINISQEKYKQEKNVINHPSKEDICATSCEEAV
Guineapig    MKLFIFTCLLAVALA KHKSEQQSSSEESVSISQEKFKD-KNMDTISSEETICASLCKEAT
Camel        MKFFIFTCLLAVVLA KHEMDQGSSSEESINVSQQKFKQVKKVAIHPSKEDICSTFCEEAV
Llama        MKFFIFTCLLAVALA KHEMDQGSSSEESINVSQQKLKQVKKVAIHPSKEDICSTFCEEAV
             **:*********.:* *:  :: ****: : : *:  *: * :     :* :*:: *:*.

Cow           RNANEEE--------YSIGSSSEESAEVATEEVKITVDDKHYQKALNEINQFYQK--FPQ
Sheep         RNADEEE--------YSIRSSSEESAEVAPEEVKITVDDKHYQKALNEINQFYQK--FPQ
Goat          RNANEEE--------YSIRSSSEESAEVAPEEIKITVDDKHYQKALNEINQFYQK--FPQ
W.Bufallo     RNANEEE--------YSIGSSSEESAEVATEEVKITVDDKHYQKALNEINQFYQK--FPQ
Horse         RNINEMESAKFPTEVYSSSSSSEESAKFPTEREEKEVEEKHHLKQLNKINQFYEKLNFLQ
Donkey        RNINEMESAKFPTEVYSSSSSSEESAKFPTEREEKEVEEKHHLKQLNKINQFYEKLNFLQ
Pig           RNIKEVG--------YASSSSSEESVDIPAENVKVTVEDKHYLKQLEKISQFYQK--FPQ
Guineapig     KNTPKMA--------FFSRSSSEEFADIHR---------------ENKKDQLYQKWMVPQ
Camel         RNIKEVE--------S---------AEVPT---------------ENKISQFYQKWKFLQ
Llama         RNIKEVE--------S---------VEVPT---------------ENKISQFYQKWKFLQ
              :*  :                    ...                :: .*:*:*  . *

Cow           YLQYLYQGPIVLNPWDQVKRNAVPIT-PTLNR---------EQLSTSEENSKKTVDMEST
Sheep         YLQYLYQGPIVLNPWDQVKRNAGPFT-PTVNR---------EQLSTSEENSKKTIDMEST
Goat          YLQYPYQGPIVLNPWDQVKRNAGPFT-PTVNR---------EQLSTSEENSKKTIDMEST
W.Bufallo     YLQYLYQGPIVLNPWDQVKRNAVPIT-PTLNR---------EQLSTSEENSKKTVDMEST
Horse         YLQALRQPRIVLTPWDQTKTGDSPFI-PIVNT---------EQLFTSEEIPKKTVDMEST
Donkey        YLQALRQPRIVLTPWDQTKTGASPFI-PIVNT---------EQLFTSEEIPKKTVDMEST
Pig           YLQALYQAQIVMNPWDQTKTSAYPFI-PTVIQSGEELSTSEEPVSSSQEENTKTVDMESM
Guineapig     YNPDFYQRPVVMSPWNQIYTRPYPIVLPTLGKEQISTIEDILKKTTAVESSSSSSSTEKST
Camel         YLQALHQGQIVMNPWDQGKTRAYPFI-PTVNTEQLSISEEST-EVPTEE---------ST
Llama         YLQALHQGQIVMNPWDQGKTMVYPFI-PTVNTEQLSISEEST-EVPTEENSKKTVDTEST
              *     *  :*:.**:*      *:  * :                 : *         *

Cow           EVFTKKTKLTEEEKNRLNFLKKISQRYQKFALPQYLKTVYQHQKAMKPWIQPKTKV---I
Sheep         EVFTKKTKLTEEEKNRLNFLKKISQYYQKFAWPQYLKTVDQHQKAMKPWTQPKTNA---I
Goat          EVFTKKTKLTEEEKNRLNFLKKISQYYQKFAWPQYLKTVDQHQKAMKPWTQPKTNA---I
W.Bufallo     EVITKKTKLTEEDKNRLNFLKKISQHYQKFTWPQYLKTVYQYQKAMKPWTQPKTKV---I
Pig           EEFTKKTELTEEEKNRIKFLNKIKQYYQKFTWPQYIKTVHQKQKAMKPWNHIKTNSYQII
Guineapig     DVFIKKTKMDEVQKLIQSLLNIIHEYSQKAFWSQTLEDVDQYLKFVMPWNHYNTNADQVD
Horse         EVVTEKTELTEEEKNYLKLL-----YYEKFTLPQYFKIVRQHQTTMDPRSHRKTNSYQII
Donkey        EVVTEKTELTEEEKNYLKLLNKINQYYEKFTLPQYFKIVHQHQTTMDPQSHSKTNSYQII
Camel         EVFTKKTELTEEEKDHQKFLNKIYQYYQTFLWPEYLKTVYQYQKTMTPWNHIKRYF----
Llama         EVFTKKTELTEEEKDHQKFLNKIYQYYQTFLWPEYLKTVYQYQKTMTPWNHIKRYF----
              : . :**:: * :*   .:*        :.    : :: * *   . : *  : :

Cow           PYVRYL-
Sheep         PYVRYL-
Goat          PYVRYL-
W.Bufallo     PYVRYL-
Pig           PNLRYF-
Guineapig     ASQERQA
Horse         PVLRYF-
Donkey        PVLRYF-
Camel         -------
Llama         -------
```

Figure 1.5. A multiple sequence alignment of 10 α_{s2}-casein sequences. The 15 amino acid long signal peptide sequence is highlighted in bold and italics. An asterisk (*) indicates positions which have a single, fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties. A period (.) indicates conservation between groups of weakly similar properties.

1.4.3.3. β-casein

Bovine β-casein is a major component of casein proteins and it is the most hydrophobic casein. It does not contain cysteine residues but is rich in proline residues. The bovine β-casein sequence consists of 209 amino acid residues (Greenberg et al., 1984). Figure 1.6 shows a multiple sequence alignment of several β-casein sequences that are available in the data bank. Like α_{s1}- and α_{s2}-casein, β-casein possesses a 15 amino acid residue long signal peptide that is also highly conserved (Farrell et al., 2004).

In solution, β-casein forms detergent like micelle aggregates and this is due to its amphipathic nature. Plasmin, a native milk enzyme can hydrolyse β-casein forming three β-casein fragments which are known as γ_1- , γ_2- , and γ_3-casein (Eigel et al., 1984). Plasmin targets the Lys-X and Arg-X bonds and acts primarily on the N-terminal moiety of β-casein (Jolivet et al., 2000). Bovine β-casein exists in one fully phosphorylated form containing 5 phosphates. However, milk of other species have several phosphoforms of β-casein with different numbers of phosphate groups attached to serine or threonine residues. Equine β-casein has seven phosphorylation sites (Girardet et al., 2006), ovine β-casein has six (Mamone et al., 2003) and human β-casein has up to five phosphorylated sites (Poth et al., 2008).

The phosphorylation of β-casein typically occurs along a single major phosphorylation site that is located near the N-terminal (Ginger and Grigor, 1999). The self-association of β-casein is micelle-like, and both ionic strength and temperature increase the quantity of polymer present as well as the degree of association. This in effect reduces its cleavage by chymosin at high temperature (Swaisgood, 1993). Nine genetic variants of β-casein exist in bovine milk, albeit their distinction by gel electrophoresis is complicated (Martin et al., 2013). CD and FTIR spectral analysis have estimated that β-casein has approximately low levels of α-helix (15 %) and intermediate levels of turn-like structure (29 %) and β-sheet (30 %) (Michael Byler et al., 1988).

```
Cow              MKVLILACLVALALA RELEELNVPGEIVESLS------SSEESITRINK-KIEKFQSEEQ
Sheep            MKVLILACLVALALA REQEELNVVGETVESLS------SSEESITHINK-KIEKFQSEEQ
Goat             MKVLILACLVALAIA REQEELNVVGETVESLS------SSEESITHINK-KIEKFQSEEQ
W.Buffalo        MKVLILACLVALALA RELEELNVPGEIVESLS------SSEESITHINK-KIEKFQSEEQ
Human            MKVLILACLVALALA RE---------TIESLS------SSEESITEY-KQKVEKVKHEDQ
Horse            MKILILACLVALALA REKEELNVSSETVESLSSNEPDSSSEE--------KLQKFKHEGQ
Donkey           MKILILACLVALALA REKEELNVSSETVESLSSNEPDSSSEESITHINKEKVQKFKHEGQ
Elephant         MKVFILACLVAFALG REKEEIIV---------------STEESVTQVNKQKPEGVKHEEQ
Pig              MKLLILACFVALALA RAKEELNASGETVESLS------SSEESITHISKEKIEKLKREEQ
Mouse            MKVFILACLVALALA RET---------D--------SISSEESVEHINE-KLQKVNLMGQ
Rat              MKVFILACLVALALA REKDAFTVSSETG--------SISSEESVEHINE-KLQKVKLMGQ
Camel            MKVLILACLVALALA REKEEFKTAGEALESIS------SSEESITHINKQKIEKFKIEEQ
Llama            MKVLILACLVALALA REKEEFKTAGEAVESIS------SSEESITHINKQKIEKFKIEEQ
                 **::****:**:*:. *                      *:**        * : .:   *

Cow              QQTEDELQDKIHPFAQTQSLVYPFP--GPIHNS-LPQNIPPLTQTPVV--VPPFLQPEVM
Sheep            QQTEDELQDKIHPFAQAQSLVYPFT--GPIPNS-LPQNILPLTQTPVV--VPPFLQPEIM
Goat             QQTEDELQDKIHPFAQAQSLVYPFT--GPIPNS-LPQNILPLTQTPVV--VPPFLQPEIM
W.Buffalo        QQMEDELQDKIHPFAQTQSLVYPFP--GPIPKS-LPQNIPPLTQTPVV--VPPFLQPEIM
Human            QQGEDEHQDKIYPSFQPQPLIYPFV--EPIPYGFLPQNILPLAQPAVV---LPVPQPEIM
Horse            QQREVERQDKISRFVQPQPVVYPYA--EPVPYAVVPQSILPLAQPPI----LPFLQPEIM
Donkey           QQREVEHQDKISRFVQPQPVVYPYA--EPVPYAVVPQNILPLAQPPI----VPFLQPEIM
Elephant         -QREDEHQNKIQPLFQPQPLVYPFA--EPIPYTVFPPNAIPLAQPIVV---LPFPQPEVK
Pig              QQTENERQNKIHQFPQPQPLAHPYT--EPIPYPILPQNILPLAQVPVV---VPLLHPEVM
Mouse            LQAEDVLQAKVHSSIQSQPQAFPYAQAQTISCNPVPQNIQPIAQPPVVPSLGPVISPELE
Rat              VQSEDVLQNKFHSGIQSEPQAIPYA--QTISCSPIPQNIQPIAQPPVVPTVGPIISPELE
Camel            QQTEDEQQDKIYTFPQPQSLVYSHT--EPIPYPILPQNFLPPLQPAVM---VPFLQPKVM
Llama            QQTEDEQQDKIYTFPQPQSLVYSHT--EPIPYPILPQNFLPPLQPAVM---VPFLQPKVM
                  * *   * *.    * :     .     :    .* .  *  *  :     *.  *::

Cow              GVSKVKEAMAPKHKEMPFPKYPV-EPFTESQSLTL-TDVENLHLPLPLLQSWMHQPHQPL
Sheep            GVPKVKETMVPKHKEMPFPKYPV-EPFTESQSLTL-TDVEKLHLPLPLVQSWMHQPPQPL
Goat             GVPKVKETMVPKHKEMPFPKYPV-EPFTESQSLTL-TDVEKLHLPLPLVQSWMHQPPQPL
W.Buffalo        GVSKVKEAMAPKHKEMPFPKYPV-EPFTESQSLTL-TDVENLHLPLPLLQSWMHQPPQPL
Human            EVPKAKDTVYTKGRVMPVLKSPT-IPFFDPQIPKL-TDLENLHLPLPLLQPLMQQVPQPI
Horse            EVSQAKETILPKRKVMPFLKSPI-VPFSERQILNP-TNGENLRLPVHLIQPFMHQVPQSL
Donkey           EVSQAKETLLPKRKVMPFLKSPI-VPFSERQILNP-TNGENLRLPVHLIQPFMHQVPQSL
Elephant         QLPEAKEITFPRQKLMSFLKSPV-MPFFDPQIPNLGTDLENLHLPLPLLQPLRHQLHQPL
Pig              KDSKAKETIVPKRKGMPFPKSPA-EPFVEGQSLTL-TDFEVLS--LPLLQSLMHQIPQPV
Mouse            SFLKAKATILPKHKQMPLLNSETVLRLINSQIPSL-ASLANLHLPQSLVQL-LAQVVQAF
Rat              SFLKAKATVLPKHKQMPFLNSETVLRLFNSQIPSL--DLANLHLPQSPAQL-QAQIVQAF
Camel            DVPKTKETIIPKRKEMPLLQSPV-VPFTESQSLTL-TDLENLHLPLPLLQSLMYQIPQPV
Llama            DVPKTKEIVIPKRKEMPLLQSPL-VPFTESQSLTL-TDLENLHLPLPLLQSLMHQIPQPV
                    :.*     : : * . :      : : *  .   .   *       *    *  * .

Cow              PPTV-MFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV-
Sheep            PPTV-MFPPQSVLSLSQPKVLPVPQKAV--PQRDMPIQAFLLYQEPVLGPVRGPFPILV-
Goat             SPTV-MFPPQSVLSLSQPKVLPVPQKAV--PQRDMPIQAFLLYQEPVLGPVRGPFPILV-
W.Buffalo        PPTV-MFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV-
Human            PQTL-ALPPQPLWSVPQPKVLPIPQQVVPYPQRAVPVQALLLNQELLLNPTHQIYPVTQP
Horse            LQTL-MLPSQPVLSPPQSKVAPFPQPVVPYPQRDTPVQAFLLYQDPRLGPTGELDPATQP
Donkey           LQTL-MLPSQPVLSPPQSKVAPFPQPVVPYPQRDTPVQAFLLYQDPQLGLTGEFDPATQP
Elephant         AQTP-VLP----LPLSLPKVLPVPQQVIPYPQRGRPIQNLQLYEEPLLDPTRKIYPVAQP
Pig              PQTP-MFAPQPLLSLPQAKVLPVPQQVVPFPQRDMPFQALLLYQDPLLGPLQGFYPVPQP
Mouse            PQTH-LVSSQTQLSLPQSKVLYFLQQVAPFLPQDMSVQDLLQYLELL-NPTVQFPATPQH
Rat              PQTPAVVSSQPQLSLPQSKSQYLVQQLAPLFQQGMPVQDLLQYLDLLLNPTLQFLATQQL
Camel            PQTP-MIPPQSLLSLSQFKVLPVPQQMVPYPQRAMPVQAVLPFQEPVPDPVRGLHPVPQP
Llama            PQTP-MIPPQSLLSLSQFKVLPVPQQMVPYPQRAMPVQALLPFQEPIPDPVRGLHPVPQP
                   *   .           *   . *       :   .* .    :   .
```

```
Cow             ----------
Sheep           ----------
Goat            ----------
W.Buffalo       ----------
Human           LAPVHNPISV
Horse           IVAVHNPVIV
Donkey          IVPVHNPVIV
Elephant        LAPVYNPVAV
Pig             VAPVYNPV—
Mouse           SVS-V-----
Rat             HSTSV-----
Camel           LVPVIA----
Llama           LVPVIA----
```

Figure 1.6. A multiple sequence alignment of 12 β-casein protein sequences. The 15 amino acid long signal peptide sequence is indicated by bold and italics. An asterisk (*) indicates positions which have a single, fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties. A period (.) indicates conservation between groups of weakly similar properties.

1.4.3.4. κ-casein

Bovine κ-caseins, the most studied milk protein, consists of 169 amino acid residues (Eigel et al., 1984). The primary structure of κ-casein displays its amphipathic nature and thereby it's dual role, which is to interact via hydrophobic interactions with the other caseins and consequently provide a hydrophilic and negatively charged surface on the micelle to stabilize the colloidal suspension in milk (Horne, 1998). Additionally, as opposed to other caseins, κ-casein does not bind calcium extensively and thus it is not sensitive to calcium precipitation (Swaisgood, 1993). Of all the proteins of the casein family, κ-caseins are the only proteins that have been conclusively shown to be glycosylated (Swaisgood, 1993). The post translational glycosylation by short oligosaccharide chains occurs at one or more of the threonine sites (Ginger and Grigor, 1999). Figure 1.7 shows the multiple sequence alignment of κ-casein from several mammalian species.

κ-casein is a target for hydrolysis by the aspartate protease, chymosin or rennin (Miyoshi et al., 1976). The destabilization of a casein micelle occurs when chymosin cleaves the hydrophilic and flexible C-terminal part, specifically between residues Phe 105 and Met 106 of κ-caseins in ruminants or Phe-Leu and Phe-Ile in other animals, thus separating the two distinct domains of the κ-casein molecule known as the para κ-casein and the macropeptide (Eigel et al., 1984). The two peptides are distinct from each other; the *N*-terminal domain carries a net positive charge, is very hydrophobic and interacts strongly with the other casein molecules (Farrell et al., 2004). The C-terminal domain carries a net negative charge and contains a prevalence of polar residues, the two domains are attached by a peptide that carries a net positive charge and is conserved in most species. Interestingly, horse milk has very little κ-casein content and the curdling property of its milk by rennet has been shown to be very limited (Iametti et al., 2001). Unlike the calcium sensitive caseins, κ-casein has a 21 amino acid residues long signal peptide (Ginger and Grigor, 1999). In bovine milk, the two major and commonly known genetic variants of κ-casein are termed A and B, these variants differ from each other in amino acid substitutions on the sequence (Farrell et al., 2004). The secondary structure of κ-casein has been investigated by CD and FITR Spectral analysis. The analysis indicates that κ-casein has relatively low content of α-helix, approximately 15 %. Additionally, the levels of β-sheets and turn-like structures were approximated to be 30 % and 25 % respectively (Michael Byler et al., 1988).

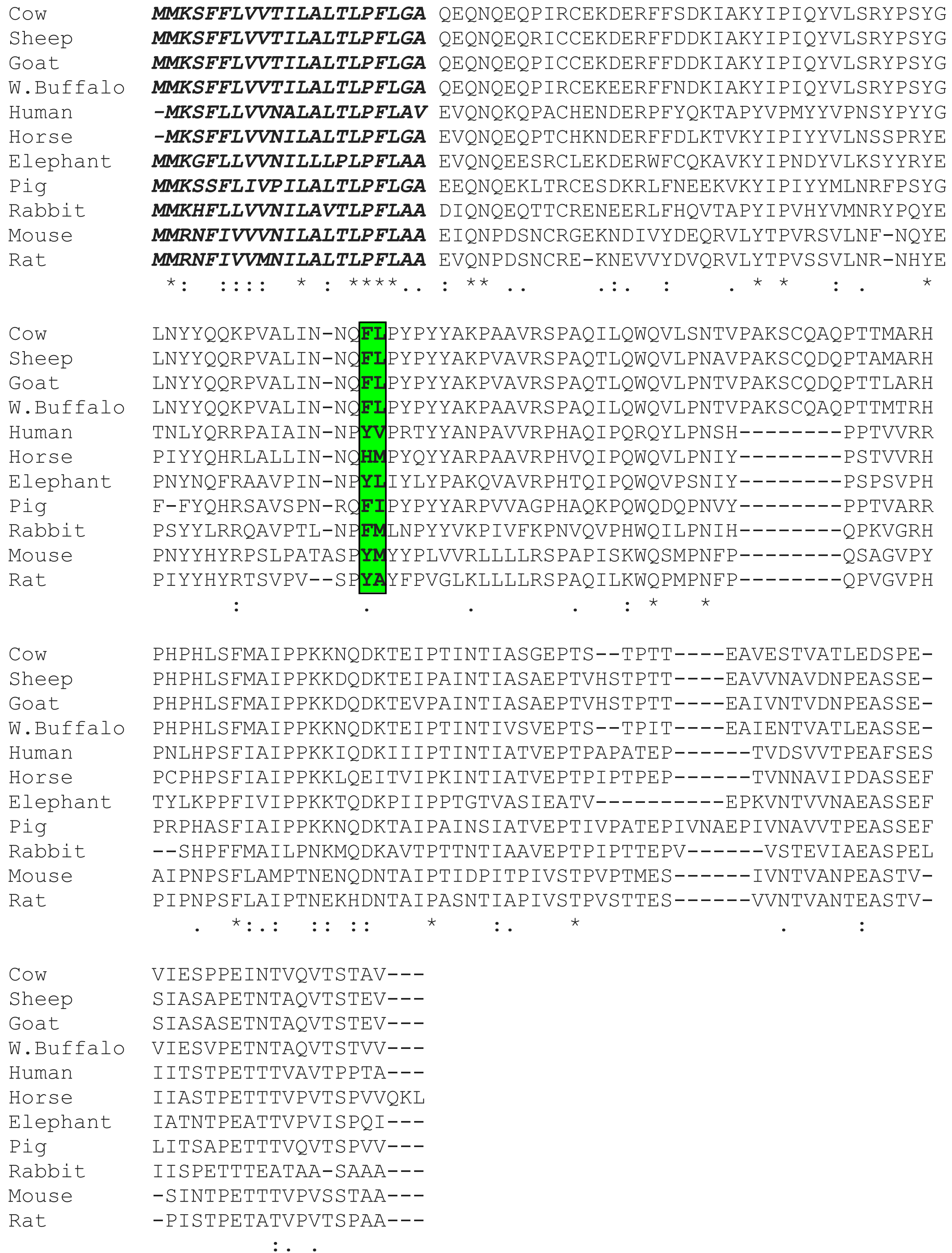

Figure 1.7. A multiple sequence alignment of 11 κ-casein sequences. The 21 amino acid long signal peptide sequence is indicated by bold and italics. The position of the chymosin cleavage site is represented by a region colored in green and a rectangular block. An asterisk (*) indicates positions which have a single, fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties. A period (.) indicates conservation between groups of weakly similar properties.

1.5. Casein micelle

Caseins in milk exist as large colloidal supramolecular aggregates which form thermodynamically stable complexes with nanoclusters of amorphous calcium and phosphates known as the casein micelles (De Kruif, 1999). Bovine casein micelles are spherical in shape and range in diameter between 150-200 nm (Dalgleish and Corredig, 2012). Casein micelles convert milk into a free flowing low viscosity liquid and additionally provide the means to transport high levels of precipitation prone calcium and phosphate in the mammary gland. The functions of caseins as part of the casein micelle in bio-mineralisation and protein supply are the most understood of its functions (Holt, 2015). The sequestration of calcium phosphate by the casein micelle is also important in preventing pathological calcification in the mammary gland ducts.

Although the casein micelle has been a subject of intensive research over the past few decades, the details of its structure at molecular level remain debatable and elusive (Qi, 2007). The difficulty in unraveling the structure of casein micelles is heightened by their relatively large size which precludes a direct and explicit structure determination (Holt and Sawyer, 1988; Phadungath, 2005). Chemical and physical studies on casein micelles has focused on features such as their size, properties and composition, as a result, a number of conflicting models have been proposed to depict bovine casein micelle structure (McMahon and McManus, 1998). These models fall into three general categories: internal structure models, coat-core models and subunit models. For each category, the original models were first

proposed in the 1960's and were either modified or abandoned as subsequent researchers revealed additional data about casein micelles.

Waugh and Nobel in 1965 proposed the first coat-core model which was based on the solubility of casein in Ca^{2+} solutions (Phadungath, 2005). The coat-core models suggest that the exterior and interior of casein micelles are composed of different proteins (McMahon and McManus, 1998). The core of the micelle is formed by α_{s1}- and β-caseins whereas the surface is covered by κ-casein as shown in Figure 1.8. The size of the micelle is limited by κ-caseins which also prevents precipitation of the caseinate (Phadungath, 2005). The latest model in this category was proposed by Paquin and coworkers in 1987 who described a casein core composed of α_{s1}-caseins and colloidal calcium phosphate whereas β- casein is bound by hydrophobic interactions, using experimental data obtained from two proteins from EDTA-dissociated casein micelles. The micelle core is surrounded by complex particles of α_{s1}- , α_{s2-} and higher amounts of κ-casein.

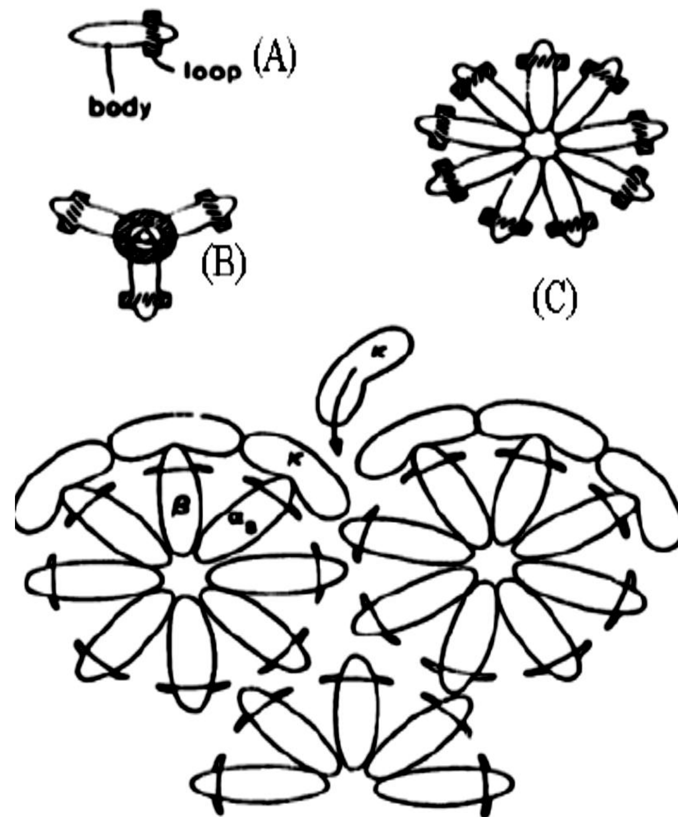

Figure 1.8. The coat-core model of casein micelle structure. A) represents the monomer of $α_{s1}$- and β-casein with charged loop, B) represents a tetramer and C) core polymer with $α_{s1}$- and β-casein. Source: (Phadungath, 2005).

The second category of models features several models which describe a casein micelle as composed of subunits hence the term subunit models, as shown in Figure 1.9. The first of these models was described by Morr in 1967 and was based on data obtained from the effects of urea and oxalate treatment on disrupted casein micelles (Phadungath, 2005). The subunit models in general describe a model that has a rough surface and is spherical in shape. In addition, the micelle is formed by smaller units (submicelles) ranging in diameter between 12-15nm, the subunits are composed of a mixture of caseins ($α_s$-, β-casein). These small units are bound together by calcium phosphate clusters, in this way, they aggregate to form bigger micelles with κ-casein located on the exterior of the micelles (Walstra, 1999).

Moreover, the negatively charged C-terminals of hairy layer κ-casein prevent the micelles from further aggregation by steric and electrostatic repulsion. Other research contributors to this group of models include Slattery and Evard in 1973, Schmidt and Payens in 1976 and Walstra in 1984 (Phadungath, 2005). Evidence for and against the submicelle models exist. Electron microscopy provides the most compelling direct evidence of the spherical shape of casein micelles whereas X-ray scattering and diffraction studies show an internal structure of micelles that support the presence of submicelles (Walstra, 1999). On the contrary, some electron microscopic evidence has shown presence of micellar calcium phosphate (MCP) rather than calcium phosphate cluster linkages. Moreover, proteolytic digestion of casein in skim milk yields a precipitate that consists primarily of MCP and peptides instead of calcium phosphate clusters.

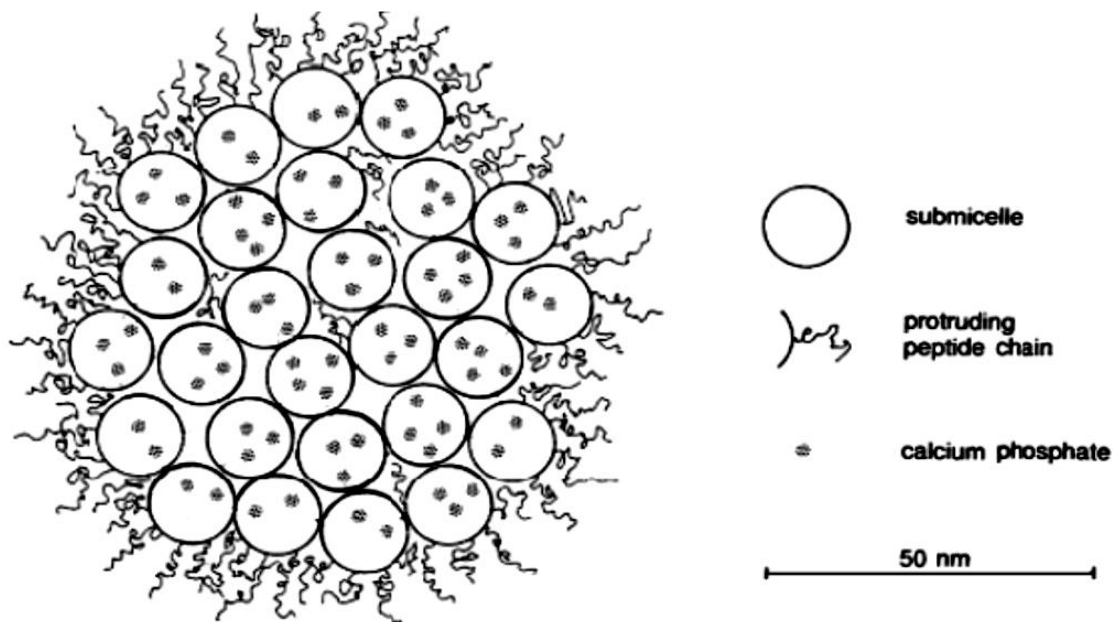

Figure 1.9. The submicelle model of casein micelle structure (Walstra, 1999)

The final category of models, the internal structure models, is based on experimental data obtained from isolated caseins constituents that affect or influence the formation of the internal structure of casein micelles (Phadungath, 2005). Rose in 1969 proposed the first internal structure model. Rose proposed that β- casein monomers self-associate into chain-like polymers and attach α_{s1}-casein molecules. Subsequently, κ-casein interacts with α_{s1}-casiens, forming small aggregates. Colloidal calcium phosphate acts as a cross-linker stabilizing agent of the aggregate network. More recently, a dual bonding model, which is a modification of the Rose model has been proposed. The model is based on the ability of individual caseins to self-associate in solution due to their amphipathic nature (Horne, 1998).

Self-association of caseins is driven by hydrophobic interactions whereas electrostatic repulsive interactions on the other hand are important in limiting polymerization and therefore micelle growth. The conformation of α_{s1}- and β-caseins, when they are adsorbed at hydrophobic interfaces, form a train-loop-train and a tail-train structure, respectively (Figure 1.10) (Phadungath, 2005). The self-association of caseins makes it possible for polymerization to occur. Colloidal calcium phosphates are considered to be one of the linkages between casein micelles and neutralizing agents of the negative charge of the phosphoserine residues. By binding to those residues; electrostatic repulsion is reduced, and the hydrophobic interaction between caseins is still dominant, resulting in more associations of proteins. Unlike other caseins, κ-caseins can only interact hydrophobically and acts as a propagation terminator, because they do not have a phosphoserine cluster to bind calcium and

also not another hydrophobic point to prolong the chain. The dual bonding model for the casein micelle structure is depicted in Figure 1.10.

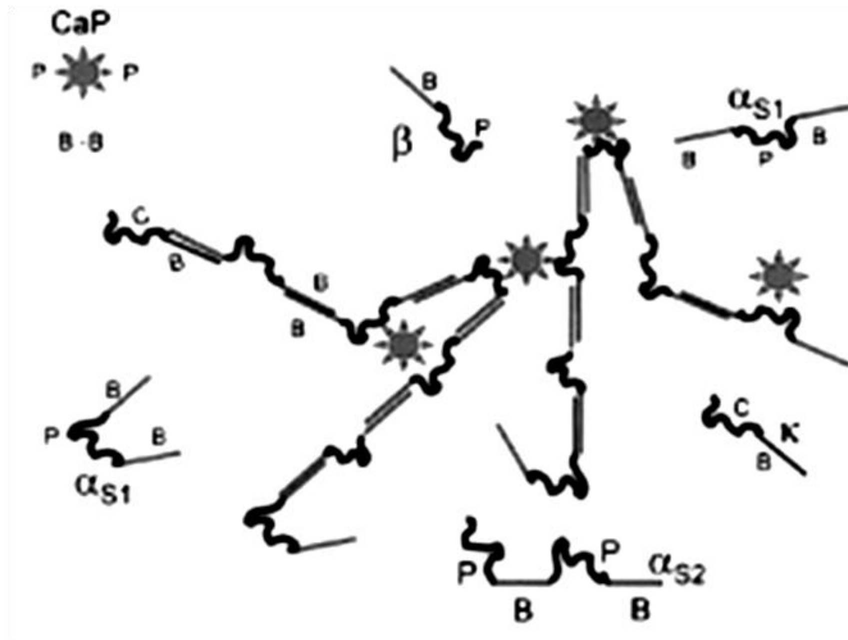

Figure 1.10. The dual bonding model of casein micelle structure. Calcium nanoclusters are represented CaP. Hydrophobic and hydrophilic chains of α- and β-caseins are indicated by B and P and respectively. The negatively charged C-terminal of κ-casein is indicated by C. (Horne, 1998)

The bovine casein micelle structure is the only casein micelle structure that has been studied in detail (Holt et al., 2013). It was apparent from this extensive research on bovine casein micelles that they are so easily perturbed by most of the usual methods used in protein structure determination and therefore very little could be deduced with absolute certainty. Moreover, the use of 2% glutaraldehyde in microscopic preparation can induce protein substructure, thereby introducing possible misinterpretation of microscopic images (De Kruif et al., 2012). However, electron and atomic force microscopy have also provided acceptable structural

evidence of casein micelles structure which supports both the nanocluster and submicelle models.

The variability among species in the number of casein genes that are expressed and their relative proportions and sequence divergence, suggest that there is much more to learn from non-bovine casein micelles (Holt et al., 2013). Furthermore, β-casein has been found to be a principal casein in human as well as African elephant milk (Madende et al., 2015), suggesting that the presence of α_{s1}- and α_{s2}-caseins is not a prerequisite for casein micelle formation. The size and appearance of casein micelles in milk varies from one species to the other.

Human casein micelles are amongst the smallest micelles (micelle average diameter 64 nm), whereas horse micelles appear to be larger with diameters of approximately 255 nm (Potočnik et al., 2011). A comparison of the casein composition of milk and micelle size thereof is shown in Table 1.1. The comparison of casein micelle sizes in literature is complicated because different methods (microscopy, SANS, centrifugation) and milk in different states (powdered milk vs fresh milk) have been used to study the micelles (Dalgleish et al., 2004). Because of the reasons above, there is room for errors in size determination and therefore the casein micelle sizes recorded in literature may not be entirely accurate. Investigating these micelles using a single standard method with milk in a fresh state may circumvent the above complication.

Table .1.1. Casein composition and micelle sizes of sheep, cow, African elephant, human and horse milk

Species	α_{S1}-casein %	α_{S2}-casein %	β-casein %	κ-casein %	Micelle size nm
Sheep	50[b]	+	40[b]	10[b]	210[b]
Cow	38[c]	10[c]	40[c]	12[a]	182[b]
African elephant	-	-	89[b]	11[a]	N/A
Human	3[c]	-	70[c]	27[c]	64[b]
Horse	40-60[b]	Trace	40-50[b]	4-7[b]	255[b]

The size comparison of casein micelles from literature is drawn with the sizes observed using fresh and frozen milk ([a]Madende et al., 2015; [b]Potočnik et al., 2011; [c]Qi, 2007)

1.6. African elephant milk

As alluded to earlier, milk composition differs from one species to the other. African elephant milk has a very unique composition compared to milk of other mammalian species (Osthoff et al., 2007). The first comprehensive study of African elephant milk was conducted by McCullagh and Widdowson (1970) where milk samples from 30 African elephant cows were collected post mortem and analyzed. The lactation stages of the sampled African elephant cows spanned between 2 and 36 months. The analysis showed that, on average, milk of African elephant milk constituted of 5.1 % protein, 9.3 % fat and 3.6 % lactose. The concentration of protein and fat increased, whilst lactose concentration decreased with advancing lactation. In comparison to other mammalian species milk, the mineral content of African elephant milk was similar to cow's milk, with a slight difference in potassium levels, which was higher in African elephant milk. Interestingly, unlike other milk, African

elephant milk fat contained high proportions of capric acid, which also increased as lactation progressed. The first study on African elephant milk drawn from a living African elephant was conducted by Osthoff et al., (2005). This study provided details of protein and sugar content of the African elephant milk. The level of lactose decreased from 52.5 to 11.8 g kg^{-1} milk, whilst the oligosaccharide (galactosyllactose) content increased from 11.8 to 15.2 g kg^{-1} milk during lactation (Osthoff, 2012).

Subsequent studies, that involved more accurate proteomic methodologies, such as 2D PAGE and tandem mass spectrometry (MS/MS) have shown that African elephant milk is devoid of α-caseins and contain very high levels of β-casein compared to κ-casein (Madende et al., 2015). This makes African elephant milk composition unique compared to milk of other mammalian species especially the casein composition and therefore warrants further investigation.

1.7. Analysis of mammary gland products

Omics is a term used to describe disciplines such as genomics, proteomics and lipidomics, which involve a comprehensive analysis of biomolecule components (Casado et al., 2009). Proteomics, a term first coined in the early 1990s, refer to the study of the total complement of proteins that are expressed by a genome, including a variety of post translational modifications that occur. Additionally, proteomics studies focuses mainly on protein identification, structure, interaction and their role in physiological functions. The proteome of an organism differs from cell to cell

depending on a distinct set of genes that is expressed at that particular time. Unlike the more dynamic proteome, an organism's genome is more or less static (Manso et al., 2012). Proteomics provides an opportunity for the analysis of hundreds of proteins simultaneously in complex mixtures via different techniques which can include high resolution two dimensional polyacrylamide gel electrophoresis (2D PAGE) coupled with versatile mass spectrometry (MS). The application of proteomic methodologies enables the acquisition of previously unattainable data.

1.7.1. Proteomics of milk

Proteomics studies have been applied to samples with varied origin including nutritionally relevant protein foods such as milk (Manso et al., 2012). This in turn has allowed the characterization of food components as well as their nutritional, biological and functional relevance including the study of protein conformation and protein interactions. The study of milk proteins has been undertaken for over 50 years (O'Donnell et al., 2004). However, in the last decade, great interest has developed towards the study of milk by proteomics (Roncada et al., 2012). A variety of research techniques have been employed in order to gain a multifaceted picture addressing the multiplicity and complexity nature of milk. A great deal of this complexity arises from abundant post translational modifications (PTMs) such as phosphorylation and glycosylation, as well as proteolysis, protein interaction and the presence of abundant genetic variants (O'Donnell et al., 2004). As a result of these PTMs, some proteins in milk naturally exist as isoforms and therefore are characterized by a great deal of heterogeneity. Figure 1.11 depicts a typical proteomic process in the characterization of milk proteins.

In dairy production, the important areas within proteomics with great potential for application are PTM and differential expression analysis. The application of proteomics in milk has been limited to a few mammalian species. Bovine and human milk are among the few species whose milk proteins have been studied thoroughly via high resolution 2D PAGE coupled with MS (Roncada et al., 2012). There are several strategies applicable to study proteomics of milk, using raw milk as the starting sample. Generally these techniques are applicable to all types of milk.

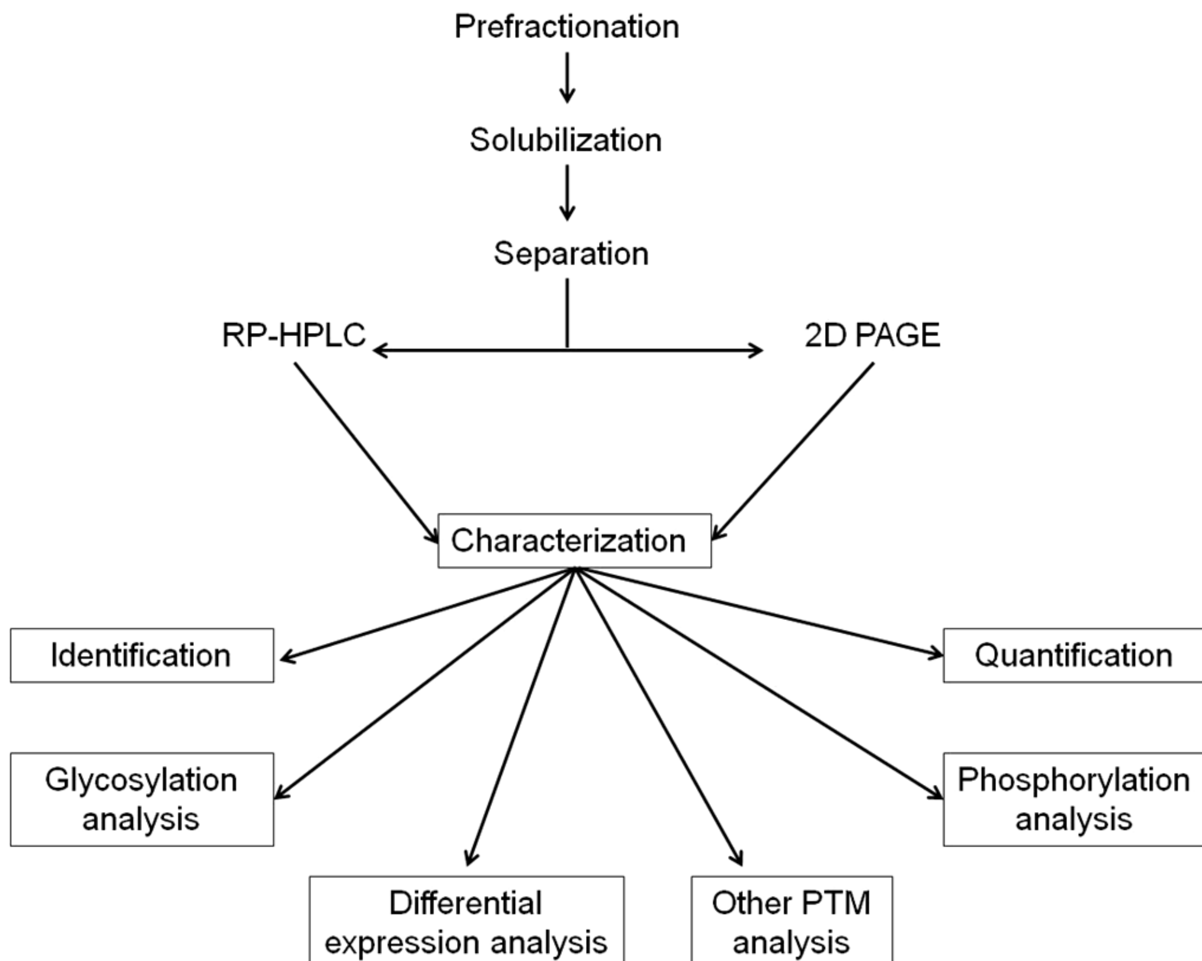

Figure 1.11. A schematic presentation of a typical proteomic process. The pathway illustrates the precipitation of proteins in the sample, the separation methods and several characterization techniques that are applicable. Source: (O'Donnell et al., 2004).

1.8. Protein structure prediction

Protein structure prediction, primarily based on amino acid sequence and homologous structures, has progressed significantly in recent years, owing to the explosion of sequence and structural information as well as advances in computational tools (Al-lazikani et al., 2001). Protein structure modeling aims to predict a structure of a particular protein from its amino acid sequence, the accuracy of the model is comparable to the best results obtained experimentally (Krieger et al., 2003). Structure models are useful in determination of protein function, rational protein design, structure-based drug design and many other applications. In the cases where proteins are too large for NMR analysis, which is limited to proteins with molecular weights in the range of less than 40-60 kDa, or in instances where a protein that requires structural analysis cannot be crystallized for X-ray crystallography analysis, an alternative option to obtain structural information will be protein modeling (Deschamps, 2010).

Homology modeling is largely based on two major observations, the first being that the amino acid sequence of a protein uniquely determines its structure. The second observation being that the protein structure is more stable during evolution and changes much slower than the associated sequence. This means that similar sequences adopt practically identical structures and distantly related sequences might also fold into a similar structure, only if they are in the safe mode of protein structure determination software as determined by Rost (1999). Much of the success in homology modeling is attributed to the explosive increase in sequences stored in the Protein Data Bank (PDB) as well as increased developments in recombinant

DNA technology, together with advances in bioinformatics and data analysis tools (Kelly et al., 2005). In homology modeling, the comparative model usually mimics the conformation of the parent structure and often adapts its features (Elbegdorj et al., 2013). Structure prediction can become challenging in cases where there is no direct template for the sequence query. This is the case for casein proteins which cannot be crystallized and hence accurate X-ray crystallography templates are unavailable.

Generally, structure prediction is a step-wise process that involves mainly six stages: identification of the template; alignment of the target sequence to the template structure; building of the initial model based on the template; loop and side chain modeling; model refinement and finally model evaluation (Petrey and Honig, 2005). Homology modeling can be a powerful and useful approach in many applications. Unfortunately, no formula currently exists that exploits the reliability of a structure model. To a certain extent, tests such as the control modeling tests can be done in an effort to evaluate the reliability of a model.

As X-ray crystallography and Nuclear magnetic resonance spectroscopy remain the best methods for accurate protein structure determination (Kundrotas et al., 2008), homology modeling is becoming a method of choice of template based structural prediction of proteins. This is due to the vast growth in databases of 3D structural templates, which are useful in homology modeling with some databases harbouring over 10 000 entries (Kundrotas et al., 2008). In light of the above, homology modeling can be used to determine the structure of milk caseins and exploit the role they play in micelle structure formation.

Homology modeling has been previously applied to determine the secondary and ultimately the 3D structure of bovine caseins. The earliest attempt at bovine casein structure modeling was by (Holt and Sawyer, 1988). The secondary structure of bovine caseins (α_{s1}-, α_{s2}- and β-casein), was predicted by employing PREDICT (a secondary structure prediction software suite). Four types of secondary structure were predicted, these include: α-helix, β-strand, β-turn and random structures using the complete primary sequence of caseins as the starting material. The bovine α_{s1}-casein model consists largely of both α-helices and random structures with a very small segment of predicted β-turn. The predicted α_{s2}-casein model was mostly composed of random structures and α-helices; only a single smaller segment was predicted to adopt the β-turn conformation. Like the bovine α_{s1}-casein structure model, the bovine β-casein model composed of longer lengths of predicted α-helices and random structure, although some smaller segments were predicted to adopt the β-strand and β-turn conformations. Unlike the α_{s1}-, α_{s2} and β-casein models, the predicted bovine κ-casein model was composed of several segments of all the four conformations (α-helix, β-strand, β-turn and random structures) although the random structure conformation was more predominant.

Following on the work by Holt and Sawyer (1988), molecular modeling studies on bovine caseins and the casein micelle have since been done (Kumosinski et al., 1991; 1993; 1994; Farrell et al., 2001). In these studies, the casein 3D structure models were constructed using sequence-based prediction algorithms in conjunction with secondary structural information obtained from Raman spectroscopy. These

models were energy minimized and molecular modeling techniques were used to discern segments of hydrophobicity and hydrophilicity as well as the functional, chemical and biochemical properties of amongst the caseins. Most of the studies on casein homology models have focused on bovine caseins and not caseins of other mammals. It is therefore paramount to revisit structure prediction of bovine caseins using the latest and more accurate structure prediction tools. Moreover, a comparative study of bovine casein models with casein models of other mammalian species may shed more light into the structure of caseins.

1.9. Comparative genomics

Comparative genomics is an emerging field in biology that involves a comparison of genome sequences of different species (Touchman, 2010). With the continuing improvements in high-throughput genomic sequencing, ever-expanding sequence databases and new advances in post-sequencing software programs, comparative genomics presents an opportunity to solve genetic puzzles such as deducing the mechanisms and history of genome evolution (Chain et al., 2003).

Comparative genomics can involve the whole genomes or comparison of discrete segments of genomes by aligning homologous DNA sequences from different species (Touchman, 2010). By comparing genes across several species using computer-based tools, genomic features that have been preserved in multiple species can be deduced. The data obtained from comparing two genomes is directly linked to the phylogenetic distance between the two species. Phylogenetic distance

measures the extent of separation between two organisms or their genomes on an evolutionary scale. The further two organisms are from each other on a Phylogenetic tree, the less the sequence similarity detected between them.

Several methodologies and resources can be used in comparative genomics to solve particular problems. One such resource is the Ensembl comparative genomics reference set that facilitates comprehensive and reproducible analysis of up to date chordate genome data (Herrero et al., 2016). Ensembl computes pairwise and multiple whole-genome alignments from which large-scale synteny, per-base conservation scores and constrained elements are obtained. Gene alignments are used to define Ensembl protein families, gene trees and homologies for both protein-coding and non-coding RNA genes.

A few comparative studies on caseins have been done in the past, albeit mostly at protein level (Ginger and Grigor, 1999; Holt, 2015). The studies have shown that the primary casein sequences are very divergent. This is mainly due to exon skipping and posttranslational processing events. It will be of interest to study gene sequences of caseins across a wide variety of mammalian species to determine if this diversity can be observed as well as the degree of diversity thereof.

1.10. Discussion and conclusions

The unique nature, high nutritional value and suitability as a raw material for production of other products have made milk a major item of human diet. Milk

contains a wide array of proteins that provide a number of biological activity, some as complete peptides or more common as short peptides encrypted in the protein sequences. To competently serve their functions, milk proteins, specifically caseins, form complexes with large amounts of amorphous calcium and phosphate known as casein micelles. For several decades, the elucidation of the exact structure of a casein micelle has been an area of intense research and debate thus several experimentally based models have been proposed for its structure.

A great deal of scientific knowledge of caseins and casein micelles is derived primarily from farm animals, particularly the cow genus *Bos*. As a consequence, casein micelle research has a biocentric bias. It is known and acceptable that no single species can provide a complete and sufficient model for another and that comparative studies have a potential for provision of a wealth of knowledge, the research community should not be species-centric. A deep understanding of milk proteomics of other mammalian species, especially those whose proteome show unique characteristics, could therefore shed some light into the structure and mechanisms of less understood milk particles such as the casein micelle. Such studies also have a potential for discovery of novel applications of milk proteins in food systems.

1.11. Aims of the study

In the previous proteomic study of African elephant milk (Madende et al., 2015), several observations were made. Firstly, no α-caseins could be detected. Secondly, β-casein content was much higher compared to very little κ-casein content and finally, β-casein of African elephant milk appeared to exist in several isoforms. As such, these observations formed the fundamental basis of the current research under the following aims:

1. To study a natural casein micelle that does not contain α-caseins and has a relatively low level of κ-casein, African elephant milk being the subject

2. To determine the extent of phosphorylation of β-casein of African elephant milk

3 To predict the structure of caseins by homology modeling in order to explain the casein micelle

4 To investigate the evolutionary development of African elephant casein genes within the mammalian kingdom

1.12. REFERENCES

Al-lazikani, B., J. Jung, Z. Xiang, and B. Honig. 2001. Protein structure prediction. *Curr Opin Chem Biol*. 5:51–56.

Anahory, T., H. Dechaud, R. Bennes, P. Marin, N.J. Lamb, and D. Laoudj. 2002. Identification of low-abundance proteins of bovine colostral and mature milk using two-dimensional electrophoresis followed by microsequencing and mass spectrometry. *Electrophoresis*. 23:1153–1160. doi:10.1002/1522-2683(200204)23:7/8<1153::AID-ELPS1153>3.0.CO;2-Y.

Anderson, S.M., M.C. Rudolph, J.L. McManaman, and M.C. Neville. 2007. Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! *Breast Cancer Res*. 9:204. doi:10.1186/bcr1653.

Bijl, E., R. De Vries, H. Van Valenberg, and T. Huppertz. 2014. Factors in fl uencing casein micelle size in milk of individual cows : Genetic variants and glycosylation of k -casein. *Int Dairy J*. 34:135–141. doi:10.1016/j.idairyj.2013.08.001.

Brennan, A.J., J.A. Sharp, C. Lefevre, D. Topcic, A. Auguste, M. Digby, and K.R. Nicholas. 2007. The Tammar Wallaby and Fur Seal: Models to Examine Local Control of Lactation1. *J Dairy Sci*. 90:E66–E75. doi:10.3168/jds.2006-483.

Burgoyne, R.D., and J.S. Duncan. 1998. Secretion of milk proteins. *J Mammary Gland Biol Neoplasia*. 3:275–286.

Capuco, A. V, and R.M. Akers. 2009. The origin and evolution of lactation. *J Biol*. 8:37. doi:10.1186/jbiol139.

Casado, B., M. Affolter, and M. Kussmann. 2009. OMICS-rooted studies of milk proteins, oligosaccharides and lipids. *J Proteomics*. 73:196–208. doi:10.1016/j.jprot.2009.09.018.

Chain, P., S. Kurtz, E. Ohlebusch, and T. Slezak. 2003. An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges. *Brief Bioinform*. 4:105–123. doi:NO_DOI.

D'Alessandro, A., A. Scaloni, and L. Zolla. 2010. Human Milk Proteins : An interactomics and updated functional overview. *J Proteome Res*. 9:3339–3373.

Dalgleish, D.G., and M. Corredig. 2012. The structure of the casein micelle of milk and its changes during processing. *Annu Rev Food Sci Technol*. 3:449–467. doi:10.1146/annurev-food-022811-101214.

Dalgleish, D.G., P.A. Spagnuolo, and H. Douglas Goff. 2004. A possible structure of the casein micelle based on high-resolution field-emission scanning electron microscopy. *Int Dairy J*. 14:1025–1031. doi:10.1016/j.idairyj.2004.04.008.

De Kruif, C.G. 1999. Casein micelle interactions. *Int Dairy J*. 9:183–188.

De Kruif, C.G., T. Huppertz, V.S. Urban, and A. V. Petukhov. 2012. Casein micelles and their internal structure. *Adv Colloid Interface Sci*. 171-172:36–52. doi:10.1016/j.cis.2012.01.002.

Deschamps, J.R. 2010. X-ray crystallography of chemical compounds. *Life Sci*. 86:585–589. doi:10.1016/j.lfs.2009.02.028.

Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrell, V.R. Harwalkar, R. Jenness, and R.M. Whitney. 1984. Nomenclature of proteins of cow's milk: fifth revision. *J Dairy Sci*. 67:1599–1631. doi:10.3168/jds.S0022-0302(84)81485-X.

Elbegdorj, O., R.B. Westkaemper, and Y. Zhang. 2013. A homology modeling study toward the understanding of three-dimensional structure and putative pharmacological profile of the G-protein coupled receptor GPR55. *J Mol Graph Model*. 39:50–60. doi:10.1016/j.jmgm.2012.10.005.

Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk--sixth revision. *J Dairy Sci*. 87:1641–74. doi:10.3168/jds.S0022-0302(04)73319-6.

Farrell, H.M., E.L. Malin, E.M. Brown, and P.X. Qi. 2006. Casein micelle structure : What can be learned from milk synthesis and structural biology ? *Curr Opin Colloid Interface Sci*. 11:135–147. doi:10.1016/j.cocis.2005.11.005.

Farrell, H.M., E.D. Wickham, J.J. Unruh, P.X. Qi, and P.D. Hoagland. 2001.

Secondary structural studies of bovine caseins: Temperature dependence of ??-casein structure as analyzed by circular dichroism and FTIR spectroscopy and correlation with micellization. *In* Food Hydrocolloids. 341–354.

Fox, P.F., and A. Brodkorb. 2008. The casein micelle: Historical aspects, current concepts and significance. *Int Dairy J*. 18:677–684. doi:10.1016/j.idairyj.2008.03.002.

Fox, P.F., and P.L.H. Mcsweeney. 1998. Dairy chemistry and biochemistry. *In* dairy chemistry and biochemistry. 478.

Ginger, M.R., and M.R. Grigor. 1999. Comparative aspects of milk caseins. *Comp Biochem Physiol Part B*. 124:133–145.

Girardet, J.M., L. Miclo, S. Florent, D. Molle, and J.L. Gaillard. 2006. Determination of the phosphorylation level and deamidation susceptibility of equine β-casein. *Proteomics*. 6:3707–3717. doi:10.1002/pmic.200500728.

Goldman, A.S. 2012. Evolution of Immune Functions of the Mammary Gland and Protection of the Infant. *Breastfeed Med*. 7:132–142. doi:10.1089/bfm.2012.0025.

Green, B., K. Newgrain, and J. Merchant. 1980. Changes in milk composition during lactation in the tammar wallaby (Macropus eugenii). *Aust J Biol Sci*. 33:35–42.

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

Heid, H.W., and T.W. Keenan. 2005. Intracellular origin and secretion of milk fat globules. *Eur J Cell Biol*. 84:245–258. doi:10.1016/j.ejcb.2004.12.002.

Herrero, J., M. Muffato, K. Beal, S. Fitzgerald, L. Gordon, M. Pignatelli, A.J. Vilella, S.M.J. Searle, R. Amode, S. Brent, W. Spooner, E. Kulesha, A. Yates, and P. Flicek. 2016. Ensembl comparative genomics resources. *Database*. 2016:1–17. doi:10.1093/database/bav096.

Holt, C. 2015. Casein and casein micelle structures, functions and diversity in 20 species. *Int Dairy J*. 60:2–13. doi:10.1016/j.idairyj.2016.01.004.

Holt, C., J.A. Carver, H. Ecroyd, and D.C. Thorn. 2013. Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J Dairy Sci*. 96:6127–46. doi:10.3168/jds.2013-6831.

Holt, C., and L. Sawyer. 1988. Primary and predicted secondary structures of the caseins in relation to their biological functions. *Protein Eng Des Sel*. 2:251–259. doi:10.1093/protein/2.4.251.

Horne, D.S. 1998. Casein Interactions : Casting light on the black boxes, the structure in dairy products. *Int Dairy J*. 6946:171–177.

Iametti, B.S., G. Tedeschi, E. Oungre, and F. Bonomi. 2001. Primary structure of kappa-casein isolated from mares' milk. *J Dairy Res*. 68:53–61.

Jolivet, P., I. Queiroz Macedo, M. Wu, and J.C. Meunier. 2000. Plasmin digestion of bovine β-casein dephosphorylated with one protein phosphatase type 2A purified from Yarrowia lipolytica. *Lait*. 80:517–526.

Kawasaki, K. 2009. The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Dev Genes Evol*. 219:147–157. doi:10.1007/s00427-009-0276-x.

Kawasaki, K., A.G. Lafont, and J.Y. Sire. 2011. The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol Biol Evol*. 28:2053–2061. doi:10.1093/molbev/msr020.

Kelly, S.M., T.J. Jess, and N.C. Price. 2005. How to study proteins by circular dichroism. *Biochim Biophys Acta - Proteins Proteomics*. 1751:119–139. doi:10.1016/j.bbapap.2005.06.005.

Krieger, E., S.B. Nabuurs, and G. Vriend. 2003. Homology modeling. *Struct Bioinforma*. 857:507–508. doi:10.1007/978-1-61779-588-6.

Kumosinski, T.F., E.M. Brown, and H.M. Farrell Jr. 1994. Predicted energy-minimized as1-casein working model. *ACS Symp Ser*. 576:1,368–391.

Kumosinski, T.F., E.M. Brown, and H.M. Farrell. 1991. 3-Dimensional molecular modeling of bovine caseins - alpha-S1-casein. *J Dairy Sci*. 74:2889–2895. doi:10.3168/jds.S0022-0302(91)78470-1.

Kumosinski, T.F., E.M. Brown, and H.M. Farrell. 1993. Three-dimensional molecular modeling of bovine caseins: an energy-minimized beta-casein structure. *J Dairy Sci*. 76:931–45. doi:10.3168/jds.S0022-0302(93)77420-2.

Kundrotas, P.J., M.F. Lensink, and E. Alexov. 2008. Homology-based modeling of 3D structures of protein-protein complexes using alignments of modified sequence profiles. *Int J Biol Macromol*. 43:198–208. doi:10.1016/j.ijbiomac.2008.05.004.

Lang, S.L.C., S.J. Iverson, and W.D. Bowen. 2005. Individual variation in milk composition over lactation in harbour seals (Phoca vitulina) and the potential consequences of intermittent attendance. *Can J Zool*. 1531:1525–1531. doi:10.1139/Z05-149.

Larson, B.L. 1979. Biosynthesis and secretion of milk proteins: a review. *J Dairy Res*. 46:161–174. doi:10.1017/S002202990001699X.

Lefèvre, C.M., J.A. Sharp, and K.R. Nicholas. 2010. Evolution of lactation: ancient origin and extreme adaptations of the lactation system. *Annu Rev Genomics Hum Genet*. 11:219–238. doi:10.1146/annurev-genom-082509-141806.

Lemay, D.G., D.J. Lynn, W.F. Martin, M.C. Neville, T.M. Casey, G. Rincon, E. V Kriventseva, W.C. Barris, A.S. Hinrichs, A.J. Molenaar, K.S. Pollard, N.J. Maqbool, K. Singh, R. Murney, E.M. Zdobnov, R.L. Tellam, J.F. Medrano, J.B. German, and M. Rijnkels. 2009. The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biol*. 10:R43. doi:10.1186/gb-2009-10-4-r43.

Lönnerdal, B. 2007. Trace element transport in the mammary gland. *Annu Rev Nutr*. 27:165–177. doi:10.1146/annurev.nutr.27.061406.093809.

Madende, M., G. Osthoff, H.-G. Patterton, H.E. Patterton, P. Martin, and D.J. Opperman. 2015. Characterization of casein and alpha lactalbumin of African elephant (Loxodonta africana) milk. *J Dairy Sci*. 98:8308–18. doi:10.3168/jds.2014-9195.

Mamone, G., S. Caira, G. Garro, A. Nicolai, P. Ferranti, G. Picariello, A. Malorni, L. Chianese, and F. Addeo. 2003. Casein phosphoproteome: Identification of

phosphoproteins by combined mass spectrometry and two-dimensional gel electrophoresis. *Electrophoresis*. 24:2824–2837. doi:10.1002/elps.200305545.

Manso, M. A; Leonil, J; Jan, G; Gagnaire, V. 2012. Application of proteomics to the characterisation of milk and dairy products. *J Proteomics*. 75:4259–4274. doi:10.1016/j.idairyj.2004.07.021.

Martin, P., C. Cebo, and G. Miranda. 2013. Interspecies comparison of milk proteins: Quantitative variability and molecular diversity. *In* Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects, 4th Edition. 387–429.

Mather, I.H., and T.W. Keenan. 1998. Origin and secretion of milk lipids. *J Mammary Gland Biol Neoplasia*. 3:259–273. doi:10.1023/a:1018711410270.

McCullagh, K.G., and E.M. Widdowson. 1970. The milk of the African elephant. *Br J Nutr*. 24:109–117.

McMahon, D.J., and W.R. McManus. 1998. Rethinking Casein Micelle Structure Using Electron Microscopy. *J Dairy Sci*. 81:2985–2993. doi:10.3168/jds.S0022-0302(98)75862-X.

Michael Byler, D., H.M. Farrell, and H. Susi. 1988. Raman Spectroscopic Study of Casein Structure. *J Dairy Sci*. 71:2622–2629. doi:10.3168/jds.S0022-0302(88)79855-0.

Miyoshi, M., C.-H. Yoon, F. Ibuki, and M. Kanamori. 1976. The Characterization of Rennin Action on κ-Casein Using CM-Cellulose. *Agric Biol Chem*. 40:347–352. doi:10.1080/00021369.1976.10862056.

Neville, M.C., J. Morton, and S. Umemura. 2001. Lactogenesis: The Transition from Pregnancy to Lactation. *Pediatr Clin North Am*. 48:35–52. doi:10.1016/S0031-3955(05)70284-4.

O'Donnell, R.O., J.W. Holland, H.C. Deeth, and P. Alewood. 2004. Milk proteomics. *Int Dairy J*. 14:1013–1023. doi:10.1016/j.idairyj.2004.04.004.

Oftedal, O.T. 2002. The origin of Lactation as a Water Source for Parchment Shelled Eggs. *J Mammary Gland Biol Neoplasia*. 7.

Oftedal, O.T. 2012. The evolution of milk secretion and its ancient origins. *Animal*. 6:355–368. doi:10.1017/S1751731111001935.

Osthoff, G. 2012. Elephant milk nova 2012. *In* Elephants: ecology, behaviour and conservationcology, behaviour and conservation. D. Maya, Aranovich ; Olivier, editor. Nova Science Publishers, New York. 97–117.

Osthoff, G., H.O. De Waal, A. Hugo, M. de Wit, and P. Botes. 2005. Milk composition of a free-ranging African elephant (Loxodonta africana) cow during early lactation. *Comp Biochem Physiol A Mol Integr Physiol*. 141:223–9. doi:10.1016/j.cbpb.2005.05.039.

Osthoff, G., M. de Wit, A. Hugo, and B.I. Kamara. 2007. Milk composition of three free-ranging African elephant (Loxodonta africana africana) cows during mid lactation. *Comp Biochem Physiol B Biochem Mol Biol*. 148:1–5. doi:10.1016/j.cbpb.2007.02.015.

Petrey, D., and B. Honig. 2005. Protein Structure Prediction : Inroads to biology. *Mol Cell*. 20:811–819. doi:10.1016/j.molcel.2005.12.005.

Phadungath, C. 2005. Casein micelle structure : a concise review. *J Sci Technol*. 27:201–212.

Poth, A.G., H.C. Deeth, P.F. Alewood, and J.W. Holland. 2008. Analysis of the human casein phosphoproteome by 2-D electrophoresis and MALDI-TOF/TOF MS reveals new phosphoforms. *J Proteome Res*. 7:5017–5027. doi:10.1021/pr800387s.

Potočnik, K., V. Gantner, K. Kuterovac, and A. Cividini. 2011. Mare’s milk: composition and protein fraction in comparison with different milk species. *Mljekarstvo*. 61:107–113.

Qi, P.X. 2007. Studies of casein micelle structure: the past and the present. *Lait*. 87:363–383. doi:10.1051/lait:2007026.

Rijnkels, M. 2002. Multispecies comparison of the casein gene loci and evolution of casein gene family. *J Mammary Gland Biol Neoplasia*. 7:327–345. doi:10.1023/A:1022808918013.

Roncada, P., C. Piras, A. Soggiu, R. Turk, A. Urbani, and L. Bonizzi. 2012. Farm animal milk proteomics. *J Proteomics*. 75:4259–4274. doi:10.1016/j.jprot.2012.05.028.

Rost, B. 1999. Twilight zone of protein sequence alignments. *Protein Eng*. 12:85–94. doi:10.1093/protein/12.2.85.

Shennan, D.B. 2008. Calcium transport by mammary secretory cells: mechanisms underlying transepithelial movement. *Cell Mol Biol Lett*. 13:514–25. doi:10.2478/s11658-008-0020-y.

Shennan, D.B., and M. Peaker. 2000. Transport of milk constituents by the mammary gland. *Physiol Rev*. 80:925–51.

Svensson, M., A. Håkansson, A.K. Mossberg, S. Linse, and C. Svanborg. 2000. Conversion of alpha-lactalbumin to a protein inducing apoptosis. *Proc Natl Acad Sci U S A*. 97:4221–4226. doi:10.1073/pnas.97.8.4221.

Swaisgood, H.E. 1993. Review and update of casein chemistry. *J Dairy Sci*. 76:3054–61. doi:10.3168/jds.S0022-0302(93)77645-6.

Touchman, J. 2010. Comparative Genomics. *Comp Genomics Nat Educ Knowl 3(10)13*.

Vanselow, J., W. Yang, J. Herrmann, H. Zerbe, H-J. Schuberth, W. Petzl, W. Tomek, and H.-M. Seyfert. 2006. DNA-remethylation around a STAT5-binding enhancer in the alphaS1-casein promoter is associated with abrupt shutdown of alphaS1-casein synthesis during acute mastitis. *J Mol Endocrinol*. 37:463–77. doi:10.1677/jme.1.02131.

Walstra, P. 1999. Casein sub-micelles: do they exist? *Int Dairy J*. 9:189–192. doi:10.1016/S0958-6946(99)00059-X.

CHAPTER 2

INTERSPECIES COMPARISON OF CASEIN MICELLES BY HIGH RESOLUTION FIELD EMISSION SCANNING ELECTRON MICROSCOPE

2.1. INTRODUCTION

Caseins of bovine milk are defined as the group of phosphoproteins that precipitate from raw skim milk by acidification to pH 4.6 at 20 °C (Eigel et al., 1984). This group of proteins consists of mainly four gene products known as α_{s1}-, α_{s2}-, κ- and β-Casein (Farrell et al., 2004). Caseins in milk exist as colloidal aggregates of individual caseins with calcium and phosphate, known as casein micelles (Walstra, 1999). Casein micelles convert milk into a free flowing, low viscosity liquid, and also provide the means for transportation of high levels of calcium and phosphate in the mammary gland (De Kruif, 1999; Holt et al., 2013). The exact structure of casein micelles has been a subject of extensive study over several decades (Holt et al., 2013).

A number of methods have been utilized in an effort to unravel the structure of casein micelles, these include: small angle neutron scattering (SANS) or small angle X-ray scattering (SAXS), scanning electron microscopy (SEM) and transmission electron microscopy (TEM) (Dalgleish and Corredig, 2012; McMahon and McManus,

1998). Electron microscopy can be used to elucidate the structure of biological molecules. Most of the electron microscopy studies on casein micelles have been focused on bovine caseins, while most data on casein micelles of other mammalian species was acquired with other techniques (Park, 2007; Qi, 2007). Furthermore, casein micelle structure studies on bovine milk have not always been on fresh milk, some researchers utilized reconstituted processed powdered milk or skim milk (Dalgleish et al., 2004; De Kruif et al., 2012). During the processing of milk, casein micelles can be affected by changing their structure, shape, size and the overall casein composition (Hillbrick et al., 1999).

In our current research, casein micelles from bovine, sheep, horse, human and African elephant milk were examined using an Extreme-resolution Analytical Field Emission Scanning Electron Microscope (FE-SEM). The milk samples were specifically selected for examination due to their relative differences in their casein composition and proportions (Poth et al., 2008; Armaforte et al., 2010; Potočnik et al., 2011; Madende et al., 2015). Despite these variations in casein composition, these species competently form casein micelles. According to literature, micelle sizes and in some cases shape, differs markedly among the five species chosen for this study. Because each casein component plays a role in micelle formation, stability and maintenance, we compared the possible structural differences these factors may exert on the casein micelle, particularly its shape and surface appearance. The effect of freezing on casein micelles was also investigated particular the effects on shape, size and surface appearance.

The aim of this chapter is to study a natural casein micelle that does not contain α-caseins and has a relatively low level of κ-casein with African elephant milk being the subject for comparison with casein micelles from other species, which differ in composition of casein types.

2.2. Materials and methods

2.2.1. Sample preparation

The research was approved by the ethical boards of the University of the Free State for animal- (project no. UFS-AED2016/0106) and human subjects (project no. ECUFS NR 193/2015). Milk from five mammalian species was examined. Fresh milk samples were collected from cow (fresh unpasteurized milk, Dairy Corporation, Bloemfontein, South Africa), sheep (Patria farm, Smithfield, South Africa) and horse (Private horse breeder, Bloemfontein, South Africa), African elephant (Knysna Elephant Ranch, Knysna, South Africa), and human. Immediately after milking, the milk samples were placed on ice to limit chemical and physical changes before SEM preparations were conducted. A part of the fresh milk was frozen at -20 °C for three months in order to prepare frozen and thawed milk for SEM. Frozen milk was thawed at 39 °C, and allowed to cool to ambient temperature at 22 °C before similar SEM preparations with that of fresh milk samples were conducted. All the SEM preparations were carried out on three different samples of each species.

2.2.2. SEM preparation

A method by Dalgleish et al. (2004) was adapted with a few modifications. Carbon rods were polished to obtain a shiny surface. A drop of fresh or frozen and thawed milk was mounted onto the smooth and polished end of carbon rods and casein micelles were allowed to adhere to the substrate for five minutes. Excess milk was then carefully removed with pieces of clean filter paper. A few drops of phosphate buffer (pH 7) were placed onto the sample and removed after two minutes and replaced by the primary fixative, a 2 % glutaraldehyde solution (Merck, Darmstadt, Germany), prepared in Sorensen's phosphate buffer. After an hour the primary fixative was removed, the sample was rinsed with the phosphate buffer and the secondary fixative, 1 % osmium tetroxide (Merck, Darmstadt, Germany) in phosphate buffer, was added for 1 hour. Following removal of the secondary fixative, the sample was rinsed three times with phosphate buffer and then dehydrated using a graded ethanol (EtOH) series: 70 %, 90 %, 95 % and 100 % EtOH (elapsed time per solution was five minutes). The 95 % EtOH step was repeated twice whereas the 100 % EtOH step was repeated three times. Following dehydration, the samples were critically point dried and mounted onto SEM stubs. The samples were stored in a desiccator at room temperature until imaging.

2.2.3. SEM imaging

SEM stubs with mounted samples were transferred to the SEM for imaging. No metal coating was done prior to imaging as this obscures proteins from view. Imaging was performed with a JSM-7800F Extreme-resolution Analytical Field Emission SEM (FE-SEM). Acceleration voltage of 5 kv was maintained. The digital

acquisition of the mages was done using the JSM-7800F FE-SEM inbuilt image capture camera.

2.3. Results

The casein micelle structures of the different species as viewed by FE-SEM are shown in Figs 2.1–2.7 and an inter-species comparison of the micelle composition and sizes is given in Table 2.1.

2.3.1. Cow milk casein micelles

The electron micrographs in Figures 2.1 and 2.2 show cow milk casein micelles. These micelles resemble electron micrographs of casein micelles as reported by Dalgleish et al. (2004). The micelles are also interconnected by threads of a protein smear, and many small micelles are trapped in this smear. Dalgleish et al. (2004) described this as partially disintegrated micelles. In the aforementioned study, spray dried cow skim milk was utilized as the sample whereas in our study fresh milk was used. Cow casein micelles appear spherical in shape and their surface appears rough with cylindrical particles seemingly bound on the surface. The average diameter of the casein micelles is approximately 220 nm which is less than the diameter range of 300-350 nm reported by Dalgleish et al. (2004).

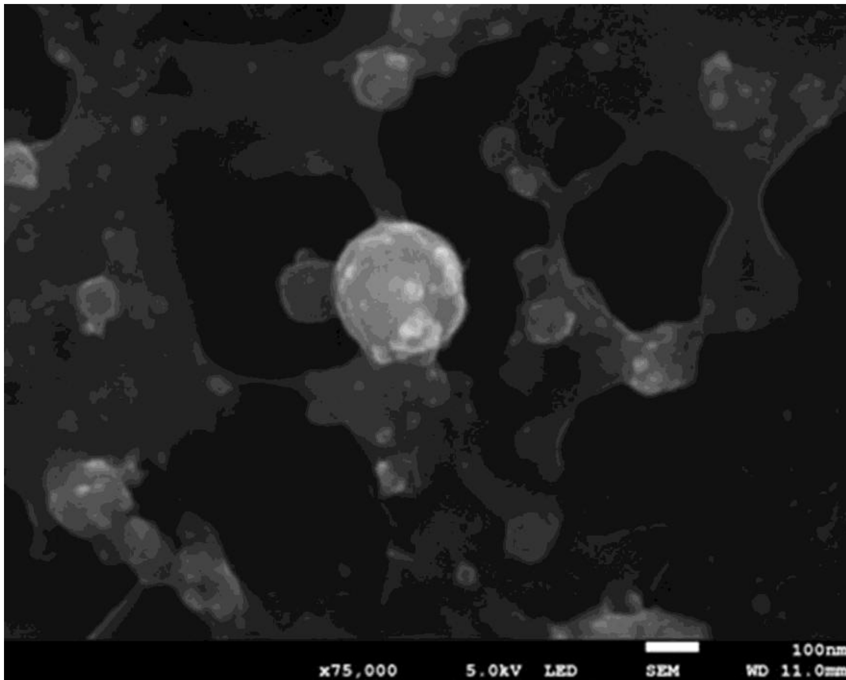

Figure 2.1. Scanning electron micrograph of cow casein micelles. Micelles are interconnected by a smear or coating and threads of a protein network.

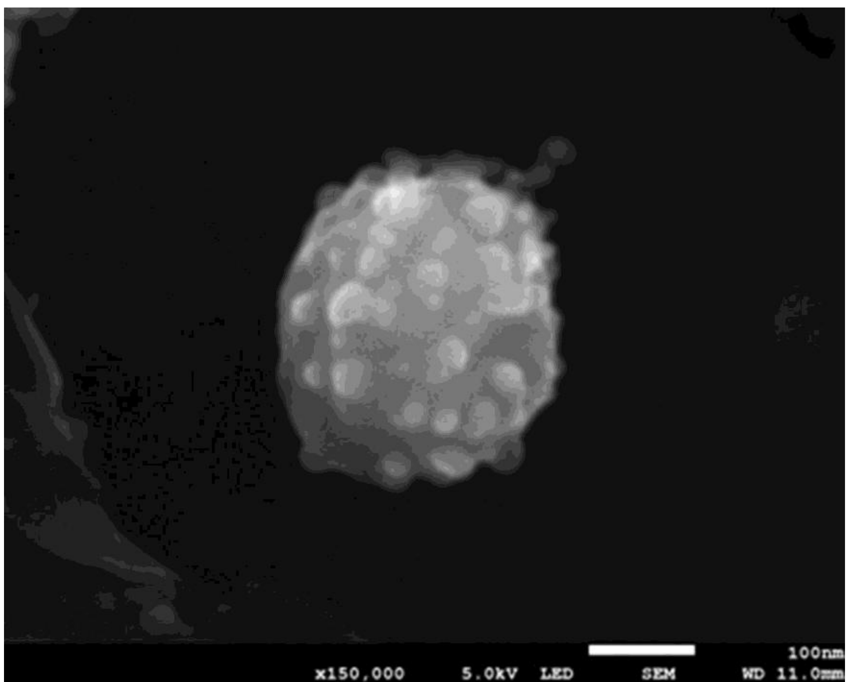

Figure 2.2. Scanning electron micrograph of a single cow casein micelle.

2.3.2. Sheep milk casein micelles

The electron micrograph in Figure 2.3 shows sheep milk casein micelles. In comparison to casein micelles of African elephant, cow, human and horse milk, sheep casein micelles were observed to be the smallest. The diameter of sheep casein micelles is in the range of 60–160 nm. Sheep and cow are both ruminants, they also contain all four major caseins (α_{s1}-, α_{s2}-, κ- and β-caseins) in approximately similar proportions. However literature reports that they differ in average micelle diameter where sheep casein micelles (193 nm) are slightly larger than cow micelles (180 nm) (Park, 2007). In contrast, our microscopic work on fresh milk shows that sheep casein micelles are much smaller than cow micelles and appear spherical in shape. Small protruding structures are visible on some of the large micelles. Other than observed for the cow casein micelles, the sheep micelles may stick together, but are not interconnected by threads of a protein smear.

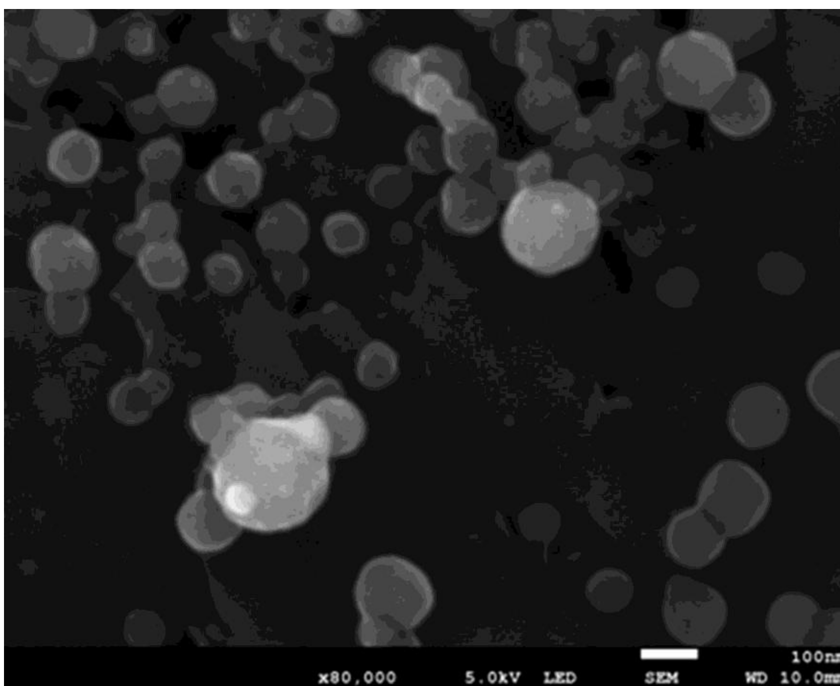

Figure 2.3. Scanning electron micrograph of sheep casein micelles. The micelles appear small and clustered.

2.3.3. Human milk casein micelles

Figure 2.4 shows the electron micrograph of human milk casein micelles. These appear spherical in shape and their surface also appears rough. Like cow and African elephant casein micelles, some structures that resemble tubular protrudences are also evident on the surface the micelles. In comparison to the previously discussed casein micelles, human milk casein micelles are in the same order of size as their counterparts in horse milk. Human milk casein micelles are in the size range of 100–1500 nm. Very few threads of a protein smear or network is obvious.

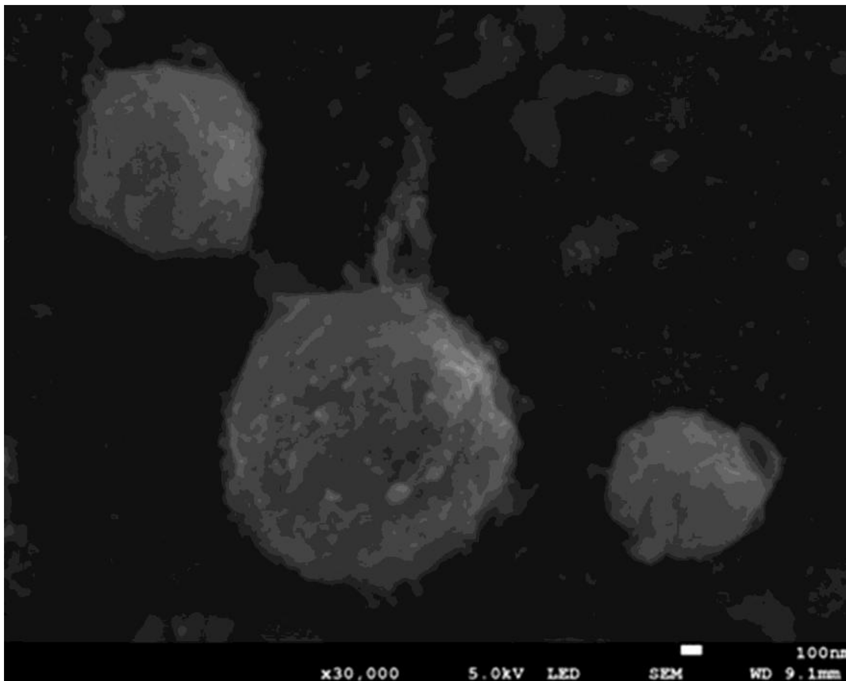

Figure 2.4. Scanning electron micrograph of human casein micelles. Some micelles seem to be connected by strands of protein.

2.3.4. Horse milk casein micelles

Figure 2.5 shows an electron micrograph of horse milk casein micelles. Of the casein micelles investigated in our study, horse casein micelles were comparable in size to sheep micelles ranging between 150-300 nm. Literature places horse casein micelles among some of the smallest (average diameter 255 nm) (Malacarne et al., 2002), the sizes observed in the present study are close to literature observations. In addition to the small size, horse casein micelles appeared very smooth on the surface with no tubular protrudences observed on the surface. The micelles are interconnected by threads of a protein smear that also entraps a large number of other small micelles.

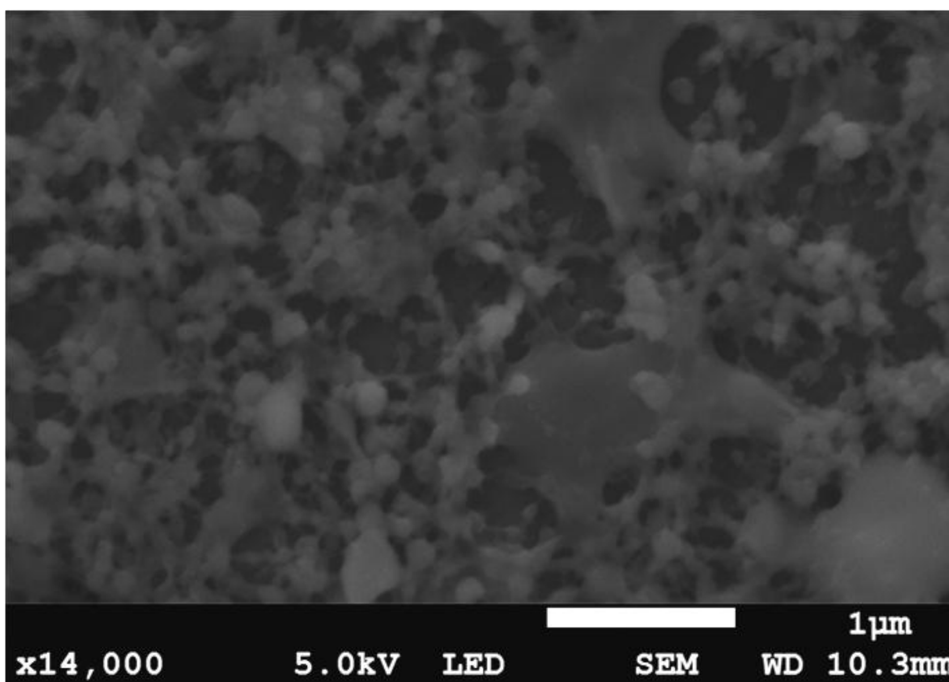

Figure 2.5. Scanning electron micrograph of horse casein micelles of different sizes. Micelles are interconnected by a smear or coating of a protein network.

2.3.5. African elephant milk casein micelles

The electron micrograph of African elephant milk casein micelles in Figure 2.6 shows a large micelle surrounded by a large number of small micelles that are interconnected by a smear and threads of a protein network. The casein micelles of African elephant ranged in diameter between 200–700 nm, and the small micelles that are entrapped in the protein network are 50-150 nm. Some tubule-like structures (diameter in the range 20-40nm) seem to protrude from the inside of the large micelles giving the micelle an uneven surface appearance.

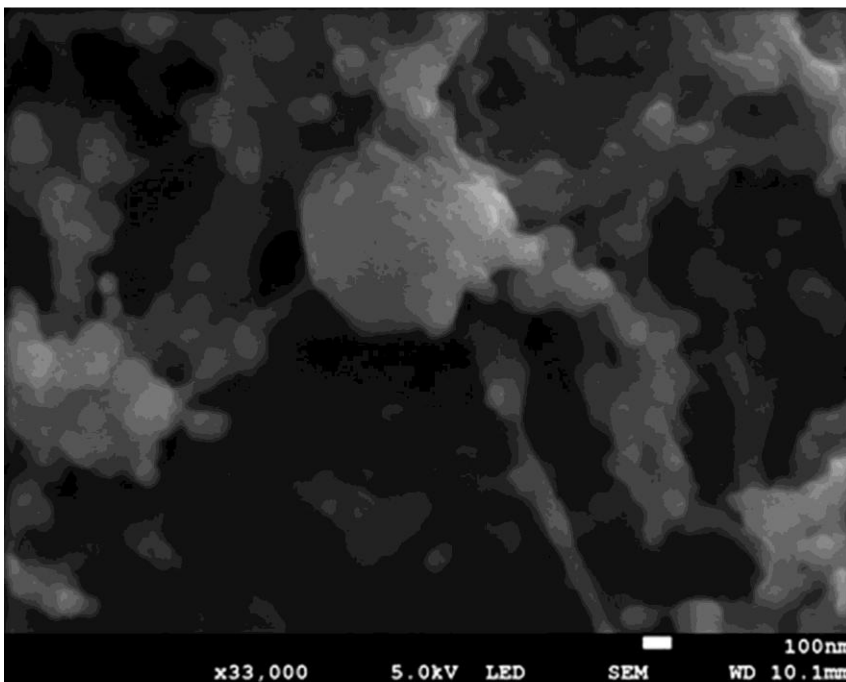

Figure 2.6. Scanning electron micrograph of African elephant casein micelles fixed on a carbon substrate. The background is composed of micelle particles that may have disintegrated from the main micelle whereas a smear or coat like material forms strands between the main micelle and the broken off particles.

2.3.6. Frozen milk casein micelles

The effect of freezing on casein micelles was investigated by freezing fresh milk samples for three months before SEM examination. Figure 2.7a, Figure 2.7b, Figure 2.7c, Figure 2.7d and Figure 2.7e depict casein micelles of cow, sheep, horse, African elephant and human milk respectively, after freezing. In comparison to their fresh milk counterparts, all the casein micelles except for sheep micelles increased in size (see Table 1) and their surface appearance was also altered. The protein smear that connected the casein micelles of cow, horse, human and African elephant of fresh milk preparation was not visible when frozen and thawed. Instead, a large number of smaller and free micelles were visible. Unlike sheep micelles, very clear surface appearance differences could be observed for cow, horse, human and African elephant casein micelles. Cow micelles appeared to have lost the previously rough micelle surface of their fresh milk counterparts, the micelles appeared smoother. Human and African elephant micelles had much more and prominent structures protruding from their surface, a lot more than what was observed in fresh milk. Horse micelles appeared loose, the material that appeared to be coating and causing strands between the horse micelles in fresh milk seem to be nonexistent. However horse micelles still appear smooth on the surface even after the freezing effect.

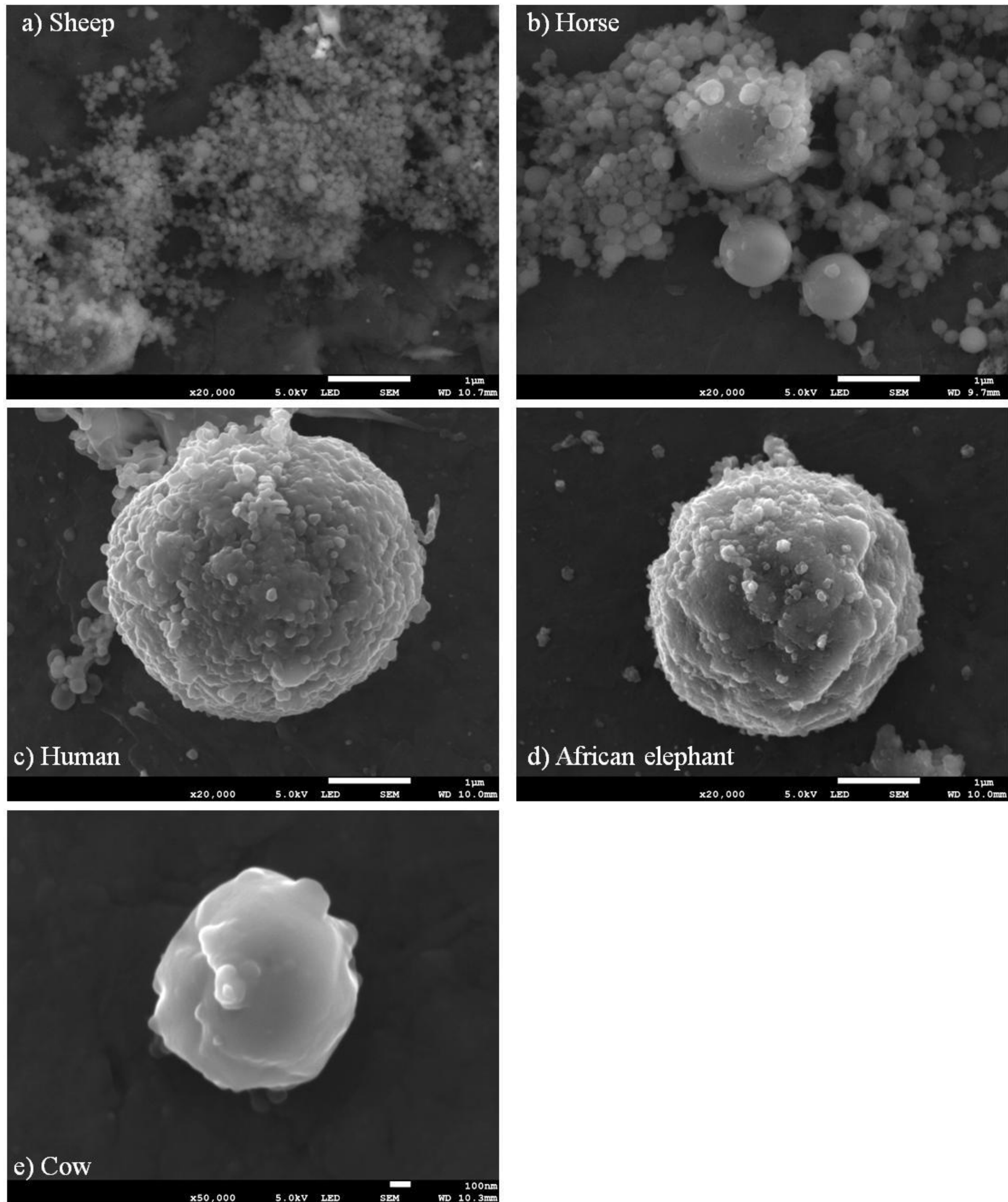

Figure 2.7. Scanning electron micrograph of a) sheep, b) horse, c) human, d) African elephant and e) cow casein micelles of milk samples that where frozen and stored at -20° C for three months and thawed at 39°C before preparation for microscopy.

Table 2.1. Casein composition and micelle sizes of sheep, cow, African elephant, human and horse milk

Species	α_{S1}-casein %	α_{S2}-casein %	β-casein %	κ-casein %	Micelle size nm (Literature)	FE-SEM micelle size nm (fresh)	FE-SEM micelle size nm (frozen)
Sheep	50[b]	+	40[b]	10[b]	210[b]	50-200	50 - 160
Cow	38[c]	10[c]	40[c]	12[c]	182[b]	50-250	1000-2300
African elephant	-	-	89[a]	11[a]	N/A	350-700	2500-3100
Human	3[c]	-	70[c]	27[c]	64[b]	100-1500	1800-2800
Horse	40-60[b]	Trace	40-50[b]	4-7[b]	255[b]	150-300	155-400

The size comparison of casein micelles from literature is drawn with the sizes observed using fresh and frozen milk ([a]Madende et al., 2015; [b]Potočnik et al., 2011; [c]Qi, 2007)

2.4. Discussion

Cow milk proteins have been used as a standard for the study of milk proteins from other mammalian species, it is therefore of no surprise that cow casein micelles have been studied more intensely than casein micelles of other mammalian species. In order to standardize the method for this study, cow micelles were investigated and it was observed that they resembled those observed by Dalgleish et al. (2004) in both shape and surface appearance using SEM as the technique of choice. Size comparison showed that the cow caseins in our study were on average smaller than those observed in the aforementioned study, this could be due to the state of milk used, i.e. fresh vs powdered skim milk, where processing can potentially modify micelle size. Literature reports even smaller average micelle sizes for cow milk (Qi,

2007). Nevertheless, other literature reported a much smaller casein micelle size (approximately 182 nm) of cow milk (Malacarne et al., 2002; Park, 2007), this could be a result of different techniques used in examining the micelles. Sheep micelles were observed to be the smallest in comparison to the rest of the micelles in this study, including cow casein micelles. Although both sheep and cow are ruminants, the micelle sizes differ significantly. This difference could be attributed to the small but significant differences in the proportion of the casein types as well as their respective total casein content of which sheep milk has almost double the casein content of cow milk (Potočnik et al., 2011).

The cow casein micelles were observed to be interconnected by threads of, what Dalgleish et al. (2004), described as a protein smear, possibly originating from partially disintegrated micelles. It is possible that the disintegration could be due to micelles that collapse during the coating of the stubs. Almost no protein smear was observed in the electron micrographs of sheep casein micelles.

Human casein micelles were observed to be much larger than those of the ruminants. The large size of micelles could possibly be attributed to the presence of low levels of κ-casein, which is critical in limiting micellar growth. In addition to the possible limited role of κ-casein, the abundant presence of β-casein in human milk may also be responsible for a rather large size of the micelles. Furthermore, β-casein is also thought to occupy the core of the micelle where it interacts via hydrophobic interactions with other calcium sensitive caseins and thereby increase the overall volume of casein micelles (Horne, 1998). The proportion of α_s-caseins is

independent of the size of micelles. However, α_s-caseins are highly phosphorylated proteins and therefore facilitate the binding of excess amounts of colloidal calcium phosphate and also interact via hydrophobic interactions with β-casein. Despite the absence of α_{s2}-casein and the limited presence of α_{s1}-casein in human milk, the abundant β-casein can determine the size of the micelles (Liao et al., 2011). Human milk is similar to horse milk in composition, particularly the protein and lactose content but differs considerably to cow milk (Malacarne et al., 2002). This observation also alludes to the influence that protein content and composition may have on the nature of casein micelles.

Horse micelles were observed to be smooth surfaced compared to casein micelles of the other four species investigated in this study. The dual bonding model suggests that κ-casein is located on the outside of the micelle where it stabilizes the micelle by electrostatic repulsion and steric stabilization in addition to limiting micellar growth (Horne, 1998). Because horse milk has very little κ-casein (Sharifi Rad et al., 2013), this may explain the smooth surface of the observed horse casein micelles as opposed to the rough surface of cow casein micelles. Furthermore, horse milk composition is very different from cow milk composition and therefore the effects of such differences in composition are further illuminated by the differences in appearance of casein micelles from micrographs obtained. This suggests that casein micelle structure and appearance might be dependent on the casein composition of milk.

Until now, there have not been any studies on African elephant casein micelles reported. The spherical shape and rough surface was similar to that of cow micelles although the background of the micelles and the sizes were different. These differences could possibly be attributed to the differences in casein composition between cow and African elephant milk. African elephant lacks α_{s}-caseins and contains very high levels of β-casein relative to κ-casein, the ratio being approximately 8.5:1 (Madende et al., 2015). Cow milk on the other hand contains all four caseins α_{s1}-, α_{s2}-, κ- and β-caseins in the ratio 3.8:1.0:1.2:4.0 respectively (Qi, 2007). According to the internal structure models, casein micelle growth is directed by κ-casein (Horne, 1998); this could help to explain why African elephant casein micelles are much bigger since κ-casein is present in very limited amounts. The background of casein micelles on the electron micrographs also appears to be covered by disintegrated micelle particles. This may suggest that African elephant casein micelles are not very stable compared to micelles of for example cow, which could perhaps also be attributed to the differences in casein composition.

The casein micelles of horse, human as well as African elephant were observed to be highly interconnected by threads of protein smear (Dalgleish et al., 2004) which could be due to micelles that collapse during the coating of the stubs. It therefore seems as if the casein micelles of these three species are more prone to collapse, either because they are more porous and entrap more water, or because the interaction between the casein proteins and casein types is not very stable, or both.

With the exception of sheep casein micelles, all the other micelles in this study increased in size following freezing. The protein smear that connected the casein micelles of cow, horse, human and African elephant of fresh milk preparation was not visible when frozen and thawed, but a large number of smaller and free micelles took its place. This may be evidence that the smear could indeed have been the result of collapsed micelles in which the proteins have not been denatured and aggregated by freezing. It is known that freezing may denature proteins, specifically milk proteins (Nakanishi and Itoh, 1970). Casein micelles are open, dynamic, hydrated and highly porous molecules that can be easily perturbed during processing (Marchin et al., 2007).

Slow freezing and thawing physically distorts micelles and may result in coagulation and moisture uptake by micelles, consequently causing the increase in size and change in micelle surface appearance. Destabilization of micelles by freezing is caused by an increase in calcium and decrease in pH due to calcium phosphate precipitation (Fox and Brodkorb, 2008). The crystallization of lactose also heightens micelle destabilization. The extent of changes that occur to micelles after freezing also points to the degree of micelle stability of the respective milks. African elephant milk tends to coagulate easily upon freeze-thawing in addition to being less stable to ethanol precipitation compared to cow milk. Ethanol stability of African elephant milk averages at 62 % (Madende et al., 2015) compared to that of cow milk which averages at 75 % (Chavez et al., 2004). This poor stability of casein micelles is also evident on the microscopic images. Interestingly, no changes could be observed on sheep casein micelles after freezing (Figure 2.7). Although these micelles are small in size and could therefore obscure any surface changes to the micelle, the size

range remained constant in comparison with fresh milk casein micelles. It is interesting that it was reported that sheep milk keeps well compared to cow milk and therefore can be stored for longer without undergoing too many biological changes (Park, 2007).

2.5. Conclusions

The gross composition of milk varies markedly amongst mammalian species as does the casein content and composition of casein types. A microscopic examination of casein micelles from fresh cow, sheep, horse, human and African elephant milk showed that these micelles adopted a more spherical shape but differed considerably in size and surface appearance. Using high resolution SEM without any coating to the micelles before viewing, sheep micelles appeared the smallest whereas horse and human milk micelles were the largest. The micelles also appeared to have distinct differences on the surface. African elephant casein micelles have tubule like structures that appear to be protruding from the inside of the micelle. The same was observed for human and cow casein micelles with the main differences being that African elephant protruding tubules are much bigger but fewer in numbers. On the contrary, similar tubular protrudences on the surface of human and cow casein micelles are numerically more but are smaller in size. Horse casein micelles appeared smooth on the surface.

Although our observations suggest that casein micelle appearance and structure might be species specific, the explanation of the differences would most probably be on a biochemical level, specifically the total casein content, the proportions of the

individual casein types and the presence and or absence of some of the casein types. The colloidal calcium and phosphate, and the calcium phosphate nanoclusters may also have an effect on the overall casein micelle structure.

The effects of freezing were also apparent from the microscopic images. Sheep milk casein micelles were more resilient to destabilization by freezing and storage at -20 °C. The casein micelles from the other species increased in size and changed in physical appearance during freezing, and the protein smear that connected the casein micelles of cow, horse, human and African elephant of fresh milk preparation was replaced by a large number of smaller and free micelles also points to structural changes during processing.

From the microscopic images, it is apparent that milk composition particularly the proportions of casein types, may play a major role in determining casein micelle size and appearance. Interestingly, casein micelles sequestrate large amounts of calcium and phosphate to enable their safe transport in milk from mother to the neonate. For this process to occur successfully, caseins have phosphorylation sites that bind phosphates and consequently minerals such as calcium. African elephant milk is devoid of α-caseins but has large amounts of β-casein and very small quantities of κ-casein. Since African elephant milk contains casein micelles, β-casein becomes an interesting candidate for study, in particular its phosphorylation pattern which is crucial for calcium binding in the absence of α-caseins. The characterisation of African elephant β-casein will be dealt with in the following chapter.

2.6. REFERENCES

Armaforte, E., E. Curran, T. Huppertz, C.A. Ryan, M.F. Caboni, P.M. O'Connor, R.P. Ross, C. Hirtz, N. Sommerer, F. Chevalier, and A.L. Kelly. 2010. Proteins and proteolysis in pre-term and term human milk and possible implications for infant formulae. *Int Dairy J*. 20:715–723. doi:10.1016/j.idairyj.2010.03.008.

Chavez, M.S., L.M. Negri, M.A. Taverna, and A. Cuatrín. 2004. Bovine milk composition parameters affecting the ethanol stability. *J Dairy Res*. 71:201–6. doi:10.1017/S0022029904000172.

Dalgleish, D.G., and M. Corredig. 2012. The structure of the casein micelle of milk and its changes during processing. *Annu Rev Food Sci Technol*. 3:449–467. doi:10.1146/annurev-food-022811-101214.

Dalgleish, D.G., P.A. Spagnuolo, and H. Douglas Goff. 2004. A possible structure of the casein micelle based on high-resolution field-emission scanning electron microscopy. *Int Dairy J*. 14:1025–1031. doi:10.1016/j.idairyj.2004.04.008.

De Kruif, C.G. 1999. Casein micelle interactions. *Int Dairy J*. 9:183–188.

De Kruif, C.G., T. Huppertz, V.S. Urban, and A.V. Petukhov. 2012. Casein micelles and their internal structure. *Adv Colloid Interface Sci*. 171-172:36–52. doi:10.1016/j.cis.2012.01.002.

Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrell, V.R. Harwalkar, R. Jenness, and R.M. Whitney. 1984. Nomenclature of proteins of cow's Milk-fifth revision. *J Dairy Sci*. 67:1599–1631. doi:10.3168/jds.S0022-0302(84)81485-X.

Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer,

C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk-sixth revision. *J Dairy Sci*. 87:1641–74. doi:10.3168/jds.S0022-0302(04)73319-6.

Fox, P.F., and A. Brodkorb. 2008. The casein micelle: Historical aspects, current concepts and significance. *Int Dairy J*. 18:677–684. doi:10.1016/j.idairyj.2008.03.002.

Hillbrick, G.C., D.J. Mcmahon, and W.R. Mcmanus. 1999. Microstructure of indirectly and directly heated ultra-high-temperature (UHT) processed milk examined using transmission electron microscopy and immunogold labelling. *LWT- Food Sci Technol*. 494:486–494. doi:10.1006/fstl.1999.0586.

Holt, C., J.A. Carver, H. Ecroyd, and D.C. Thorn. 2013. Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J Dairy Sci*. 96:6127–46. doi:10.3168/jds.2013-6831.

Horne, D.S. 1998. Casein Interactions : Casting light on the black boxes, the structure in dairy products. *Int Dairy J*. 6946:171–177.

Liao, Y., R. Alvarado, B. Phinney, and B. Lönnerdal. 2011. Proteomic characterization of specific minor proteins in the human milk casein fraction. *J Proteome Res*. 10:5409–5415. doi:10.1021/pr200660t.

Madende, M., G. Osthoff, H. Patterton, H.E. Patterton, P. Martin, and D.J. Opperman. 2015. Characterization of casein and alpha lactalbumin of African elephant (Loxodonta africana) milk. 8308–8318.

Malacarne, M., F. Martuzzi, A. Summer, and P. Mariani. 2002. Protein and fat composition of mare ' s milk : some nutritional remarks with reference to human and cow ' s milk. *Int Dairy J*. 12:869–877.

Marchin, S., J.L. Putaux, F. Pignon, and J. Léonil. 2007. Effects of the environmental factors on the casein micelle structure studied by cryo transmission electron microscopy and small-angle X-ray scattering/ultrasmall-angle X-ray scattering. *J Chem Phys*. 126:1–10. doi:10.1063/1.2409933.

McMahon, D.J., and W.R. McManus. 1998. Rethinking casein micelle structure using electron microscopy. *J Dairy Sci*. 81:2985–2993. doi:10.3168/jds.S0022-0302(98)75862-X.

Nakanishi, T., and T. Itoh. 1970. Studies on the changes of the milk casein by various treatments part X. *J Agric Chem Soc Japan*. 44:118–123. doi:10.1271/nogeikagaku1924.44.118.

Park, Y.W. 2007. Rheological characteristics of goat and sheep milk. *Small Rumin Res*. 68:73–87. doi:10.1016/j.smallrumres.2006.09.015.

Poth, A.G., H.C. Deeth, P.F. Alewood, and J.W. Holland. 2008. Analysis of the human casein phosphoproteome by 2-D electrophoresis and MALDI-TOF/TOF MS reveals new phosphoforms. *J Proteome Res*. 7:5017–5027. doi:10.1021/pr800387s.

Potočnik, K., V. Gantner, K. Kuterovac, and A. Cividini. 2011. Mare's milk: composition and protein fraction in comparison with different milk species. *Mljekarstvo*. 61:107–113.

Qi, P.X. 2007. Studies of casein micelle structure: the past and the present. *Lait*. 87:363–383. doi:10.1051/lait:2007026.

Walstra, P. 1999. Casein sub-micelles: do they exist? *Int Dairy J*. 9:189–192. doi:10.1016/S0958-6946(99)00059-X.

Sharifi Rad, J., M. Hoseini Alfatemi, and M. Sharifi Rad. 2013. Horse Milk; the Composition, Equine Milk Proteins, Milk Allergy and Homology between Mammal Species with Horse. *Br Biomed Bull*.

CHAPTER 3

ELUCIDATION OF AFRICAN ELEPHANT BETA CASEIN PHOSPHORYLATION STATE

3.1. INTRODUCTION

Caseins (α_{s1}-, α_{s2}-, κ- and β-casein) in milk assemble into large colloidal aggregates with mineral calcium and phosphate to form structures known as casein micelles (Sorensen et al., 2003; Farrell et al., 2004). Like most proteins, caseins naturally undergo post-translational modification (PTM) which is important for protein function and formation of protein complex (Parker et al., 2010). Phosphorylation and glycosylation are the two most pervasive and naturally occurring PTMs that affect caseins where α- and β-caseins are predominantly phosphorylated whereas κ-casein is glycosylated. The presence of multiple phosphorylation sites enable proteins to adapt multiple functions depending on the occupied phosphorylation site (Thingholm et al., 2009). In addition, phosphorylation modification is reversible and therefore changes in protein activity can be altered by phosphorylation and dephosphorylation in response to cellular stimuli (Stamm, 2008).

Phosphorylation of caseins is initiated by a variety of mammary gland specific casein kinases and occurs at one or more of the threonine or serine residues that comprise

part of the motif Ser/Thr-X-Y (X represents any amino acid whereas Y represents an acidic amino acid) (Sorensen et al., 2003; Thingholm et al., 2009). Phosphorylated serines or threonines can also become determinants of phosphorylation of other serines (Ginger and Grigor, 1999). The above is illustrated in Figure 3.1 which depicts the phosphorylation pattern of sheep, human, cow and horse β-caseins. Caseins have one or more multiple phosphorylation sites, typically a string of serine and/or threonine residues in the peptide sequence (Ginger and Grigor, 1999).

```
Cow          RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHPFAQTQSLVY
Sheep        REQEELNVVGETVESLSSSEESITHINKKIEKFQSEEQQQTEDELQDKIHPFAQAQSLVY
Human        ---------RETIESLSSSEESITE-YKKVEKVKHEDQQQGEDEHQDKIYPSFQPQPLIY
Horse        REKEELNVSSETVESLSSNEEDSSS-EEKLQKFKHEGQQQREVERQDKISRFVQPQPVVY
                       * :*****.* . :   :*::*.: * *** * * ****    * * ::*

Cow          PFPGPIHNS-LPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPF
Sheep        PFTGPIPNS-LPQNILPLTQTPVVVPPFLQPEIMGVPKVKETMVPKHKEMPFPKYPVEPF
Human        PFVEPIPYGFLPQNILPLAQPAV-VLPVPQPEIMEVPKAKDTVYTKGRVMPVLKSPTIPF
Horse        PYAEPVPYAVVPQSILPLAQPPI--LPFLQPEIMEVSQAKETILPKRKVMPFLKSPIVPF
             *:  *:  . :**.* **:*  :   *. ***:* * :.*:::  * : **. * *  **

Cow          TESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKAVP
Sheep        TESQSLTLTDVEKLHLPLPLVQSWMHQPPQPLPPTVMFPPQSVLSLSQPKVLPVPQKAV-
Human        FDPQIPKLTDLENLHLPLPLLQPLMQQVPQPIPQTLALPPQPLWSVPQPKVLPIPQQVVP
Horse        SERQILNPTNGENLRLPVHLIQPFMHQVPQSLLQTLMLPSQPVLSPPQSKVAPFPQPVVP
              : *  . *: *:*:**: *:*  *:*  * :  *: :* * : *  * ** *.** .*

Cow          YPQRDMPIQAFLLYQEPVLGPVRGPFPIIV-----------
Sheep        -PQRDMPIQAFLLYQEPVLGPVRGPFPILV-----------
Human        YPQRAVPVQALLLNQELLLNPTHQIYPVTQPLAPVHNPISV
Horse        YPQRDTPVQAFLLYQDPRLGPTGELDPATQPIVAVHNPVIV
              ***  *:**:** *:  *.*.    *
```

Figure 3.1. Multiple sequence alignment of sheep (Mamone et al., 2003), cow (Farrell et al., 2004), human (Greenberg et al., 1984) and horse (Girardet et al., 2006) β-casein. The phosphorylated amino acid residues are colored in yellow. An asterisk (*) indicates positions which have a single, fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties. A period (.) indicates conservation between groups of weakly similar properties. Rectangular blocks around amino acid residues show the phosphorylation motif S/T-X-Y.

Bovine, human and equine β-caseins are well characterized β-caseins, with bovine β-casein being used as reference to characterize β-caseins of other mammals (Greenberg et al., 1984; Girardet et al., 2006). Bovine β-casein exists in one completely phosphorylated form with 5 phosphate groups per molecule (Eigel et al., 1984; Farrell et al., 2004). In contrast, the human β-casein homolog occurs as a multi-phosphorylated protein with up to 5 phosphate groups per molecule (Greenberg et al., 1984). Like human β-casein, equine β-casein is composed of multi-phosphorylated isoforms having between 3 and 7 phosphate groups per molecule (Girardet et al., 2006). Unlike α-caseins that have more than one multi-phosphorylation region over the extended amino acid sequence, β-caseins typically have a single major multi-phosphorylation region which is situated near the N-terminus (Ginger and Grigor, 1999). African elephant milk lacks α-caseins and contains β-casein as the major protein (Madende et al., 2015).

The detailed characterization of African elephant β-casein has not been done and its phosphorylation pattern has not been elucidated. There are three splice variants of the African elephant β-casein gene in the elephant genome database (ENSLAFG00000014755) (http://www.ensembl.org). However in a study by (Madende et al., 2015), a fourth African elephant β-casein sequence, originating through exon skipping (Figure 3.7), was proposed. A multiple sequence alignment of the four possible β-casein sequences in African elephant milk is shown in Figure 3.2.

```
EleBeta1     MKVFILACLVAFAL-GRE---------TVENLSS-------SEESVTQVNKQKPEGVKHE
EleBeta2     MKVFILACLVAFAL-GREKEEIIVSTETVENLSS-------SEESVTQVNKQKPEGVKHE
EleBeta3     MKVFILACLVAFAL-GREKEEIIVSTETVENLSSSEIRQFYSEESVTQVNKQKPEGVKHE
EleBeta4     MKVFILACLVAFALGGREKEEIIVST----------------EESVTQVNKQKPEGVKHE
             ************** ***                        :*****************

EleBeta1     EQQR-EDEHQNKIQPLFQPQPLVYPFAEPIPYTVFPPNAIPLAQPIVVLPFPQPEVKQLP
EleBeta2     EQQREEDEHQNKIQPLFQPQPLVYPFAEPIPYTVFPPNAIPLAQPIVVLPFPQPEVKQLP
EleBeta3     EQQR-EDEHQNKIQPLFQPQPLVYPFAEPIPYTVFPPNAIPLAQPIVVLPFPQPEVKQLP
EleBeta4     EQQR-EDEHQNKIQPLFQPQPLVYPFAEPIPYTVFPPNAIPLAQPIVVLPFPQPEVKQLP
             **** *******************************************************

EleBeta1     EAKEITFPRQKLMSFLKSPVMPFFDPQIPNLGTDLENLHLPLPLLQPLRHQLHQPLAQTP
EleBeta2     EAKEITFPRQKLMSFLKSPVMPFFDPQIPNLGTDLENLHLPLPLLQPLRHQLHQPLAQTP
EleBeta3     EAKEITFPRQKLMSFLKSPVMPFFDPQIPNLGTDLENLHLPLPLLQPLRHQLHQPLAQTP
EleBeta4     EAKETIFPRQKLMSFLKSPVMPFFDPQIPNLGTDLENLHLPLPLLQPLRHQLHQPLAQTP
             ****  ******************************************************

EleBeta1     VLPLPLSLPKVLPVPQQVIPYPQRGRPIQNLQLYEEPLLDPTRKIYPVAQPLAPVYNPVA
EleBeta2     VLPLPLSLPKVLPVPQQVIPYPQRGRPIQNLQLYEEPLLDPTRKIYPVAQPLAPVYNPV-
EleBeta3     VLPLPLSLPKVLPVPQQVIPYPQRGRPIQNLQLYEEPLLDPTRKIYPVAQPLAPVYNPVA
EleBeta4     VLPLPLSLPKVLPVPQQVIPYPQRGRPIQNLQLYEEPLLDPTRKIYPVAQPLAPVYNPVA
             ***********************************************************

EleBeta1     V
EleBeta2     -
EleBeta3     V
EleBeta4     V
```

Figure 3.2. Multiple sequence alignment of the four possible sequences of β-casein in African elephant milk. The putative phosphorylation sites are colored in green. An asterisk (*) indicates positions which have a single, fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties. A period (.) indicates conservation between groups of weakly similar properties. The phosphorylation motif S/T-X-Y is indicated by rectangular blocks.

Because phosphorylation of caseins plays an important role in binding calcium phosphate and consequently casein micelle formation and maintenance, it is of key interest to characterize and obtain detailed knowledge on the phosphorylation profile of β-casein in African elephant. One of the many challenges of characterizing phosphopeptides is their presence in low abundance, requiring enrichment of the sample to circumvent this obstacle (Steen et al., 2002; Parker et al., 2010). In the present study, the phosphorylation pattern of African elephant β-casein was investigated for the first time. The four splice variants mentioned above, were all

taken into account. A series of experimental techniques were employed which include a combination of mass spectrometric (MS) analysis on both a Thermo orbitrap and Sciex triple TOF 6600, 2D electrophoresis (2D PAGE), chromatography and phospho-enrichment. PEAKS DB, a complete software package for proteomics mass spectrometry data analysis was also utilized. These experimental and computational proteomic strategies have been used successfully in the past for the characterization of phosphorylation patterns of β-caseins in bovine, human and horse milk (Girardet et al., 2006; Poth et al., 2008; Li et al., 2012).

3.2. Materials and methods

3.2.1. Sample preparation

African elephant milk samples were obtained from Knysna Elephant Ranch, Knysna, South Africa. Immediately after milking, milk samples were kept and transported on ice for 8 hours before storage at -20 °C until used. Frozen milk samples were thawed in a water bath at 39 °C and skimmed by centrifugation (9000 × g at 20 °C for 15 seconds) Protein precipitation was done with acetic acid. Briefly, 600 µl of skimmed African elephant milk was diluted with an equal volume of water and vortexed for 5 seconds. 30 µl of 10 % acetic acid (Merck, South Africa) and 1N acetic acid (Merck, South Africa) were added to the diluted milk in succession with 20 minutes of incubation at room temperature allowed in between the acetic acid additions. The mixture was centrifuged at 10 000 rpm in an Eppendorf centrifuge (Bio-Rad, South Africa) for 5 minutes. After centrifugation the casein containing pellet was freeze dried and stored until use.

3.2.2. RP-HPLC fractionation

RP-HPLC analysis was carried out on a Shimadzu Prominence instrument (Shimadzu, Johannesburg, South Africa) fitted with a dual wavelength UV/Vis detector and a Phenominex Jupiter C18 analytical column (250 x 4.6 mm, 5 μm, Separations, Johannesburg, South Africa). Column temperature was set to 40 °C and 80 μg of sample was loaded on the column. Elution was performed with a water/acetonitrile containing 0.1 % trifluoro acetic acid gradient starting with 30 % acetonitrile and increasing to 41 % over 55 minutes. The column was re-equilibrated for 10 minutes at 30 % acetonitrile between runs. The flow rate was 1.1 ml/min and UV detection carried out at 214 nm.

3.2.3. Electrophoresis separation

SDS PAGE analysis was carried out according to Laemmli (1970) with a Mini-Protein II dual slab system (Bio-Rad Laboratories, South Africa). The stacking gel composed of 4 % polyacrylamide in 1 M Tris-HCl buffer (pH 6.8) in the presence of 10 % SDS and the resolving gel composed of 12 % polyacrylamide in 1.5 M Tris-HCl buffer (pH 8.8) in the presence of 10 % SDS. HPLC protein fraction samples (0.9 mg/ml) were solubilized in 65.8 mM Tris-HCl buffer (pH6.8) in the presence of 2.1 % SDS, 355 mM β-mercaptoethanol, 26.3 % glycerol, and 0.01 % bromophenol blue. Proteins were stained with coomassie brilliant blue (CBB) R250. Urea PAGE was performed with a Mini-Protein II dual slab system (Bio-Rad Laboratories, South Africa) according to a method by Creamer (1991) with some modifications (Arzu Kavaz, 2012). Beta casein samples (0.9 mg/ml) were solubilised in 62 mM Tris-HCl buffer, pH 7.6, containing 8M urea, 0.2 % β-mercaptoethanol, and 0.05 % bromophenol

blue. A sample volume of 20 µl was loaded onto the gel. Proteins were stained with CBB R250. 2D PAGE was carried out according to the method by O'Farrell (1975). The first dimension separation (isoelectric focusing) was performed on 7 cm immobilized pH gradient (4-7) in a PROTEAN i12 IEF system (Bio-Rad, Johannesburg, South Africa). The following IEF program: 250 V for 20 minutes, 4,000 V for 2 hours, and linear gradient to 10,000 V was applied. The second dimension separation (SDS PAGE) was performed on a 12 % (wt/vol) acrylamide:bis-acrylamide [19:1 (wt/wt)]. Proteins were stained with CBB R250.

3.2.4. Enzymatic dephosphorylation

Dephosphorylation of protein sample by alkaline phosphatase was carried out according to a modified method by Atar et al (2003). 1.2 U of bovine intestinal alkaline phosphatase (EC 3.1.3.1) (Sigma-Aldrich, South Africa) and 4 µl of 10X dephosphorylation buffer (50 mM/l Tris, 100 mM/l NaCl, 10 mM/l $MgCl_2$, and 1 mM/l DTT; pH 7.9) were added to 10 µl of β-casein (0.9 mg/ml) substrate and incubated for 3 hours at 30 °C (1 U of alkaline phosphatase hydrolyzes 1 nM of p-nitrophenylphosphate per minute at 30 °C; pH 8.5). Reactions were stopped by adding 4 µl of 1X sample buffer (62 Mm Tris, 8 M Urea, 0.4 % HCl v/v, 0.2 % 2-mercaptoethanol and 0.15 % w/v bromophenol blue in 10 % EtOH) and boiling for 2 minutes.

3.2.5. LC MS/MS (Orbitrap) analysis of 2D gel spots

Individual protein spots on 2D gels were identified by excising from the gel and subsequently subjected to LC MS/MS by using the method described by (Piersma et al., 2013). Mass spectrometry analysis was performed on an Orbitrap (XLS, Thermo Scientific, Bremen, Germany) instrument according to the method by Marx et al., (2013). In the latter method, data were acquired using the Xcalibur Software package (Thermo Scientific). The precursor ion scan MS spectra (m/z 400–2,000) were acquired at a resolution of R = 60,000 with the number of accumulated ions being 1 x 10^6. The 20 most intense ions were identified and fragmented by collision-induced dissociation (CID) in the linear ion trap (number of accumulated ions 1.5 x 10^4). The lock mass option (polydimethylcyclosiloxane; m/z 445.120025) enabled accurate mass measurement in the MS mode. Thermo Proteome Discoverer 1.3 (Thermo Scientific) was used to identify proteins by automated database searching, using the Mascot search engine (Matrix Science, London, UK) of all tandem mass spectra against the Swiss-Prot 57.15, National Center for Biotechnology Information (NCBI) mammalian and Uniprot elephant databases (version 119, March 6, 2013). Carbamidomethyl cysteine was set as fixed modification, and oxidized methionine, *N*-acetylation, and deamidation (NQ) were selected as variable modifications. The precursor and fragment mass tolerance were set to 10 mg/kg and 0.8 Da, respectively, and 2 missed tryptic cleavages were allowed. Proteins were considered as positively identified if at least 2 tryptic peptides per protein could be matched, and at a Mascot significance threshold of $P < 0.05$.

3.2.6. In-liquid digestion and phospho-enrichment

In liquid digestion was done with both trypsin and chymotrypsin. Freeze dried African elephant β-casein was re-suspended in 50 mM ammonium bicarbonate (NH_4HCO_3)

and 5 M urea to a final concentration of 10 mg/ml. For reduction, 25 µl of the re-suspended sample solution was reduced by adding 1 M dithiothreitol (DTT) to a final concentration of 20 mM and incubated at 37 °C for 1 hour. Following reduction, alkylation of sample was done by adding 1 M Iodoacetamide to a final concentration of 60 mM and incubated in the dark room for 30 minutes. Digestion of the sample with trypsin and chymotrypsin was done by first diluting the sample 5X with 50 mM NH_4HCO_3. Samples were transferred to LoBind tubes, for trypsin digestion 20 µl of ReSyn trypsin was added to a sample concentration of 15 mg/ml whereas for chymotrypsin the sample concentration per 20 µl ReSyn Chymotrypsin was 10 mg/ml. The sample was washed with NH_4HCO_3 and the supernatant was removed. NH_4HCO_3 buffer was added and the solution was mixed at 70 rpm in an Intellimixer set at 37 °C for 1 hour (trypsin digestion) and 25 °C for 30 minutes (chymotrypsin digestion). To each digest, 10 µl of formic acid to a final volume of 1 % was added and mixed gently. The supernatant was transferred to fresh 0.5 ml LoBind tubes and analysed with SDS-PAGE before storage at -20 °C until analysis.

Phosphopeptide enrichment was performed with a KingFisher Flex magnetic particle-processing robot. Deep-well 96-well plates were assigned to each of the eight carousel positions. Individual positions were loaded with the following: in position 1 the elute (sample in 1.25 M NH_4OH); in position 2 the wash buffer 2 (sample in 10 % acetonitrile (ACN), 0.2 % trifluoroacetic acid (TFA)); in position 3 the wash buffer 1 (80 % ACN, 1 % TFA); in position 4 the equilibration buffer (80 % ACN, 5 % TFA, 1 M GA); in position 5 the phosphopeptide binding (40 µl digest, 160 µl 80% ACN, 5% TFA, 1M glycolic acid (GA)); in position 6 the equilibration (80 % ACN, 5 % TFA, 1M GA) and in position 7 the microspheres (2.5 µl beads, 197.5 µl

80 % ACN/ 5 % TFA/1 M GA). The 8th position was left empty. A volume of 500 μl of the relevant buffer was added to each well, except for the sample binding and elution steps, where only 200 μl of sample and elution buffer were used. Following each enrichment cycle, phosphopeptide fractions were collected in 0.5 ml LoBud tubes and freeze dried. Samples were resuspended in 50 μl 0.1 % TFA, vortexed for 30 seconds, centrifuged for 10 minutes at 14 000 rpm and transferred into 0.5 ml LoBind tubes prior to analysis by LC-MS/MS.

3.2.7. LC MS/MS (Triple TOF) analysis of phospho-enriched peptides

The stock sample (10 mg/ml) was diluted to 1 mg/ml using 50 mM ammonium bicarbonate. Following dilution, the mixture was centrifuged at 14 000 rpm for 10 minutes and transferred to 0.5 ml LoBind eppendorf. For intact LCMS analysis, 1 ul of the sample was loaded, on a Phenomenex C4 column (1 mm x 150 mm) at 0.2 ml/min. The sample protein was eluted using linear acetonitrile gradient (10-90 %) of solution B. Solution A constituted of 0.1 % formic acid (FA) and solution B constituted of ACN in 0.1 % FA. The TOF MS spectra was set in the range between 700 and 2000 m/z and collected using an AB Sciex 6600 TripleTOF mass spectrometer with TurboIon source installed. The multiply charged series were deconvoluted by the Bayesian Protein Reconstruct tool of BiopharmaView 1.0 using a mass range of 10-40 kD and signal to noise threshold of 5.

3.2.8. Mass spectrometry data analysis

Following MS analysis, the data obtained was imported into and analysed using PEAKS DB software. The software identifies peptides from a sequence database with MS/MS data (Zhang et al., 2012). The sequence database consisted of all casein sequences as well as a large number of contaminant sequences. The main steps of the PEAKS DB software proceed as follows: Initially *de novo* sequencing is performed for each input spectrum. This is followed by protein short listing where *de novo* sequencing tags are used to find approximate matches in the protein sequence database. Following protein short listing, peptide short listing is done, which involves matching peptides of the protein short listing to the MS/MS spectra. A precise scoring function is then used to score the top 512 peptides from the peptide short listing. Finally, the high confidence peptides identified through the above steps are used to infer the proteins.

In the case of post translational modifications (PTMs), PEAKS uses two methods for the localization of phosphorylation modification, the minimal ion intensity (≥5%) and the minimal Ascore (>20). Both methods are a measure of the likelihood that modification occurs where it is reported instead of other possibilities. The minimal ion intensity method requires that fragmentation before and after the modified amino acid residue must be validated by major fragment ions with at least 5% ion intensity. The minimal Ascore on the other hand is a probability-based score that measures the probability of correct modification based on the presence and intensity of site-determining ions in the MS/MS .An Ascore of 20 (P = 0.01) should result in the site being localized with 99 % certainty (Beausoleil et al., 2006).

3.3. Results

3.3.1. Determination of β-casein sequence

African elephant milk protein precipitate was separated by C18 RP-HPLC. From previous work by Madende et al (2015), it had been observed that African elephant milk is highly abundant in β-casein and therefore the largest chromatogram peak (Figure 3.3) from the casein mixture was collected. The identity and purity of the recovered fraction was assessed with SDS-PAGE (Figure 3.3 insert) and MS analysis confirmed the identity of the single band to be African elephant β-casein.

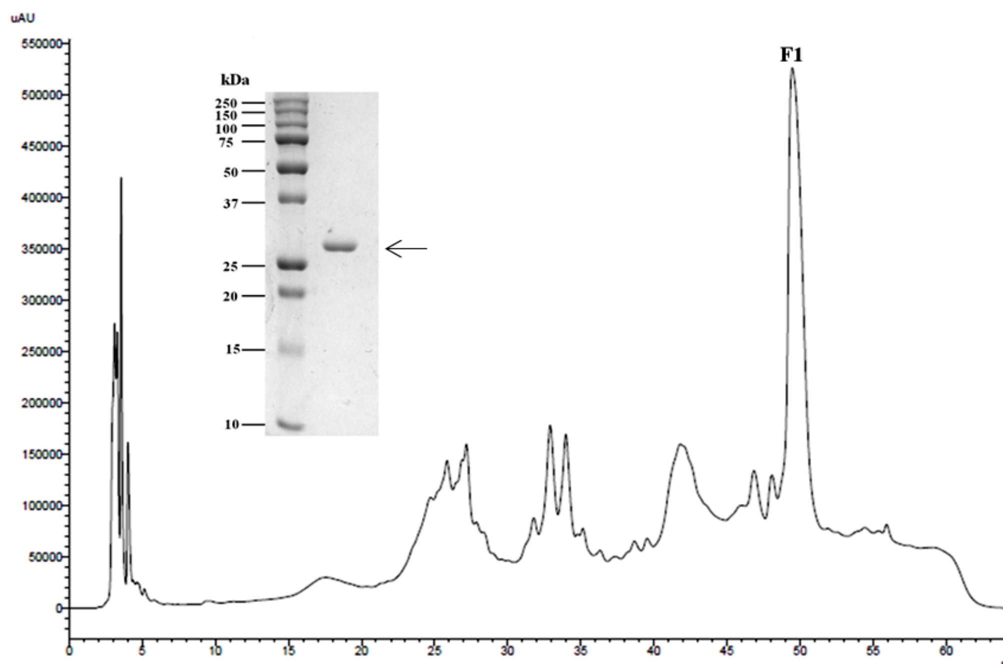

Figure 3.3. RP-HPLC fractionation of African elephant milk protein precipitate. Pure β-casein was recovered from the fraction labeled F1. A single band was observed (arrow) on SDS PAGE (insert) at ~28 kDa showing that it is pure African elephant β-casein.

3.3.2. Determination of β-casein phosphorylation

The presence of African elephant β-casein isoforms was investigated by Urea PAGE and 2D PAGE. Two dimensional separation of pure African elephant β-casein revealed up to five spots (Figure 3.4). The five spots had different isoelectric points and MS analysis confirmed that they were all African elephant β-casein.

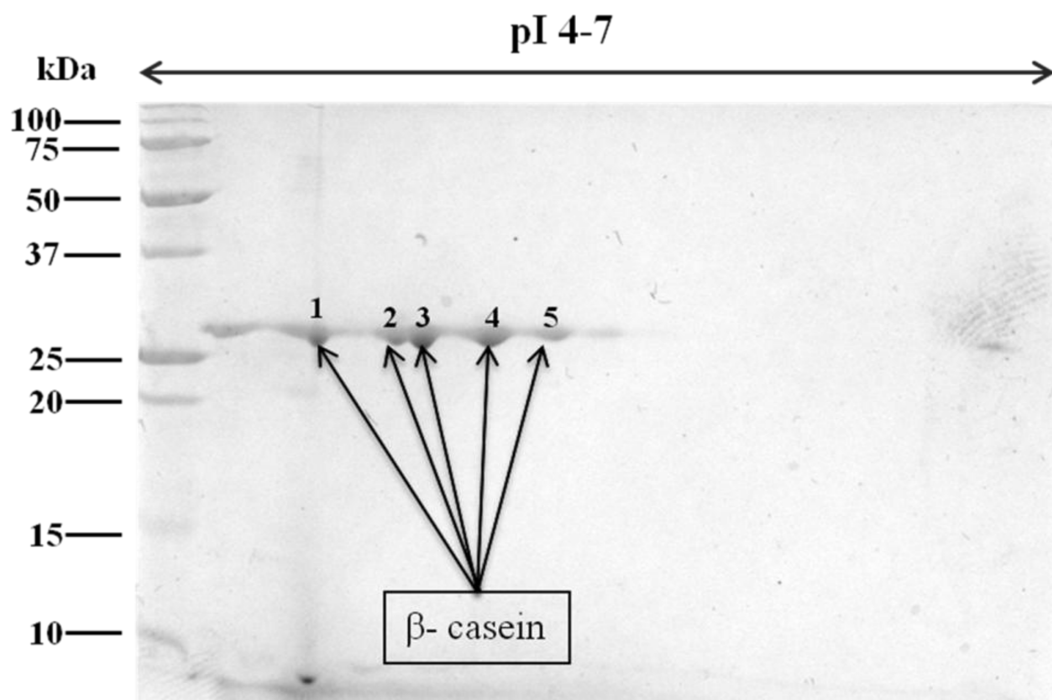

Figure 3.4. 2D PAGE separation of African elephant β-casein. Up to 5 spots can be visualized on the gel.

Analysis of pure African elephant β-casein by Urea PAGE showed four separate bands (Figure 3.5). Urea PAGE is most commonly used in the separation of caseins due to its high resolving capability compared to SDS PAGE and can distinguish between casein isoforms (Creamer, 1991; McSweeney and Fox, 1997). The bands

differed in intensity; two of the bands appeared much larger and darker whereas the other two appeared light and smaller. To confirm the type of post translational modification of African elephant β-casein isoforms, a dephosphorylation step with alkaline phosphatase was undertaken. Figure 3.5 shows a comparison of Urea PAGE separation of native and dephosphorylated β-casein. The protein bands profile of native and dephosphorylated β-casein Urea PAGE profiles were different.

Compared to its native counterpart, dephosphorylated β-casein lower bands appeared to be reduced in intensity whereas an upper band with shorter migration increased in intensity. It was also observed that a new protein band (band 0), which was not initially present in the native β-casein, can now be observed after the dephosphorylation step. This may be a deamidation effect which occurs more readily under alkaline conditions, resulting in an mass increase of proteins (Girardet et al., 2006). Given that alkaline phosphatase was used for the dephosphorylation step, its optimum catalytic conditions are also conducive for deamidation.

Deamidation of caseins involves the release of the amide functional group in the side chain of asparagine or glutamine converting it to aspartic acid and glutamic acid respectively. Deamidation is often measured by the release of ammonia and can result in generation of isoforms (Girardet et al., 2006).

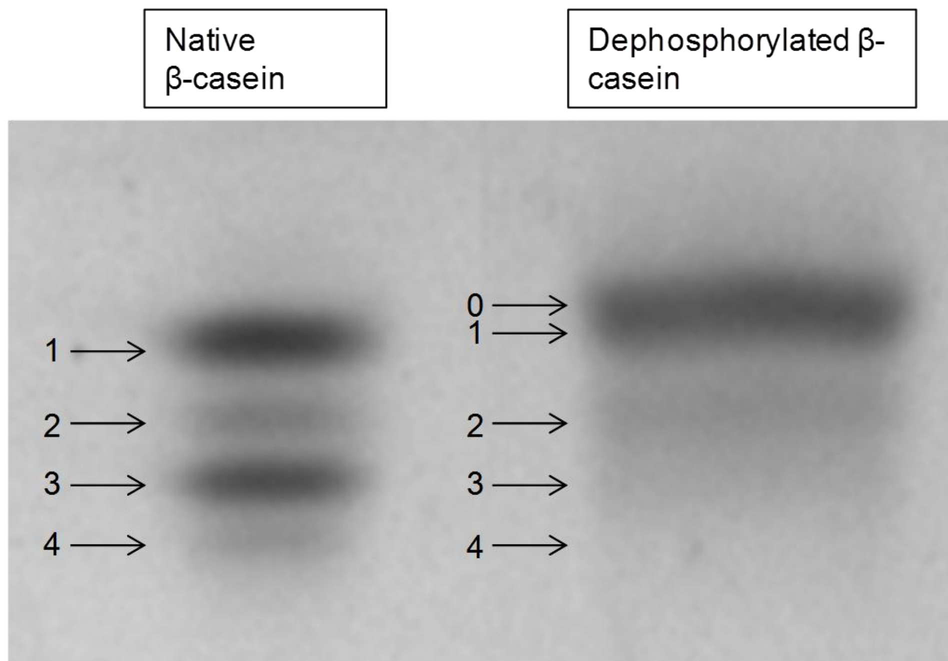

Figure 3.5. Urea-PAGE separation of native and dephosphorylated African elephant β-casein. The arrows indicate the bands of interest.

The exact amino acid sequence (splice variant) of African elephant β-casein in the African elephant milk sample, utilized in this study, was determined by evaluating the molecular mass of the intact β-casein via LC MS. The deconvoluted mass spectrum of the intact β-casein is shown in Figure 3.6. Two main peaks were observed with average masses of 22895.11Da and 22974.85Da. Interestingly, none of the intact protein masses matched the average molecular masses of sequences 1-4 in Figure 1.1 which were calculated without the signal peptide (first 15 amino acid residues). The average molecular masses were as follows: sequence 1 (23699.57Da), sequence 2 (24687.63Da), sequence 3 (25652.75Da) and sequence 4 (23781.78Da). Caseins are often subjected to exon skipping (Martin et al., 2013), meaning that there was a possibility of another splice variant. Several possibilities of exon skipping were tested, working from sequence 4 (Madende et al., 2015) which lacks exon 4 (TVENLSSSEIRQFYSE), exon 5 (ESVTQVNK) was deleted.

This resulted in a short length β-casein splice variant named sequence 5. Figure 3.7 shows a schematic representation of how sequence 5 was obtained. Sequence 5 is 200 amino acid long and has an average molecular mass of 22895.79Da, which matches the first peak of the intact protein analysis (Figure 3.6). The molecular masses of the sequences are calculated with the exclusion of the signal peptide sequence (exon 1), since it does not form part of the mature protein (Ginger and Grigor, 1999). The first and second peaks on the intact protein analysis differ by 79.74Da which is indicative of phosphorylation by a single phosphate (1P), thus suggesting that sequence 5 may be singly phosphorylated.

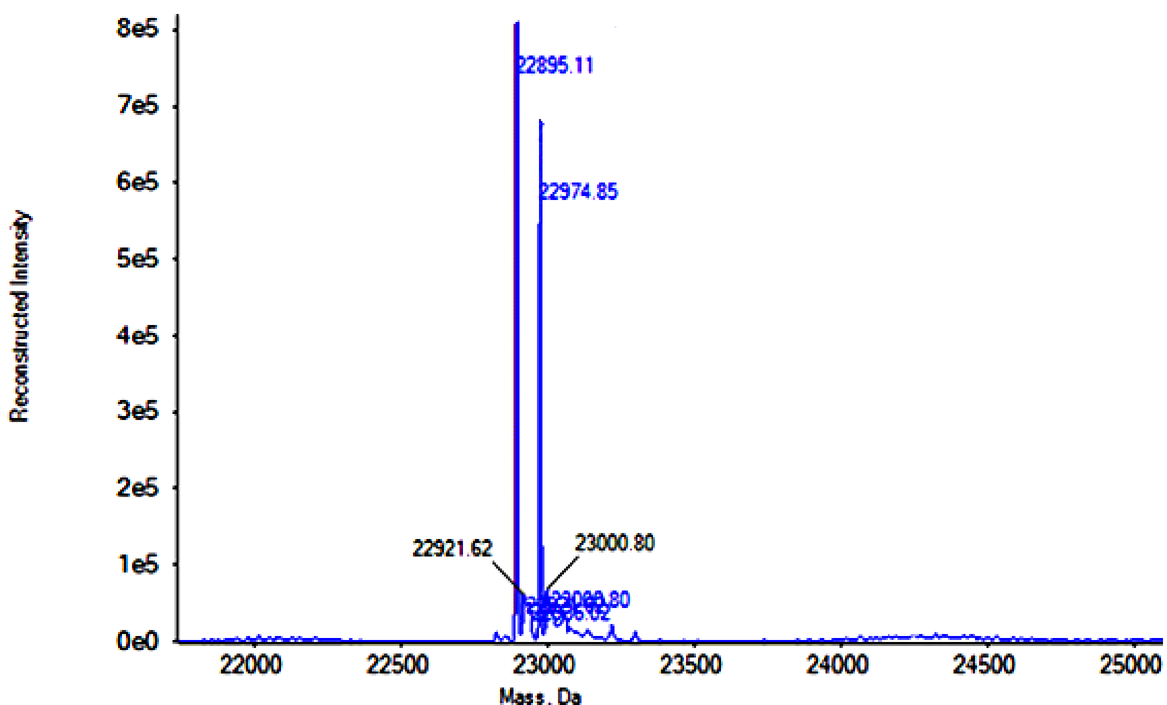

Figure 3.6. Reconstructed mass spectrum of African elephant β-casein intact protein.

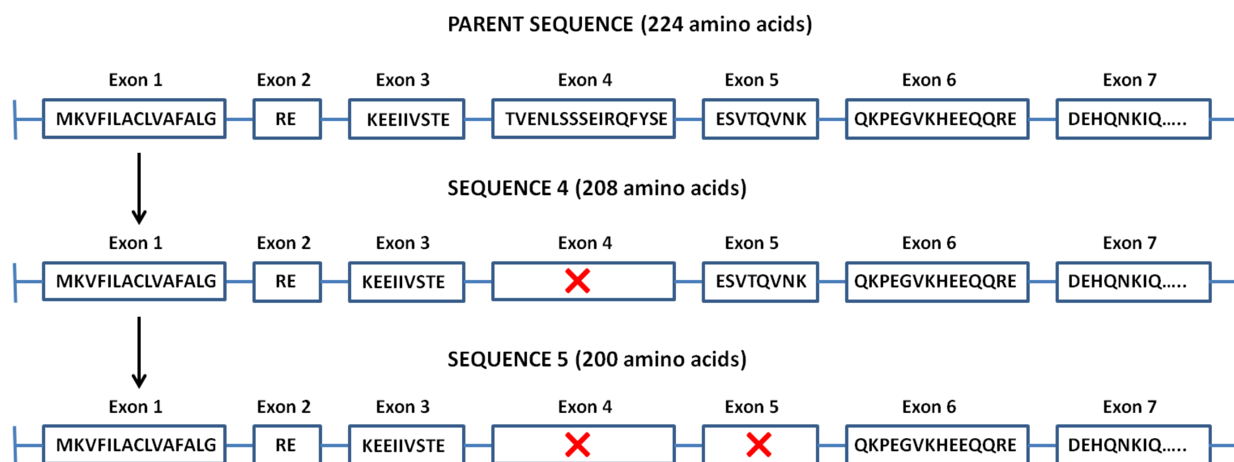

Figure 3.7. A schematic representation of the steps followed in the construction of sequence 5 splice variant. The average molecular weights of the sequences excluding exon 1 (signal peptide) are as follows: Parent sequence (25652.75Da), sequence 4 (23781.76Da) and sequence 5 (22895.79Da). The deleted exons 4 and 5 in sequences 4 and 5 are depicted by red crosses.

Since the intact protein mass pointed towards sequence 5 as the sequence in the sample, the next step was to localize the exact site of phosphorylation by analysis of tryptic peptides of African elephant β-casein by TOF MS/MS. The MS/MS data was analysed in PEAKS. Both the minimal Ascore and minimal ion intensity methods were used to localize phosphorylation modification sites. The Ascore is a probability based score that measures the probability of correct phosphorylation site localization based on the presence and intensity of site-determining ions in MS/MS spectra with a certainty of 99 %. This approach maximizes data sensitivity by efficiently distracting incorrect peptide spectral matches (Beausoleil et al., 2006). Using this approach, five phosphorylation sites were identified with confidence at the maximum Ascore of 500. The phosphorylation pattern of sequence 5 is shown in Figure 3.8a. Phosphorylation of sequence 5 is localized at Ser 9, Thr 10, Thr 138, Ser 146 and Thr 181. The phosphorylation sites are located at both the N- and C-terminus.

Because the minimal Ascore uses the assumption that a modification (phosphorylation) assigned to a peptide is present and assigns a default Ascore value of 1000 to it, it was imperative that we manually process and refine the data taking into account the known phosphorylation motifs of β-casein (Ser/Thr-X-Y where X represents any amino acid whereas Y represents an acidic amino acid) as well as its location. In this way, Thr 10, Thr 138, Ser 146 and Thr 181 phosphorylation sites were all eliminated. The remaining phosphorylation site (Ser 9 is depicted in Figure 3.8b.

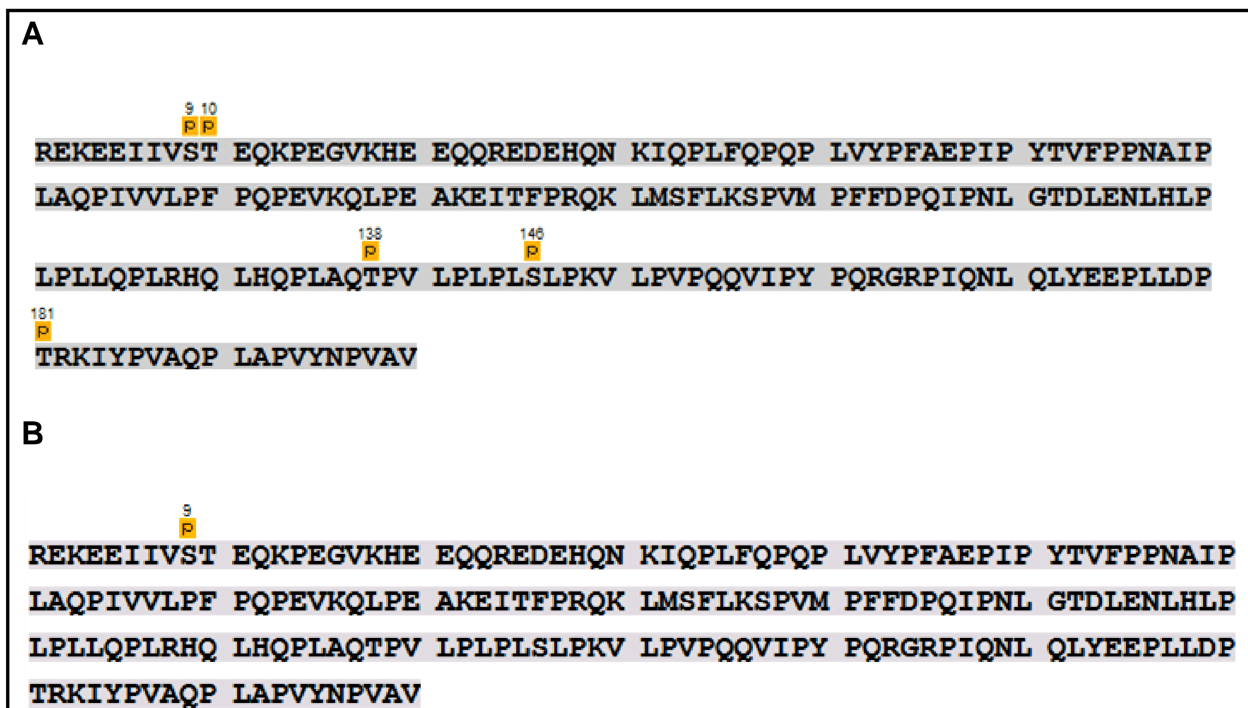

Figure 3.8. Sequence 5 showing position of phosphorylation without the signal peptide. (A) The numbered phosphorylation sites positioned (orange icons) above the protein sequence indicate the sequence position where phosphorylation was identified with high confidence in PEAKS using the minimal Ascore method. (B).The numbered phosphorylation site (orange icon) positioned above the protein sequence indicates the sequence position where phosphorylation was identified with high confidence after manually screening the identified phosphorylation sites.

Alternatively, the minimum fragment ion intensity method was also used to localize phosphorylation modification sites of sequence 5. This method requires that fragmentation before and after the modified amino acid has to be validated by major fragment ions. A single phosphorylation site was identified with confidence at the minimum relative ion intensity of 5 % in CID as indicated by a pair of fragmentation y ions (y_{16}(-98) and y_{17}(-98)) in Figure 3.8. The phosphorylation site is located along the sequence [1]REKEEIIV**s**TEQKP[14], where the amino acid residue that is phosphorylated is represented in bold and lower case. The minimal ion intensity data was similar to the minimal Ascore data (after manually screening phosphorylation sites) and therefore confirmed the position of phosphorylation on Ser 9. The MS/MS spectrum of the modified residue is shown in Figure 3.9.

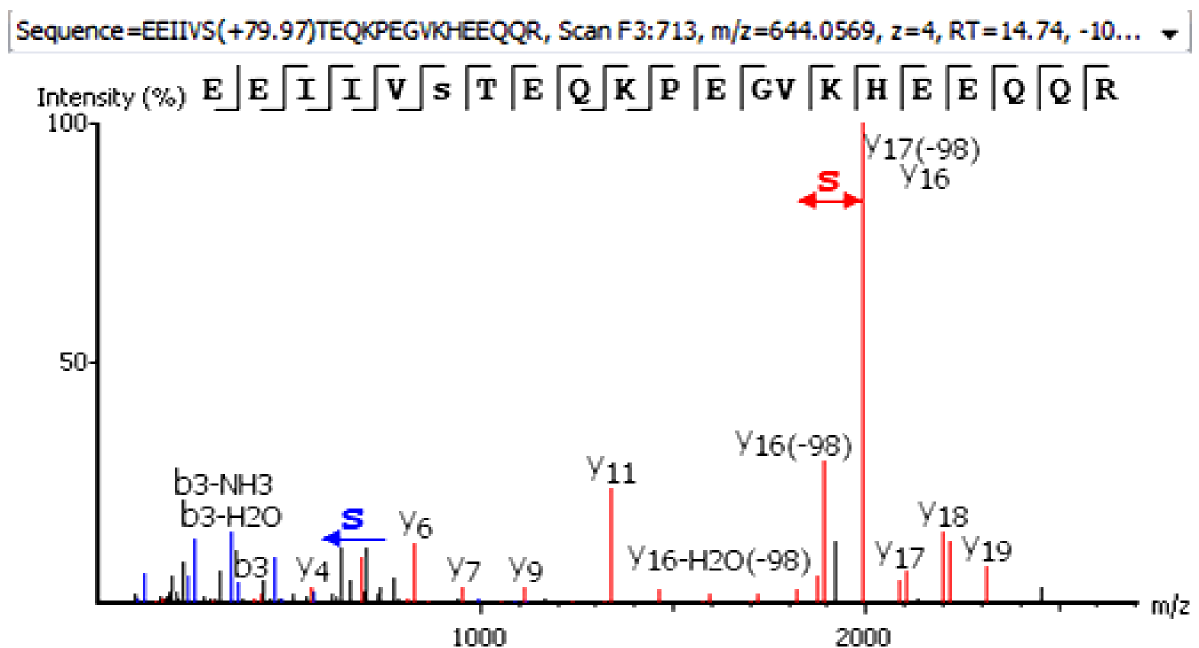

Figure 3.9. MS/MS spectrum of phosphorylated sequence [4]EEIIV**s**TEQKPEGVKHEEQQ[22]. Fragmentation occurs before and after the serine residue as indicated by the y_{16}(-98) and y_{17}(-98) ions respectively. A loss of a doubly charged phosphate (2[H_3PO_4]) in the MS/MS spectrum is represented by -98. This indicates confident localization of a phosphorylation modification at 5 % minimal ion intensity.

3.4. Discussion

The intact protein analyzed via LC-MS shows the presence of two main peaks that have molecular weights of 22895.11Da and 22974.85Da, with a mass difference of approximately 80Da (indicative of phosphorylation modification). Moreover, it was determined by average molecular weight calculation that the difference between these peaks correspond to the non-phosphorylated as well as the singly phosphorylated African elephant β-casein forms. Sequence 4 which is 208 amino acids long (excluding the signal peptide) and has a molecular weight of 23781.76Da was used to elucidate the sequence of a possible fifth sequence. Exon skipping is a well known occurrence in caseins (Martin et al., 2013), as a result, it is possible that sequence 4 (Madende et al., 2015) has some exons deleted. When exon 5, which codes for the peptide sequence ESVTQVNK, is removed (Figure 3.7), the result is a short length 200 amino acid long protein with an average molecular mass of 22895.79Da. This molecular mass corresponded to that observed for the first MS peak of the intact protein, being 22895.11Da (Figure 3.6). In the same way, the second peak of Figure 3.5 therefore corresponded to the singly phosphorylated (+80Da) form of African elephant β-casein. It would be interesting to find out whether a “short length” sequence of β-casein, in addition to its full length counterpart is present in African elephant milk. This phenomenon was reported for equine milk where a short length multi-phosphorylated β-casein (94 amino acid residues) together with the full length variant was observed. In that case, an internal truncation within exon 7 of the full length homolog (226 amino acid residues) occured, resulting in a short length variant (Miclo et al., 2007).

It was possible to show from the present study that native African elephant β-casein exists in multi-phosphorylated isoforms. This was initially seen from multiple bands on Urea PAGE and multiple spots on 2D PAGE separations of pure African elephant β-casein. Interestingly, there were five separate and distinct spots on 2D PAGE (Figure 3.3), whereas Urea PAGE (Figure 3.4) only showed four bands. The differences in the number of isoforms that were visualized on the gels could be a result of the differences in the sensitivity of the two methods. Nevertheless, it was shown that African elephant β-casein has up to five isoforms according to electrophoresis separation of the β-casein purified by RP-HPLC.

The natural existence of β-casein in the form of several multi-phosphorylated isoforms has been previously demonstrated in other mammals such as human (Poth et al., 2008) and equine (Girardet et al., 2006). Calcium sensitive caseins such as β-casein are notorious for phosphorylation during post translation modification. To determine if this was also the case with African elephant β-casein, native African elephant β-casein was enzymatically dephosphorylated with alkaline phosphatase before Urea PAGE separation. After dephosphorylation, the lower protein bands of African elephant β-casein on the electrophoretogram disappeared or were reduced. This confirmed that the protein was phosphorylated.

In light of the above, phosphorylation modification was targeted for during MS analysis, where multiple experiments were done on different MS instruments. Phosphorylation localization of African elephant β-casein showed that phosphorylation is localized at Ser 9, Thr 10, Thr 138, Ser 146 and Thr 181 using the

Ascore method. This data explains the electrophoresis results where up to 5 isoforms of African elephant β-casein could be identified. The 5 spots on the 2D PAGE gel (Figure 3.4) may represent 1P, 2P, 3P, 4P and 5P phosphoforms of African elephant β-casein. However, the intact MS analysis does not support these isoforms. Moreover, phosphorylation of β-caseins of most mammals (cow, goat, sheep, horse, human, mouse, rat etc) usually occur at a string of serine residues near the N-terminus according to the motif Ser/Thr-X-Y (X represents any amino acid whereas Y represents an acidic amino acid) (Ginger and Grigor, 1999). In contrast, the phosphorylation sites (specifically Thr 138, Ser 146 and Thr 181), do not comply with this and may therefore not be considered as correct phosphorylation sites of African elephant β-casein.

As a result, the localized phosphorylation sites were further processed manually by removing peptide entities that showed singular success which skew the scoring assignment and instead focusing on the peptides that had multiple phosphorylation assignments of the same amino acid. Through an elimination process, only Ser 9 was deemed a more accurate phosphorylation site. Phosphorylation of Ser 9 was further confirmed by the minimal ion intensity method. The presence of fragmentation ions before and after the Ser 9 (Figure 9) in the form of y_{16}(-98) and y_{17}(-98) is an indication that Ser 9 is modified by phosphate group. In PEAKS, a minimum of 5 % ion intensity is acceptable for confident modification site localization, the fragment ion intensity observed for Ser 9 phosphorylation modification was 26 %. Apart from Ser 9, none of the modification sites suggested by the minimal Ascore methods had fragmentation ions above 5 %.

The single phosphorylation site accounts for the second peak of the intact MS analysis that has a molecular weight of 22974.85Da. In this way, the two peaks in Figure 3.7 can be interpreted as representing 0P and 1P forms of African elephant β-casein. Although the MS data reveals only a single phosphorylation site for African elephant β-casein, electrophoresis evidence cannot be completely disregarded. From the data of 2D PAGE of whole African elephant milk several isoforms of β-casein were detected (Madende et al. 2015).

It may further be derived that only two of these occur in large quantity, which is confirmed by Urea PAGE separation of purified African elephant β-casein (Figure 3.5). Excision of protein spots from 2D gels proved to provide too little material for the phosphorylation studies. Although the RP-HPLC purified protein contained all the isoforms, only the unphosphorylated and singly phosphorylated were detected (Figure 3.7). Further studies are therefore required to identify other phosphorylated isoforms of African elephant β-casein

3.5. Conclusions

The characterization of African elephant β-casein by gel electrophoresis show up to five phosphoforms whereas the MS analysis reveals confident phosphorylation at a single site (Ser 9). MS studies reveal the presence of a short length African elephant β-casein that is 200 amino acid long and has exon 4 and exon 5 deleted. Other phosphoforms of this protein may have evaded detection due to their presence in

small quantities. Elucidation of African elephant β-casein phosphorylation state also sheds light on the casein micelle structure and calcium binding in African elephant milk. While the large size of African elephant casein micelle observed by FE-SEM can be attributed to the abundance of β-casein with consequently limited formation of calcium phosphate nanoclusters, may also play a major role.

Since all caseins have a role to play in casein micelle formation and that protein function is linked to its structure, the next chapter deals with structure modeling of caseins. Caseins can not be crystallized for X-ray crystollography analysis in order to directly determine their structure, homology modeling can thus be used as an alternative to predict the structure of caseins. Such models can be useful in explaining the structural aspects of the casein micelle. As such, the following chapter deals with structure modeling of caseins.

3.6. REFERENCES

Arzu Kavaz. 2012. Comparison of quality characteristics of Çökelek and Lor cheeses. *African J Biotechnol*. 11. doi:10.5897/AJB11.4230.

Atar, D., L.H. Madsen, R. Labugger, and J.E. VanEyk. 2003. Molecular modifications of troponin I and T detected in serum from patients with acute myocardial infarction. *Ugeskr Laeger*. 165:116–120. doi:10.1161/01.CIR.102.11.1221.

Beausoleil, S.A., J. Villén, S.A. Gerber, J. Rush, and S.P. Gygi. 2006. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*. 24:1285–1292. doi:10.1038/nbt1240.

Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrell, V.R. Harwalkar, R. Jenness, and R.M. Whitney. 1984. Nomenclature of proteins of cow's milk-fifth revision. *J Dairy Sci*. 67:1599–1631. doi:10.3168/jds.S0022-0302(84)81485-X.

Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk-sixth revision. *J Dairy Sci*. 87:1641–74. doi:10.3168/jds.S0022-0302(04)73319-6.

Ginger, M.R., and M.R. Grigor. 1999. Comparative aspects of milk caseins. *Comp Biochem Physiol Part B*. 124:133–145.

Girardet, J.M., L. Miclo, S. Florent, D. Molle, and J.L. Gaillard. 2006. Determination of the phosphorylation level and deamidation susceptibility of equine β-casein. *Proteomics*. 6:3707–3717. doi:10.1002/pmic.200500728.

Greenberg, R., M.L. Groves, and H.J. Dower. 1984. Human beta-casein. Amino acid sequence and identification of phosphorylation sites. *J Biol Chem*. 259:5132–8.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227:680–685. doi:10.1038/227680a0.

Li, S. shan, J. qi Wang, H. yang Wei, Y. xin Yang, D. pan Bu, L. ying Zhang, and L. yun Zhou. 2012. Identification of bovine casein phosphorylation using titanium dioxide enrichment in combination with nano electrospray ionization tandem

mass spectrometry. *J Integr Agric*. 11:439–445. doi:10.1016/S2095-3119(12)60029-X.

Madende, M., G. Osthoff, H.-G. Patterton, H.E. Patterton, P. Martin, and D.J. Opperman. 2015. Characterization of casein and alpha lactalbumin of African elephant (Loxodonta africana) milk. *J Dairy Sci*. 98:8308–18. doi:10.3168/jds.2014-9195.

Mamone, G., S. Caira, G. Garro, A. Nicolai, P. Ferranti, G. Picariello, A. Malorni, L. Chianese, and F. Addeo. 2003. Casein phosphoproteome: Identification of phosphoproteins by combined mass spectrometry and two-dimensional gel electrophoresis. *Electrophoresis*. 24:2824–2837. doi:10.1002/elps.200305545.

Martin, P., C. Cebo, and G. Miranda. 2013. Interspecies comparison of milk proteins: Quantitative variability and molecular diversity. *In* Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects, 4th Edition. 387–429.

Marx, I.J., N. van Wyk, S. Smit, D. Jacobson, M. Viljoen-Bloom, and H. Volschenk. 2013. Comparative secretome analysis of Trichoderma asperellum S4F8 and Trichoderma reesei Rut C30 during solid-state fermentation on sugarcane bagasse. *Biotechnol Biofuels*. 6:172. doi:10.1186/1754-6834-6-172.

McSweeney, P.L.H., and P.F. Fox. 1997. Chemical methods for the characterization of proteolysis in cheese during ripening. *Lait*. 77:41–76. doi:10.1051/lait:199713.

Miclo, L., J. Girardet, A.S. Egito, D. Mollé, P. Martin, and J. Gaillard. 2007. The primary structure of a low- M r multiphosphorylated variant of â -casein in equine milk. *Proteomics*. 7:1327–1335. doi:10.1002/pmic.200600683.

O'Farrell, P.H. 1975. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*. 250:4007–21. doi:10.1016/j.bbi.2008.05.010.

Parker, C.E., V. Mocanu, M. Mocanu, N. Dicheva, and M.R. Warren. 2010. Mass Spectrometry for Post-Translational Modifications.

Piersma, S.R., M.O. Warmoes, M. de Wit, I. de Reus, J.C. Knol, and C.R. Jiménez. 2013. Whole gel processing procedure for GeLC-MS/MS based proteomics. *Proteome Sci*. 11:17. doi:10.1186/1477-5956-11-17.

Poth, A.G., H.C. Deeth, P.F. Alewood, and J.W. Holland. 2008. Analysis of the human casein phosphoproteome by 2-D electrophoresis and MALDI-TOF/TOF MS reveals new phosphoforms. *J Proteome Res*. 7:5017–5027. doi:10.1021/pr800387s.

Sorensen, E.S., L. Moller, M. Vinther, T.E. Petersen, and L.K. Rasmussen. 2003. The phosphorylation pattern of human αs1-casein is markedly different from the ruminant species. *Eur J Biochem*. 270:3651–3655. doi:10.1046/j.1432-1033.2003.03755.x.

Stamm, S. 2008. Regulation of alternative splicing by reversible protein phosphorylation. *J Biol Chem*. 283:1223–7. doi:10.1074/jbc.R700034200.

Steen, H., A. Pandey, J.S. Andersen, and M. Mann. 2002. Analysis of tyrosine phosphorylation sites in signaling molecules by a phosphotyrosine-specific immonium ion scanning method. *Sci STKE*. 2002:pl16. doi:10.1126/stke.2002.154.pl16.

Thingholm, T.E., O.N. Jensen, and M.R. Larsen. 2009. Analytical strategies for phosphoproteomics. *Proteomics*. 9:1451–1468. doi:10.1002/pmic.200800454.

Zhang, J., L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G.A. Lajoie, and B. Ma. 2012. PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics*. 11:M111.010587–M111.010587. doi:10.1074/mcp.M111.010587.

CHAPTER 4

STRUCTURE MODELING OF CASEIN PROTEINS

4.1. INTRODUCTION

Protein structure prediction can be described as the inference of a protein's 3D structure from its amino acid sequence (Al-lazikani et al., 2001). It involves the prediction of a protein's folding, secondary and tertiary structure using the primary structure as the starting point. Progress in protein structure prediction has been largely due to an influx in genome sequencing and improvement in tools thereof. Experimental structure determination remains the best way to determine high resolution protein structures. However, computational structure prediction methods can provide valuable information for those sequences whose protein structure cannot be determined experimentally (Baker and Sali, 2001). Many proteins, including caseins, are either too large for nuclear magnetic resonance spectroscopy or cannot be crystallized for X-ray crystallography thus protein/homology modeling becomes the method of choice for structure determination (Holt and Sawyer, 1988; Krieger et al., 2003). Homology modeling involves structure prediction of proteins based on the structure of a homologous sequence (including sequence identity between two sequences as low as 30 %) (Tramontano, 1998).

The recent advances in homology modeling contribute immensely to understanding the relationship between protein structure and function (Xiang, 2006). The detection of distant homologues, and aligning sequences with template structures added to modeling of loops and side chains have contributed to reliable prediction of protein structure, which was not possible even several years ago. During evolution, protein structure is more stable and changes much slower than the associated sequence (Sander and Schneider, 1991).

Structure prediction of caseins has been attempted in the past (Holt and Sawyer, 1988). In the study referred to above, κ-casein mature peptide was deemed to have very little α-helical structure but several regions of the β-turn. β-casein structure was predicted to compose of α-helical and loop (β-turn conformation) structures. The secondary structure prediction α_{s1}-casein was observed to contain α-helical conformation of the signal peptide and β-turns on the conserved phosphorylation sites. The α_{s2}-casein conformation was predicted to be composed of a combination of β-turns and α-helices. Following the work by (Holt and Sawyer, 1988), more structure modeling studies in the form of molecular modeling on bovine caseins and the casein micelle have since been done (Kumosinski et al., 1991; 1993; 1994; Farrell et al., 2001).

In these studies, the casein 3D structure models were constructed using sequence-based prediction algorithms in conjunction with secondary structural information obtained from Raman spectroscopy. It is important to note that most of the studies on casein homology models have focused specifically on bovine caseins. Moreover,

the predicted structures are relatively old and as such it is imperative to revisit these bovine models using latest and more accurate structure prediction tools.

As such, this chapter aims to determine the structure of cow, sheep, horse, human and African elephant caseins using the latest version of the web based I-TASSER homology modeling tool. Construction of 3D models for structure prediction using I-TASSER has been successfully applied in the past (Zhang et al., 2015; Quan et al., 2016). The predicted casein secondary structure models will be compared to the SEM images of the respective species in order to establish a possible link between secondary structure of caseins and appearance of casein micelles when examined by the microscope. African elephant casein micelles are of particular interest and therefore the relation of the predicted models to its SEM image would be discussed in detail.

4.2. Materials and methods

4.2.1. Homology modeling

The α_{s1}-, α_{s2}-, β- and κ-casein primary amino acid sequences from cow, horse, sheep, African elephant and human were retrieved from Ensembl genome browser to perform homology modeling (http://www.ensembl.org/index.html). For homology modeling of casein proteins, the iterative threading assembly refinement (I-TASSER) protein structure and function prediction online tool was used (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). Casein primary sequences were submitted to the I-TASSER online server, where automated full-length 3D protein structures were generated (Zhang, 2008). Protein structure model generation

turnover was approximately 48 hours. The I-TASSER suite pipeline consists of mainly four steps, threading template identification, iterative structure assembly simulation, model selection and refinement, and structure-based function annotation (Zhang, 2008). LOMETS software program is used to thread the query against a non-redundant structure library to identify templates. Following query-to-template alignments, a full length model is constructed by reassembling the aligned fragments from templates. The unaligned region is built from scratch by *ab initio* folding. Final atomic structure models are constructed from the low-energy conformations by a two-step atomic-level energy minimization approach. Global model correctness is assessed by the confidence score, which is based on the significance of threading alignments and the density of structure clustering; the residue-level local quality of the structural models and B-factor of the target protein are evaluated by ResQ. For function prediction, the structure models with the highest confidence scores are matched against the BioLiP5 database of ligand-protein interactions to detect homologous function templates.

4.3. Results

4.3.1. Alpha caseins

Caseins cannot be crystallized and as a result homology modeling using the amino acid sequence as the starting point is an alternative option to predict a protein's secondary structure. The homology models shown in Figures 4.1–4.17 were all predicted with I-TASSA and analysed in YASARA. For each casein sequence, five different models are predicted; these are ranked according to their quality from number 1 to number 5. The quality of each predicted model is measured by the

confidence score (C-score) which is typically between -5 for the lowest confidence model and 2 for a highest confidence model. However, in some cases although rare, the number 1 ranked model may not have the highest C-score compared to the rest of the models. In the current study, only the highest ranked models were chosen and presented for structure comparison. Figures 4.1-4.4 represent homology models of cow, sheep, horse and human α_{s1}-caseins. The C-scores of the highest ranked α_{s1}-casein homology models were as follows: cow (-3), sheep (-2.16), horse (-2.88) and human (3.39). The cow α_{s1}-casein model is predominantly alpha helices distributed throughout the model length. Additionally, the structure also contains two smaller beta strands sections located near the N-terminus whereas the rest of the structure is composed of turns and random coils. Longer lengths of turns are located at the C-terminus. The rest of the α_{s1}-casein homology models are structurally similar to cow α_{s1}-casein models. The structures are all dominated by longer lengths of α-helices but unlike the cow α_{s1}-casein model, they lack any beta strand conformation.

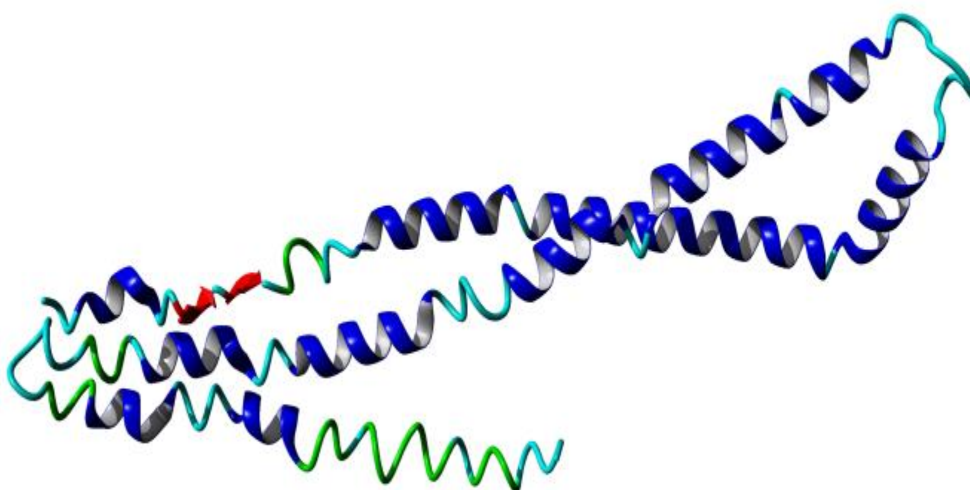

Figure 4.1. Homology model of cow α_{s1}-casein with a C-score of -3. The N-terminus end is located at the upper chain whereas the C-terminus end is located at the lower chain of the

model. The colour codes for the secondary structure elements are as follows: α-helices are blue, β-strands are red, turns are green and random coils are cyan.

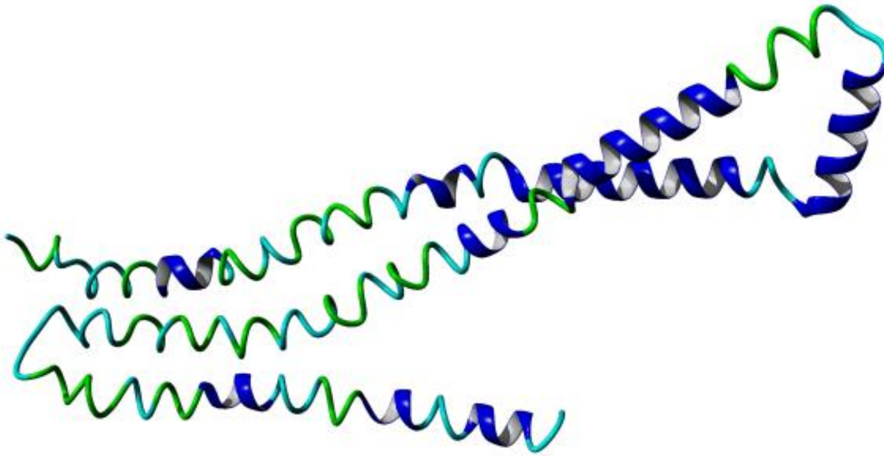

Figure 4.2. Homology model of sheep α_{s1}-casein with a C-score of -2.16. The N-terminus end is located at the upper chain whereas the C-terminus end is located at the lower chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

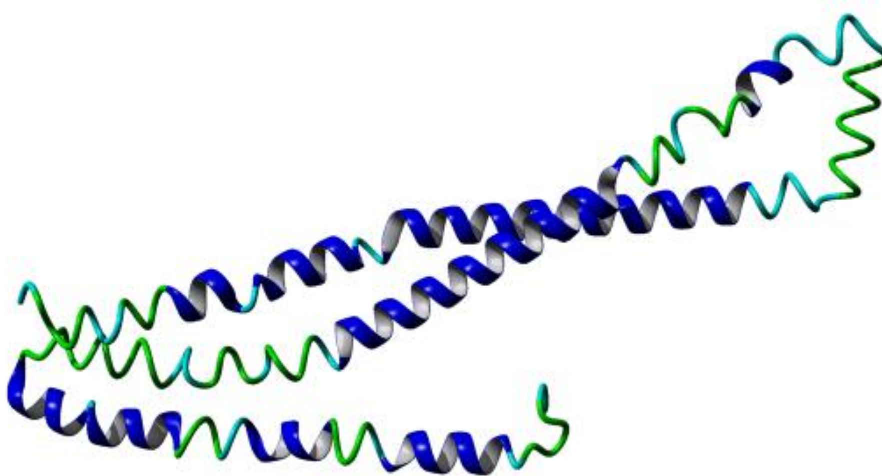

Figure 4.3. Homology model of horse α_{s1}-casein with a C-score of -2.88. The N-terminus end is located at the upper chain whereas the C-terminus end is located at the lower chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

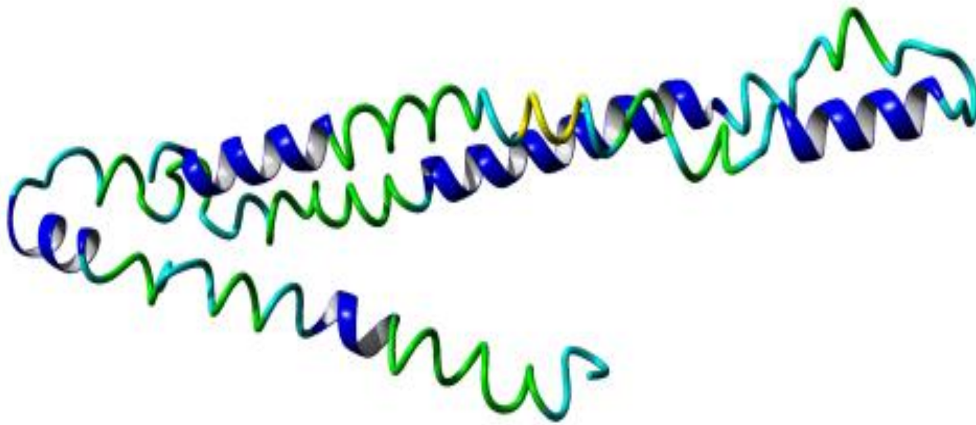

Figure 4.4. Homology model of human α_{s1}-casein with a C-score of -3.39. The N-terminus end is located at the upper chain whereas the C-terminus end is located at the lower chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

African elephant and human milk lacks α_{s2}-caseins, as a result, three homology models of cow, sheep and horse α_{s2}-caseins with C-scores of -2, -3.23 and -3.2 respectively are presented in Figures 4.5–4.7. The cow α_{s2}-casein homology model does not have any β-strands and has shorter lengths of α-helices compared to the α_{s1}-casein structure model. The homology model is composed of approximately equal lengths of alpha helices and random coils with turns. The N-terminus begins with a short length of turns followed by longer lengths α-helices with more turns and random coils inserted between the α-helices. The C-terminus end also composes of longer sections of both random coils and turns. The homology models of sheep and horse α_{s2}-casein are similar in structure compared to the cow α_{s2}-casein model. However, the former models fold into two chains whereas the latter model folds into three chains.

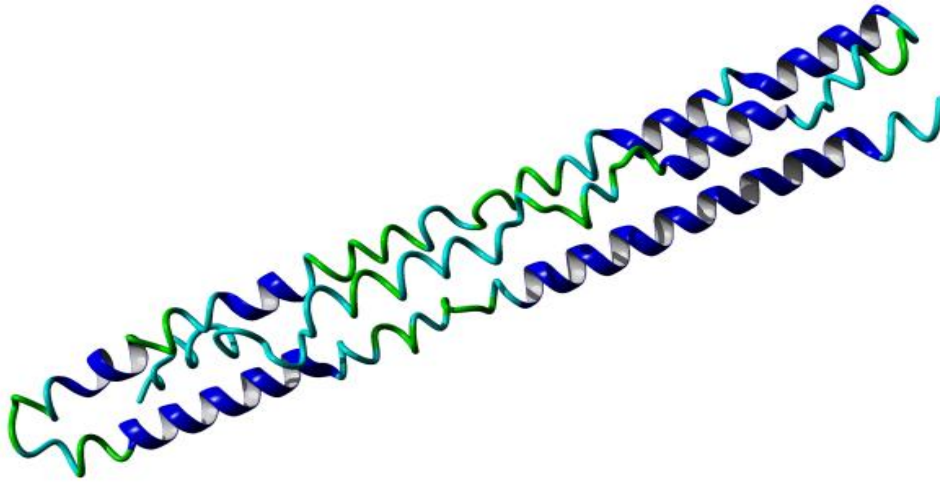

Figure 4.5. Homology model of cow α_{s2}-casein with a C-score of -2. The N-terminus end is located at the upper chain whereas the C-terminus end is located at the lower chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

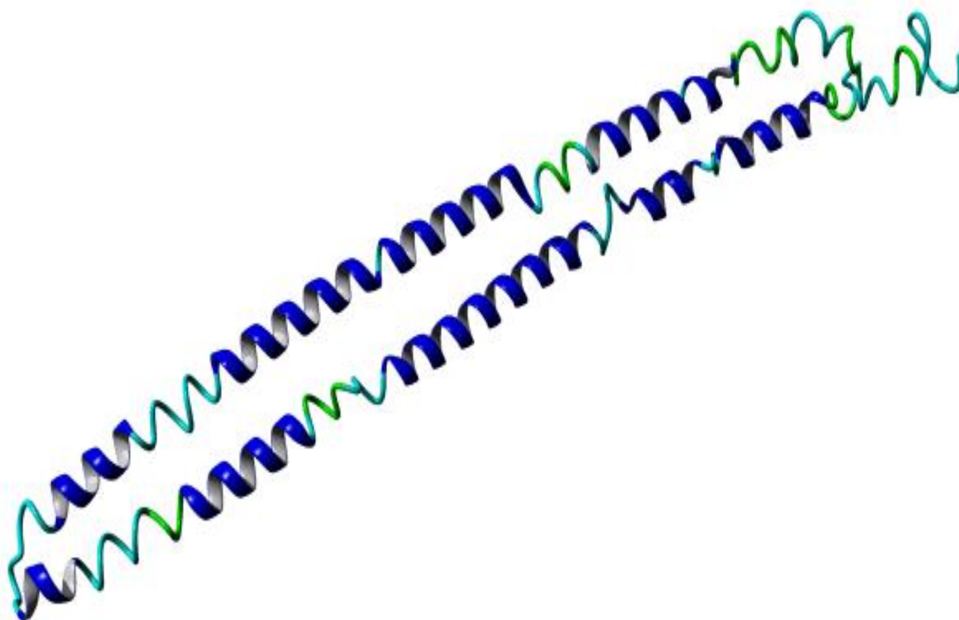

Figure 4.6. Homology model of sheep α_{s2}-casein with a C-score of -3.23. The N-terminus end is located at the upper chain whereas the C-terminus end is located at the lower chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

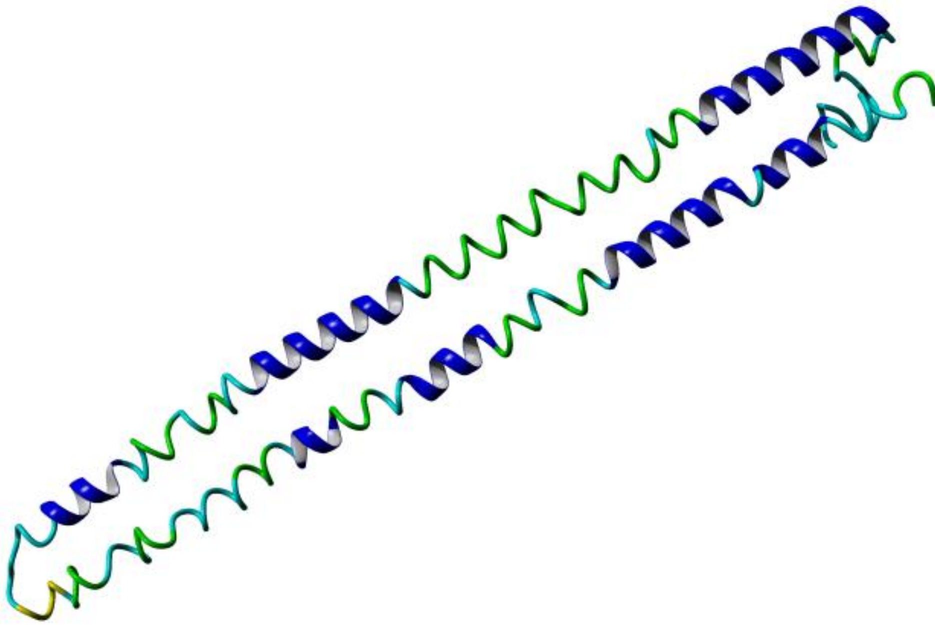

Figure 4.7. Homology model of horse α_{s2}-casein with a C-score of -3.2. The N-terminus end is located at the upper chain whereas the C-terminus end is located at the lower chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

4.3.2. Beta caseins

The homology models of cow, sheep, horse, human and African elephant β-casein are shown in Figures 4.8 – 4.12. The C-scores of the aforementioned models were determined as -1.98, -2.78, -4.09, -3.1 and -2,78 respectively. Interestingly, cow β-casein homology model shows that, only a small section of the sequence located at the N-terminus folds into an α-helix whereas the rest of the structure is in the form of a combination of both random coils and turns. Sheep, human and African elephant β-casein models are structurally similar to the cow casein model, however, the African elephant model has two additional beta strands that are absent in cow, human and sheep β-casein models. Unlike the rest of the β-casein models, the horse β-casein homology model does not have any α-helices in its structure but also has a total of two beta strands.

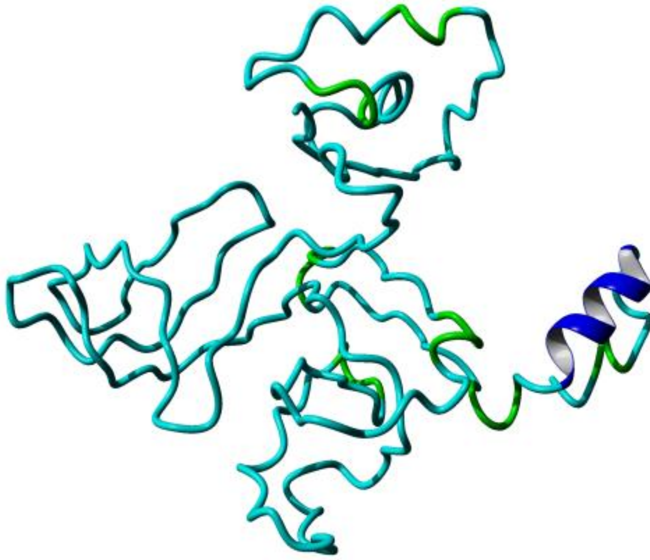

Figure 4.8. Homology model of cow β-casein with a C-score of -1.98. The N-terminus end is located at the middle-upper chain whereas the C-terminus end is located at the far right chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

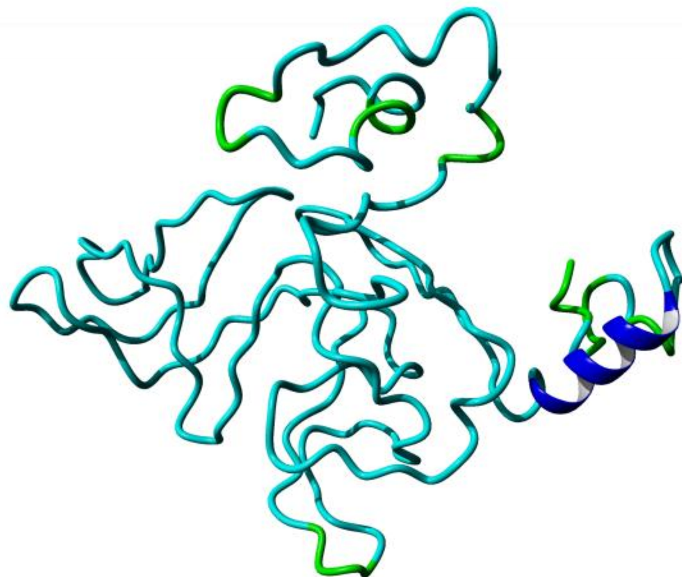

Figure 4.9. Homology model of sheep β-casein casein with a C-score of -2.78. The N-terminus end is located at the middle-upper chain whereas the C-terminus end is located at the far right chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

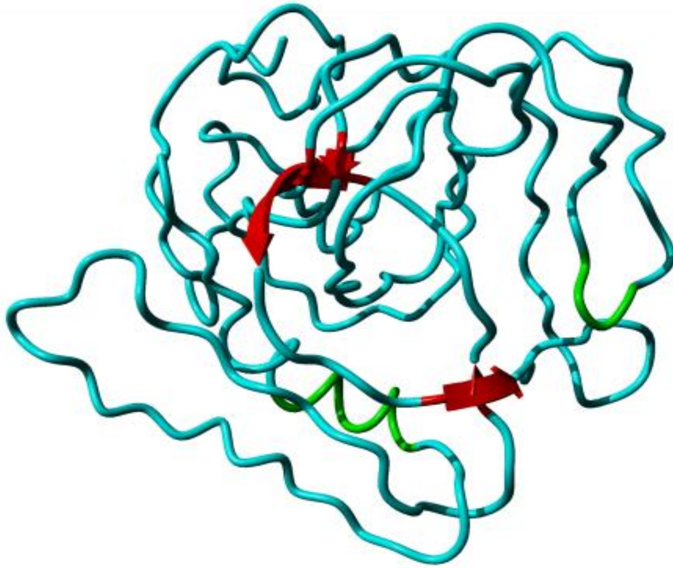

Figure 4.10. Homology model of horse β-casein casein with a C-score of -4.09. The N-terminus end is located at the middle-upper chain whereas the C-terminus end is located towards the lower right section of model. The colour codes for the secondary structure elements are as follows: β-strands are red, turns are green and random coils are cyan.

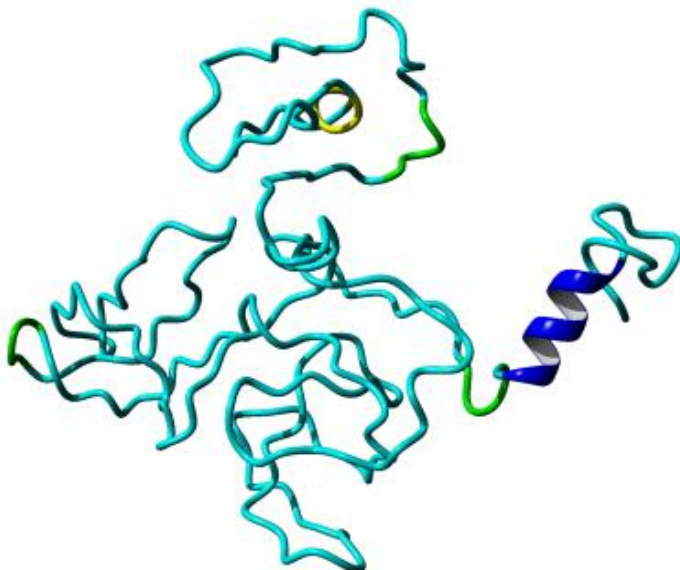

Figure 4.11. Homology model of human β-casein casein with a C-score of -3.1. The N-terminus end is located at the middle-upper chain whereas the C-terminus end is located at the far right chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

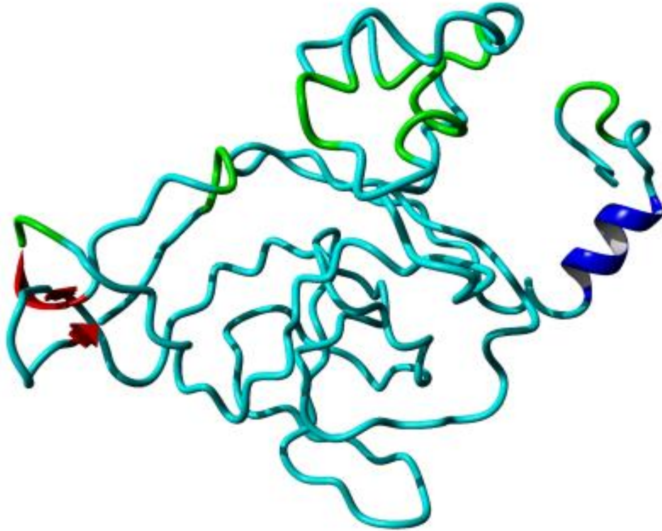

Figure 4.12. Homology model of African elephant β-casein casein with a C-score of -2.78. The N-terminus end is located at the middle-upper chain whereas the C-terminus end is located near the mid section of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, β-strands are red, turns are green and random coils are cyan.

4.3.3. Kappa caseins

The homology models of cow, sheep, horse, human and African elephant κ-casein are depicted in Figures 4.13–4.17. The C-scores of the models were as follows: cow (-3.58), sheep (-3.14), horse (-3.84), human (-3.45) and African elephant (-4.01). The cow κ-casein model comprises of larger sections of undefined secondary structure and turns. Only a short section of the model folds into an α-helix which is located near the N-terminus. The sheep κ-casein homology model is structurally similar to the cow κ-casein homology model. The human κ-casein model is unique in that it is composed of larger sections of beta strands in addition to random coils, turns and an α-helix. The horse κ-casein model also composes of beta strands although most of its structure is predominantly random coils. African elephant κ-casein model is largely composed of random coils and turns without any α-helices or beta strands.

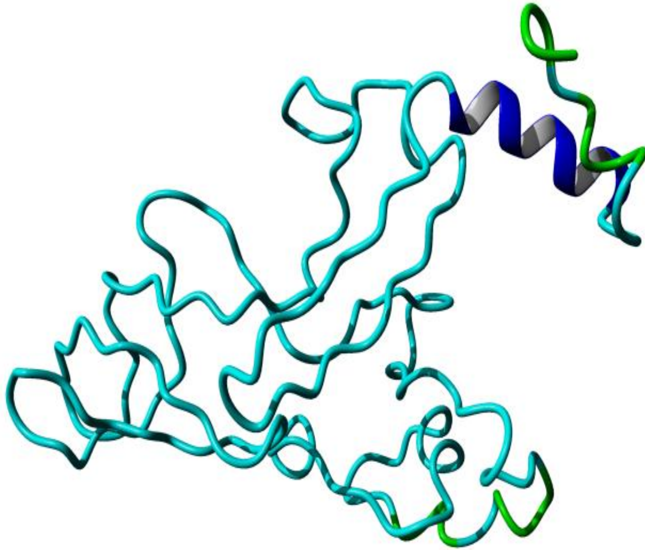

Figure 4.13. Homology model of cow κ-casein with a C-score of -3.58. The N-terminus end is located at the top section of the model whereas the C-terminus end is located near the left midsection of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

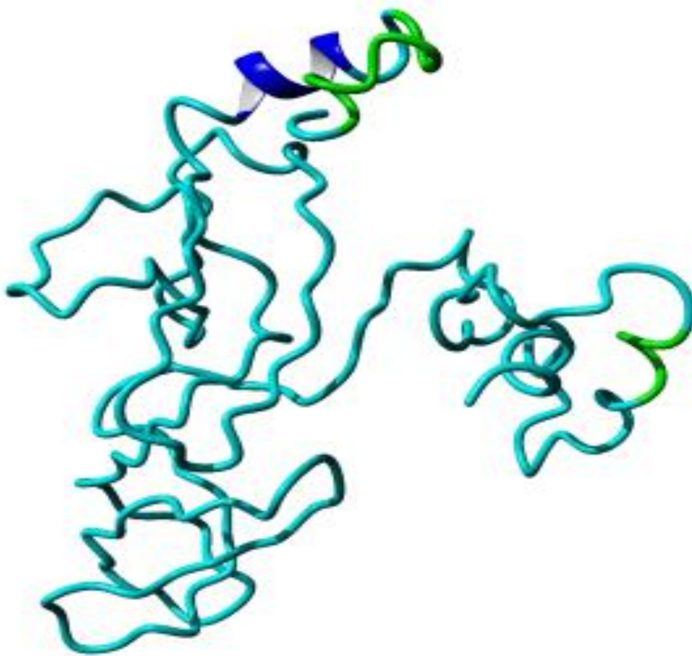

Figure 4.14. Homology model of sheep κ-casein with a C-score of -3.14. The N-terminus end is located at the middle-upper chain whereas the C-terminus end is located at the far right chain of the model. The colour codes for the secondary structure elements are as follows: α-helices are blue, turns are green and random coils are cyan.

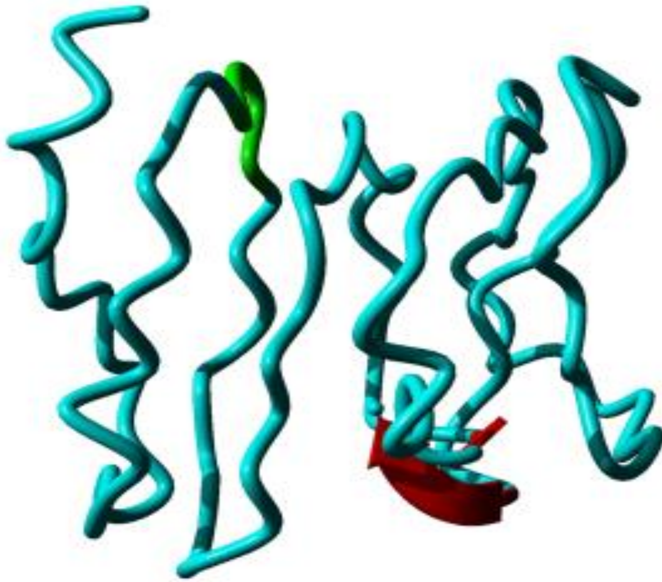

Figure 4.15. Homology model of horse κ-casein with a C-score of -3.84. The N-terminus end is located towards the left of the model whereas the C-terminus is located towards the right side of the model. The colour codes for the secondary structure elements are as follows: β-strands are red, turns are green and random coils are cyan.

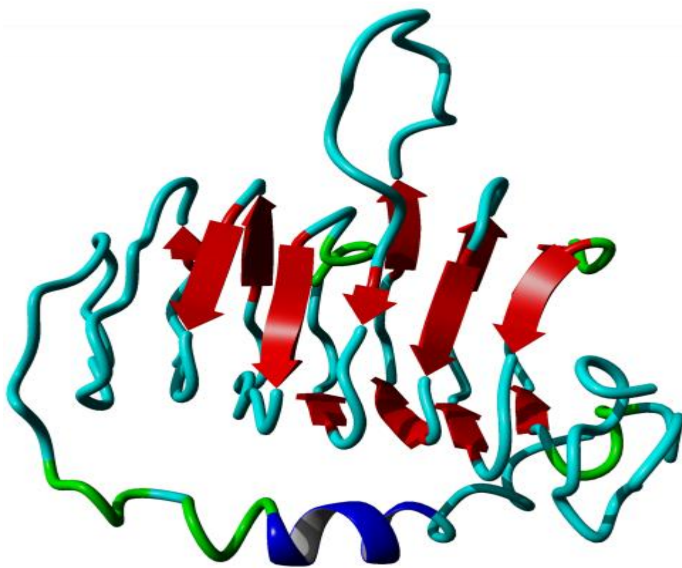

Figure 4.16. Homology model of human κ-casein with a C-score of -3.45. Both the N-terminus end and the C-terminus ends are located towards the right of the model. The C-terminus end is slightly positioned lower than the N-terminus end. The colour codes for the secondary structure elements are as follows: α-helices are blue, β-strands are red, turns are green and random coils are cyan.

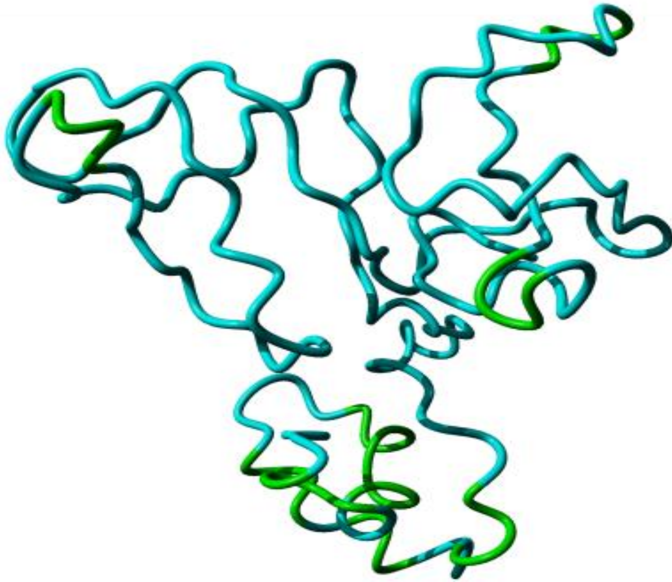

Figure 4.17. Homology model of African elephant κ-casein with a C-score of -4.01. The N-terminus end is located at the bottom of the model whereas the C-terminus is located at the mid-section of the model. The colour codes for the secondary structure elements are as follows: turns are green and random coils are cyan.

4.4. Discussion

Homology modeling is a powerful tool that can be used to predict the secondary structure of caseins. In most cases in biological systems, the structure of proteins is related to their function. Moreover, proteins that share the same structure usually perform the same function. The first attempt at bovine casein structure modeling was done by (Holt and Sawyer, 1988) over two decades ago. With the advancement in computational biology, more powerful and accurate tools for structure modeling are now freely available.

Homology modeling of cow caseins using I-TASSA shows that α-caseins (α_{s1}- and α_{s2}-caseins) secondary structure is dominated by α-helices whereas the secondary structure of β- and κ-caseins is dominated by undefined secondary structure in the form of random coils as well as turns. Interestingly, Holt and Sawyer (1988), through structure prediction observed that the signal peptide of bovine α_{s1}-casein at the N-terminus folds into an alpha helix followed by a β-sheet conformation at the major phosphorylation site. The structure then continues with another α-helical conformation followed by regions of no secondary structure at the C-terminus. Apart from the C-terminus conformation, or lack thereof, the described model remarkably resembles the structure model of bovine α_{s1}-casein predicted in our study.

The C-terminus of bovine α_{s1}-casein predicted in the current study shows mostly α-helix conformations. These differences could be a result of the advent of expanded protein structure databases and much more accurate protein structure prediction tools. Circular dichroism or Raman spectral analysis indicates the presence of approximately 14 % α-helix, 40 % β-sheets and 24 % turn-like structures in α_{s1}-casein (Michael Byler et al., 1988). In addition, the α_{s1}-casein 3D molecular models predicted by Kumosinski et al., (1991) also shows limited presence α-helix, approximately 14 %. This data contrast homology modeling data in the current study which shows over 70 % α-helix and no β-sheets. This difference may be attributed to temperature and medium dependence of circular dichroism or Raman spectroscopy which affect the final conformations obtained.

As highlighted above, the model for bovine α_{s2}-casein predicted in the current study is dominated by α-helix conformation with the presence of smaller lengths of random coils. In contrast, Holt and Sawyer (1998) studies predicted a model that is mostly dominated by random coils and turns although some sections appear in the form of α-helices. However, the positions in which the α-helices are predicted resembles the same positions in our bovine α_{s2}-casein structure model, particularly the N-terminus region and the middle sections. Similar to homology models in the current study, circular dichroism and FTIR spectral analysis indicates that there is an increased level of α-helix conformation (30-40 %) in addition to approximately 20 % turn-like structures and 20 % β-sheet in bovine α_{s2}-casein (Farrell et al., 2004).

The homology models of α-caseins predict that their structure is predominantly α-helices. The presence of α-helices has been shown to stabilise micelles and since α-caseins are mostly strategically located at the core of the micelle, where they are responsible for stabilizing the micelle (Horne, 1998), they are thus dominated by α-helix conformations.

With regards to bovine β-casein structure models, the proposed model by Holt and Sawyer (1988) and our model only have the α-helix conformation near the N-terminus in common. The former model consists of many sections that fold into α-helices, β-strands, a few turns and relatively longer lengths of random coils. In contrast the latter model consists of over 80 % of the entire length adopting random coil conformation with a few turns and only a single α-helix conformation. Like the cow β-casein model, the cow κ-casein structure model predicted in the current study

lacks β-strand conformations and has very long lengths of random coils (over 75 %). The opposite is true for the homology model by Holt and Sawyer (1988), where the model has an equal balance in length for defined secondary structures (β-strands, α-helices and turns) and random coils. However, both models show the presence of an α-helix located near the N-terminus.

The β-casein structure models of cow, human and sheep show high homology, while those of horse and African elephant show little homology. These data differences in structure can be attributed to the diverse differences in primary structure of the β-caseins of the latter two species. The diverse nature of casein sequences is driven by evolutionary events such as exon skipping, mutations and to a certain extent posttranslational modifications (Martin et al., 2013). This may also suggest that these β-casein molecules may not interact with other caseins via similar mechanisms during casein micelle formation as proposed for cow casein micelles (Horne, 2008). The presence of greater lengths of random coils may also encourage the formation of a rather open and flexible casein micelle that can also form thermodynamically stable complexes with calcium and phosphate. CD and FTIR spectral analysis have estimated that β-casein has approximately low levels of α-helix (15 %) and intermediate levels of turn-like structure (29 %) and β-sheet (30 %) (Farrell et al., 2004). This data also agrees to some extent with homology modeling data in the current study where there is less α-helix and more of turn-like structures.

Kappa caseins play a very important role in stabilizing micelles by steric stabilization which in turn prevent coagulation (Horne, 1998). Homology models of κ-caseins are

structurally distinct from each other. Interestingly, even the N-terminus, where high homology is observed between the species' signal peptides, structural differences are still observed. While the sheep and cow κ-casein models N-terminus begin with turns followed by α-helix, horse, human and African elephant models, begin with random coil structures. Interestingly, CD and FITR Spectral analysis indicate that κ-casein has relatively low content of α-helix (10 %) and high levels of both β-sheets (20 %) and turn-like structures (35 %) (Michael Byler et al., 1988), similar to observations in our current study using homology modeling for κ-casein.

With regards to structure models of caseins observed above and their link to biological function, it may be derived that α-caseins have a very much defined secondary structure in the form of long lengths of α-helices and turns. These conformations are favourable for, and could play a critical role in stabilizing casein micelle structure. Moreover, these defined secondary structure conformation are more likely to be located at the core of the micelle making it more compact and may therefore result in a casein micelle that is smaller in size.

The β-casein structure models predicted in this study showed interesting structure conformations. They display shorter segments of α-helices compared to α-caseins. Furthermore, β-casein is the most hydrophobic casein and is located at the core of the micelle; its structure model supports this due to presence of longer lengths of random coils and turns. Casein micelles with large amounts of β-casein are more likely to adopt a larger size and a more open structure.

The homology models of cow and sheep κ-casein are similar and therefore may be involved in the same function. However, the structure models of human, horse and African elephant are different and thus based on these structural differences; they may not function in similar ways. Located at outer areas of a casein micelle, the role of κ-casein is to prevent aggregation of casein micelles by steric stabilization (repulsion between negative charges) (Creamer et al., 1998). In order to achieve the above function, flexibility of the structure in the form of random coils, is necessary. Random coils do not have a defined secondary structure and therefore allows other interactions between amino acid residues to occur which can hence introduce flexibility during protein fold, this is critical for κ-casein role where it covers the outside of the micelle and also to expose the negatively charged C-terminus (Parry et al., 2008; Lupas and Gruber, 2005). The homology model of κ-casein fulfils this criterion.

A comparison of sheep and cow casein models show that there is a high degree of similarity in secondary structure. It would be interesting if it could be conclude by X-ray crystallographic methods. This similarity in structure would suggest that casein micelles of cow and sheep milk are expected to have the same structure and size. However, this was not the case as observed by SEM studies in chapter 2 (Figures 2.1-2.3). A possible explanation why the caseins of homologous primary structure and seemingly similar secondary structure would form different micelles might be due to the differences in the ratio of the individual caseins (Table 2.1).

Following the comparison of cow and sheep casein structure models and casein micelle structure, the large differences in predicted secondary structures of horse, human and African elephant caseins may be expected to affect the structure of their respective casein micelles. The α-caseins of all species under study seem to display very similar secondary structures of helices and turns, which may result in a similar stabilization of the core of the casein micelles. Since it was shown that African elephant milk does not contain α-caseins (Madende et al., 2015), it might therefore be predicted that a less stable casein micelle should be formed. This seems to be the case, as the casein micelles observed by SEM are much larger than that of the other four species under study (Figure 2.6 and Table 2.1).

Human milk casein micelles were also observed to be larger by SEM (Figure 2.4 and Table 2.1). The absence of α_{s2}-casein and the low amounts of α_{s1}-casein (Table 2.1) with their stabilizing function seem to be the reason. It is, however, possible that the difference in casein micelle size and structure cannot solely be ascribed to the absence or presence of the α-caseins. The differences in the structure of the β- and κ-caseins may also play a role. The predicted secondary structures of human and African elephant κ-caseins differ from each other and also differ from that of the cow and sheep proteins.

While the predicted secondary structures of the horse α-caseins resemble that of cow and sheep. The amounts of α-caseins are also very similar (Table 2.1). It may therefore be predicted that the casein micelles of horse milk would be stabilized in a similar way. However, the casein micelle sizes of horse milk, observed by SEM, are

slightly larger than that of cow and sheep, but not as large as that of human and African elephant (Figure 2.4 and Table 2.1). The explanation might be two-fold. Firstly the κ-casein is present at lower levels (Table 2.1). Secondly, although the amounts of β-casein present in the casein micelle is similar compared to that of cow and sheep, the secondary structure is predicted to differ, which might exert itself in a somewhat looser casein micelle structure.

4.5. Conclusions

Although it cannot be claimed that the real structure of the caseins are as predicted in this study, it may be possible that the structure of the α-caseins may be similar between species, and would play a role in stabilizing the casein micelle structure. Lesser amounts of α-caseins in the casein micelle seem to result in a looser structure and larger micelle size. The role of β- and κ-casein structure is not that clear, however, it seems that an absence of α-caseins, or low amounts thereof, and the stabilization function, needs compensation by the β- and κ-casein. This is the case in human and especially African elephant milk, and their β- and κ-caseins seem to adapt conformations that are very different from that of cow and sheep, and also horse.

While the focus of this chapter was on the structure of caseins and their possible role in the formation of the casein micelles, the phosphorylated sites and cross linking by calcium phosphate nanoclusters on micelle structure (Holt and Sawyer, 1988) was

not taken into account. The existence of cow β-casein in a fully phosphorylated 5P form and presence of highly phosphorylated α-caseins could mean the formation of strong nanoclusters. Horse β-casein is composed of multi-phosphorylated isoforms having between 3 and 7 phosphate groups per molecule (Girardet et al., 2006) and may form strong nanoclusters as well. African elephant lacks α-caseins and from chapter 3, its β-casein exists in the 1P form, which suggests that it may possibly form weaker nanoclusters and therefore contribute to a more open micelle structure.

The following chapter deals with the distribution of casein genes across mammalian species. Comparative genomics of casein genes can give an indication of the casein genes that have been lost or gained through evolution and ultimately which caseins could be crucial in the formation of casein micelles.

4.6. REFERENCES

Al-lazikani, B., J. Jung, Z. Xiang, and B. Honig. 2001. Protein structure prediction. *Curr Opin Chem Biol*. 5:51–56.

Baker, D., and A. Sali. 2001. Protein structure prediction and structural genomics. *Science*. 294:93–96. doi:10.1126/science.1065659.

Creamer, L.K., J.E. Plowman, M.J. Liddell, M.H. Smith, and J.P. Hill. 1998. Micelle stability: kappa-casein structure and function. *J Dairy Sci*. 81:3004–3012. doi:10.3168/jds.S0022-0302(98)75864-3.

Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk-sixth revision. *J Dairy Sci*. 87:1641–74. doi:10.3168/jds.S0022-0302(04)73319-6.

Farrell, H.M., E.D. Wickham, J.J. Unruh, P.X. Qi, and P.D. Hoagland. 2001. Secondary structural studies of bovine caseins: Temperature dependence of ??-casein structure as analyzed by circular dichroism and FTIR spectroscopy and correlation with micellization. *In* Food Hydrocolloids. 341–354.

Girardet, J.M., L. Miclo, S. Florent, D. Molle, and J.L. Gaillard. 2006. Determination of the phosphorylation level and deamidation susceptibility of equine β-casein. *Proteomics*. 6:3707–3717. doi:10.1002/pmic.200500728.

Holt, C., and L. Sawyer. 1988. Primary and predicted secondary structures of the caseins in relation to their biological functions. *Protein Eng Des Sel*. 2:251–259. doi:10.1093/protein/2.4.251.

Horne, D.S. 1998. Casein Interactions : Casting light on the black boxes, the structure in dairy products. *Int Dairy J*. 6946:171–177.

Horne, D.S. 2008. Casein micelle structure and stability. *In* Milk proteins. H.S. T. Abby, B. Mike, editor. Academic press, San Diego, CA. 133–162.

Krieger, E., S.B. Nabuurs, and G. Vriend. 2003. Homology modeling. *Struct Bioinforma*. 857:507–508. doi:10.1007/978-1-61779-588-6.

Kumosinski, T.F., E.M. Brown, and H.M. Farrell Jr. 1994. Predicted energy-minimized as1-casein working model. *ACS Symp Ser*. 576:1,368–391.

Kumosinski, T.F., E.M. Brown, and H.M. Farrell. 1991. 3-Dimensional molecular modeling of bovine caseins - alpha-S1-casein. *J Dairy Sci*. 74:2889–2895. doi:10.3168/jds.S0022-0302(91)78470-1.

Kumosinski, T.F., E.M. Brown, and H.M. Farrell. 1993. Three-dimensional molecular modeling of bovine caseins: an energy-minimized beta-casein structure. *J Dairy Sci*. 76:931–45. doi:10.3168/jds.S0022-0302(93)77420-2.

Lupas, A.N., and M. Gruber. 2005. The structure of alpha-helical coiled coils. *Adv Protein Chem*. 70:37–78. doi:10.1016/S0065-3233(05)70003-6.

Madende, M., G. Osthoff, H.-G. Patterton, H.E. Patterton, P. Martin, and D.J. Opperman. 2015. Characterization of casein and alpha lactalbumin of African elephant (Loxodonta africana) milk. *J Dairy Sci*. 98:8308–18. doi:10.3168/jds.2014-9195.

Martin, P., C. Cebo, and G. Miranda. 2013. Interspecies comparison of milk proteins: Quantitative variability and molecular diversity. *In* Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects, 4th Edition. 387–429.

Michael Byler, D., H.M. Farrell, and H. Susi. 1988. Raman Spectroscopic Study of Casein Structure. *J Dairy Sci*. 71:2622–2629. doi:10.3168/jds.S0022-0302(88)79855-0.

Parry, D.A.D., R.D.B. Fraser, and J.M. Squire. 2008. Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J Struct Biol*. 163:258–69. doi:10.1016/j.jsb.2008.01.016.

Quan, L., Q. Lv, and Y. Zhang. 2016. STRUM: Structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics*. 32:2936–2946. doi:10.1093/bioinformatics/btw361.

Sander, C., and R. Schneider. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*. 9:56–68. doi:10.1002/prot.340090107.

Tramontano, A. 1998. Homology modeling with low sequence identity. *METHODS A Companion to Methods Enzymol*. 14:293–300. doi:10.1006/meth.1998.0585.

Xiang, Z. 2006. Advances in homology protein structure modeling. *Curr Protein Pept Sci*. 7:217–27. doi:10.2174/138920306777452312.

Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 9:40. doi:10.1186/1471-2105-9-40.

Zhang, W., J. Yang, B. He, S.E. Walker, H. Zhang, B. Govindarajoo, J. Virtanen, Z. Xue, H. Bin Shen, and Y. Zhang. 2015. Integration of QUARK and I-TASSER for Ab Initio Protein Structure Prediction in CASP11. *Proteins Struct Funct Bioinforma*. 76–86. doi:10.1002/prot.24930.

CHAPTER 5

COMPARATIVE GENOMICS OF CASEIN GENES

5.1. INTRODUCTION

With the advent of powerful genome sequencing tools and consequently an extensive number of sequenced genomes, new opportunities have arisen for dissecting gene regulation and molecular evolution (Rijnkels et al., 2003). Comparative genomics is one such opportunity which allows the discovery of new genes and aid in the identification of functional components. In comparative genomics, two or more genomes are compared in a large-scale holistic approach to discover the differences and similarities between the individual genomes (Wei et al., 2002). Over 17 000 fully sequenced and draft sequences of archaea, bacteria, and eukaryote genomes with a wealth of sequence data are freely available for public use (www.ncbi.nlm.nih.gov/genome) and with comparative genomics, there is a potential to gain major scientific insights about gene gain or loss, species origins, mammal orders and survival (Wei et al., 2002).

Casein genes (α_{s1}-, α_{s2}-, β- and κ-casein) have evolved from members of a group of secreted calcium phosphate binding phosphoproteins gene, specifically the ODAM gene (Kawasaki et al., 2011). In milk, caseins and large amounts of colloidal calcium

and phosphate form aggregates that are known as casein micelles (Walstra and Jenness, 1984).

The formation of casein micelles is very critical in the safe transport of high concentrations of calcium and phosphate from the lactating mother to the neonate via milk (Holt et al., 2013). Consideration the above, it appears that some mammalian species are devoid of other casein types, for example elephant milk lacks α-caseins, although all of these species milk contain casein micelles (Martin et al., 2013; Madende et al., 2015). The bovine casein micelle models stipulates the importance of each of the casein types for casein micelle formation, effectively meaning that these functions could be lost in mammalian species that do not have some of these caseins.

This observation has prompted the investigation of the distribution of casein genes across several mammalian species by comparative genomics. Several comparative studies have been done on caseins in the past, albeit mostly at protein level (Ginger and Grigor, 1999; Holt, 2015). Comparing casein genes (presence/absence and gene sequences) may shed light into the possible functional aspects of casein genes and their gene products particularly with reference to their importance in the formation of casein micelles in milk especially in those species where some of the caseins are absent.

5.2. Materials and methods

5.2.1 Comparative genomics

Ensembl genome browser tool of comparative genomics was utilized in the comparison of casein genes across mammalian species (Herrero et al., 2016). In summary, Ensembl provides comprehensive evidence-based annotation of all supported genome sequences. The gene annotations across all species provided by Ensembl gene build are automatically integrated. Gene trees were constructed from all casein genes available and the data were used to extract homologs (orthologs and paralogs). Using LastZ and its predecessor BlastZ tools, the synteny mappings from pair-wise alignments of species whose mammalian genomes are not too fragmented were derived. For this study, casein gene comparison was focused on eutherian (placenta) mammals.

5.3. Results

For all the comparative genomics data presented in this section, the cow gene was selected as the query gene for the location of homologous genes in other mammalian species and therefore is highlighted in red as a result.

5.3.1. α_{s1}-casein

Alpha s1 casein is one of the least represented genes across eutherian mammalian species. In total, there are only 22 homologs of the α_{s1}-casein (CSN1S1) gene, of which 11 of the homologs are primates and rodents whereas 3 of the homologs are laurasitherian mammals (super order of Laurasia originating mammals) (Figure 5.1). The rest of the homologs, 8 in total are eutherian mammals which include cow,

human and sheep as the notable examples. Noted as the species of interest in this study, African elephant does not have the CSN1S1 gene, although its closest relative the hyrax does. The gene sequence alignment representation of the 22 α_{s1}-casein homologs is depicted in Figure 5.1. The alignment shows several gaps in the sequences. Most of the gap positions are consistent with each mammalian group or sub-tree, for example the primates have 3 large gaps (indicated in white), that are consistent amongst the gene sequences.

Figure 5.1 also underlines the disparate nature of CSN1S1 gene sequences, most of the regions in the sequence alignment share only 33-66 % sequence homology and none of the regions have over 66-100 % sequence homology. The α_{s1}-casein sequences also vary in length from one to the other. As an example, the cow gene is translated into a 214 amino acid long protein, whereas the pig gene is translated into a 206 amino acid long protein. All the 22 homologs of the α_{s1}-casein gene are orthologous meaning they all have a common ancestral gene and were separated through a speciation event. Orthologous gene products often retain the same function in the new species.

Figure 5.2 depicts the CSN1S1 gene gain or loss in a radial view. The gene gain or loss figure shows that the CSN1S1 gene has been lost in a number of ancient mammalian species such as elephant and armadillo. Interestingly, this gene which has been predicted as the oldest of the casein gene has been lost even between the closely related species. The classical example is between the vervet monkey, olive baboon and the macaque. Although these 3 primates branched from the same α_{s1}-

casein ancestral gene, only the vervet monkey retained the α_{s1}-casein gene. This aforementioned pattern between closely related species is observed consistently throughout the gene gain or loss tree. However, it must be mentioned that only genomes sequences that are not too fragmented were considered for this study. In the case of the absence of CSN1S1 gene in elephant and armadillo, several genome databases have been consulted and in all the cases, the CSN1S1 gene was absent..

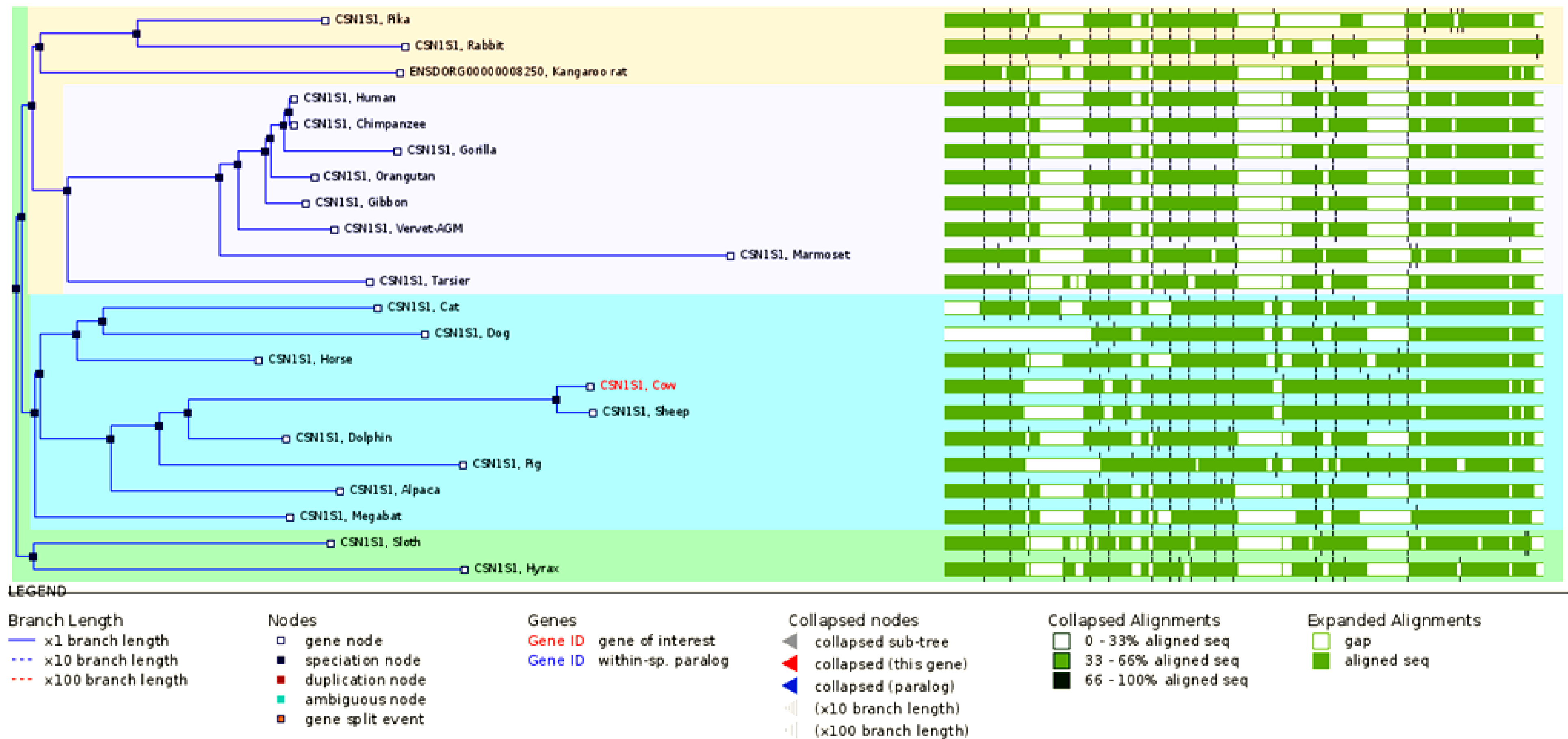

Figure 5.1. Gene tree of the relationship between 22 α_{s1}-casein gene sequences and their sequence alignment.

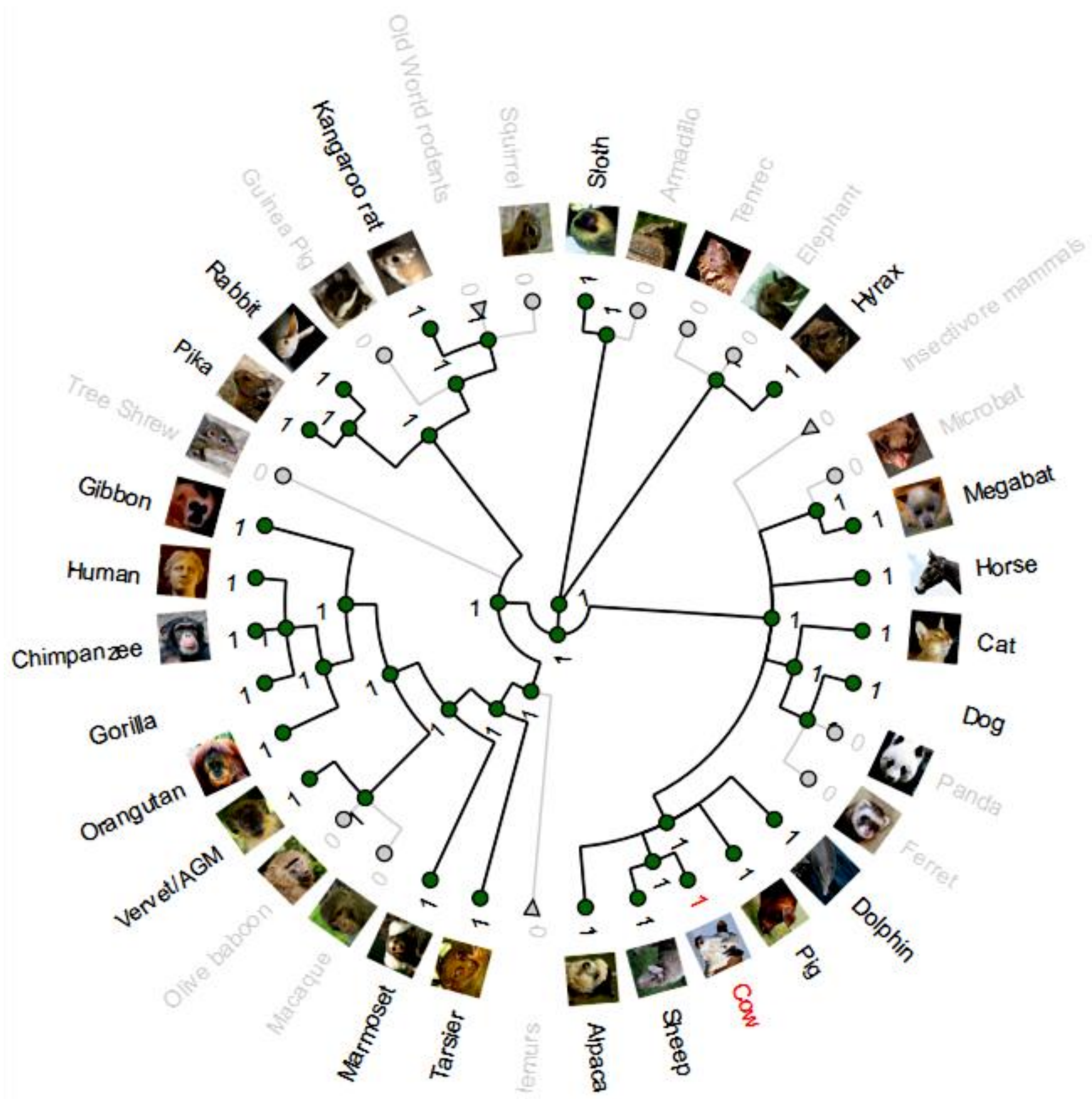

Figure 5.2. The radial gene gain or loss representation of α_{s1}-casein gene amongst several mammalian species. The green circular nodes indicate the presence of the gene whereas the grey nodes indicate its loss or absence.

5.3.2. α_{s2}-casein

Like CSN1S1 gene, the α_{s2}-casein (CSN1S2) gene is also minimally represented. Only 13 homologs can be observed in Figure 5.3. The placental or eutherian mammals are the most represented with 9 homologs whereas the rodents and

rabbits family only have four representing members. Interestingly the rat and mouse are unusual in possessing a CSN1S2-like casein gene copy which is represented in the alignment as csn1s2b. Unlike the CSN1S1gene, the CSN1S2 gene family has paralogs present in addition to the orthologs. In cats specifically, the CSN1S2 gene has undergone a duplication event through the course of evolution, resulting in a CSN1S2 gene that shares a common ancestor with other α_{s2}-casein homologs. In most cases, paralogous gene products usually perform different functions in the same species. Interestingly, African elephant lacks both α-casein genes, its close relatives such as the hyrax only lack the CSN1S2-like casein gene which developed later than all the other casein genes. The hyrax retained the CSN1S1 gene which is the oldest of the casein genes. The absence of both α-casein genes appears to be unique to African elephant, although the gene loss or gain plots show absence of both α-casein genes in the squirrel, this is not exactly accurate. The squirrel does have the CSN1S1 gene but lacks the CSN1S2-casein gene, these data have not been taken into account on the gene gain or loss plots due to the high fragmentation of the squirrel gene sequence.

There are also several gaps in the sequences as shown by the sequence alignment. Sequence comparison across all the 13 species indicates that there is between 33-66% homology between the sequences. Figure 5.3b further shows that there is increased homology among CSN1S2 genes sequences that are in the same subgroup. For example, when the placental mammals' subgroup is considered, the sequence homology increases to between 66-100 % (as indicated by dark green shaded areas on the alignment). Furthermore, Figure 5.3b also highlights the high homology and conserved nature of the signal peptide which is located at the N-

terminus region. The sequence length also varies considerably from one species to the other as highlighted by the gaps in the sequences, this further increases variability among orthologous gene products. Some gaps are much larger making the sequences shorter (armadillo) whereas other gaps in the sequence are relatively smaller making the sequences longer (sheep).

Csn1s2b, Rat
Csn1s2b, Mouse
ENSOCUG00000023469, Rabbit
ENSCJAG00000034512, Marmoset
Csn1s2, Guinea Pig
ENSOARG00000010683, Sheep
ENSBTAG00000005005, Cow
ENSSSCG00000009263, Pig
ENSAMEG00000005747, Panda
ENSFCAG00000023649, Cat
ENSMLUG00000014268, Microbat
ENSFCAG00000027298, Cat
ENSDNOG00000030588, Armadillo

LEGEND

Branch Length
x1 branch length
x10 branch length
x100 branch length

Nodes
gene node
speciation node
duplication node
ambiguous node
gene split event

Genes
Gene ID gene of interest
Gene ID within-sp. paralog

Collapsed nodes
collapsed sub-tree
collapsed (this gene)
collapsed (paralog)
(x10 branch length)
(x100 branch length)

Collapsed Alignments
0 - 33% aligned seq
33 - 66% aligned seq
66 - 100% aligned seq

Expanded Alignments
gap
aligned seq

Figure 5.3a. Gene tree of the relationship between 13 α_{s2}-casein gene sequences and their sequence alignment.

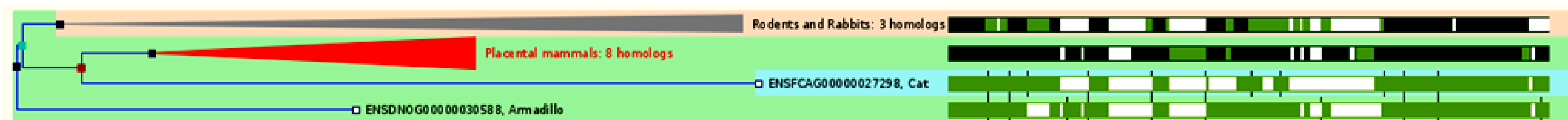

Figure 5.3b. Gene tree of the relationship between 13 α_{s2}-casein gene sequences and their sequence alignment. The sequences have been grouped into their subgroups

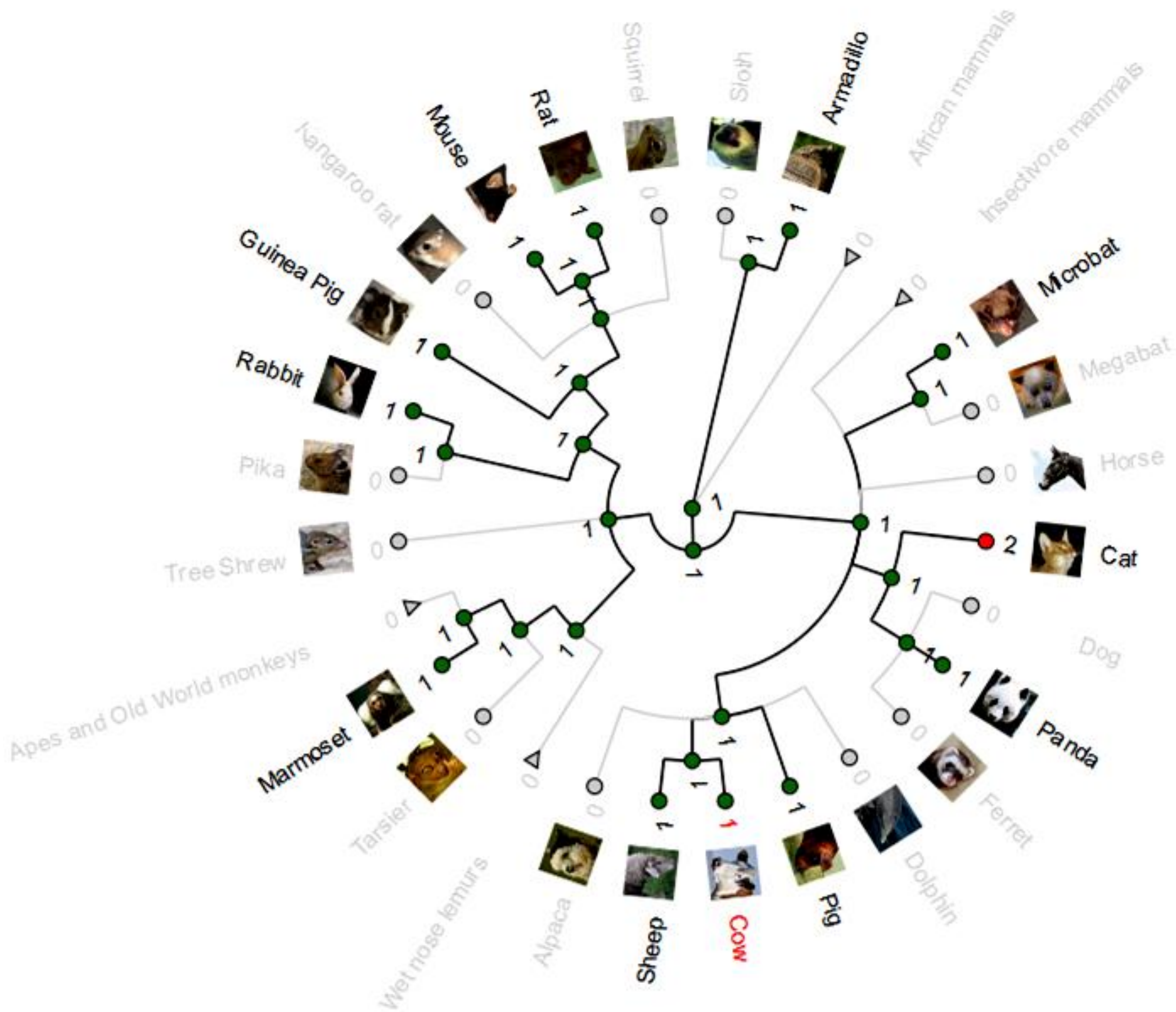

Figure 5.4. The radial gene gain or loss representation of α_{s2}-casein gene amongst several mammalian species. The green nodes indicate gene presence; grey nodes represent gene loss and red nodes represent presence of paralogues.

5.3.3. β-casein

The β-casein (CSN2) encoding gene is much more common among mammalian species compared to both α_{s1}- and α_{s2}-casein encoding genes. The gene tree and alignment in Figure 5.5a shows 40 homologues of the CSN2 gene. The primates and rodents are the most represented species with up to 19 β-casein homologous genes. Interestingly, the squirrel and microbat genomes show the presence of paralogs of

the β-casein gene as depicted by duplication nodes (red coloured) in both Figures 5.5a and 5.5b. Paralogs are a consequence of evolution through gene duplication resulting in two active sets of genes whose products usually assume different functions although having a common ancestral gene.

Like CSN1S1 and CSN1S2 genes, the sequence alignment across all the species shows a higher degree of divergence with between 33-66 % sequence homology. However, the opposite is true for a sequence alignment of species that are in the same subgroup, where a much higher sequence homology of between 66-100% is observed (Figure 5.5b). Several gaps also exist in the sequence alignment with most gaps consistent throughout the alignment. As mentioned before, such gaps increase the heterogeneity of casein genes and their products. The CSN2 gene is more conserved among the closely related species as depicted in Figure 5.6. In addition to retaining the CSN2 gene, it appears that more genes have been gained through duplication events over the course of evolution as illustrated by red nodes.

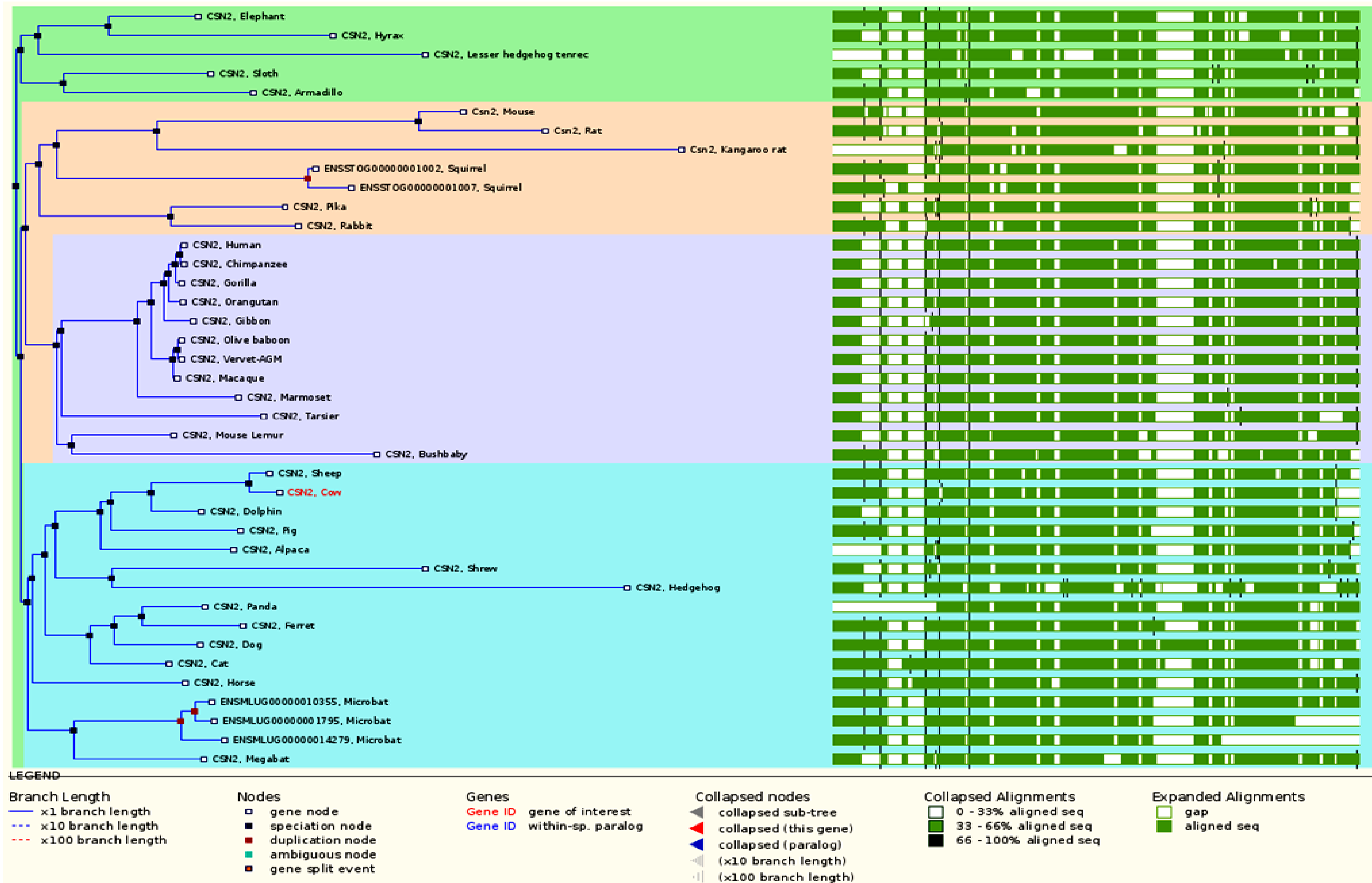

Figure 5.5a. Gene tree of the relationship between 40 β-casein gene sequences and their sequence alignment.

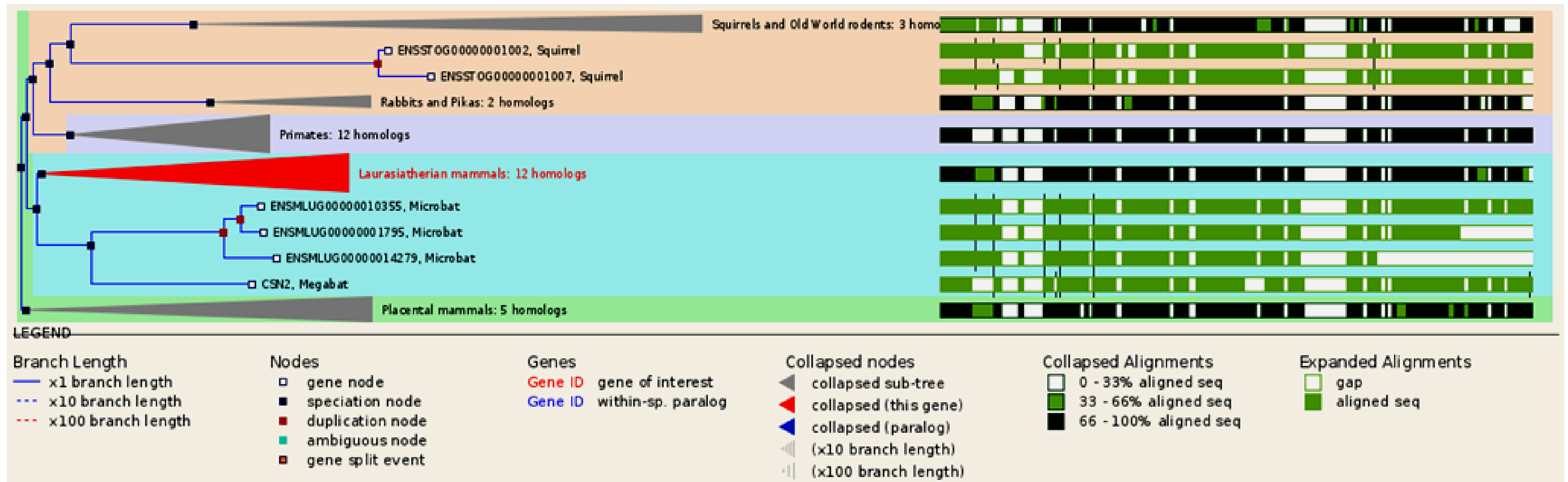

Figure 5.5b. Gene tree of the relationship between 40 β-casein gene sequences and their sequence alignment. The sequences have been further grouped into their subgroups

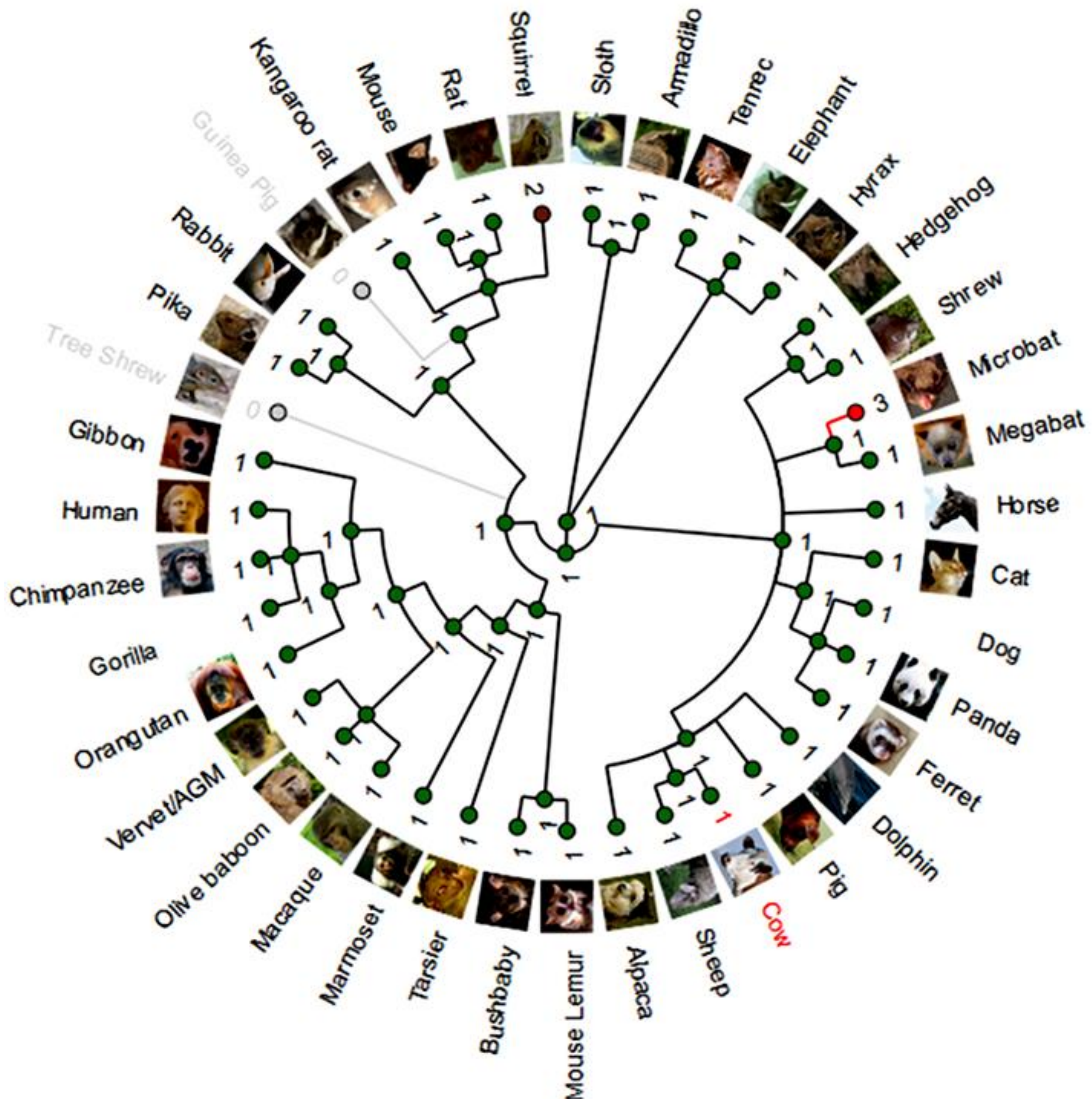

Figure 5.6. The radial gene gain or loss representation of β-casein gene amongst several mammalian species. The green nodes indicate gene presence; grey nodes represent gene loss and red nodes represent presence of paralogues.

5.3.4. κ-casein

The κ-casein (CSN3) gene is the most studied casein gene and as a result this gene presents interesting comparative genomics. Unlike α_{s1}-, α_{s2}- and β-casein encoding gene products that are calcium sensitive, the gene product of κ-casein is soluble in

calcium (Ginger and Grigor, 1999). Figure 5.7 shows a CSN3 gene tree relationship between 35 mammalian species. All homologs presented on the gene tree are also orthologs meaning they are as a result of a speciation event rather than a duplication event. The gene tree members are dominated by primates and rodents with 18 members, bats are the least represented with only 2 sequences.

A look at these homologous gene sequences shows a great deal of divergence. As was noted for the CSN1S1, CSN1S2 and CSN2 genes, the sequences mostly share between 33-66 % (light green colour) homology when compared together. Increased sequence homology is observed when sequences are compared within a subgroup. Gaps in the sequences are also common amongst CSN3 gene sequences, although these gaps are much more consistent throughout the sequences. In addition, the sequence comparison also reveals the differences in length of each gene. The horse CSN3 gene appears to be the shortest of the 35 genes. It is important to note that the length of genes do not necessarily reflect the length of its gene products. Events such as exon skipping often result in short length gene products that are shorter than their full length counterparts.

Figure 5.8 shows the gene gain or loss relationship of CSN3 genes from a variety of species. The figure also suggests that some mammalian species such as the wallaby and squirrel have lost the κ-casein gene, this is not entirely correct. The CSN3 gene is present in these species but the sequences are highly fragmented and therefore were omitted from the gene gain or loss plot. The genome sequence of the squirrel is complete but the sequence is of very low quality and therefore several

mistakes are expected from the genome. Nevertheless, it is clear from the figure that most mammalian species have returned the CSN3 gene. The same observation was noted for CSN2 gene tree.

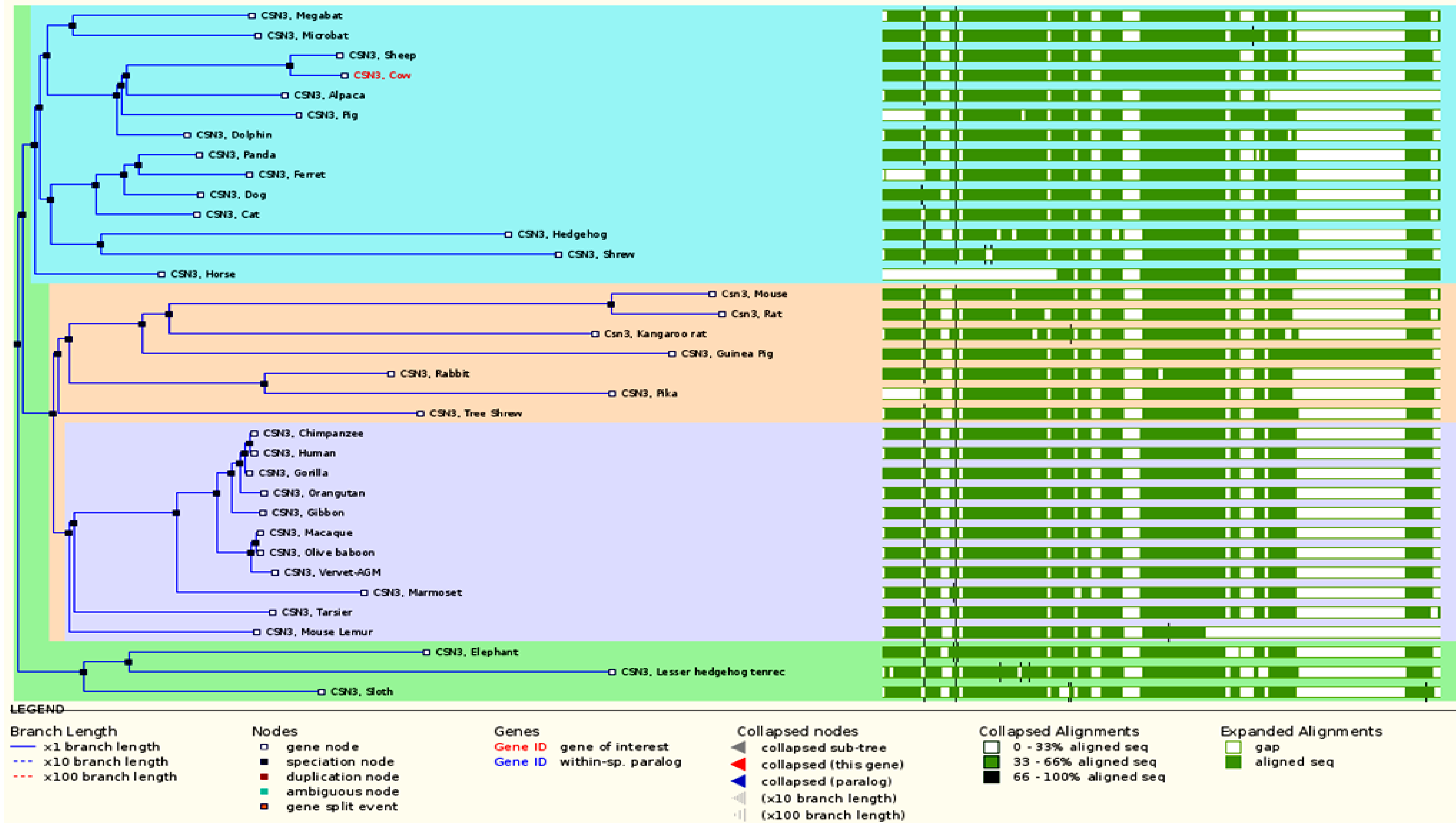

Figure 5.7. Gene tree of the relationship between 35 κ-casein gene sequences and their sequence alignment

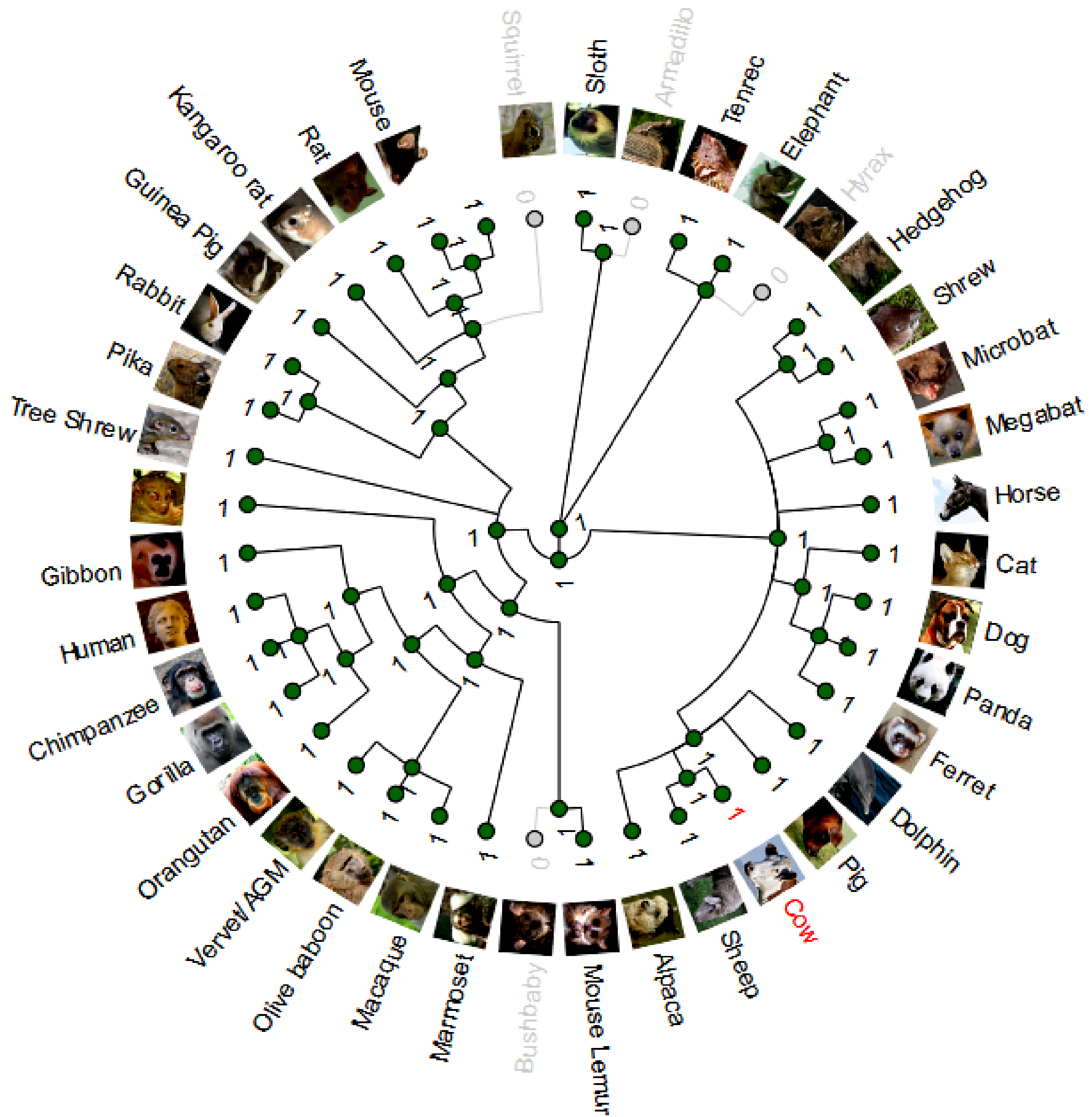

Figure 5.8. The radial gene gain or loss representation of κ-casein gene amongst several mammalian species. The green nodes indicate gene presence and the grey nodes represent gene loss.

5.4. Discussion

A number of comparative studies of caseins have been previously conducted at protein level with the more recent study conducted by (Holt, 2015). Because no single organism can adequately describe the functionality of the other, comparative studies are of paramount importance if more light is to be shed on debatable concepts and models. Caseins are rapidly evolving genes and this was evident in this study of comparative genomics of casein genes across all mammalian species whose sequence data is available and is not too fragmented. Of the 4 casein genes, the α-casein genes are the less represented group. In contrast, most mammalian species do possess CSN2 and CSN3 genes.

With regards to alignment of casein genes sequences, it is clear that casein gene sequences (α_{s1}-, α_{s2}-, β- and κ-caseins) are very diverse although very closely related species sequences show increased homology. Apart from the signal peptide sequence that is highly conserved, the rest of the mature peptide sequence is very diverse. Moreover, gaps in the sequences that are introduced to maximise the multiple alignment, also contribute to the diverse nature of casein gene sequences since they highlight the conserved and non conserved sequence regions. The non-homologous nature of casein gene sequences could be related to the rather less specific function of their gene products in casein micelle assembly. In addition to gene sequence differences, further divergence and variability of caseins is introduced by events such as exon skipping which occur during processing of primary transcripts (Martin et al., 2013).

In addition to the presence of orthologos, paralogos are also a common feature amongst casein genes, specifically the CSN1S2 and CSN2 gene family. Paralogs occur due to duplication events leading to the same species having more than one pair of the particular casein gene. The gene gain or loss figures also highlight the absence of α-casein encoding genes in most ancient mammalian species such as the African elephant and armadillo whereas in most modern mammals such as cow and horse the gene is present. It is interesting to note that according to the casein micelle models, all four caseins have a role to play in the formation of a casein micelle (Horne, 1998). However, it appears from the comparative genomics of casein genes that both ancient and modern mammalian species have both CSN2 and CSN3 genes as a common feature and therefore suggesting that these two genes and their gene products may have a much bigger and important role to play in casein micelle formation. Human milk lacks α_{s2}-casein, it may be possible that its role in casein micelle formation could be shifted to α_{s1}-casein which is capable of forming disulfide-linked heteromultimers with κ-casein (Martin et al., 2013). This demonstrates that caseins can be multifunctional with regards to micelle formation and as a result the presence of all four (sometimes five) caseins may not be a prerequisite for casein micelle formation in milk.

The gene comparison data shows a lot of differences with regards to the presence or absence of casein genes among mammalian species. Some of these differences are rather extreme, for example, the gene gain or loss tree shows that the squirrel only has β-casein and the rest of the caseins are absent. However, it is of paramount importance to note that comparative genomics data is as good as the quality of the genome databases (Muller et al., 2003). Sequencing errors that are carried over to

the actual genome database may be misleading and therefore result in inaccurate interpretation of results. In addition, gene data is also dependent on databases and the quality of sequencing and therefore some of the genes shown to be absent from gene tree because of incompleteness of the genome database or an omission error.

The evolution of casein genes follows the order: CSN1S1, CSN2, CSN1S2 and CSN3 (Martin et al., 2013) and as shown in Figure 1.1 in chapter 1, the development of CSN1S2 gene occurred less than 147.7 million years ago (MYA). Interestingly, the sloth and armadillo which are ancient mammals lack α_{s2}- and α_{s1}-casein encoding genes respectively. Figure 5.8, which is an improved version of Figure 1.1, shows their location on the gene tree. Assuming that their genome sequences are complete and without errors, then the sloth clearly did not develop the CSN1S2 gene and the armadillo lost the CSN1S1 gene during evolution.

The above highlights the rapidly evolving nature of casein genes which may be linked to the specific nutritional requirements and adaptability of mammals. The African elephant is also an ancient mammal and its genome database shows that it lacks both α_{s2}- and α_{s1}-casein encoding genes. It appears that the CSN1S2 gene developed between 147.7 and 91 MYA and some species did not develop it for example the sloth. The CSN1S1 gene has also been lost for example in the elephant and armadillo. It appears the β-casein gene is the more conserved of the ancient genes whereas the κ-casein gene (the last casein gene to develop) has been acquired by most if not all mammalian species. This possibly highlights the importance that these genes have in casein micelle formation and maintenance.

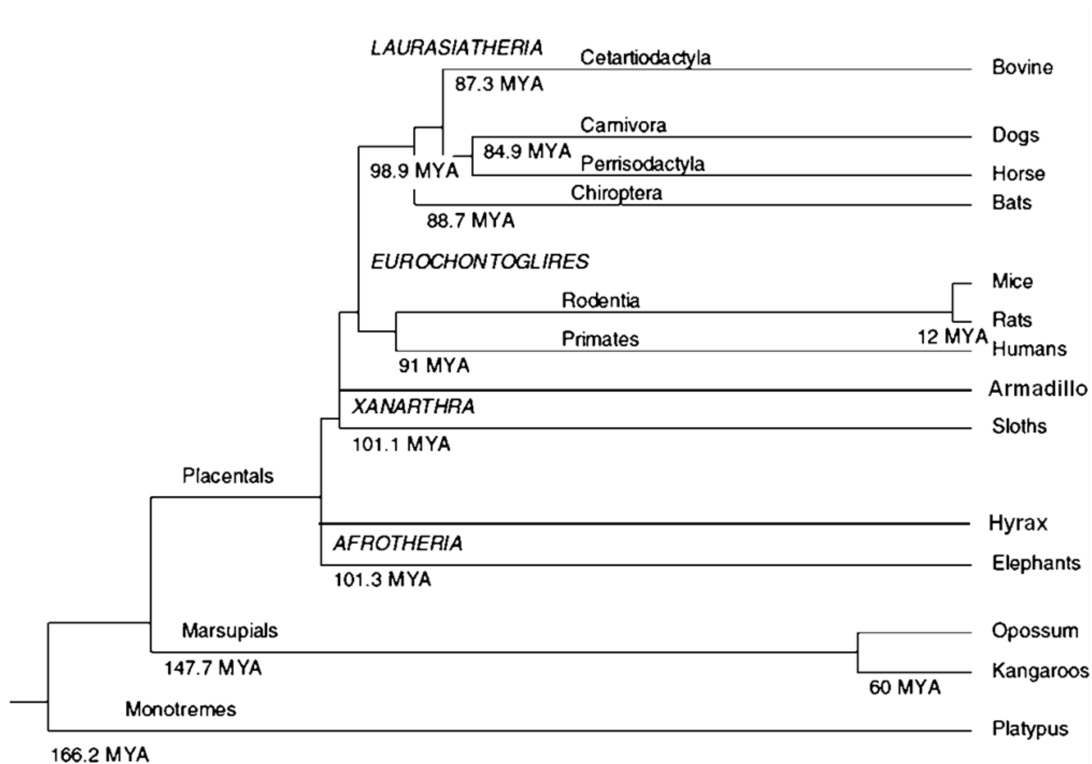

Figure 5.8. Splitting topology and divergence of representative mammals with the addition of hyrax and armadillo. The time of origin of each major branch is represented in million years ago (MYA).

5.5. Conclusions

Caseins remain some of the most studied milk proteins. Most dairy foods owe their properties to the characteristics of caseins and casein micelles. Caseins gene sequences are very diverse from each other and these genes have evolved with time resulting in more than one copy of the same gene in some species (paralogs). The diverse gene sequences and absence of some casein genes in a number of mammalian species that contain casein micelles suggest different mechanisms of casein micelle formation, opposed to those described for bovine casein micelles where all four caseins are present. The genes encoding α-caseins are absent in

most mammalian species, in contrast, genes encoding β- and κ-caseins are widely distributed amongst mammals and this may possibly suggest that latter genes products have a more significant role to play in milk, particularly in the assembly and mineral (calcium and phosphate) sequestration of the casein micelle.

5.6. REFERENCES

Ginger, M.R., and M.R. Grigor. 1999. Comparative aspects of milk caseins. *Comp Biochem Physiol Part B*. 124:133–145.

Herrero, J., M. Muffato, K. Beal, S. Fitzgerald, L. Gordon, M. Pignatelli, A.J. Vilella, S.M.J. Searle, R. Amode, S. Brent, W. Spooner, E. Kulesha, A. Yates, and P. Flicek. 2016. Ensembl comparative genomics resources. *Database*. 2016. doi:10.1093/database/bav096.

Holt, C. 2015. Casein and casein micelle structures, functions and diversity in 20 species. *Int Dairy J*. 60:2–13. doi:10.1016/j.idairyj.2016.01.004.

Holt, C., J.A. Carver, H. Ecroyd, and D.C. Thorn. 2013. Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J Dairy Sci*. 96:6127–46. doi:10.3168/jds.2013-6831.

Horne, D.S. 1998. Casein Interactions : Casting Light on the Black Boxes , the Structure in Dairy Products. *Int Dairy J*. 6946:171–177.

Kawasaki, K., A.G. Lafont, and J.Y. Sire. 2011. The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol Biol Evol*. 28:2053–2061. doi:10.1093/molbev/msr020.

Madende, M., G. Osthoff, H. Patterton, H.E. Patterton, P. Martin, and D.J. Opperman. 2015. Characterization of casein and alpha lactalbumin of African elephant (Loxodonta africana) milk. 8308–8318.

Martin, P., C. Cebo, and G. Miranda. 2013. Interspecies comparison of milk proteins: Quantitative variability and molecular diversity. *In* Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects, 4th Edition. 387–429.

Muller, H., F. Naumann, and J.C. Freytag. 2003. Data Quality in Genome Databases. *In* Proceedings of the Conference on Information Quality. 269–284.

Phadungath, C. 2005. Casein micelle structure : a concise review. *J Sci Technol*. 27:201–212.

Rijnkels, M., L. Elnitski, W. Miller, and J.M. Rosen. 2003. Multispecies comparative

analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics*. 82:417–432. doi:10.1016/S0888-7543(03)00114-9.

Walstra, P., and R. Jenness. 1984. Dairy chemistry and physics. *New York Wiley*. 467.

Wei, L., Y. Liu, I. Dubchak, J. Shon, and J. Park. 2002. Comparative genomics approaches to study organism similarities and differences. *J Biomed Inform*. 35:142–150. doi:10.1016/S1532-0464(02)00506-3.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSIONS

Milk is a complete and yet very complex biological fluid that has co-evolved with mammalian species and serves as a primary and essential source of nutrients for the neonate (Yang et al., 2013). The gross composition of milk varies from one species to the other (Fox and Mcsweeney, 1998). Milk proteins are grouped into caseins and whey proteins, and play an important function in the provision of essential amino acids for muscular tissue development (Phadungath, 2005). Caseins are phosphoproteins that are synthesised in the mammary gland during lactation (Kawasaki et al., 2011). Caseins are grouped into mainly four gene products termed:α_{s1}-, α_{s2}-, β- and κ-casein (Farrell et al., 2004). Caseins are assembled into large colloidal complexes with calcium and phosphate to form casein micelles (Horne, 1998). The formation of casein micelles in milk is crucial, in that it enables the sequestration of the otherwise precipitation prone large amounts of calcium and phosphate, which can be toxic to the mammary cell (Holt et al., 2013).

Caseins undergo post-translational modification (PTM) which result in a high degree of heterogeneity (Holland, 2008). The best known casein PTMs are phosphorylation of α_s- and β-caseins and glycosylation of κ-casein. These modifications are critical for the assembly, maintenance and stability of casein micelles. The exact structure of

a casein micelle is still an area of debate, as a result, several models have been put forward to describe its structure in bovine milk (Phadungath, 2005). The models emphasise on the functions of each of the caseins in the assembly and stability of casein micelles (Horne, 2008). The stability of casein micelles is influenced by a steric layer of κ-casein that exerts electrostatic repulsions between micelles and therefore preventing their coagulation (Tuinier et al., 2002). Moreover, the α_s- and β-caseins form the core of the micelle and interact via hydrophobic interactions while binding calcium and phosphate (Horne, 1998). Most experimental studies on casein micelles were done on cow caseins and casein micelles (Holt, 2015). Although cow milk proteins are used as a standard in the study of proteins of other mammals (Farrell et al., 2004), there is also diversity in casein composition and the ratio of casein types that are present in milk of other species (Potočnik et al., 2011). As an example, African elephant milk is devoid of all α_s-caseins but contain high levels of β-casein and lower levels of κ-casein whereas human milk also lacks α_{s2}-casein (Martin et al., 2013; Madende et al., 2015).

The aforementioned observations prompted a study of caseins and casein micelles in milk of non-cow origin, where the milk casein composition varies from that of cow milk. The first aim was to investigate the impact of casein composition variations on the nature of casein micelles using high resolution field emission scanning electron microscopy (FE-SEM). The variations in casein composition of fresh milk for this study were as follows: cow (contains all four caseins), horse (contains very little κ-casein), sheep (all four caseins present), human (α_{s2}-casein absent) and African elephant (α_{s1}- and α_{s2}-casein absent). Moreover, fresh milk was also frozen in order to demonstrate the effect of processing (freezing) on casein micelles.

In terms of casein micelle shape, all the casein micelles investigated appeared spherical but differed in surface appearance and size under high resolution SEM. The microscopic images of cow caseins were similar to those observed by Dalgleish et al. (2004), using a similar microscopic technique although the milk sample was in the form of reconstituted powdered milk. The fresh milk cow micelles had average diameters of 50-200 nm. The micelle surface appeared rough with structures that seem to be protruding from the inside to the outside of the micelle. In contrast, fresh horse milk casein micelles, where κ-casein content is very low appeared smooth on the surface, and they were observed to be slightly larger in size (average diameter 150-300 nm).

Of all the fresh milk caseins investigated in our study, sheep casein micelles were the smallest (50-200 nm) whereas human casein micelles had a larger range of micelle diameter (100-1500 nm). The surface of sheep micelles was slightly obscured due to their small size whereas human casein micelles had a rough surface resembling that of cow milk. African elephant casein micelles were also second largest in size (350-700 nm) and their surface also appeared rough. The elephant micelles were surrounded by structures that resembled disintegrated micelles which could mean that African elephant casein micelles are less stable and easily perturbed.

The above observations suggest that casein micelles appearance and structure could be species specific rather than universal. It is more likely that the total casein content and the proportions of the individual casein types, among other factors, have an impact on micelle overall appearance and structure of casein micelles. The absence or presence of α-caseins seems to be a determinant of casein micelle size.

Except for sheep casein micelles, all the casein micelles increased in size after freezing, the surface appearance was also slightly different with the protruding structures on micelle surfaces seemingly more pronounced. The freezing of milk and the observed structural changes on casein micelles shows that the treatment may have an impact on casein micelle structure and their integrity. It is therefore paramount that studies on casein micelles should be carried out on fresh milk, and not on frozen milk, as inaccurate conclusions might be drawn.

The second aim of this study was to determine the phosphorylation state of African elephant β-casein. Since African elephant milk casein content is predominantly β-casein and also that post translational phosphorylation of casein plays an important role in casein micelle formation and maintenance, it was imperative to characterise its phosphorylation state. Firstly the sequence of African elephant β-casein was determined by means of intact protein mass determination through LC MS. The mass did not match any of the three spliced variants that are inferred from the gene sequence or the fourth sequence proposed by Madende et al. (2015) and subsequently a fifth sequence was derived from the molecular mass of the intact β-casein. The sequence was determined to be a short length (200 amino acids long)

protein that had peptide sequences coded for by exons 4 and exon 5 truncated. Following the primary sequence determination, the phosphorylation pattern was also determined. Electrophoresis examination of pure African elephant β-casein by 2D dephosphorylation and Urea PAGE showed that there are up to 5 phosphoforms of African elephant β-casein. However, through phospho-enrichment of tryptic and chymotryptic digests and numerous LC MS/MS analyses, a single phosphorylation site at Ser9 was identified. Considering the above and LC MS data, African elephant milk appears to be dominated by unphosphorylated and singly phosphorylated β-casein forms, while the multi-phosphorylated isoforms may be present in low abundance. Compared to human, sheep, horse and cow milk β-caseins with five, six, seven and five phosphorylation sites respectively (Ginger and Grigor, 1999; Mamone et al., 2003; Girardet et al., 2006; Poth et al., 2008), African elephant β-casein has the least phosphorylation sites.

The third aim of this study was to determine the secondary structure of caseins by structure modeling, in order to predict their possible structure based effects in casein interactions and subsequent micelle formation. In biological systems, the structure of proteins is usually linked to their function (Krieger et al., 2003; Krieger et al., 2012), therefore determining the structure of proteins may give insight into their function. Using I-TASSA (a structure modeling software program), the structure models of cow, sheep, horse, human and African elephant were modeled.

Homology models of α_{s1}- and α_{s2}-caseins composed predominantly of α-helices and were very similar amongst the species under investigation. Because of the

dominance of α-helices in the secondary structure of α-caseins, it may suggest that the α-caseins may play a crucial role in maintaining a compact but not rigid micelle since they are located at the core of a casein micelle (Holt and Sawyer, 1988; Horne, 1998). In contrast, the secondary structure of β- and κ-caseins is dominated by random coils. A comparison of homology models of cow, human and sheep β-casein show high homology, while those of horse and African elephant show little homology. These differences suggest that these β-casein molecules may not interact with other caseins via similar mechanisms during casein micelle formation as proposed for cow casein micelles.

Homology models of κ-caseins were observed to be structurally distinct from each other. The main function of κ-casein is to stabilise casein micelles by steric stabilization. The difference in κ-casein models may also suggests that κ-caseins in milk of the compared species function differently. Although it cannot be claimed that the exact structure of the caseins are as predicted in this study, it may be likely that the caseins with similar predicted structures between species also assume a similar structure in the respective casein micelles, while the opposite accounts for the caseins with different predicted structures.

The final aim of this study was to investigate the distribution of casein genes in the mammalian kingdom by doing comparative genomics. As stated before, not all mammals have all four caseins in their milk although all milk that has been studied so far shows the presence of casein micelles (Martin et al., 2013; Madende et al., 2015). The α_{s1}-casein encoding gene is the first casein gene that developed whereas

the κ-casein encoding gene was the last (Martin et al., 2013). The comparative genomics data of casein genes shows that there are more mammalian species with CSN2 and CSN3 genes than those with CSN1S1 and CSN1S2 genes. This suggests that CSN2 and CSN3 genes and their gene products may have a much bigger and important role to play in casein micelle formation.

Human milk lacks α_{s2}-casein, it may be possible that its role in casein micelle formation could be shifted to α_{s1}-casein which is capable of forming disulfide-linked heteromultimers with κ-casein (Martin et al., 2013). This demonstrates that caseins can be multifunctional with regards to micelle formation and as a result the presence of all four (sometimes five) caseins may not be a prerequisite for casein micelle formation in milk.

Considering the order of casein genes development, it appears that African elephant lost the CSN1S1 gene and did not develop the CSN1S2 gene. In addition to gene loss, some mammalian species have paralogs of CSN2 and CSN1S2 genes which also reiterates earlier findings that caseins are rapidly evolving genes (Kawasaki et al., 2011). Since the casein micelle is central to this study, the above could mean that CSN2 and CSN3 genes are more important in the formation of casein micelles than α-casein genes and the former have evolved into being multifunctional and therefore possibly nullifying the role of α-caseins in casein micelle formation.

From the alignment of casein gene sequences, it was also apparent that these sequences are very divergent. The diverse nature of casein gene sequences is as a result of events such as exon skipping as well as rapid evolution. It appears that homology among casein sequences is not preserved and therefore it may suggest that the functional properties of caseins are not limited to a specific primary sequence.

SEM studies have shown that African elephant milk casein micelles are much larger compared to cow casein micelles. The difference in appearance seems to be related to the casein proportions and properties of the respective mammals. African elephant milk does not have any α-caseins which were predicted to adopt α-helical secondary structure. This structure conformation may influence cow micelles to be slightly more compact and reduce their size. In contrast, African elephant milk micelles, in which α-caseins are absent, may have a looser structure and therefore appear larger under microscopic examination.

The divergent gene sequences of African elephant β- and κ-caseins also influence their secondary structure which is different from that of cow β- and κ-caseins. This also appears to contribute to the overall larger size of African elephant casein micelles. Finally, African elephant milk is dominated by unphosphorylated and singly phosphorylated β-casein. This may limit phosphate binding, which results in limited formation of nano-clusters between β-casein molecules, which could further influence the size of African elephant casein micelles, compared to fully phosphorylated β-casein casein of cow milk.

6.1. Future research

Since casein micelles are complex structures, future work could include more sensitive microscopic examinations, such as confocal microscopy, where the effect of fixation methods is eliminated. This may give a more accurate comparison of casein micelles from different mammalian species' milk. Furthermore, the investigation of casein micelles of non-eutherian mammals, the monotremes and marsupials, may shed more light on the casein micelle structure.

Further experiments on the elucidation of African elephant β-casein PTM to locate other phosphorylation sites would involve isolation and MS analysis of individual spots on gels. Finally, as genome databases and comparative genomics tools increase and are improved, comparative genomics combined with structure modeling may contribute to an understanding of casein micelle formation and stability.

6.2. REFERENCES

Dalgleish, D.G., P.A. Spagnuolo, and H. Douglas Goff. 2004. A possible structure of the casein micelle based on high-resolution field-emission scanning electron microscopy. *Int Dairy J*. 14:1025–1031. doi:10.1016/j.idairyj.2004.04.008.

Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk-sixth revision. *J Dairy Sci*. 87:1641–1674. doi:10.3168/jds.S0022-0302(04)73319-6.

Fox, P.F., and P.L.H. Mcsweeney. 1998. Dairy chemistry and biochemistry. 1542. 478 pp.

Ginger, M.R., and M.R. Grigor. 1999. Comparative aspects of milk caseins. *Comp Biochem Physiol Part B*. 124:133–145.

Girardet, J.M., L. Miclo, S. Florent, D. Molle, and J.L. Gaillard. 2006. Determination of the phosphorylation level and deamidation susceptibility of equine β-casein. *Proteomics*. 6:3707–3717. doi:10.1002/pmic.200500728.

Holland, J.W. 2008. Post translational modifications of caseins. *In* Milk Proteins. 107–132.

Holt, C. 2015. Casein and casein micelle structures, functions and diversity in 20 species. *Int Dairy J*. 60:2–13. doi:10.1016/j.idairyj.2016.01.004.

Holt, C., J.A. Carver, H. Ecroyd, and D.C. Thorn. 2013. Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J Dairy Sci*. 96:6127–46. doi:10.3168/jds.2013-6831.

Holt, C., and L. Sawyer. 1988. Primary and predicted secondary structures of the caseins in relation to their biological functions. *Protein Eng Des Sel*. 2:251–259. doi:10.1093/protein/2.4.251.

Horne, D.S. 1998. Casein Interactions : Casting light on the black boxes , the structure in dairy products. *Int Dairy J*. 6946:171–177.

Horne, D.S. 2008. Casein micelle structure and stability. *In* Milk proteins. H.S. T.

Abby, B. Mike, editor. Academic press, San Diego, CA. 133–162.

Kawasaki, K., A.G. Lafont, and J.Y. Sire. 2011. The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol Biol Evol*. 28:2053–2061. doi:10.1093/molbev/msr020.

Krieger, E., S.B. Nabuurs, and G. Vriend. 2003. Homology modeling. *Struct Bioinforma*. 857:507–508. doi:10.1007/978-1-61779-588-6.

Krieger, E., S.B. Nabuurs, and G. Vriend. 2012. Homology Modeling. *Homol Model Methods Protoc*. 857:419. doi:10.1007/978-1-61779-588-6.

Madende, M., G. Osthoff, H.-G. Patterton, H.E. Patterton, P. Martin, and D.J. Opperman. 2015. Characterization of casein and alpha lactalbumin of African elephant (Loxodonta africana) milk. *J Dairy Sci*. 98:8308–18. doi:10.3168/jds.2014-9195.

Mamone, G., S. Caira, G. Garro, A. Nicolai, P. Ferranti, G. Picariello, A. Malorni, L. Chianese, and F. Addeo. 2003. Casein phosphoproteome: Identification of phosphoproteins by combined mass spectrometry and two-dimensional gel electrophoresis. *Electrophoresis*. 24:2824–2837. doi:10.1002/elps.200305545.

Martin, P., C. Cebo, and G. Miranda. 2013. Interspecies comparison of milk proteins: Quantitative variability and molecular diversity. *In* Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects, 4th Edition. 387–429.

Phadungath, C. 2005. Casein micelle structure : a concise review. *J Sci Technol*. 27:201–212.

Poth, A.G., H.C. Deeth, P.F. Alewood, and J.W. Holland. 2008. Analysis of the human casein phosphoproteome by 2-D electrophoresis and MALDI-TOF/TOF MS reveals new phosphoforms. *J Proteome Res*. 7:5017–5027. doi:10.1021/pr800387s.

Potočnik, K., V. Gantner, K. Kuterovac, and a Cividini. 2011. Mare's milk: composition and protein fraction in comparison with different milk species. *Mljekarstvo*. 61:107–113.

Tuinier, R., C. Rolin, and C.G. de Kruif. 2002. Electrosorption of pectin onto casein

micelles. *Biomacromolecules*. 3:632–638. doi:10.1021/bm025530x.

Yang, Y., D. Bu, X. Zhao, P. Sun, J. Wang, and L. Zhou. 2013. Proteomic analysis of cow, yak, buffalo, goat and camel milk whey proteins: Quantitative differential expression patterns. *J Proteome Res*. 12:1660–1667. doi:10.1021/pr301001m.

SUMMARY

The exact structure of casein micelles still remains a debated subject. While most of the experimental work on cow caseins and casein micelles has provided a wealth of data, data of caseins and casein micelles of non-bovine origin provide a new insight into the structure of casein micelles. Microscopic examination of cow, sheep, horse, human and African elephant milk casein micelles show that the respective casein micelles are all spherical in shape but differ in size as well as surface appearances. Human casein micelles were the largest of the casein micelles whereas sheep casein micelles were the smallest. Apart from their smaller size, sheep micelles also had a smooth surface compared to a rough surface observed on the rest of the casein micelles. African elephant casein micelles were the second largest of the five casein micelles compared. It may be derived that, although casein micelle shape and size seem to be species specific, the differences observed may be a result of the differences in total casein content, the proportions of the individual casein types and the presence and or absence of some of the casein types.

The elucidation of African elephant β-casein phosphorylation state by LC MS/MS, showed the presence of a single phosphorylation site at Ser9. In contrast, electrophoresis analysis showed that there are up to five phosphoforms of African elephant β-casein. The LC MS/MS also showed that the presence of a short length African elephant β-casein that is 200 amino acids long and that the gene sequences coded for by exons 4 and 5 have been truncated.

Homology modeling of cow, sheep, horse, human and African elephant caseins showed that the secondary structure of α-caseins predominantly consist of α-helices, whereas the secondary structure of β- and κ-caseins is dominated by random coils. Alpha caseins give micelles a slightly compact structure whereas random coils result in a more open and larger size of micelles. These structural differences of caseins could possibly explain the varied size of casein micelles in milk. Comparative genomics of casein genes across mammalian species shows that several mammalian species are devoid of CNS1S1 and CSN1S2 genes. Considering the evolution of the casein gene locus organization, it appears that the CNS1S1 gene has been lost whereas the CSN1S2 gene has not been gained or developed in these species. In contrast, the CSN2 and CSN3 genes have been preserved and gained respectively, in most mammalian species. This suggests that these genes have a more important role in casein micelle formation and consequently the sequestration of large amounts of calcium and phosphate. Evidence from this study suggests that studying of non-cow caseins may shed more light on the casein micelle structure.

Keywords: African elephant; Casein micelle; Phosphorylation; Mass spectrometry; Comparative genomics; Structure modeling.

OPSOMMING

Die presiese struktuur van kaseïenmiselle bly ,n debateerbare onderwerp. Terwyl die meeste eksperimentele werk op bees kaseïene en kaseïenmiselle ,n magdom data gelewer het, dra data van kaseïene en kaseïenmiselle van nie-bees oorsprong by tot insig in die stuktuur van kaseïenmiselle. Mikroskopiese ondersoeke van kaseïenmiselle van bees, skaap, perd, mens en Afrikaanse olifant wys dat die onderskeie kaseïenmiselle almal sferies in vorm is, maar verskil in grootte en oppervlakvoorkoms. Menslike kaseïenmiselle was die grootste, terwyl dié van skaap die kleinste was. Buiten die kleiner grootte, het skaapkaseïenmiselle ,n gladde oppervlak vergeleke met die growwe oppervlak van die ander. Kaseïenmiselle van Afrikaanse olifant was die tweede grootste van die vyf bestudeerdes. Daar kan afgelei word dat, alhoewel die vorm en grootte van kaseïenmiselle spesiespesifiek voorkom, die waargenome verskille ook die resultaat mag wees van verskille in totale kaseïeninhoud en/of die afwesigheid van sekere kaseïentipes.

Die opklaring van Afrikaanse olifant β-kaseïen fosforileringstatus deur LC MS/MS, het die teenwoordigheid van ,n enkele fosforileringspunt by Ser9 aangedui. In kontras het elektroforetiese analise voorgetsel dat to vyf fosfo-forme van die afrikaanse olifant β-kaseïen mag voorkom. Die LC MS/MS het ook gewys dat die Afrikaanse olifant β-kaseïen 200 aminosure lank is en dat die geengebiede wat deur eksons 4 en 5 kodeer word, verkort is.

Homologie modelering van bees, skaap, perd en Afrikaanse olifant kaseïene dui daarop dat die sekondêre struktuur van α-kaseïene hoofsaaklik bestaan uit α-helikse, terwyl die sekondêre struktuur van β- en κ-kaseïene gedomineer word deur willekeurige struktuur. Alfa kaseïene gee aan miselle ‚n gedeeltelik stewige struktuur, terwyl willekeurige struktuur lei tot ‚n oop struktuur en groter miselle. Hierdie strukturele verskille mag moontlik die variasie in kaseïenmiselgroottes in melk verklaar.

Vergelykende genomika van kaseïengene oor soogdierspesies heen wys dat sekere soogdiespesies nie die CNS1S1 en CSN1S2 gene besit nie. Wanneer die ewolusie van die rangskikking van die kaseïengeenlokusse in ag geneem word, kom dit voor asof die CNS1S1 geen verlore geraak het, terwyl die CSN1S2 geen nie in hierdie spesies ontwikkel het nie. Daarteenoor was die CNS2 en CSN3 gene beskerm of verwerf in die meeste soogdierspesies. Dit stel voor dat hierdie gene ‚n belangriker rol speel in die vorming van kaseïnmiselle en gevolglike sekwestrasie van groot hoeveelhede kalsium en fosfaat. Bewyse uit hierdie studie stel voor dat die bestudering van nie-koei kaseïene meer lig mag werp op die struktuurkennis van kaseïenmiselle.

Sleutelwoord: Afrikaanse olifant; kaseïenmisel; fosforilering; massaspektrometrie; vergelykende genomika, struktuurmodelering