A proteomic study of African elephant milk: Inter-species comparisons and proteome dynamics

Moses Madende

Submitted in accordance with the requirements for the degree

Magister Scientiae

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-supervisors: Prof. H-G Patterton

: Dr. H. E. Patterton

January 2014

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to the following:

God, most high, for his grace, wisdom and understanding he granted me throughout this study. May all the glory, honor and praise be directed to Him.

- My supervisor Prof. G. Osthoff, my co-supervisors, Prof. H-G. Patterton and Dr. E. Patterton. I am extremely grateful and indebted to them for their expert, sincere and valuable guidance, support and encouragement in this research.
- Dr. G. Kemp and Dr. S. Smit for their assistance with mass spectrometry analysis as well as Dr. D. Oppermann for assistance with structure modeling.
- My parents and the rest of my dear family. It is in this support structure and their unconditional love that I thrived whenever I found myself wanting. And also for the myriad of ways they have supported me to find and realize my potential, I will forever be grateful.
- National Research Foundation (NRF) and UFS Cluster, Advanced Biomolecular Research, for partial funding of the project.

I also place on record, my sense of gratitude to my friends and laboratory colleagues, who knowingly and unknowingly, led me to the understanding of some subtle challenges in my work as well as their continual inspiration.

This thesis is dedicated to my family!

CONTENTS

Chapter 2: Separation of African elephant milk proteins by 1D Electrophoresis and partial amino acid sequence determination by LC-MS/MS.

Chapter 3: Purification of African elephant milk proteins by 2D Electrophoresis and amino acid sequence determination by LC-MS/MS and Orbitrap MS.

Chapter 1

Literature review

1.1. Motivation

Milk has co-evolved with mammals and serves as a complete, complex and key element suited for specific nutritional requirements of the neonate (Casado *et al*., 2009; D'Allesandro *et al*., 2010). In addition to the role of milk as an enhancer of growth and development, immunoglobulins present in milk carry out a protective role, whilst enzymes and binding or carrier proteins aid in digestion, and hormones in milk impact normal growth of the neonate (Roncada *et al*., 2012). Breast milk can be further described as a functional food; that is, it contains health promoting compounds such as oligosaccharides, as well as proteins that impact on physiological functions (Casado *et al*., 2009). In essence, milk is an extremely complex fluid and comparative studies of bioactive, nutritional and protective compounds in milk, even between a few mammalian species, can be a daunting task (McClellan *et al.*, 2008). Milk from a number of mammalian species has been investigated and it was concluded that there are marked differences in terms of the nutritional composition; this diversity is as a result of uniqueness of the nutritional and physiological requirements of each species (Casado *et al*., 2009). In order to improve production and to enhance quality and safety of milk containing foods, great interest has developed over the last decade to study the proteome of milk. Different research techniques have been combined to explore genetic aspects, molecular pathways, and cellular functions involved in milk production, quality and safety, in order to gain a multifaceted picture addressing the complexity and multiplicity components of milk (Roncada *et al.*, 2012). Furthermore, a number of substantiated, or potentially bioactive milk proteins, have been identified and many more remain to be identified, either as intact proteins or derived peptides encrypted in the sequence of milk proteins (Gobbetti *et al*., 2002).

In the past few years, reports dealing with the identification of the low abundant milk proteins have increased (Gagnaire *et al*., 2009). These findings are useful for characterization of pathways and mechanisms that occur during lactation and give information on the biological activity and functionality of these important proteins (Roncada *et al.*, 2012). Indeed, it remains one of the biggest challenges in dairy science to provide the consumers and the food industry with the basis for these bioactive milk compounds before their inclusion as ingredients in functional foods, especially in infant formulas.

The advent of two dimensional polyacrylamide gel electrophoresis (2D PAGE) and mass spectrometry (MS), which are employed in the proteomics approach of protein research, has allowed the unraveling and attainment of previously unknown information on milk proteins (Claverol *et al*., 2003; Holland *et al*., 2006). This study therefore aims to investigate African elephant milk proteins using the proteomics approach, in order to gain more knowledge on the more abundant caseins and the less abundant whey proteins. In addition to obtaining information about the African elephant milk proteome, this study will also contribute to the knowledge of milk proteins in general.

1.2. Phylogenetic aspect of mammals

Mammals are distinctive in that they possess specialized mammary glands that produce milk which is intended to feed the neonate (Simpson, 1945; McKenna & Bell, 1997; Fox & McSweeney, 1998). The two main divisions or sub-classes of Mammalia are Prototheria and Theria. Prototheria consists of only one order, the Monotremata (egg laying mammals) for example the duck-billed platypus and four echidna species. Milk composition of Monotremes tends to change progressively during the different stages of lactation, which is important to meet the neonatal growth requirements at different stages of development (Fox & McSweeney, 1998).

Unlike Monotremes, the offspring of Theria are born live and fall into two main infraclasses, the Placentalia and Marsupialia (Fox & McSweeney, 1998). The Theria neonate rely solely on milk as their source of nutrition and the milk composition progressively changes throughout the long lactation stage to meet the nutritional requirements of offspring. In terms of composition of the individual nutritional elements, milk of marsupials generally tends to increase in protein and lipid content while on the other hand carbohydrate levels drop. Milk of placentalian species remains more or less constant in terms of composition of nutrients (Fox & McSweeney, 2002).

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 (Larsen, 1985; Anderson *et al*., 2007; Lonnerdal, 2007; Shennon, 2008). Figure 1.1 depicts the five major milk biosynthesis and secretion pathways in the secretory epithelial cells into the alveoli lumen where milk components collect.

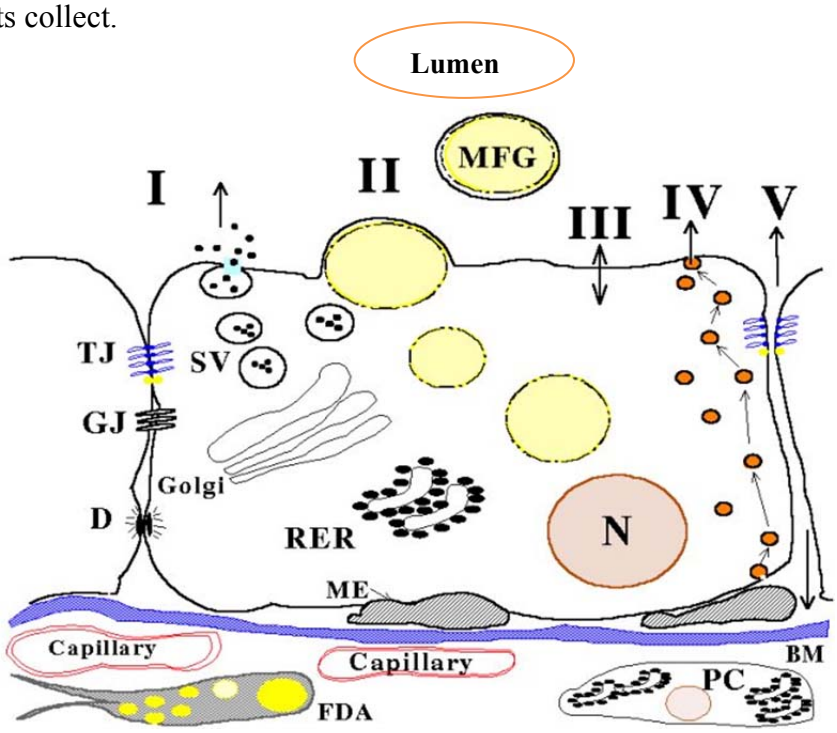

Figure 1.1. Cellular mechanisms for biosynthesis and secretion of milk components in the lactating mammary tissue (alveolar cell) as described in the text (Neville, 1995).

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 & 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.

In the nucleus of the cell, gene expression starts with transcription by RNA polymerase where a copy of the DNA template is synthesized in the form of mRNA which contains the base uracil instead of thiamine in DNA. The mRNA which is now the blueprint for the protein and determines the amino acid sequence of the protein it codes for moves to the cytosol and the endoplasmic reticulum (ER) where translation occurs to form a polypeptide with the aid of tRNA on the ribosome. 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 (Singh *et al*., 2008). 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). Also transported via this pathway is lactose, some minerals and water (Shennon & Peaker, 2000). The major milk proteins which include α_{s1}-, α_{s2}-, β-, κ-caseins, α-lactalbumin (α-LA), β-lactoglobulin and immunoglobulins 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.

The co- and posttranslational modification of proteins then occurs in the ER lumen, as well as by the proteolytic removal of the signal peptides. 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 may occur (Burgoyne & Duncan, 1998). This occurs during the transportation to and inside the Golgi apparatus. 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 & 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 & 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

Milk proteins consist of a wide variety of large and small proteins which are mostly mammary gland derived (Bequette *et al*., 1998). Proteins in milk are mainly found in the aqueous phase, either in the soluble or colloidal state, as well as in the lipid phase (Larsen, 1985). Milk proteins are generally categorized into four main groups: caseins, whey, milk fat globule membrane (MFGM) and peptones (D'Alessandro *et al*., 2010). Bovine caseins represent about 80 % of the total milk proteins; they compose of α_{s1}-, α_{s2}-, β- and κ- caseins (Yamada *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-LA and β-lactoglobulin (β-LG) are the main whey proteins (Fox & McSweeney, 1998). Because of their relative abundance, whey and casein proteins have been widely studied by a combined effect of MS and electrophoresis approaches. Peptone, that is low molecular weight peptides and proteins in the 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).

Some proteins are present in milk at low abundance, these proteins are mainly blood derived, and thus they are taken up without further processing into milk via a transcellular or paracellular route (Shennon & Peaker, 2000). Additionally these proteins are mainly immune related for example immunoglobulins, serum albumin and lactoperoxidase, which are taken up by passive diffusion and/or active transport into the mammary cell. A number of factors affect the synthesis of milk proteins in the mammary tissue and hence the level of individual proteins in milk; these factors include diet, endocrine control, milking frequency, stage of lactation, mastitis and lastly breed differences.

1.4.1. Caseins

Caseins in milk exist as colloidal aggregates known as the casein micelles which are aggregations of caseins and mineral calcium and phosphates (Walstra, 1999). The 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.

Caseins exert properties and sequences that are different from each other. This is a consequence of varying levels of post-translational phosphorylation of serine (or threonine) and/or glycosylation of threonine residues, mutational changes in casein genes, proteolysis by indigenous milk proteases or oxidation of cysteine to disulphide bonds (Swaisgood, 1992; Martin *et al*., 2003). Inability to crystallize caseins made it challenging to determine their structure using X-ray crystallography (Swaisgood, 1992).

Caseins are not globular in structure, thus they lack strategically placed cysteine residues that stabilize the structure of globular proteins (Swaisgood, 1992). The key to their structure stabilization is formation of intermolecular disulphide bonds between κ-casein molecules. There are two universal casein properties in milk, these include the chymosin cleavage site in κ-casein which is critical for coagulation of casein micelles, and the phosphorylation sites important for proper bonding of the caseins to hydrated calcium phosphate entities present in casein micelles.

Bovine α_{s1}-casein contains 199 amino acids and does not have any cystein residues in its sequence (Martin *et al.*, 2003). This component of caseins has the highest net negative charge in neutral pH buffer with only monovalent cations present (Farrell *et al.*, 2004). It 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.*, 2003). 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 casein and contains a very acidic region between residues 38 and 78 that is responsible for calcium binding (Farrell *et al.*, 2004). Circular dichroism and Raman spectral analysis indicate the presence of about 14 % α-helix, 40 % β-sheet and 24 % turn-like structures. 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.

Bovine α_{s2}- caseins, the most highly and variably phosphorylated of the calcium sensitive caseins was the last bovine casein to be sequenced and it consists of 207 amino acids (Martin *et al*., 2003). In addition, this group of caseins is also the least hydrophobic (Farrell *et al*., 2004). It occurs in milk in several forms with phosphorylation ranging from 10-13 phosphate groups (Eigel *et al*., 1984). 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 sequence comparison (Ginger & Grigor, 1999). The majority of the protein occurs with 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 as well as polymers with κ-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).

β-casein is a major component of casein proteins and it is the most hydrophobic casein, furthermore, it does not contain a cysteine and the sequence consists of 209 amino acid residues (Greenberg *et al*., 1984; Martin *et al*., 2003). In solution, β-casein forms detergent like micelle aggregates and this is due to its amphipathic nature (Farrell *et al*., 2004). In addition, bovine β-casein consists of six proteins with similar amino acid sequence only differing in the number of phosphate groups attached to the serine residue, which ranges from 0-5. Hydrolysis of β-casein by plasmin at regions (105-106/107-108) yields γ-casein, which is not originally present in milk during synthesis (Eigel, 1977; Swaisgood, 1992). The self association of β-casein is micelle-like, and both ionic strength and temperature increase the quantity of polymer present and the degree of association, this in effect reduces its cleavage by chymosin at high temperature.

Of all the proteins of the casein family, κ-caseins are the only proteins that are glycosylated (Swaisgood, 1992). The post translational glycosylation by short oligosaccharide chains occurs at one or more of the threonine sites (Farrell *et al.*, 2004). Bovine κ-caseins, the most studied milk protein, consists of 169 amino acids and is the target of chymosin. 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 (Martin *et al*., 2003). Additionally, as opposed to other caseins, κ-casein does not bind calcium extensively and thus it is not sensitive to calcium precipitation.

The destabilization of the 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 (Delfour *et al*., 1965; Jolle`s *et al*., 1968; Hennighausen & Sippel, 1982). The two domains 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. 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.

1.4.2. The casein micelle

There are four gene products in the milk of most mammalian species that, after posttranslational modification, give rise to α_{s1}-, α_{s2}-, β- and κ-caseins (Farrell *et al*., 2004). Caseins in fresh milk exist as spherical particles composing of many protein molecules and calcium phosphate with a diameter size range between 15 and 1000 nm which are ultimately known as the casein micelle with a very open, dynamic and highly hydrated structure (Rollema, 1992). Caseins are evenly distributed within the micelle with the exception of κ-caseins, which are mostly found on the surface of the micelle. Almost all of the casein proteins that exist in bovine milk are incorporated into the casein micelles (Fox & McSweeney, 1998). Figure 1.2 shows the structure of the casein micelle and distribution of caseins in the micelle.

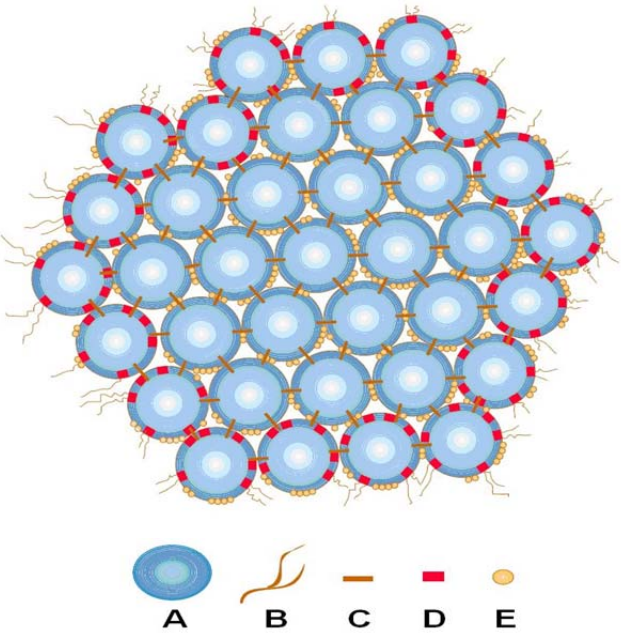

Figure 1.2. The casein micelle, A-submicelles, B-hair like structures, C-calcium phosphate molecule, D-κ casein and E-phosphate groups (Rollema, 1992).

Isolation of whole casein is achieved by acidic destabilization of the micelle suspension or by size fractionation which can give rise to pure caseins (Farrell *et al*., 2004). Complete disintegration of the casein micelles can be achieved by utilizing high levels of urea or strong calcium sequestrates such as ethylenediaminetetraacetic acid (EDTA) (Horne, 2008).

Micelle size remains constant with milk storage; furthermore the size is not changed on cooling or pasteurization. Fractionation of casein micelles can be achieved by subjecting to different levels of centrifugation, where-after the largest micelles are found in the pellet. Fractionation experiments have shown that the proportion of α_{s1}- and α_{s2}-caseins is not dependent on micelle size and κ-casein content is inversely proportional to micelle size (Horne, 2008). Calcium content of micelles per mole of casein are consistent with the increasing content of κ-casein and the compensating decrease in β-casein fraction with decreasing micelle size. The κ-casein component resides on the micelle surface, where it controls micelle surface area and hence size, other caseins and colloidal calcium phosphate and the other caseins are found primarily in the micelle core, contributing to volume.

The stability of the casein micelle strongly influences the technological properties of milk. The κ-casein molecules are thought to provide a steric stabilizing layer, with their hydrophilic C-terminal peptides protruding into the aqueous phase (Holt & Horne, 1996). Proteolysis at the Phe105-Met106 bond releases the hydrophilic peptide of κ-casein, which directly results in curd formation. Three models are at the fore front of explaining the structure of the casein micelle namely the sub-micelle model, the Holt model and the dual binding model.

The sub-micelle models suggest that the sub-micelles are linked by calcium phosphate (Rollema, 1992; Walstra, 1999). Production of the of κ-casein coat is thought to be as a result of two subpopulations of the subunits that are created, one rich in κ-casein found at the outer reaches and the other devoid of κ-casein forming the micelle core. The shortcoming of this model is that it does not explain the assembly of the two subunits and the driving force for creation of subunits of different composition. The Holt model suggests that calcium phosphate nano-clusters dispersed through the

micelle are the nodes for micellar assembly. They interact with the phosphoserine clusters of the casein molecules. Since individual α_{s1}- and α_{s2}-caseins have more than one phosphoserine, cross-links and networks are possible, which results in a micro-gel particle. The model provides an explanation for the heterogeneity in appearance and scattering behavior of the micelle. The short coming of this model is that there is no inherent mechanism that limits growth which could continue to lead to a giant macro-gel.

The dual-binding model suggests that micellar assembly and growth takes place by a polymerization process involving two distinct forms of bonding, which are hydrophobic interactions of the caseins and bridging across calcium phosphate nano-clusters (Horne, 1998). Central to the model is that micellar stability is maintained by an excess of hydrophobic attraction over electrostatic repulsion. This model also suggests that individual caseins behave and interact by self-association, thus interaction of their hydrophobic regions (Figure 1.3). The α_{s1}-casein molecules form a chain polymer giving a worm like chain whereas β-caseins give rise to detergent like micelles as shown in Figure 1.4. Phosphoserine negative charges are neutralized by incorporation into calcium phosphate nanoclusters thereby allowing entrance of more α_{s1}-caseins to enter and provide further options for cross-linking to other polymerization paths, β-caseins forms detergent-like micelles utilizing their hydrophobic regions and phosphoserine cluster, thus forming further polymer links and ultimately chain growth.

In this dual binding model, κ-caseins are central to assembly and structure of the casein micelle. They link into, and extend, the growing polymer chain via N-terminal blocks, whereas chain termination occurs at the C-terminal with no phosphoserine cluster, thus leaving the micelle with an outer κ-casein layer. The dual binding model shows assembly, growth and lastly termination of micellar growth (Figure 1.5), which is a shortcoming in other models.

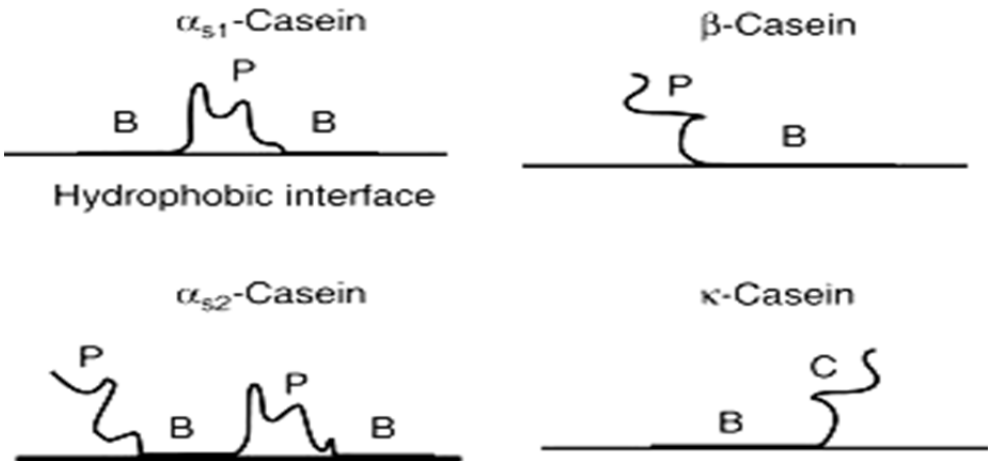

Figure 1.3. Properties of individual caseins that enable self-association as well as interaction with other caseins (Horne, 1998).

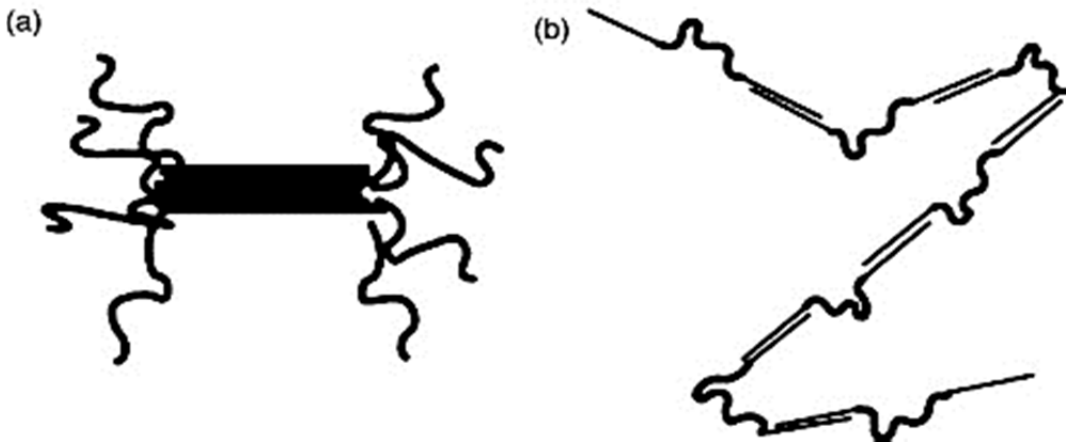

Figure 1.4. Self-association of caseins, (a) detergent like micelle of β-caseins and (b) worm like micelle of α_{s1}-caseins (Horne, 1998).

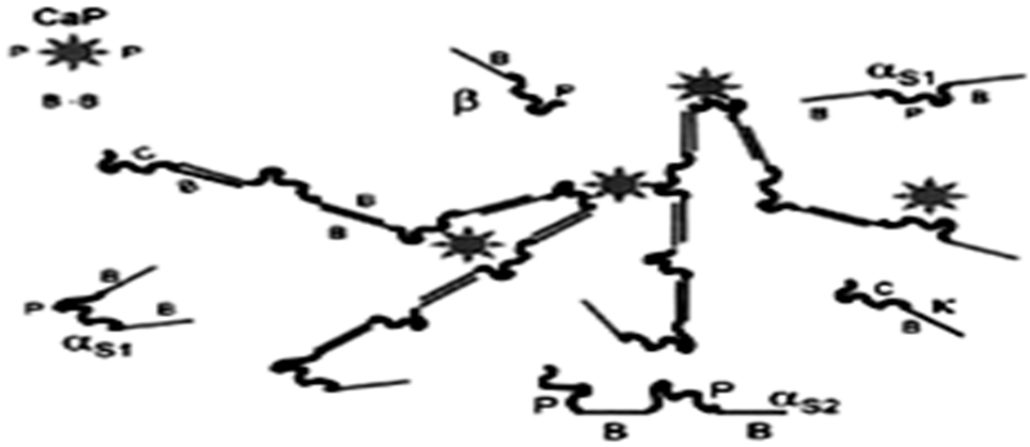

Figure 1.5. Assembly of the micelle through hydrophobic interaction and calcium phosphate molecule that paste the caseins together (Horne, 1998).

This model also satisfies the appearance and scattering behavior shown by the native micelle, as well as destabilization of the micelle after κ-casein hydrolysis by chymosin, the first step in cheese making. In addition, this model imposes no requirements at amino acid level or defined secondary structure, the requirements solely being that caseins are amphiphilic and the majority possesses phosphoserine. Initially it was thought that mare's milk does not contain κ-casein. However, it still has micelles and sterical stability. Chain termination was thought to be fulfilled by dephosphorylated β-caseins (Horne, 1998). Absence of a phosphoserine cluster would thus prevent its entrance into chain polymerization. It is thought that the same occurs in human milk, which also contains very little κ-casein (Roncada *et al*., 2012). It has however been discovered recently that mare's milk does contain κ-caseins, albeit at very low levels compared to human and cow's milk (Iametti *et al*., 2001; Malacarne *et al*., 2002). In terms of micelle size, mare's has larger micelles than cow and human milk.

The native states of α_{s1}-, α_{s2}-, β-, κ-caseins are the states in which they exist when fully immersed in the micelle. The hydrophobic regions of all the caseins are intermingled, with the κ-casein molecules close to each other. Additionally, the very acidic and very basic regions are bound to hydrated calcium phosphate in a micelle.

1.4.3. Whey proteins

Caseins in bovine milk can be isoelectrically precipitated at pH 4.6; the resultant protein fraction that remains in solution is referred to as the whey proteins (O'Donnell *et al*., 2004). In addition to being globular in structure, whey proteins have a more organized secondary as well as tertiary structure. Whey proteins provide a wide variety of nutritional, biological and food functional attributes with the main constituents being alpha lactalbumin (α-LA) and beta lactoglobulin (β-LG) (Chatterton *et al*., 2006). Serum albumin, α-LA, β-LG and immunoglobulins account for over 95 % of the non-casein proteins.

α-LA was first isolated over 70 years ago and soon became a model for early development of methods for investigation of the chemical and biophysical properties of proteins because of its ease to isolate and relative abundance (Brew *et al*., 1967). It constitutes 3.4 % of total protein in bovine milk and 20 % of whey proteins, with a concentration of 1-1.5 g l^{-1} (Chatterton *et al*., 2006). However, in human milk, α-LA is the most abundant whey protein in terms of quantity comprising of 21-34 % of whey proteins, depending on the stage of lactation.

The concentration of α-LA in mature human milk is approximately 2.447 g l^{-1}. Sequence alignment has shown that there is approximately 76 % amino acid homology between bovine and human α-LA (Seivers *et al*., 2011); a similar homology also exists between human α-LA and α-LA in milks of other mammalian species (Pike *et al*., 1996). Similarly α-LA and lysozyme C have amino acid sequence elements in common, furthermore the two proteins have a similar 3-D structure, and in addition, a similar gene structure (Piotte *et al*., 1997). An amino acid sequence identity of up to 40 % has been observed between hen egg white lysozyme and α-LA.

Divergence evolution of lysozyme and α-LA is shown in Figure 1.6. Approximately 500 million years ago, the progenitor gene split into two genes, the insect lysozyme gene, and the gene that later split to give rise to the lysozyme C and α-LA genes 300 million years later. The α-LA gene later split into genes that code for a variety of α-LA from different species 50 million years later.

All α-LA contain a tightly bound Ca^{2+} which is important for conformational stability and structure (Kronman *et al*., 1981; Brew *et al*., 1967). In addition to elucidation of the amino acid sequence of α-LA and comparison between species, X-ray crystallography structures of α-LA from goat, cow, human, baboon and mouse have been solved. These structures are similar to each other and also share structural similarity to lysozyme C (Pike *et al*., 1996).

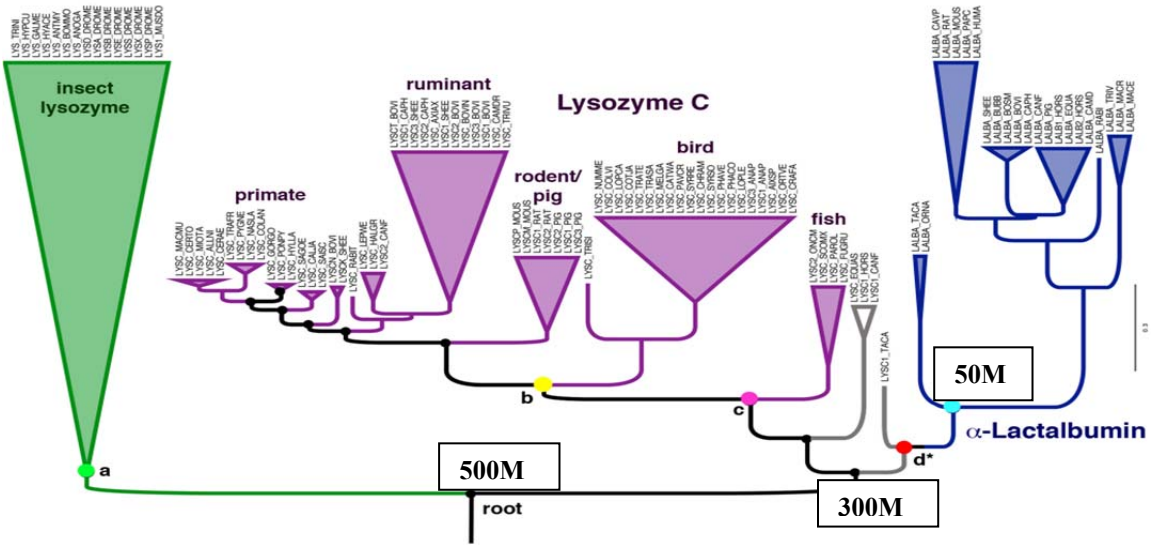

Figure 1.6. Divergence evolution of lysozyme and α-LA (theobold.brandeis.edu).

Bovine α-LA is initially synthesized as a 142 amino acid pre protein, which is cleaved to produce a 123 amino acid containing mature α-LA with a molecular weight of 14.128 kDa (Brew *et al*., 1967). Additionally, other mammalian mature α-LA amino acid sequences range from 121 to 140 amino acids. Amino acid sequence differences between mammalian species tend to be directly related to evolutionary divergence between species, for example α-LAs from eutherian and monotreme mammals have only less than 50 % amino acid identity.

The enzyme α-LA is specific to milk and the mammary gland and has high homology to lysozyme. α-LA is synthesized in the endoplasmic reticulum, it makes its way to the Golgi apparatus where, together with β-galactosyltransferase, initiates lactose synthesis, and the substrates for lactose synthesis are also present in the Golgi apparatus (Burgoyne & Duncan, 1998). Lactose, together with milk proteins, is then carried by the Golgi vesicles as they bud off from the Golgi apparatus to the plasma membrane for secretion. Galactosyltransferase is localized on the Golgi apparatus and is membrane bound. Thus lactose synthesis occurs on the luminal side of the Golgi body. Table 1.1 shows the enzyme catalyzed reactions that occur from the cytosol to the Golgi apparatus in lactose synthesis.

Table 1.1. Enzyme catalyzed reactions in lactose synthesis (Holden *et al*., 2003).

	Reaction	Enzyme catalyzed step	Product
1	Glucose + ATP	i →	Glucose-6-P + ATP
2	Glucose-6-P	ii →	Glucose-1-P
3	Glucose-1-P +UTP	iii →	UDP-glucose + PP_i
4	UDP-glucose	iv →	UDP-galactose
5	UDP-galactose + glucose	v →	Lactose + UDP

(i)hexokinase, (ii) phosphoglucomutase, (iii) UDP-glucose pyrophosphorylase (iv) UDP-galactose-4-Epimerase, (v) lactose synthetase; UDP, uridine diphosphate; UTP, uridine triphosphate.

After synthesis, lactose lacks the ability to move out of the Golgi apparatus on its own, movement is facilitated together with proteins via the secretory vesicles into the pool of milk inside the alveolar vesicle (Larsen, 1985). Presence of lactose inside the vesicle makes it hypertonic, so that water will be drawn from the cytosol into the secretory vesicle, thereby making lactose the major determinant of milk volume. The other important role of lactose in milk is its contribution as the readily available energy source for the neonate.

α-LA is present in almost all mammalian species except the Cape fur seal, California sea lion and primitive mammals (Martin *et al*., 2003; Urashima *et al*., 2001). The enzyme has been assigned many roles, the main one of course being its active involvement in the synthesis of lactose by the lactose synthase complex. Two proteins are involved in lactose synthesis and these are, β-1,4-galactosyltransferases 1 (β-1,4-GT1) and α-LA. In the mammary tissue, the lactose synthase becomes active when the substrate specificity of β-1,4-GT1 is directed towards glucose and this is done by the modulator enzyme α-LA (Nicholas *et al*., 1981; Ramakrishnan *et al*., 2001). The mammary gland lacks glucose-6-phosphatase, an enzyme that is required for glucose synthesis, thus for lactose to be synthesized glucose is extracted into the mammary tissue from the blood (Juhn *et al*., 1980).

Galactosyltransferases are enzymes that compose a family of peptides that are involved in the synthesis of carbohydrates that are bound on glycoproteins and glycolipids (Ramakrishnan *et al*., 2001). These enzymes catalyze the transfer of galactose (Gal) to sugar moieties from UDP-galactose. The milk galactosyltransferase β-1,4-GT1 is responsible for the transfer of galactose to N-acetylglucosamine (GlcNAc) to form β-1,4-linked galactosylated glycan. Furthermore, it is this particular galactosyltransferase (β4Gal-T1) whose specificity is altered by α-LA from Glc-NAc to glucose.

In the presence of α-LA, β-1,4-GT1 is approximately 30 fold faster in the transfer of galactose to glucose than in its absence. Galactosyltransferase can synthesize lactose on its own, without the involvement of α-LA, at very high concentrations of glucose. However, α-LA was found to increase the affinity of β-1,4-GT1 for glucose by bringing down the K_m for glucose to approximately 1mM which is within the physiological range of the glucose concentration in the cell (Ramakrishnan *et al*., 2001). In addition to glucose, the interaction between α-LA and β-1,4-GT1 occurs in the presence of N-acetylglucosamine, the cofactor Mn^{2+} and UDP-galactose.

Mutational studies have shown that Phe 31, His 32, Leu 110 are important residues of α-LA that are involved in glucose binding. The other residues involved in α-LA function are Gln 117 and Trp 118, which are involved in protein-protein stabilization of the lactose synthase complex. Figure 1.7 shows the synthesis of lactose in the mammary tissue by β-1,4-GT1 and α-LA complex with Mn^{2+} acting as a cofactor for this enzymatic reaction.

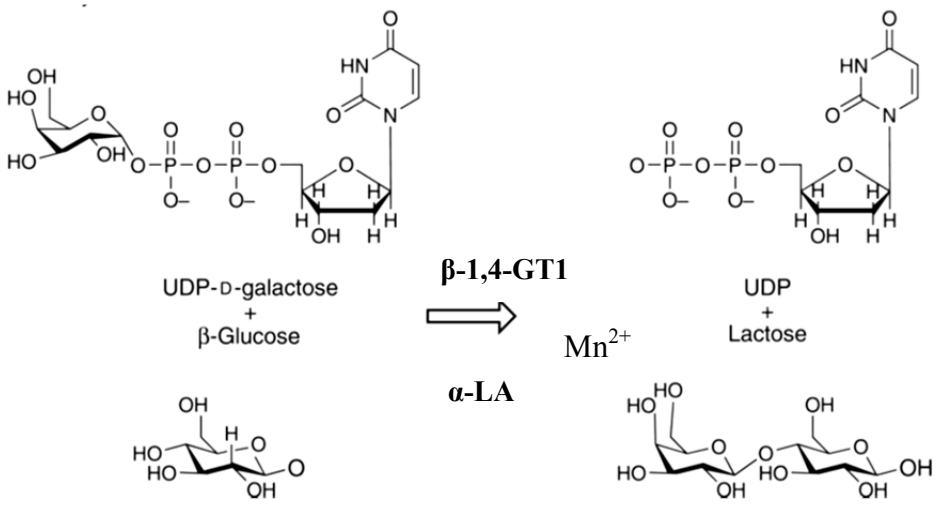

Figure 1.7. Lactose synthesis, α-LA regulates β-1,4-GT1 action (Ramakrishnan *et al*., 2001).

Important amino acid residues in β-1,4-GT1 that are involved in the interaction with α-LA for lactose synthesis are largely hydrophobic, these include Phe 280, Tyr 286, Gln 288, Tyr 289, Phe 360, and Ile 363. Figure 1.8 depicts the interaction of the two proteins with each other to form the lactose synthase complex and consequent selectivity of glucose as the substrate of the enzymatic reaction.

The amino acids that are involved in interaction of α-LA with β-1,4-GT1 are almost similar to those that are involved in the lytic function of lysozyme, the difference being their positioning on the peptide chain and some amino acid substitution (Piotte *et al*., 1997). Information about the difference in active sites of lysozyme and α-LA as obtained by amino acid sequencing has shown that α-LA does not have lytic activity (McKenzie & White, 1991). However, experimental data has demonstrated slight lytic activity of α-LA in various mammalian sources. Comparison of α-LA from different species gives an overview of the molecular mechanism of the specifier activity (Pike *et al*., 1996). The refined crystal structures of guinea-pig α-LA (GPLA), goat α-LA (GOLA) and bovine α-LA (mLA) are similar to those for human α-LA (HLA) and baboon milk α-LA.

Figure 1.8. X-ray structure depicting the interaction between bovine β-1,4-GT1 and mouse α-LA. The amino acid residues that are involved in the interaction are also highlighted as well as the substrate glucose and a manganese cofactor (Ramakrishnan & Qasba, 2007).

Figure 1.9 shows the three α-LA structures superimposed on each other and clearly demonstrate the structural similarities. The structures consist of two domains, the large α-domain and the small β-domain, separated by a deep cleft. The α-domain consists of three major helices and constitute of amino acid residues 1-34 and 86-123, on the other hand the β-domain, comprises amino acid residues 35-85. Variation in secondary structure of the mentioned α-LA occurs in residues 101-110.

The four α-LA under investigation are very similar and reflect a high level of amino acid sequence identity. Variation in the polypeptide backbone occurs in two loops in

the β-domain (residues 43–47 and 62–65), the region between helices H1 and H2 (residues 13–18) and in an area (residues 105–110) adjacent to the lower end of the cleft. Other differences in conformation occur in the C-terminal tail. The calcium binding site of the α-LA of the four species under comparison is essentially identical in terms of ligand coordination and conformation (Pike *et al*., 1996). The amino acid residues involved in calcium binding are Lys 79, Asp 82, Asp 84, Asp 87 and Asp 82.

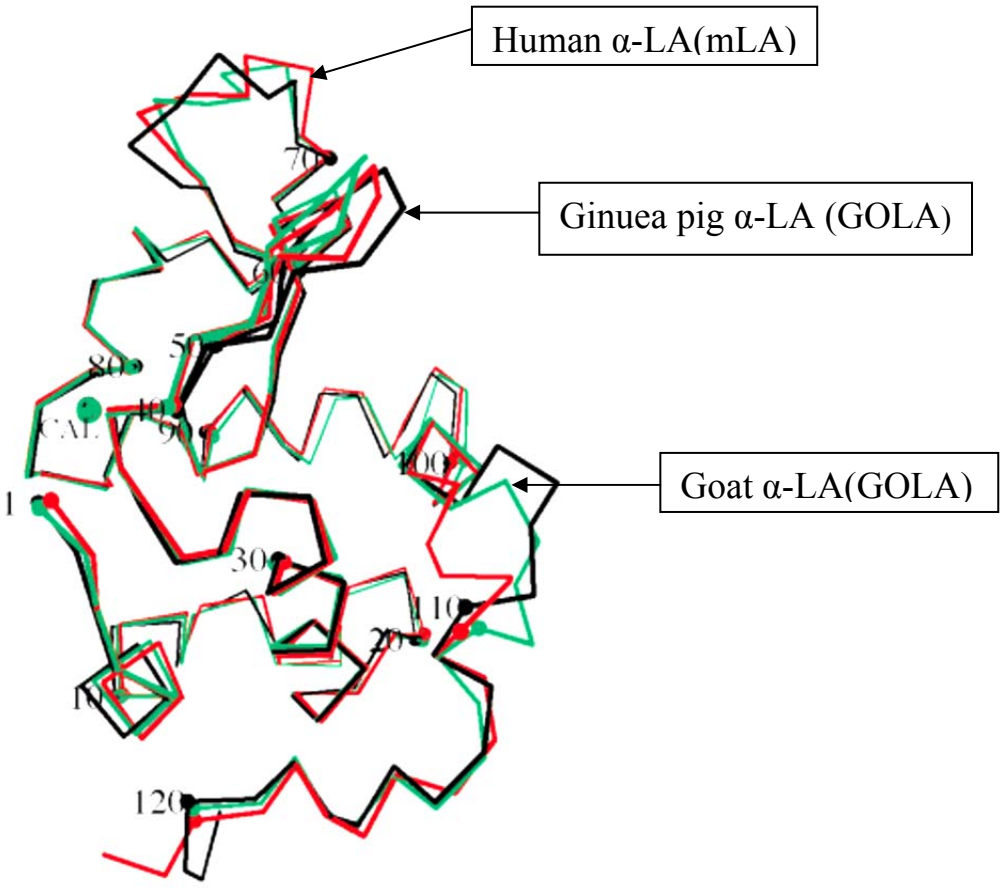

Figure 1.9. Superposition of human α-LA, guinea pig α-LA and goat α-LA, the three proteins have high structural homology (Pike *et al*., 1996).

Apart from lactose biosynthesis, other distinct biological activities of α-LA have been investigated and these include apoptosis and mammary involution (Reich & Arnould, 2007). Furthermore, α-LA has also been attributed to the inhibition of angiotensin-converting enzyme (ACE) activity, anti-microbial and anti-carcinogenic activity (Chatterton *et al*., 2006). At low concentration of calcium, α-LA binds to C18:1fatty acids, changes its conformation with consequent apoptosis in cancerous, but not

normal cells. Thus this form of α-LA has been proposed as a cancer therapeutic agent, and was named HAMLET (human alpha lactalbumin made lethal to cancer cells (Reich & Arnould, 2007). Involution of the mammary gland, a process that occurs after weaning, is also triggered by α-LA.

1.5. Carbohydrates in milk

Lactose is not the only carbohydrate in milk, despite being the major simple sugar in mammalian milk, constituting over 80 % of the total carbohydrate. Other complex carbohydrate compounds are present and these are known as oligosaccharides (Urashima *et al*., 2001). Milk of most, if not all, mammalian species contain oligosaccharides, which are described as carbohydrate compounds containing between three and ten monosaccharide units in linear or branched form. Compared to ruminant milk, such as cow, with low oligosaccharide levels of less than 0.1 g kg^{-1}, some mammalian species, such as marsupials, bears, elephants and primates, have considerably high oligosaccharide levels of greater than 0.5 g kg^{-1} (Urashima *et al.*, 2004; Osthoff *et al*., 2005). Approximately 130 oligosaccharides are present in human milk, but due to the advent of mass spectrometry and its use in analysis of milk, longer and more branched oligosaccharides have since been identified (Urashima *et al*., 2007). Oligosaccharides have been assigned many roles, especially in neonatal normal growth, namely, brain development, bactericidal effect and a prebiotic effect on the intestinal microorganisms, due to its indigestibility by humans (Kunz & Rudloff, 2006).

In the course of milk consumption by infants, lactose is hydrolyzed by β-galactosidase (lactase) to simple sugars that is glucose and galactose (Urashima *et al*., 2007). Glucose enters the circulation and is used as an energy source whilst galactose is transformed to glucose in the liver and is also used as energy source. The exact fate of enzymatic breakdown resistant oligosaccharides has not been completely unraveled (Urashima *et al*., 2001). Human milk oligosaccharides are not digestible; they enter the colon where they act as prebiotics preventing colonization of the colon by pathogenic microorganisms and promote bifidobacteria growth, which in turn reduces the incidence of diarrhea in infants.

Invasion of the intestinal mucosa by pathogenic viruses and bacteria occurs by adhesion to specific carbohydrate structures; some oligosaccharides can bind to these glycoconjugates, there-by preventing attachment of pathogens. Sialylated oligosaccharides, at physiological concentration, strongly inhibit the binding of influenza A virus and S-fimbriated enteropathogenic *E. coli* to their respective host target cells. Human milk oligosaccharides can modulate the immune system of maturing infants (Bode, 2006).

Most milk oligosaccharides, humans included, contain lactose at their reducing end; glycosyltransferases are responsible for their synthesis, acting on free lactose as the acceptor (Messer & Urashima, 2002; Urashima *et al*., 2007). The synthesis of lactose within lactating mammary gland follows the transfer of donor UDP-Gal to a glucose acceptor molecule, the reaction is catalyzed by lactose synthase complex (a complex of α-LA and β-1, 4-GT1).

The biosynthesis of oligosaccharides involve β-1, 4-GT1 which transfers UDP-Gal to non reducing GlcNAc to synthesize N-acetyllactosamine units, α-LA changes the preferred acceptor to glucose in its presence thus making it the key to lactose presence in milk. Figure 1.10 shows the synthesis of oligosaccharides in the absence of α-LA, N-acetyllactosamine is the precursor for further synthesis of a wide variety of oligosaccharides by glycosyltransferases.

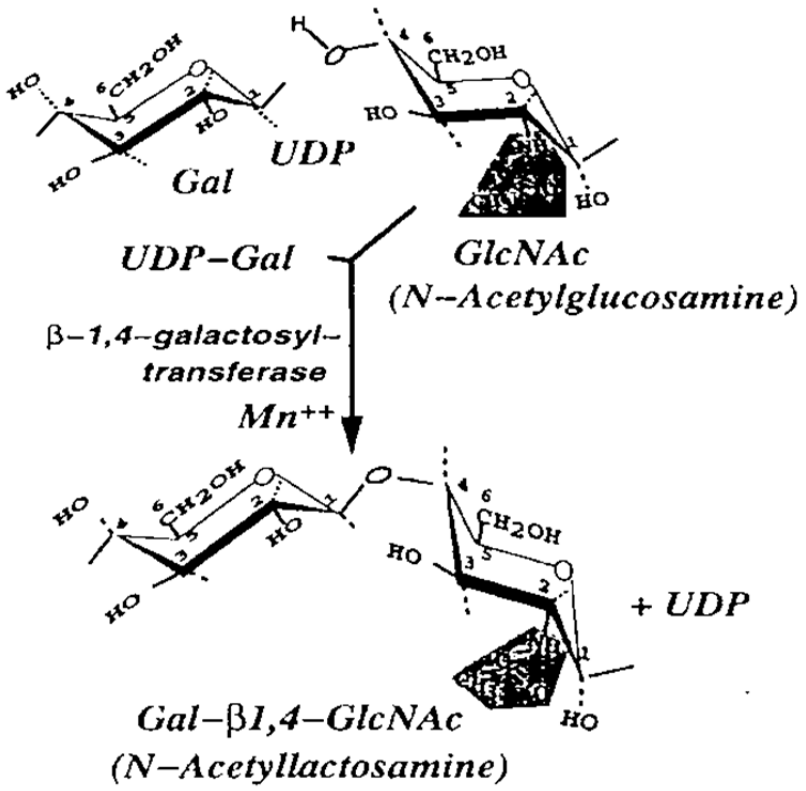

Figure 1.10. Synthesis of the oligosaccharides precursor in the absence of α-LA (N-Acetylactosamine) (Urashima *et al*., 2007)

Glycosyltransferases are very specific and their specificity is directed towards the acceptor molecule and type of bonds they form. Glycoconjugates in vertebrates differ from those found in invertebrates. Invertebrates contain β-1, 4-N-acetylgalactosaminyltransferase instead of β-1,4-GT1 of vertebrates (Ramakrishnan & Qasba, 2007).

Oligosaccharides from human milk are formed by extension of the lactose molecule by addition of Neu5Acα-2-3/2-6 residue to Gal or GlcNAc and of Fucα-1-2/1-3/1-4 to Gal, GlcNAc or a reducing Glc of the core units (Urashima *et al*., 2007). Human milk oligosaccharides are the only oligosaccharides that contain the type I branch (Gal (β 1-3) GlcNAc) residues. Additionally human milk oligosaccharides contain type II branch (Gal (β 1-4) GlcNAc), which are also present in many other oligosaccharides from other species. Table 1.2 shows the chemical structure of some oligosaccharides that have been identified in human milk.

Table 1.2. Oligosaccharides from human milk and their chemical structure (Urashima *et al*., 2007).

No.	Oligosaccharide	Structure
Neutral oligosaccharides		
1	2′-FL	Fuc(α1-2)Gal(β1-4)Glc
2	3-FL	Gal(β1-4)Glc Fuc(α1-3)
3	β3′-GL	Gal(β1-3)Gal(β1-4)Glc
4	β4′-GL	Gal(β1-4)Gal(β1-4)Glc
5	β6′-GL	Gal(β1-6)Gal(β1-4)Glc
6	LNTri II	GlcNAc(β1-3)Gal(β1-4)Glc
7	DF-L(LDFT)	Fuc(α1-2)Gal(β1-4)Glc Fuc(α1-3)
8	LNT	Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc
9	LNnT	Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc
10	LNFP I	Fuc(α1-2)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc
11	LNFP II	Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc Fuc(α1-4)
12	LNFP III	Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc Fuc(α1-3)
13	LNFP V	Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc Fuc(α1-3)
14	LNDFH I	Fuc(α1-2)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc Fuc(α1-4)
15	LNDFH II	Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc Fuc(α1-4) Fuc(α1-3)

There seems to be a significant correlation between the amount of oligosaccharides in milk and the level of α-LA in the mammary tissue as shown in (Table. 1.3). In the presence of α-LA, high lactose and low oligosaccharide levels are observed in milk, and likewise, in its absence, higher amounts of oligosaccharides, as opposed to lactose, are observed. However, in African elephant milk, a eutherian species with high levels of α-LA, there are significantly high levels of both oligosaccharides and lactose (Osthoff *et al*., 2005; Uemura *et al*., 2006). This observation poses a question, is it only the up and down regulation of alpha lactalbumin that determines inversely the level of lactose and oligosaccharides? Or is it the structure function relationship of α-LA?

Table 1.3. Lactose and oligosaccharide level in different mammalian species. (Oftedal & Iverson, 1995; Messer & Urashima, 2002; Osthoff *et al*., 2007).

Species	**Lactose %**	**Oligosaccharides %**
Prototheria		
Monotremes (No **α-LA**)	≈ 0	37
Theria		
Marsupial (No **α-LA**)	1-2	4.6-12
Eutheria (**α-LA** present)		
Seal (No **α-LA**)	0	0
Sea lion (low **α-LA**)	0.5	0.3
Cow (high **α-LA**)	4	0.1
Human (high **α-LA**)	7.3	1.2
African elephant (high **α-LA**)	0.7-5.3	1.5-2.7

1.6. Biologically active compounds in milk

Milk is intended to be a balanced, complete and only source of nutrition for the neonate. The many peptides in milk serve many biological functions of which some have not been fully deduced (Gobbetti *et al.*, 2002). Bioactive compounds are substances that consist of proteins, lipids and or carbohydrate molecules and are able to initiate a biological response such as killing bacteria, stimulating an immune response, reducing hypertension, enhancing lean body mass and reducing cancer (Kitts & Weiler, 2003). As highlighted before, bioactive compounds have the ability to initiate a biological response and a number of these compounds have been identified in milk. Most bioactive compounds occur in milk in very limited amounts, and often comprising the whey component of milk, these proteins or peptides are either blood derived or synthesized in the mammary gland (Korhonen & Pihlanto, 2006). Table 1.4 gives a summary of some of the bioactive compounds found in milk and the roles they play.

Apart from proteins that impact their physiological action directly, some only do so upon digestion (Korhonen & Pihlanto, 2006). Biologically active peptides encrypted

within protein sequences are released via enzymatic hydrolysis. Bioactive peptides have been defined as specific protein fragments that have a positive impact on body functions or conditions, and may ultimately influence health (Kitts & Weiler, 2003). The size of these active sequences varies from two to twenty amino acids and may induce multiple roles. These include antihypertensive, antioxidative, antithrombotic, hypocholesterolemic, mineral binding, anti-appetizing, antimicrobial and immunomodulatory functions.

Table 1.4.. Biologically active compounds found in milk (Korhonen & Pihlanto, 2006).

Bioactive function	*Bioactive component*	*Chemical nature*
Antibacterial	Lactoferrin, lactoperoxidase, lysozyme, defensin	Peptide
Gastrointestinal function (intestinal motility, emptying, absorption)	Casomorphin, lactorphin	Peptide
	Casoxin	Peptide
	Serophin	Peptide
	Lactoferroxin	Peptide
	β-Lactotensin	Peptide
	Albutensin	Peptide
	Caseinomacropeptide	Peptide
Cell growth and repair	Growth factors (e.g., IGF-I, EGF, TGF-α), growth inhibitory factors (MDGI, TGF-β)	Peptide
	β-Casein-derived fragments	Peptide
	Glutamylcysteine	Peptide
	Lactoferrin	Peptide
Hypertension lowering	ACE inhibitors	Peptide
	Calcium	Mineral
Mineral utilization	α_{s1}-, α_{s2}-, and β-casein-derived phosphopeptides	Peptide
Bone synthesis	Calcium	Mineral
	Hormone (PTHrP)	Peptide
Immunoregulation	Lactoferrin	
	Immunoglobulins	Peptide
	α- and β-casein-derived fragments	Peptide
	Cytokines	Peptide
	Minerals (zinc, iron, copper, selenium) and vitamins (A, β-carotene, B_6, C, E)	
Anticarcinogenic	CLA	Lipid
	Sphingolipids (sphingomyelin, ceramides, gangliosides)	Glycolipids
Increasing lean body mass	CLA	Lipid
Prebiotics/probiotics	Galacto-oligosaccharides	Carbohydrate
Atherosclerosis	CLA	Lipid

1.7. Proteomics approach

Proteomics, a term first coined in the early 1990's, can be defined as the systematic separation, identification and characterization of proteins from a common source (O' Donnell *et al*., 2004). In addition to identification of proteins and consequent determination of their role in physiological function, proteomics is also dedicated to determination of protein structures, as well as their interaction with each other (Roncada *et al*., 2012). The proteome, which differs from cell to cell, refers to the entire complement of proteins that is produced by a system or organism, including the modifications made to a particular set of proteins.

The milk proteome is extremely complex largely due to post translational modification (phosphorylation, glycosylation and proteolysis) as well as evolutionary divergence (O'Donnell *et al*., 2004). Additionally milk proteomics becomes even more complex in that low abundance proteins require some form of enrichment, so that they can be visualized against a background of the more represented caseins.

The proteomics approach becomes superior to traditional protein biochemistry in that it enables simultaneous high resolution analysis of an often complex protein mixture (O'Donnell *et al*., 2004). This attractive tool provides a unique opportunity in dairy protein science, enabling the unraveling of milk proteins and acquirement of previously unattainable data about proteins in milk (Chevalier, 2004). The application of proteomic studies on milk has been on the move for the past five years and even better and more intricate studies are being undertaken in this regard, for example, post translational modification analysis of milk proteins.

For the analysis of milk proteins, mainly two strategies have become popular, that is gel based and gel free proteomics (O'Donnell *et al*., 2004). The former involve separation of proteins by 2D polyacrylamide gel electrophoresis (2D PAGE) and analysis of the protein spots by mass spectrometry (MS), whereas the latter requires separation of proteins by liquid chromatography followed by peptide analysis by MS.

The 2D PAGE analysis of milk proteins has been increasingly favored and utilized by the food and agricultural industry for analysis of the behavior of proteins in processed foods.

In addition, liquid phase separation can also be utilized in resolving milk proteins. Milk proteins from a variety of species have been studied using 2D PAGE; these include human, bovine, goat, wallaby and mouse milk (O'Donnell *et al*., 2004). In 2D PAGE individual proteins are resolved based on both their isoelectric point (by isoelectric focusing electrophoresis; IEF) as well as their size (by SDS PAGE). Identification of proteins of interest after 2D PAGE is normally done by utilizing MS.

Matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) is normally the method of choice because of its economic feasibility and time saving. The proteomics of bovine and human milk has been studied extensively and a few dominating primary proteins have been identified, these include caseins (α_{s1}-, α_{s2}-, β- and κ-caseins), as well as whey proteins (α-LA, β-LG) (Goldfarb *et al*., 1989). Unlike caseins which mainly serve the nutrition purpose, α-LA, and β-LG are bioactive proteins that assume a more active role in biochemical reactions of the mammary gland (Madureira *et al*., 2007). In addition to high abundance proteins in milk, low abundance proteins can also be identified and this requires fractionation of the milk sample and consequent 2D PAGE and MS analysis (Yamada *et al*., 2002).

The isolation and fractionation of milk proteins is mainly governed by the starting material, its nature and concentration (O'Donnell *et al*., 2004). When whole milk is the starting material, coagulation at pH 4.6 can be used to separate whole casein from whey proteins, even though some casein proteins are not precipitated at this pH. Alternatively, enzymatic hydrolysis can be applied. Colostrum is a rich source of immunoglobulins and minor proteins. Milk derived bioactive peptides are mainly obtained from enzymatic hydrolysis of milk proteins with specific proteases (Korhonen & Pihlanto, 2006).

The separation and characterization of heterogeneous protein mixtures is undertaken by several physicochemical methods that depend on solubility as well as isoelectric points. In bovine milk in particular, there are significant differences between low abundance proteins of colostrums and mature milk and this has been explained with the fact that neonates require specific nutrients at early lactation that have special physiological relevance to their normal development (Yamada *et al*., 2002). The investigation of colostrums has significant relevance in that the data obtained maybe useful in development of infant milk formula for low birth mass infants.

1.8. African elephant milk

The artificial rearing of elephant calves after birth in captivity is often met with challenges, such as, provision of a suitable milk substitute. This is mainly due to a lack of knowledge of African elephant (*Loxodonta africana*) milk composition and as such, its unraveling is imperative (McCullagh & Widdowson, 1970). The first comprehensive study of African elephant milk was conducted by (McCullagh & Widdowson, 1970). In this study, milk samples from 30 African elephant cows were collected post mortem and analyzed. The lactation stages of the sampled African elephant cows, spanned between two 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 milks, African elephant milk fat composed of 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.

Electrophoresis was also employed in this study, which gave further insight into the composition of African elephant milk. The protein bands on the electrophoretograms showed a similar pattern of migration, in comparison to cow milk proteins. However, some of the corresponding proteins in African elephant milk were less negatively charged, particularly α-casein. The intensity of the bands on the electrophoretogram indicated that the γ- and κ-casein occur in higher amounts while the β- and α-casein occur in lower amounts. Preliminary results from experimental work that followed up on this study, using 2D PAGE and MS/MS showed the possible absence of κ-casein. The study of African elephant α-LA may provide answers to the co-existence of high levels of lactose and oligosaccharides in African elephant milk, since α-LA is actively involved in biosynthesis of lactose and oligosaccharides, directly and indirectly so, respectively (Brew *et al*., 1967). Additionally, the study of African elephant caseins may provide further insight into the casein model, with reference to the structural role of the proteins.

1.9. Protein structure prediction and Hydropathy plots

Protein structure prediction, primarily based on sequence and structure prediction, 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 that 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 (Krieger *et al*., 2003; Deschamps, 2009).

Homology modeling is relatively easy to perform. The basis of its application is attributed largely to 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 also fold into similar structure but only if they are in the safe mode 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). As such, elephant α-LA probably is more likely to structurally resemble other α-LAs.

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 & Honig, 2005). Homology modeling can be such 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, although it faces challenges such as inability to use at large scale (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 African elephant α-LA and exploit the role it plays in lactose synthesis.

Hydropathy analysis, often complements structure modeling in determination of functions of known protein structures (Damodharan & Pattabhi, 2004). The

hydrophilic and hydrophobic properties of individual amino acids in the protein sequences determine the structure and fold. (Kyte & Doolittle, 1982) composed an experiment based hydropathy scale, wherein the hydrophilic and hydrophobic properties of each of the 20 amino acids side-chains are taken into consideration. The hydropathy scale is in cooperated into a computer program that evaluates the hydrophilicity and hydrophobicity of a protein along its amino acid sequences. The application of hydropathy analysis at present include: distinguishing the interior and exterior regions of a protein; identification of β-strands in globular proteins and determining the nature of membrane proteins (Damodharan & Pattabhi, 2004). In our own research on African elephant milk, evaluation of the hydropathy of African elephant casein may be pivotal to provide structure evidence of casein micelle formation.

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. A deep understanding of milk proteomics of other mammalian species, especially those whose proteome show unique characteristics could therefore be an answer to the increased demand of the food industry for functional proteins. This knowledge can now be implemented by the food industry in the efficient and increased production of other food products via biotechnology.

The lactose synthase complex plays a major role in availability and levels of lactose in milk. α-LA seems to be the more involved protein of the two, and consequently there is a significant correlation between α-LA levels in the mammary gland and lactose or oligosaccharides content of milk.

One of the properties of milk that enables milk proteins to stay in solution at normal conditions of milk, are casein micelles. Its functional attributes and assembly

mechanism has been a topic of research for a long time. Of all the models put forward, the dual binding mechanism seems to better explain the casein micelle and the behavior of the proteins that are compose the casein micelle as well as how it functions. As far as diversity of mammals and the milks they produce is concerned, there exists a great deal of diversity in milk composition and function, quantity of milk output, length of lactation period and array of proteins at different lactation stages.

The explosive growth in databases harbouring protein sequences and structures has increased the accuracy of homology models. The models are important in the determination of protein function and have already been employed in drug design. Homology modeling will be employed in our current research to determine the structure of α-LA in African elephant milk. Hydropathy studies are also crucial as they provide information about the possible function of the protein by evaluation of the hydropathy profile of the proteins. In our research, hydropathy plots of caseins in African elephant milk may indicate how they interact with each other in casein micelles.

1.9. Aims of the study

Due to the literature findings and rigorous consideration of the previously done proteomics work on mammalian species, the following became the aims of this particular study:

1. To investigate the milk proteome of a species that differs from human and cow's milk with regards to a) content of high levels of both lactose and oligosaccharides, and b) of which the casein component differs substantially in ratio and types of casein proteins.

2. Identification and characterization of African elephant milk proteins, with emphasis on α-LA and caseins.

3. Characterization of amino acid sequence and structure of α-LA of African elephant milk.

4. Characterization of amino acid sequence and structure of caseins of African elephant milk.

1.10. References

Anderson, S. M., Rudolph, M. C., McManaman, J. L. & Neville, M. C. (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**, 204-208.

Al-lazikani, B., Jung, J., Xiang, Z. & Honig, B. (2001). Protein structure prediction. *Curr Opin Chem Biol* **5**, 51-56.

Bequette, B. J., Backwell, F. R. & Crompton, L. A. (1998). Current concepts of amino acid protein metabolism in the mammary gland of the lactating ruminant. *J Dairy Sci* **81**, 2540-2559.

Bode, L. (2006). Recent advances on structure, metabolism, andfunction of human milk oligosaccharides. *J Nutr* **136**, 2127-2130.

Brew, K., Vanaman, T. C. & Hill, R. L. (1967). Comparison of the amino acid sequence of bovine α-lactalbumin and hens egg white lysozyme. *J Biol Chem* **242**, 3747-3748.

Burgoyne, R. D. & Duncan, J. S. (1998). Secretion of milk proteins *J Mammary Gland Biol Neoplasia* **3**, 275‑286.

Casado, B., Affolter, M. & Kussmann, M. (2009). OMICS-Rooted studies of milk proteins, oligosaccharides and lipids. *J Proteomics* **73**, 196-208.

Chatterton, D. E. W., Smithers, G., Roupas, P. & Brodkorb, A. (2006). Bioactivity of β-lactoglobulin and α-lactalbumin. Technological implication for processing. *Int Dairy J* **16**, 1229-1240.

Chevalier, F., Rofidal, V., Vanova, P., Bergoin, A. & Rossignol, M. (2004). Proteomic capacity of recent fluorescent dyes for protein staining. *Phytochemistry* **65**, 1499-1506.

Claverol, S., Burlet-Schiltz, O., Gairin, J. E. & Monsarrat, B. (2003). Characterization of protein variants and posttranslational modifications: ESI-MSn analyses of intact proteins eluted from polyacrylamide gels. *Mol Cell Proteomics* **2**, 483-493.

D'Allessandro, A., Scaloni, A. & Zolla, L. (2010). Human milk proteins: An interactomics and updated functional overview. *J Proteome Res* **9**, 3339-3373.

Damodharan, L. & Pattabhi, V. (2004). Hydropathy analysis to correlate structure and function of proteins. *Biochem Biophys Res Commun* **323**, 996–1002.

Deschamps, J. R. (2009). X-ray crystallography of chemical compounds. *Life Sci* **86**, 585-589.

Delfour, A., Jolles, J., Alais, C. & Jolles, P. (1965). Caseino-glycopeptides. Characterization of a methionine residue and of the N-terminal sequence. *Biochem Biophys Res Commun* **19**, 452-455.

Eigel, W. N. (1977). Formation of γ1-A2, γ2-A2 and γ3-A caseins by in vitro proteolysis of β-casein A2 with bovine plasmin. *Int J Biochem* **8**, 187-192.

Eigel, W. N., Butler, J. E., Ernstrom, C. A., Farrell, H. M. J., Harwalkar, V. R., Jenness, R. & Whitney, R. M. (1984). Nomenclature of proteins of cow's milk: fifth revision. *J Dairy Sci* **67**, 1599-1631.

Elbegdorj, O., Westkaemper, R. B. & Zhang, Y. (2013). A homology modeling study toward the understanding of three-dimensional structure and putative pharmacological profile of G-protein coupled receptor GPR55. *J Mol Graph Model* **39**, 50-60.

Farrell, H. M., Jiminez-Flores, R., Bleck, G. T., Brown, E. M., Butler, J. E., Creamer, L. K., Hicks, C. L., Hollar, C. M., Ng-Kwai-Hang, K. F. & Swaisgood, H. E. (2004). Nomenclature of the proteins of cow's milk: sixth-revision. *J Dairy Sci* **87**, 1641-1674.

Fox, P. F. & McSweeney, P. L. H. (1998). *Dairy Chemistry and Biochemistry*. London: Chapman & Hall.

Fox, P. F. & McSweeney, P. L. H. (2002). *Advanced Dairy Chemistry,* 3rd edn. New York: Kluwer Academic.

Gagnaire, V., Jardin. J., Jan, G. & Lortal, S. (2009). Proteomics of milk and bacteria used in fermented dairy products: from qualitative to quantitative advances. *J Dairy Sci* **92**, 811-825.

Ginger, M. R. & Grigor, M. R. (1999). Comparative aspects of milk caseins. *Comp Biochem Physiol B* **124**, 133-145.

Gobbetti, M., Stepaniak, L., De Angelis, M., Corsetti, A. & Di, C. R. (2002). Latent bioactive peptides in milk proteins: proteolytic activation and significance in dairy processing. *Crit Rev Food Sci Nutr* **42**, 223-239.

Goldfarb, M. F., Savadove, M. S. & Inman, J. A. (1989). Two-dimensional electrophoretic analysis of human milk proteins. *Electrophoresis* **10**, 67-70.

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

Heid, H. W. & Keenan, T. W. (2005). Intracellular origin and secretion of milk fat globules. *Eur J Cell Biol* **84**, 245‑258.

Hennighausen, L. G. & Sippel, A. E. (1982). Characterization and cloning of the mRNAs specific for the lactating mouse mammary gland. *Eur J Biochem* **125**, 131-141.

Holden, H. M., Rayment, I. & Thoden, J. B. (2003). Structure and function of enzymes of the Leloir pathway for galactose metabolism. *J Biol Chem* **278**, 43885-43888.

Holland, J. W., Deeth, H. C. & Alewood, P. F. (2006). Resolution and characterisation of multiple isoforms of bovine κ-casein by 2-DE following a reversible cysteine-tagging enrichment strategy. *Proteomics* **6**, 3087-3095.

Holt, C. & Horne, D. S. (1996). The hairy casein micelle: evolution of the concept and its implications for dairy technology. *Neth Milk Dairy J* **50**, 85-111.

Horne, D. S. (1998). Casein interactions: casting light on the black boxes, the structure in dairy products. *Int Dairy J* **8**, 171-177.

Horne, D. S. (2008). Casein micelle structure and stability. In *Milk proteins: from expression to food,* pp. 133-162. Edited by A. Thomson, M. Boland. & H. Singh. New York: Elsevier.

Iametti, S., Tedeschi, G., Oungre, E. & Bonomi, F. (2001). Primary structure of κ-casein isolated from mare's milk. *J Dairy Res* **68**, 53-61.

Jolle`s, J., Alais, C. & Jolle`s, P. (1968). The tryptic peptide with rennin-sensitive linkage of cow's κ-casein. *Biochim Biophys Acta* 168, 591-593.

Juhn, N. J., Carrick, D. T. & Wilde, C. J. (1980). Lactose synthesis: The possibilities of regulation. *J Dairy Sci* **63**, 328-336.

Kelly, S. M., Jess, T. J. & Price, N. C. (2005). How to study proteins by circular diochroism. *Biochim Biophys Acta* **1751**, 119-139.

Kitts, D. D. & Weiler, K. (2003). Bioactive proteins and peptides from food sources. Applications of bioprocesses used in isolation and recovery. *Curr Pharm Des* **9**, 1309‑1323.

Korhonen, H. & Pihlanto, A. (2006). Bioactive peptides: Production and functionality. *Int Dairy J* **16**, 945-960.

Krieger, E., Nabuurs, S. B. & Vriend, G. (2003). Homology modeling. In *Structural Bioinformatics,* pp. 507-521. Edited by P. E. Bourne & H. Weissig. Wiley-Liss, Inc.

Kronman, M. J., Sinha, S. K. & Brew, K. (1981). Characteristics of the binding of Ca^{2+} and other divalent metal ions to bovine alpha lactalbumin. *J Biol Chem* **256**, 8582-8587.

Kundrotas, P. J., Lensink, M. F. & Alexov, E. (2008). Homology-based modeling of 3D structures of protein -protein complexes using alignments of modified sequence profiles. *Int J Biol Macromol* **43**, 198–208.

Kunz, C. & Rudloff, S. (2006). Health promoting aspects of milk oligosaccharides. *Int Dairy J* **16**, 1341–1346.

Kyte, J. & Doolittle, R. F. (1982). A Simple Method for Displaying the Hydropathic Character of a Protein. *J Mol Biol* **157**, 105–132.

Larsen, B. L. (1985). Biosynthesis and cellular secretion of milk: In *Lactation,* pp. 129-163. Edited by B. L. Larson. Ames, IA: The Iowa State University Press.

Lönnerdal, B. (2007). Trace element transport in the mammary gland. *Annu Rev Nutr* **27**, 165-177.

Madureira, A. R., Pereira, C. I., Gomes, A. M. P., Pintado, M. E. & XavierMalcata, F. (2007). Bovine whey proteins-overview on their main biological properties. *Food Res Int* **40**, 1197-1211.

Martin, P., Ferranti, P., Leroux, C., & Addeo, F. (2003). Non-bovine caseins: Quantitative variability and molecular diversity. In *Advanced Dairy Chemistry,* pp. 387-419. Edited by P. F. Fox & P. L. H. McSweeney. New York: Plenum Publishers.

Mather, I. H. & Keenan, T. W. (1998). Origin and secretion of milk lipids. *J Mammary Gland Biol Neoplasia* **3**, 259‑273.

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

McClellan, H. L., Miller, S. J. & Hartmann, P. E. (2008). Evolution of lactation: nutrition v protection with special reference to five mammalian species. *Nutr Res Rev* **21**, 97-116.

McCullagh, K. G. & Widdowson, E. M. (1970). The milk of the African elephant. *Br J Nutr* **24**, 109-117.

McKenna, M. C & Bell, C. K. (1997). *Classification of Mammals above the Species Level,* 1st edn. New York: Columbia University Press.

McKenzie, H. A & White, F. H. (1991). Lysozyme and α-lactalbumin. Structure, function and interrelationships. *Adv Protein Chem* **41**, 173-345.

Messer, M. & Urashima, T. (2002). Evolution of milk oligosaccharides and lactose. *Trends Glycosci Glycotechnol* **14**, 153-176.

Neville, M. C. (1995). Lactogenesis in women: A cascade of events revealed by milk composition. In *The Composition of Milks,* pp. 87-89. Edited by R. D. Jensen. San Diego: Academic Press.

Nicholas, K. R., Hartmann, P. E. & McDonald, B. L. (1981). Alpha-Lactalbumin and lactose concentrations in rat milk during lactation. *Biochem J* **194**, 149-154.

O'Donnell, R., Holland, J. W., Deeth, H. C. & Alewood, P. (2004). Milk proteomics. *Int Dairy J* **14**, 1013-1023.

Oftedal, O. T. & Iverson, S. J. (1995). Comparative analysis on nonhuman milks:A phylogenetic variation in the gross composition of milks. In *Handbook of milk composition.* pp. 749-789. Academic press.

Osthoff, G., de Wit, M., Hugo, A. & Kamara, B. I. (2007). Milk composition of three free-ranging African elephant (*Loxodonta africana africana*) cows during mid lactation. *Comp Biochem Physiol B* **148**, 1-5.

Osthoff, G., De Waal, H. O., Hugo, A., De Wit, M. & Botes, P. (2005). Milk composition of a free-ranging African elephant (*Loxodonta africana*) cow during early lactation. *Comp Biochem Physiol A* **141**, 223-229.

Petrey, D. & Honig, B. (2005). Protein Structure Prediction: Inroads to Biology. *Mol Cell* **20**, 811-819.

Pike, A. C. W., Brew, K. & Acharya, K. R. (1996). Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure* **4**, 691-703.

Piotte, C. P., Marshall, C. J., Hubbard, M. J., Collet, C. & Grigor, M. R. (1997). Lysozyme and α-lactalbumin from the milk of a Marsupial, the Common Brushtailed possum (*Trichosurus vulpecula*). *Biochim Biophys Acta* **1336**, 235-242.

Qasba, P. K. & Kumar, S. (1997). Molecular divergence of lysozyme and α-lactalbumin. *Crit Rev Biochem Mol Biol* **32**, 255-306.

Ramakrishnan, B., Shah, P. S. & Qasba, P. K. (2001). α-Lactalbumin (LA) stimulates milk β-1,4-Galactosyltransferase I (β4Gal-T1) to transfer glucose from UDP-glucose to N-Acetylglucosamine. *J Biol Chem* **276**, 37665-37671.

Ramakrishnan, B. & Qasba, P. K. (2007). Role of amino acid in the evolution of glycans of invertebrates and vertebrates. *J Mol Biol* **365**, 570-576.

Reich, C. M. & Arnould, J. P. Y. (2007). Evolution of Pinnipedia lactation strategies: A potential role for α-lactalbumin. *Biol Lett* **3**, 546-549.

Rollema, H. S. (1992). Casein association and micelle formation, In *Advanced Dairy Chemistry,* pp. 111-140. Edited by P. F. Fox. Essex: Elsevier.

Roncada, P., Piras, C., Soggiu. A., Turk, R., Urbani, A. & Bonizzi, L. (2012). Farm animal milk proteomics. *J Proteomics* **75**, 4259-4274.

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng* **2**, 85-94.

Shennon, D. B. (2008). Calcium transport by mammary secretory cells: mechanisms underlying trans epithelial movement. *Cell Mol Biol Lett* **13**, 514-525.

Shennon, D. B. & Peaker, M. (2000). Transport of milk constituents by the mammary gland. *Physiol Rev* **80**, 925‑951.

Sievers, F., Wilm, A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. & other authors. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol* **7**,1–6.

Simpson, G. G. (1945). The principles of classification and a classification of mammals. *Bull Am Meseum Nat Hist* **85**, 1‑350.

Singh, K., Swanson, K., Couldrey, C., Seyfert, H. M. & Stelwagen, K. (2008). Suppression of bovine α_{s1}-casein gene expression during involution of the mammary gland is associated with increased DNA methylation at a STAT5-binding site in the α_{s1}-casein promoter. *J Dairy Sci* **91**, 378-386.

Swaisgood, H. E. (1992). Chemistry of caseins. In *Advanced Dairy Chemistry,* pp. 63-110. Edited by P. F. Fox. Essex: Elsevier.

Uemura, Y., Asakuma, S., Yon, L., Saito, T., Fukuda, K., Arai, I. & Urashima, T. (2006). Structural determination of the oligosaccharides in the milk of an Asian elephant (*Elephas maximus*). *Comp Biochem Physiol B* **145**, 468-478.

Urashima, T., Saito, T., Nakamura, T. & Messer, M. (2001). Oligosaccharides of milk and colostrum in non-human mammals. *Glycoconj J* **18**, 357-371.

Urashima, T., Nakamura, T., Teramoto, K., Arai, I., Saito, T., Komatsu, T. & Tsubota, T. (2004). Chemical characterization of sialyl oligosaccharides in milk of the Japanese black bear Ursus thibetanus japonicus. *Comp Biochem Physiol B* **139**, 587-595.

Urashima, T., Asakuma, S., Oftedal, O. & Messer, M. (2007). Predominance of oligosaccharides containing the type 1 chain in human milk: possible significance. *Glycoconj J* **24**, 395-405.

Walstra, P. (1999). Casein sub-micelles: do they exist? *Int Dairy J* **9**, 189-192.

Yamada, M., Murakami, K., Wallingford. J. C. & Yuki, Y. (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.

Chapter 2

Separation of African elephant milk proteins by 1D Electrophoresis and partial amino acid sequence determination by LC-MS/MS.

2.1. Introduction

Milk is an extremely complex fluid purely designed as the sole source of nutrition for the neonate (Roncada *et al*., 2012). The study of milk proteins has been in progress for over 50 years and advances in technology have made it possible for in-depth understanding of this complex protein system (O'Donnell *et al.*, 2004). Proteomics is a term that can be used to describe the analysis of proteins, that is, their separation, identification as well as characterization. Additionally, and more importantly, proteomics enable concurrent study of a large number of proteins.

Sodium dodesulphate polyacrilamide gel electrophoresis (SDS PAGE) followed by mass spectrometry (MS) can be utilized in milk protein studies (Manso *et al*., 1994) Milk proteins of several species have been studied by SDS PAGE (Jovanovic *et al*., 2007), but application of MS has mainly been used for milk of cow and human milk. The proteins studied in great detail are the ones that occur in high abundance, such as α_{s1}-, α_{s2}-, β- and κ-caseins, as well as whey proteins, such as α-lactalbumin and β-lactoglobulin (O'Donnell *et al*., 2004). Additionally, studies of milk proteins from several other species such as goat and wallaby have also been reported (Roncada *et al*., 2012). In this study, we attempt to unravel the array of proteins in African elephant (*Loxodonta africana africana*) milk using the proteomics approach, an extremely powerful and sensitive tool for this particular purpose. In this chapter, the aim is to investigate the proteome of elephant milk using SDS PAGE-MS/MS. Since SDS PAGE separates proteins according to size only, the MS results will only give us a general idea of the array of proteins in African elephant milk. Further investigations using 2D PAGE, a much better technique in resolving proteins compared to SDS PAGE, will be reported in the following chapter.

2.2. Materials and Methods

All chemicals used during the experimental procedures were of proteomics grade or of molecular biology purity. The suppliers of the all the chemicals and bioassays are Bio-Rad Laboratories unless stated otherwise.

2.2.1. Sample preparation

Milk samples were obtained from three African elephants at different stages of lactation. The one sample was of early lactation (E1), 4days after parturition, and was obtained from the Bloemfontein Zoological Gardens (Osthoff *et al*., 2005). The other two were obtained from Welgevonden private game reserve, Vaalwater, South Africa at 12 months (E2) and 18 months lactation (E3) (Osthoff *et al*., 2007). During transporting milk samples were kept on ice. On arrival at the laboratory the milk was subdivided into smaller volumes that were frozen and kept at -20 °C. For each of the milk samples 600 µl of milk was thawed at 39 °C in a warm water bath and 600 µl distilled water was added. The mixture was vortexed and 50 µl of 100 % Trichloroacetic acid (TCA) (Merck, South Africa) was added to obtain a 4 % TCA final concentration. The following step was to vortex the mixture, leave it standing at room temperature for 10 minutes and finally centrifuged at 10 000 rpm in an Eppendorf centrifuge (Bio-Rad, South Africa) for five minutes. After centrifugation the top fat layer was removed and the supernatant discarded, while the protein containing pellet was freeze dried and stored.

In order to concentrate the whey proteins and other proteins of low abundance, sample preparation also included precipitation of high abundance proteins, specifically the caseins, from the whole milk. This was done by thawing 100 µl of milk at 39 °C in a water bath followed by the addition of 100 µl distilled water. The mixture was then vortexed and 30 µl of 10 % acetic acid added. The mixture was vortexed again and then left standing at room temperature for 10 minutes, after which 30 µl of 1 N acetic acid was added. The solution was mixed again and left standing for 10 minutes, after which 250 µl of distilled water was added to the precipitate, which was centrifuged at 10 000 rpm in an Eppendorf centrifuge (Eppendorf –Netheler- Hinz GmbH 5415, Germany) for five minutes.

After centrifugation the fat and casein containing supernatant was discarded, whereas the whey protein containing pellet was retained. The whey was reconstituted in 100 µl distilled water and mixed, followed by the addition of 200 µl of 15 % v/v TCA. The mixture was left standing for 10 minutes, followed by centrifugation at 10 000 rpm to recover the final sample in the resultant pellet.

2.2.2. Determination of protein concentration

The Bradford assay was used to determine the protein concentration of the African elephant milk protein samples (Bradford, 1976). Bovine serum albumin was used as a standard. The protein samples from the three elephants were dissolved in 100 µl of distilled water. The standard composed of duplicate aliquots of 2 mg ml^{-1} bovine serum albumin (BSA) which were added in volumes of 20, 15, 10, 5 µl into microcentrifuge tubes. To the tubes with 15, 10, and 5 µl volumes of BSA, 5, 10 and 15 µl distilled water was added respectively. To all the tubes, 1000 µl Bradford solution was added to obtain a final concentration of 2mg ml^{-1}, 1.5 mg ml^{-1}, 1.0 mg ml^{-1} and 0.5 mg ml^{-1}. The blank contained 20 µl of distilled water to which 1000 µl Bradford solution was added. The tubes were carefully mixed by inversion and left standing for 5 minutes at room temperature.

The dissolved protein samples were prepared for the assay by adding 10 µl of sample, 10 µl of distilled water and 1000 µl Bradford solution into a microcentifuge tube and mixed thoroughly. Absorbance of the standards as well as the sample was then measured at 595 nm (A_{595}) using a 1cm path length microcuvette. The standard curve of A_{595} versus protein concentration was plotted (r^2 = 0.9868) and concentrations of the three samples were determined from the slope of the standard curve.

2.2.3. Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SDS-PAGE analysis was carried out by using a modified method by (Laemmli, 1970) and utilizing a Mini-Protein II dual slab system (Bio-Rad Laboratories, South Africa).

2.2.3.1. Separating gel

A 12 % acrylamide gel was prepared with 1.5 ml of 40 % acralymide, 2.1 ml of Milli-Q pure water, 1.3 ml of 1.5 M Tris.Cl pH 8.8, 50 µl of 10 % SDS, 50 µl of 10 % ammonium persulphate (APS) and 2 µl TEMED in a 25 ml flask. The mixture with a total volume of 5 ml was swirled gently to mix and cast immediately. Isobutanol was used to layer and seal off the gel.

2.2.3.2. Stacking gel

A 4 % acrylamide stacking gel was prepared in 25 ml flask with 0.2 ml 40 % acralymide, 1.505 ml millipure water, 0.35 ml 1 M Tris.Cl pH 6.8, 20 µl 10 % SDS, 20 µl 10 % APS and 2 µl TEMED. The 2 ml mixture then was swirled gently to mix and immediately cast over the solidified separating gel. A 15 well 1cm Teflon comb was inserted and the gel was left to set.

Four milk samples were analyzed, that is, three African elephant milk samples and a bovine milk sample as a control. Amounts of 10 µg and 5 µg protein samples of bovine and the three African elephant milk samples were loaded on the gel respectively. Electrophoresis was carried out at 125 V constant until the bromophenol blue loading dye front reached the bottom of the gel.

2.2.4. In-gel protein digestion

The one dimensional SDS PAGE gels were divided into five segments per sample lane, each segment was cut into small cubes approximately 1 mm^3 in volume and transferred to a sterile 1.5 ml micro centrifuge tube. The gel pieces were then washed by adding 800 µl Milli-Q pure water and incubated for 15 minutes at room temperature. Excess water was removed after the incubation time had elapsed, 800 µl of 50 % acetonitrile (Sigma, South Africa) was added to the tubes and incubated for 15 minutes. The steps above were repeated again before the gel pieces were shrunk by

adding 200 ml of 100 % acetonitrile until the gel pieces stuck together and turned white.

Reduction and alkylation of the proteins inside the gel followed the washing step. To do this 200 µl of 10 mM dithiotreitol/0.1 M ammonium carbonate was added to the gel particles and then incubated for 45 minutes at 56 °C to allow the gel particles to swell up. Excess liquid was removed and 200 µl of 10 mg ml^{-1} (55 mM) iodoacetamide in 0.1 M ammonium carbonate was added and left standing for 30 minutes in darkness. Following reduction and alkylation, the gel particles were washed, shrunk again as described above, washed and then dried in a fume hood. For the trypsin digestion, 40 µl of re-suspension buffer was added to 2 µg trypsin (sequencing grade, Promega) to make up a 500 ng ul^{-1} solution. To 10 µl of this solution 250 µl of 100 mM ammonium carbonate was added to make the digestion solution. The digestion solution was then added to epperndoff tubes containing the gel segments.

After digestion the digest solution was transferred into clean 1.5 ml micro centrifuge tubes, the gel pieces that remained were treated with 150 µl of 5 % v/v formic acid (Merck), vortexed for 10 seconds and left standing for 15 minutes. 150 µl of 100 % v/v acetonitrile was then added to the tubes after which the mixture was vortexed for 10 seconds followed by another 15 minute standing period. The mixture was centrifuged at 10 000 rpm for 5 minutes and the supernatant transferred to the digest solution. The above processes were repeated again and the extracted peptides were mixed with the rest of the digestion product. The solution containing the extracted peptides was then dried in a Vapor Trap (Thermo Electron Corporation) and stored at -20 °C for MS analysis.

2.2.5. Mass spectrometry analysis

Following digestion by trypsin, the peptides generated were extracted for NanoLC/MS/MS analysis. Five microliters of each digest was individually injected

and concentrated on a 300 μm × 5mm Zorbax 300SB-C18 reverse phase trapping column (Agilent) and then eluted onto, and separated by, a custom packed 150 mm × 75 μm Zorbax 300SB-C18 reverse phase column. The peptides were separated and eluted from the column with a 10 to 25 % actetonitrile, containing 5 % formic acid (Merck) gradient, over 60 min. The eluting peptides were analyzed on an AB SCIEX 4000QTRAP hybrid triple quadrupole ion trap MS with a nanospray source at 350 nl min^{-1} using a Nano HPLC (Agilent model 1200).

A survey scan between 400 and 1200 Da was performed looking for eluting peptides; an enhanced resolution scan was performed on peptides to determine the charge state of each peptide before fragmenting the peptides in the collision cell. The peptide sequence information obtained from this MS/MS experiment was analyzed by an in-house Mascot server (Perkins *et al*., 1999), using the latest Swissprot database. The following parameters were used: mass tolerance of 100 ppm, a minimum of two peptides matching to the protein, carbamidomethylation of cysteine as a fixed modification, oxidation of methionine and pyroglutamylation of glutamine as variable modifications, and up to four missed cleavages allowed.

2.3. Results

SDS PAGE was carried out in order to gain an overview insight into the array of proteins present in African elephant milk and their respective abundances at three different stages of lactation, that is early, mid and late lactation as shown by E1, E2 and E3 respectively. Bovine milk was added to the experiment as a control since its proteome has been widely investigated (O'Donnell *et al*., 2004; Tay & Gam, 2011). Figure 2.1 shows the distribution of the proteins and their abundance as separated according to size only on an SDS PAGE gel. In order to identify the less abundant proteins in African elephant milk, a fractionation step was carried out to precipitate the caseins out of solution which are in high abundance so that the low abundance whey proteins can be analyzed. Figure 2.2 shows the distribution of the bands on SDS PAGE of the precipitated samples. Tables 2.1 and 2.2 show different proteins

identified from whole milk protein samples (caseins and whey) as well as the enriched samples (whey) after LC-MS/MS analysis.

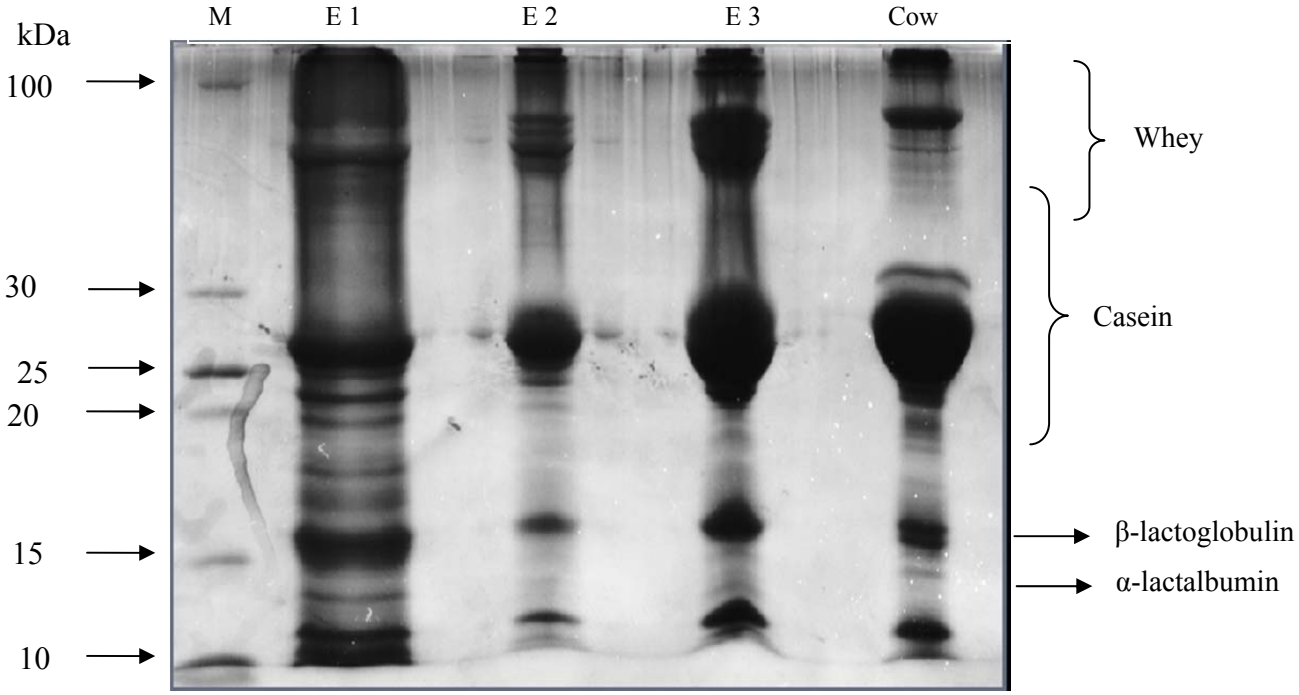

Figure 2.1 SDS PAGE electrophoretograms of three whole milk protein samples of cow and African elephant at different lactation stages (E1 = 4days, E2 = 12 months and E3 = 18 months). The positions, on which the major whey proteins and caseins, upon electrophoresis separation of cow milk proteins are situated, are indicated.

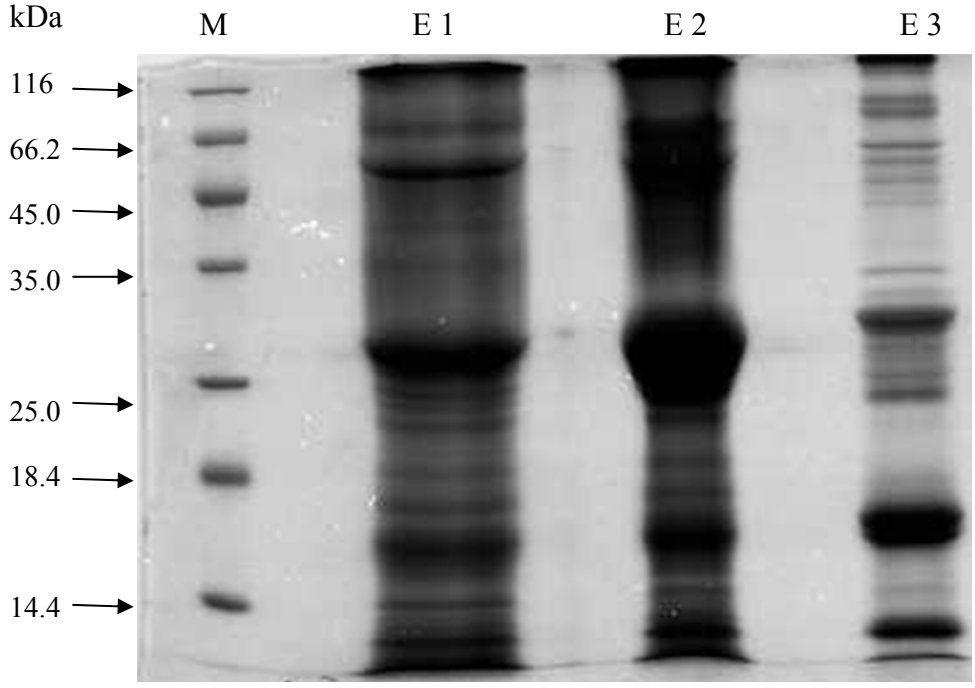

Figure 2.2. SDS PAGE of three casein precipitated elephant milk protein samples at different lactation stages (E1 = 4 days, E2 = 12 months and E3 = 18 months).

Table 2.1. MS results for SDS PAGE of whole milk proteins of African elephant at 12 months of lactation, identification by comparison of closest resemblance.

Sample	Identified proteins
E2a	Histone H4 OS=Bos taurus
	Elongator complex protein 1 OS=Homo sapiens GN=IKBKAP
E2b	Beta-lactoglobulin OS=Ovis orientalis musimon GN=LGB
	Beta-lactoglobulin OS=Bos taurus GN=LGB
E2c	Alpha-S1-casein OS=Bos taurus GN=CSN1S1
	Apolipoprotein E OS=Mus musculus GN=Apoe
	Beta-casein OS=Bos taurus GN=CSN2
	Alpha-S2-casein OS=Bos taurus GN=CSN1S2
E2d	Actin, cytoplasmic 1 OS=Bos mutus grunniens GN=ACTB
	Elongation factor 1-alpha 1 OS=Bos taurus GN=EEF1A1
	Serum albumin OS=Homo sapiens GN=ALB
	Serum albumin OS=Felis catus GN=ALB
	Fructose-bisphosphate aldolase A OS=Homo sapiens GN=ALDOA
	Alpha-S1-casein OS=Bos taurus GN=CSN1S1
	Alpha-S2-casein OS=Bos taurus GN=CSN1S2
	Lactadherin OS=Sus scrofa GN=MFGE8 PE=1
	Elongator complex protein 1 OS=Mus musculus GN=Ikbkap
E2e	Lactotransferrin OS=Sus scrofa GN=LTF
	Lactotransferrin OS=Capra hircus GN=LTF
	Probable carboxypeptidase X1 OS=Homo sapiens GN=CPXM1
	Leucine-rich repeat and IQ domain-containing protein 3 OS=Mus musculus
	Elongator complex protein 1 OS=Oryctolagus cuniculus GN=IKBKAP

1. GN = name of gene 2. OS = name of species

Table 2.2. MS results for SDS PAGE of casein precipitated milk proteins of African elephant

Sample	Proteins identified
1A	Olfactomedin-4 OS=Homo sapiens
	Xanthine dehydrogenase/oxidase OS=Mus musculus
	Serum albumin OS=Macaca fascicularis GN=ALB
	Alpha-lactalbumin OS=Sus scrofa GN=LALBA
	Lactotransferrin OS=Sus scrofa GN=LTF
1B	Serum albumin OS=Felis catus GN=ALB
	Serum albumin OS=Macaca fascicularis GN=ALB
	Lactotransferrin OS=Sus scrofa GN=LTF
	Beta-lactoglobulin OS=Bos taurus GN=LGB
1C	Serum albumin OS=Felis catus GN=ALB

1. GN = name of gene 2. OS = name of species

2.4. Discussion

One dimensional SDS PAGE was done in order to obtain a basic picture of the array of proteins that are present in African elephant milk. Figure 2.1 shows that some African elephant milk proteins occur in higher amounts than others, as shown by the size and intensity of the protein bands and light thin bands respectively. In-gel digestion, followed by MS analysis and protein database search enabled the identification of a wide variety of proteins (Table 2.1) including the more common caseins (α_{s1}-, α_{s2}-, β- and κ-caseins) and the major whey proteins (α-LA, β-LG and serum albumin) as well as some enzymes, suggesting that the African elephant milk proteome is more or less similar to that of other mammalian species, such as cow's milk, which have been extensively characterized. However, when urea PAGE was used as the separation technique, the distribution of protein bands is not similar to that of SDS PAGE separation (Osthoff *et al*., 2005; 2007). For this part of the work, the mascot server was used, specifically using the SwissProt database for protein identification. Since this database does not contain the African elephant proteins, the

partial amino acid sequences had to be compared to other mammalian peptide sequences for potential homology and hence identifications (Perkins *et al*., 1999). The implication is that identification of proteins would only be possible if there is a high degree of amino acid sequence homology of the African elephant proteins with other mammalian proteins. If sequence homology is very low, identification might be impossible, albeit it would not necessarily implicate that proteins are absent. Peptide homology was observed mainly with human, cow, pig, goat and horse milk proteins.

A further step to identify proteins that may occur in lower amounts, specifically the whey fraction, was undertaken by isoelectric precipitation of the caseins, which normally occur in higher amounts in mammalian milk (Martin *et al*., 2003; Yamada *et al*., 2002). Figure 2.2 shows an SDS PAGE gel picture from which caseins were precipitated before separation. The Figure shows additional protein bands which were not visible in Figure 2.1. This is because these proteins are in low abundance; as a result they are extensively dominated mainly by the more abundant caseins when whole milk protein is loaded.

Comparing the three milk samples of African elephant at three different lactation stages (Figure 2.1), it can be deduced that at 4 days lactation (E1), the whey proteins seem to occur at higher amounts relative to the caseins. This is also observed in most other animals, and this milk is called colostrums (Yamada *et al*., 2002). At the later stages of lactation, i.e. 12 months (E2) and 18 months (E3), the caseins occur in abundance. The cow's milk sample that serves as a reference gives an indication of the position of bands of equivalent casein and whey proteins.

In-gel digestion, followed by MS analysis and comparison of amino acid sequences in the protein database enabled the identification of a limited number of proteins (Table 2.2). Proteins that are commonly found in mammalian milk were identified; these include lactotransferrin, α-LA, β-LG, serum albumin as well as some enzymes (Yamada *et al*., 2002). Casein proteins such as α- and β-caseins were identified;

however, these proteins were matched to similar proteins in other mammalian species, specifically bovine and human milk proteins.

Of interest is that κ-casein, one of the well studied caseins due to its role in coagulation by rennet and hence important in casein micelle formation (Jollès *et al*., 1968; Ginger & Grigor, 1999), could not be identified in African elephant milk. As was mentioned above, it might be possible that κ-casein was merely not detected due to a low sequence homology with the equivalent protein of other mammals of which the sequence is available in the protein data bank. Alternatively, the κ-casein may occur at such low amounts that it is beyond the limits of detection by MS. Nevertheless, whether absent or present only at low quantities, this observation poses a number of questions about elephant milk casein micelle formation. κ-casein plays a crucial role in the formation and stabilization of the casein micelle enabling milk to occur as a free-flowing low-viscosity fluid (Holt & Horne, 1996).

Early research had shown that horse milk does not contain κ-casein but still has micelles (Visser *et al*., 1982). Sterical stability and chain termination was thought to be fulfilled by dephosphorylated β-caseins which do not contain a phosphoserine cluster, thus preventing its entrance into chain polymerization. Similar action occurs in human milk which also contains low amounts of κ-casein (Horne, 1998; Roncada *et al*., 2012). It has however been discovered recently that mare's milk does contain κ-casein, although at very low levels compared to human and cow's milk, at about 6-7 % of the total casein (Iametti *et al*., 2001, Malacarne *et al*., 2002). In terms of micelle size, mare's milk has larger micelles than cow and human milk.

With the information gained by SDS PAGE and MS in this chapter, it was shown that most of the well-known whey proteins are present in the milk of African elephant. In the following chapters the purification of α-LA would be attempted by 2-D PAGE, followed by identification of the amino acid sequence and structural studies. Since the κ-casein seems to be absent, this will have to be verified by 2-D PAGE and MS. If the κ-casein is indeed absent, this would be a unique situation in mammalian milk, and the role of the structure of the remaining caseins in micelle formation will have to be studied.

2.5. References

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248–52.

Ginger, M. R. & Grigor, M. R. (1999). Comparative aspects of milk caseins. *Comp Biochem Physiol B Biochem Mol Biol* **124**, 133-145.

Holt, C. & Horne, D. S. (1996). The hairy casein micelle: evolution of the concept and its implications for dairy technology. *Neth Milk Dairy J* **50**, 85-111.

Horne, D. S. (1998). Casein interactions: casting light on the black boxes, the structure in dairy products. *Int Dairy J* **8**, 171-177.

Iametti, S., Tedeschi, G., Oungre, E. & Bonomi, F. (2001). Primary structure of κ-casein isolated from mare's milk. *J Dairy Res* **68**, 53-61.

Jolle`s, J., Alais, C. & Jolle`s, P. (1968). The tryptic peptide with rennin-sensitive linkage of cow's κ-casein. *Biochim Biophys Acta* **168**, 591-593.

Jovanovic, S., Barac, M., Macej, O., Vucic, T. & Lacnjevac, C. (2007). SDS-PAGE analysis of soluble proteins in reconstituted milk exposed to different heat treatments. *Sensors* **7**, 371-383.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4 *Nature* **227**,680–685.

Malacarne, M., Martuzzi, F., Summer, A. & Mariani, P. (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.

Manso, M. A., Leonil, J., Jan, G. & Gagnaire, V. (2005). Application of proteomics to the characterization of milk and dairy products. *Int Dairy J* **15**, 845-855.

Martin, P., Ferranti, P., Leroux, C., & Addeo, F. (2003). Non-bovine caseins: Quantitative variability and molecular diversity. In *Advanced Dairy Chemistry,* pp. 387-419. Edited by P. F. Fox & P. L. H. McSweeney. New York: Plenum Publishers.

O'Donnell, R., Holland, J. W., Deeth, H. C. & Alewood, P. (2004). Milk proteomics. *Int Dairy J* **14**, 1013-1023.

Osthoff, G., De Waal, H. O., Hugo, A., De Wit, M.& Botes, P. (2005). Milk composition of a free-ranging African elephant (*Loxodonta africana*) cow during early lactation. *Comp Biochem Physiol A* **141**, 223-229.

Osthoff, G., de Wit, M., Hugo, A. & Kamara, B. I. (2007). Milk composition of three free-ranging African elephant (*Loxodonta africana africana*) cows during mid lactation. *Comp Biochem Physiol B* **148**, 1-5.

Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence database using mass spectrometry. *Electrophoresis* **20**, 3551-3567.

Roncada, P., Piras, C., Soggiu. A., Turk, R., Urbani, A. & Bonizzi, L. (2012). Milk proteins: Farm animal milk proteomics. *J Proteomics* **75**, 4259-4274.

Swissprot database.version 2.2.04: http://cbio.ufs.ac.za/mascot

Tay, E. P. & Gam, L. H. (2011). Proteomics of human and the domestic bovine and caprine milk. *Asia Pac J Mol Biol Biotechnol* **19**, 45-53.

Visser, S., Jenness, R. & Mullin, R. J. (1982). Isolation and characterization of β- and γ-caseins from horsemilk. *Biochem J* **203**, 131-139.

Yamada, M., Murakami, K., Wallingford, J. C. & Yuki, Y. (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.

Chapter 3

Purification of African elephant milk proteins by 2D electrophoresis and amino acid sequence determination by LC-MS/MS and Orbitrap MS.

3.1. Introduction

Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) is by far the most popular and versatile method for separation and quantification of proteins (Rabilloud *et al*., 2010). 2D PAGE alone can only achieve partial identification of proteins; new analytical strategies that involve combining 2D PAGE with specific detection systems have allowed critical characterization of milk proteins (Manso *et al*., 2005). In addition, liquid phase separation can also be utilized in resolving milk proteins (O'Donnell *et al*., 2004). Milk proteins from a variety of species have been studied using 2D PAGE; these include human, bovine, goat, wallaby and mouse milk (Aslam *et al*., 1994). In 2D PAGE, individual proteins are resolved based on both their isoelectric point (isoelectric focusing electrophoresis, IEF) as well as their size (Sodium dodecyl sulphate polyacrylamide gel electrophoresis, SDS).

Identification of proteins of interest that follows 2D PAGE separation is normally done by utilizing a mass spectrometer (MS) (Patterson & Aebersold, 1995). Comparative studies using 2D PAGE coupled to computer imaging have shown that human, murine and porcine milk proteins are completely different from those of ruminants (Goldfarb, 1999). The coupling of MS to high resolution 2D PAGE can allow the simultaneous analysis of proteins in complex mixtures (Manso *et al*., 2005). The presence of several molecular forms, great heterogeneity and presence of low abundance proteins are some of the characteristics of milk proteins and there-by require a method which shows high resolution, in this case 2D PAGE, for characterization.

Matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) is normally the method of choice because it is economically feasible and time saving to conduct (Roncada *et al*., 2002). Moreover, this method has also led to the identification of differences in the degree of phosphorylation as well as some glycosylated forms amongst casein families (Manso *et al*., 2005). Variations in molecular mass and net charge of proteins is generally brought about by modifications such as phosphorylation, deamidation, glycosylation and proteolytic cleavage thus making 2D PAGE the method of choice in separation of milk isoforms. Over the past 20 years, fascinating transformations have been witnessed in organic and biological mass spectrometry in every respect, which include application, user base, software and instrumentation (Denisov *et al*., 2012). In spite of all the developments witnessed in MS that also include Fourier Transformation Ion Cyclotron Resonance (FT ICR), the introduction of Orbitrap analysis instruments is undoubtedly one of the major developments in recent times (Hu *et al*., 2005; Makarov *et al*., 2009). Orbitrap technology is capable of ultra-high resolving power of 1 000 000 with optimum tolerance and tuning. Such high resolution allows resolution of fine isotopic structure and direct identification of some important posttranslational modifications of peptides which opens up analysis of complex protein mixtures.

In this chapter, the main objective is to separate and identify individual proteins of milk from African elephant using 2D PAGE combined with mass spectrometry and thus circumvent poor resolution, a shortcoming of SDS PAGE. Although the technique allows sequence analysis of many proteins simultaneously, the emphasis will be on the characterization of α-LA and caseins.

3.2. Materials and Methods

The materials and methods for sample preparation and protein concentration determination were described in sections 2.2.1 and 2.2.2 respectively.

3.2.1 Two-dimensional polyacrylamide gel electrophoresis of the protein samples (2D PAGE)

3.2.1.1 Isoelectric focusing electrophoresis (IEF) / First dimension separation

The 7 cm ReadyStrip IPG strips, pH 3-10 (Bio-Rad, South Africa) for the four samples (E1, E2, E3 and cow) were taken out of the -20 °C freezer and thawed at room temperature. The lyophilized rehydration powder was reconstituted by completely dissolving it in 6.1 ml nanopure water. For the first dimension of separation, 70 µg proteins per sample was mixed with rehydration buffer, incubated for 30 minutes before loading the solution on a rehydration tray. The IPG strips were stripped off the plastic coverings and carefully transferred to the rehydration tray using clean forceps. The strips were then covered with 2 ml mineral oil (Bio-Rad, South Africa) and rehydration was conducted in an IEF cell (Bio-Rad, South Africa) at 20 °C overnight.

The IEF focusing tray was prepared by covering the wire electrodes with wicks which were moistened with 8 µl nanopure water. Excess mineral oil from the rehydrated strips was removed by blotting with filter paper. The strips were then placed in IEF focusing tray wells and covered by 2 ml mineral oil. For the IEF run, the parameters were as follows. Voltage was set at 4 000 V for two hours with linear ramp and up to 10 000 V with rapid ramp overnight.

3.2.1.2. Second dimension separation / SDS PAGE

Equilibration of the IPG strips was done immediately after the IEF run was completed. The strips were placed in equilibration buffer (Bio-Rad, South Africa) (7.2 g urea; 4 ml 10 % w/v SDS; 0.66 ml 1.5 M Tris/HCl, pH 8.8; 5 ml 80 % v/v glycerol and water to make up to 20 ml). To 10 ml of the equilibration buffer, 200 mg dithiotreitol (Bio-Rad, South Africa) was added. The buffer was separated into four equal parts for the four duplicate samples and poured into petri dishes; the strips were placed into the petri dishes and incubated for 10 minutes at room temperature on a

shaker. The strips were then transferred to the iodoacetamide equilibration buffer containing 250 mg iodoacetamide, incubation was done as above.

A 12 % SDS separating gel was used for the second dimension electrophoresis. The IPG strips containing the milk proteins were first dipped in 1X SDS running buffer (Bio-Rad, South Africa) before they were applied on top of the gel using clean forceps, and overlay agarose (Bio-Rad, South Africa) was applied on top to cover and keep the strips in place during the second dimension SDS PAGE run. SDS PAGE was run at a voltage of 200 V and current of 0.03 A (4 gels) for the first 30 minutes. Current was increased to 0.06 A for the rest of the electrophoresis run until the bromophenol blue front of the overlay agarose reached the bottom of the gel. Coommassie Brilliant Blue R-250 Stain was used to stain the gels.

3.2.2. In-gel digestion of proteins

After the completion of the second dimension separation of 2D PAGE, the protein spots of interest were selected and then carefully cut out from the gel. Each spot was then cut into small cubes of approximately 1 mm^3 in volume and transferred to sterile 1.5 ml micro centrifuge tubes. The gel pieces were then washed by adding 800 µl Milli-Q pure water and incubated for 15 minutes. Excess water was removed after the incubation time had elapsed, 800 µl of 50 % acetonitrile was added to the tubes and incubated for 15 minutes. The steps above were repeated again before the gel pieces were shrunk by adding 200 ml of 100 % acetonitrile until the gel pieces stuck together and turned white.

Reduction and alkylation of the proteins trapped inside the gel matrix followed the washing step. To do this 200 µl of 10 mM dithiotreitol/0.1 M ammonium carbonate was added to the gel particles and incubated for 45 minutes at 56 °C to allow the shrunken gel pieces to swell. Excess liquid was removed and 200 µl of 10 mg ml^{-1} (55

mM) Iodoacetamide in 0.1 M ammonium carbonate was added, followed by incubation for 30 minutes in darkness. Following reduction and alkylation, the gel particles were washed and shrunk again as previously described above and dried in a fume hood immediately after washing. For the trypsin digestion, 40 µl of re-suspension buffer was added to 2 µg trypsin (Promega) to make up a 500 ng ul^{-1} solution. To 10 µl of this solution, 250 µl of 100 mM ammonium carbonate was added to constitute the digestion solution. The digestion solution was added to respective gel segments according to the intensities of the spots.

After digestion the digest solution was transferred into clean 1.5 ml micro centrifuge tubes. The gel pieces that remained were treated with 150 µl of 5 % v/v formic acid (Merck), vortexed for 10 seconds and incubated at room temperature for 15 minutes. 150 µl of 100 % v/v acetonitrile was added to the tubes after the incubation period; the mixture was vortexed for 10 seconds and incubated on the bench for 15 minutes. The mixture was centrifuged at 10 000 rpm for 5 minutes and the supernatant was transferred to the epperndorf tubes containing the digest solution. The above processes were repeated again and the extracted peptides were mixed with the rest of the digestion product. The solution containing the extracted peptides was then dried in a Vapor Trap (Thermo Electron Corporation) and later subjected to MS analysis.

3.2.3 Mass spectrometric analysis (Tandem Mass spectrometry)

Following digestion by trypsin, the peptides generated were extracted for NanoLC/MS/MS analysis. Five microliters of each digest was individually injected and concentrated on a C18 reverse phase trapping column and then eluted onto, and separated by, a custom packed C18 reverse phase column. The peptides were separated and eluted from the column with a 10 to 25 % Actetonitrile, containing 5 % formic acid gradient, over 60 min. The eluting peptides were analyzed on an AB SCIEX 4000QTRAP hybrid triple quadrupole ion trap MS with a nanospray source at 350 $nl.min^{-1}$ using an Agilent Nano HPLC.

A survey scan between 400 and 1200 Da was performed looking for eluting peptides; an enhanced resolution scan was performed on peptides to determine the charge state of each peptide before fragmenting the peptides in the collision cell. The peptide sequence information obtained from this MS/MS analysis was analyzed by an in-house Mascot server using the latest Swissprot database (Perkins *et al*., 1999). The following parameters were used: mass tolerance of 100 ppm, a minimum of two peptides matching to the protein, carbamidomethylation of cysteine as a fixed modification, oxidation of methionine and pyroglutamylation of glutamine as variable modifications, and four missed cleavages allowed.

3.2.4 Orbitrap Mass spectrometry

3.2.4.1 In-gel trypsin digestion

All gel pieces were cut into small cubes and washed twice with water followed by 50 % (v/v) acetonitrile for 10 minutes. The acetonitrile was replaced with 50 mM ammonium bicarbonate and incubated for 10 min, and repeated two more times. All the gel pieces were then incubated in 100 % acetonitrile until they turned white, after which the gel pieces were dried. Proteins were reduced with 10 mM dithiotreitol for 1 hour at 57 °C. This was followed by brief washing steps of ammonium bicarbonate followed by 50 % acetonitrile before proteins were alkylated with 55 mM iodoacetamide for 1 h in the dark. Following alkylation, the gel pieces were washed with ammonium bicarbonate for 10 minutes, followed by 50 % acetonitrile for 20 minutes, before being dried in vacuo. The gel pieces were digested with 20 µl of a 10 $ng.\mu l^{-1}$ trypsin solution at 37 °C overnight. The resulting peptides were extracted twice with 70 % acetonitrile in 0.1 % formic acid for 30 minutes, and then dried and stored at -20 °C. Dried peptides were dissolved in 5 % acetonitrile in 0.1 % formic acid (Merck) and 10 µl were injected for nano-LC chromatography.

3.2.4.2 Mass spectrometry

All experiments were performed on a Thermo Scientific EASY-nLC II connected to a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with a nano-electropsray source. For liquid chromatography, separation was performed on an EASY-Column (2 cm, ID 100µm, 5 µm, C18) pre-column followed by XBridge BEH130 NanoEase column (15 cm, ID 75 µm, 3.5 µm, C18) column with a flow rate of 300 nl min^{-1}. The gradient used was from 5-17 % B in 5 minutes, 17-25 % B in 90 minutes, 25-60 % B in 10 minutes, 60-80 % B in 5 minutes and kept at 80 % B for 10 minutes. Solvent A was 100 % water in 0.1 % formic acid, and solvent B was 100 % acetonitrile (Sigma) in 0.1 % formic acid.

The mass spectrometer was operated in data-dependent mode to automatically switch between Orbitrap-MS and LTQ-MS/MS acquisition. Data were acquired using the Xcaliber software package. The precursor ion scan MS spectra (m/z 400-2000) were acquired in the Orbitrap with resolution R = 60000 with the number of accumulated ions being 1 x 10^6. The 20 most intense ions were isolated and fragmented in linear ion trap (number of accumulated ions 1.5 x 10^4) using collision induced dissociation. The lock mass option (polydimethylcyclosiloxane; m/z 445.120025) enabled accurate mass measurement in both the MS and MS/MS modes. In data-dependent LC-MS/MS experiments, dynamic exclusion was used with 60 s exclusion duration. Mass spectrometry conditions were 1.8 kV, capillary temperature was 250 °C, with no sheath and auxiliary gas flow. The ion selection threshold was 500 counts for MS/MS and an activation Q-value of 0.25 and activation time of 10 ms were also applied for MS/MS.

3.2.4.3 Data analysis

Thermo Proteome Discoverer 1.3 (Thermo Scientific, Bremen, Germany) was used to identify proteins via automated database searching (Mascot, Matrix Science, London, UK) of all tandem mass spectra against the Swissprot 57.15, NCBI mammalian and Uniprot elephant databases. Carbamidomethyl cysteine was set as fixed modification, and oxidized methionine, N-acetylation and deamidation (NQ) were used as variable modifications. The precursor mass tolerance was set to 10 ppm, and fragment mass

tolerance set to 0.8 Da. Two missed tryptic cleavages were allowed. Proteins were considered positively identified when they were identified with at least 2 tryptic peptides per proteins, a Mascot score of more than $p<0.05$ as determined by Proteome Discoverer 1.3. Percolator was also used for validation of search results. In Percolator a decoy database was searched with a FDR (strict) of 0.02 and FDR (relaxed) of 0.05 with validation based on the q-value.

3.3. Results

Three milk protein samples of African elephant (denoted by E1, E2 and E3) were subjected to 2D PAGE and the subsequent gels are shown in Figure 3.1. The different spots on the gels represent pure proteins separated from each other by charge as well as size. The resolution is quite significant and both high and low abundant proteins can be observed.

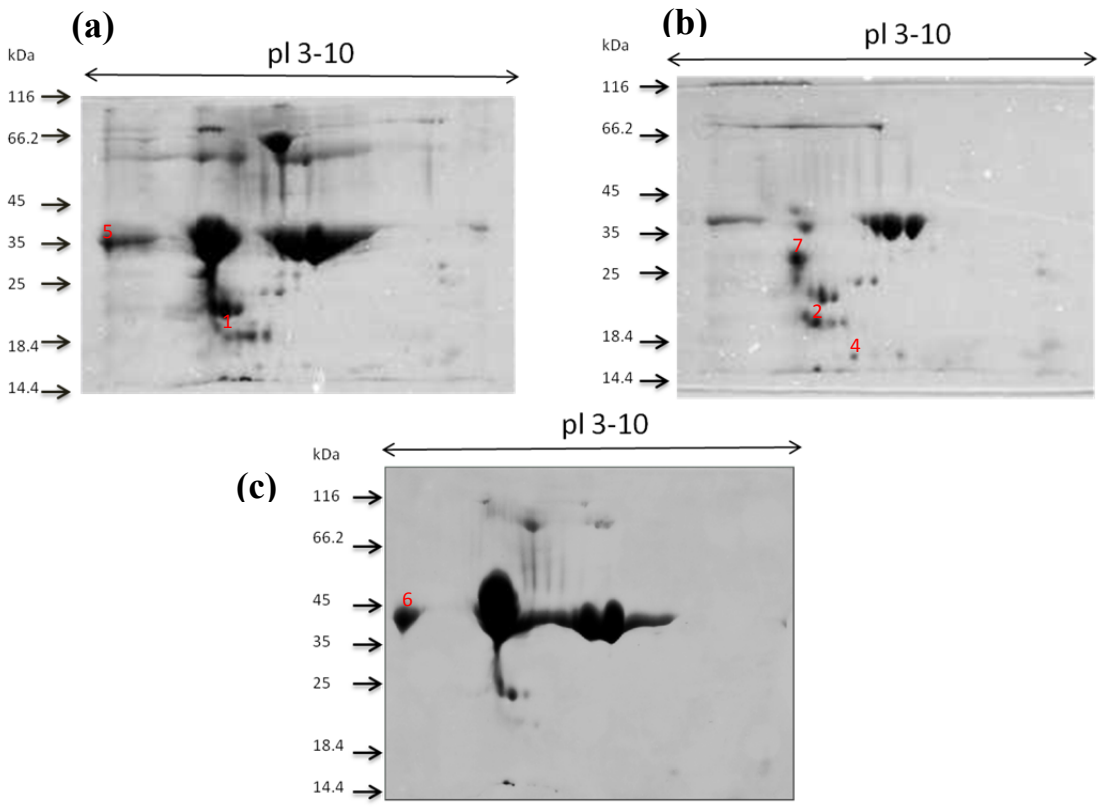

Figure 3.1. 2D PAGE gels of African elephant milk protein at (a) early lactation (4 days), (b) mid lactation (12 months) and (c) late lactation (18 months).

To concentrate the whey proteins, caseins were precipitated in order to analyze low abundance proteins (Roncada *et al*., 2012). Figure 3.2 shows two gels that represent the fractionated samples, thus most of the proteins in the samples were whey proteins. The figure represent milk at early and late lactation stages denoted E1 and E2 respectively.

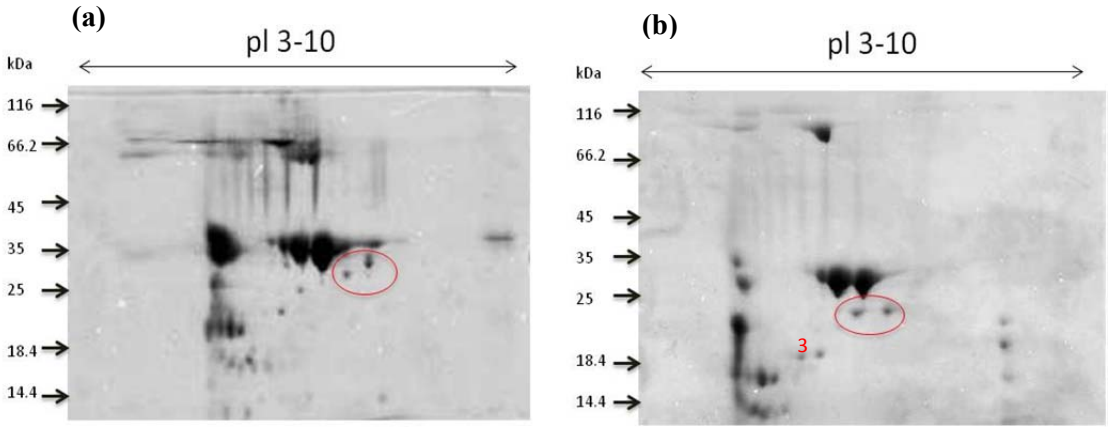

Figure 3.2. 2D PAGE gels of fractionated African elephant milk protein samples at (a) early lactation (4 days) and (b) mid lactation (12 months). Circled areas indicate spots that are only visible after fractionation. The spot number is situated just above the spot that was identified.

A 2D PAGE gel of cow sample was also carried out as a reference (O'Donnell *et al*., 2004). This was done in order to compare and validate the methods, techniques and results of the current project. Figure 3.3 shows a comparison of a 2D PAGE gel of cow's milk of the current project with one in the literature. Although a pH range between 3 and 10 was chosen for the current project compared to the pH range between 4 and 7, the similarity in migrated protein spots is obvious. Although cow's milk proteins showed better separation on a gel of pH range between 4 and 7, the pH range between 3 and 10 was chosen for the current research because of improved separation of proteins from African elephant.

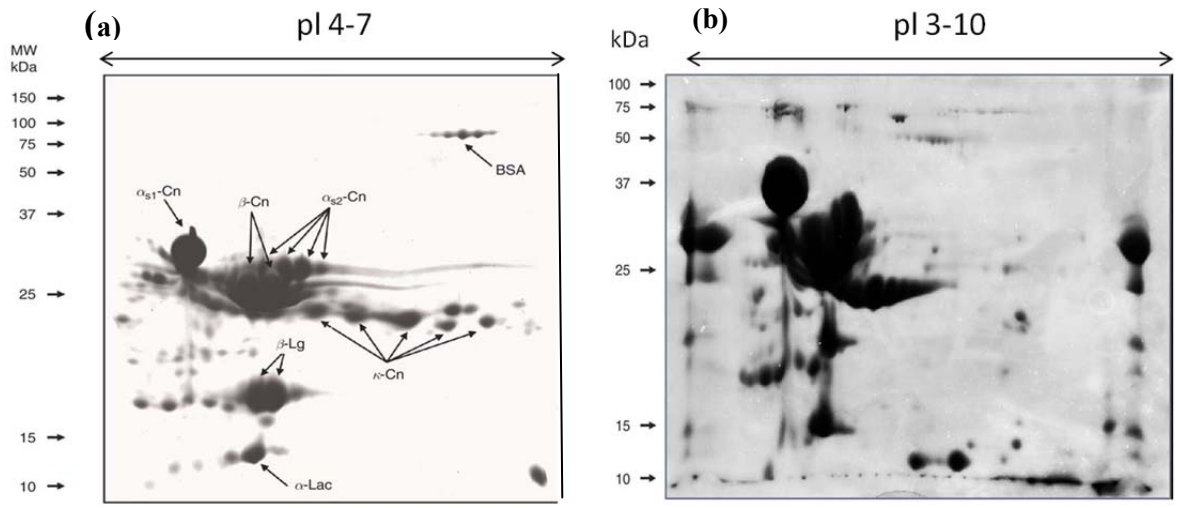

Figure 3.3. Comparison of 2D PAGE gels of cow's milk proteins according to Chevalier (2004) (a) and gel of the current project (b). In (a) the pH range was between 4 and 7 whereas in gel (b) a pH range between 3 and 10 was used.

The individual protein spots of interest were cut out from the gel, digested with trypsin and analyzed by MS. Identification of some protein spots was successful whereas some spots could not be identified. Table 3.1 shows the results of the spot identifications made from the gels. Most of the major proteins in African elephant milk that were identified from 1D PAGE gels using LC-MS/MS could not be identified from 2D PAGE.

Table 3.1. Proteins identified from spots on 2D PAGE gels shown and numbered in Figure 3.1 using LC-MS/MS

Sample/spot	Protein identified
1	Alpha-lactalbumin
2	Alpha-lactalbumin
3	Serum albumin
4	Serum amyloid A-3 protein
5	Alpha-S1-casein
6	Elongation factor 1-alpha 1
7	Serum albumin
3A2, 6A1,4A2,6A6,6A7,6B6,4B2,2B1	No Identifications

Since some of the major milk proteins could not be identified with LC-MS/MS, an orbitrap MS was utilized as it poses a number of advantages over tandem MS. These include, increased sensitivity, high resolution, less detection time, high accuracy in mass detection and finally direct identification of certain post translational modifications (Denisov *et al*., 2012). African elephant milk protein samples were used for the 2D PAGE gels; Figure 3.4 shows the 2D PAGE gel of these samples.

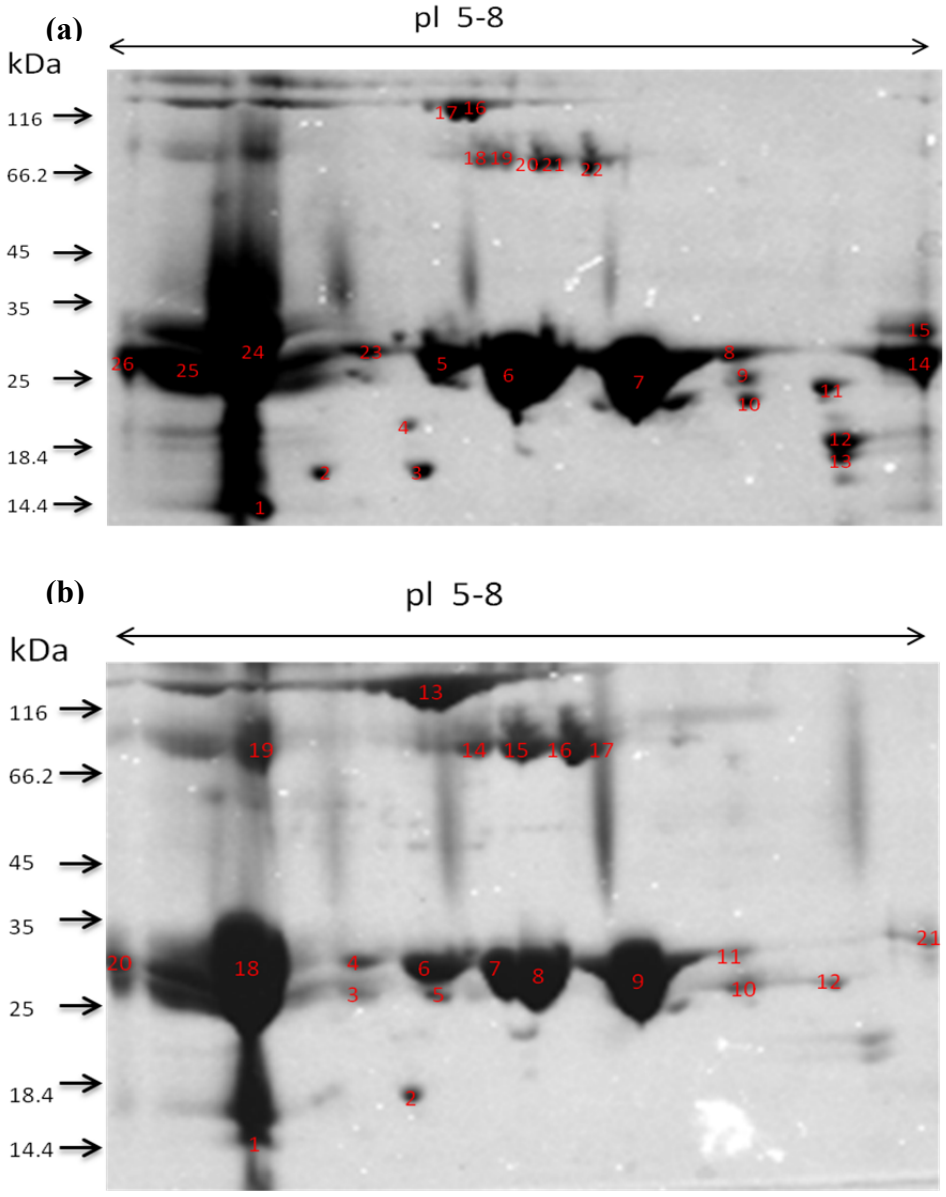

Figure 3.4. 2D PAGE gel of African elephant milk proteins, (a) milk at mid (12 months) lactation and (b) milk at late lactation (18 months).

By using In-gel digestion and subsequent orbitrap MS, several spots on the gel represented by Figures 3.4a and 3.4b were analyzed. The protein identities of the

spots on the gel are shown in table 3.2. Figures 3.5a, 3.5b and 3.5c show the amino acid sequences of African elephant α-LA, κ-casein and β-casein as obtained from the search query of the elephant database (Broad institute, elephant database). The MS data gave information such as, estimated pI, unique peptides identified, non-unique peptides identified, the peptide score, molecular weight as well as the protein length. The peptide sequences in green are unique to that of African elephant milk protein and the ones colored in red can be found in other sequences whilst the sequences in black were completed out by the program. These peptides were identified by the MS after digestion of protein spots on gels, the main milk protein groups identified are depicted by sequences below.

α-LA

AKMMSFVPLL LVGILFPAIQ AKQFTK**CELS QVLK**DIDGYA GITLPEFTCT IFHISGYDTQ
TIVNNNGSTE YGLFQISNKY WCR**DHQIPQS RNICDISCDK FLDDDLTDDM MCAK**KILDSK
GIDYW

Figure 3.5a. Amino acid sequence of African elephant α-lactalbumin.

β-casein

MKVFILACLV AFALGRETVE NLSSSEESVT QVNKQKPEGV **KHEEQQREDE HQNK**IQPLFQ
PQPLVYPFAE PIPYTVFPPN AIPLAQPIVV LPFPQPEVK**Q LPEAKEITFP R**QK**LMSFLKS**
PVMPFFDPQI PNLGTDLENL HLPLPLLQPL RHQLHQPLAQ TPVLPLPLSL PKVLPVPQQV
IPYPQRGRPI QNLQLYEEPL LDPTRKIYPV AQPLAPVYNP VAV

Figure 3.5b. Peptide sequence of African elephant β-casein.

κ-casein

MMKGFLLVVN ILLLPLPFLA AEVQNQEESR CLEKDERWFC QKAVKYIPND **YVLKSYYRYE**
PNYNQFRAAV PINNPYLIYL YPAKQVAVRP HTQIPQWQVP SNIYPSPSVP HTYLKPPFIV
IPPKKTQDKP IIPPTGTVAS IEATVEPKVN TVVNAEASSE FIATNTPEAT TVPVISPQI

Figure 3.5c. Amino acid sequence of African elephant κ-casein.

Table 2.3. Spot identifications after analysis by 2D PAGE and Orbitrap MS of Gel A and Gel B (Shown in figures 3.4a and 3.4b) of milk proteins of African elephant.

Spot	Gel A	Gel B
1	Alpha lactalbumin	Alpha lactalbumin, proSAAS
2,3,4	No I.D	No I.D
5	Beta casein	Beta casein, Serum albumin
6	Beta casein	No I.D
7	Beta casein	Beta casein
8	Beta casein	Beta casein, Suprabasin isoform 1
9	Beta casein, Serum albumin	Beta casein
10	No I.D	Beta casein
11	No I.D	Beta casein, Serum albumin
12	Kappa casein, Cofilin-1	Serum albumin
13	Peptidoglycan recognition protein	Cofilin-1
14	No I.D	Serum albumin, Beta casein
15	No I.D	Serum albumin, Beta casein
16	No I.D	Beta casein
17	Serum albumin	Serum albumin, Beta casein
18	Serum albumin, Polymeric immunoglobulin receptor	Beta casein
19	Lactotransferrin, Serum albumin	Vitamin D binding protein
20	Haptoglobin	Transport associated protein
21	Serum albumin	
22	Lactoferrin, serum albumin	
23	Serum albumin	
24,25	No I.D	

Multiple sequence alignments were done to compare the amino acid sequences (α-LA, β- and κ-casein) of African elephant milk obtained from orbitrap mass spectrometry with corresponding milk protein sequences from the databank using clustal omega version 1.2.0 (Sievers *et al*., 2011). Figures 3.6, 3.7 and 3.8 show the three alignments.

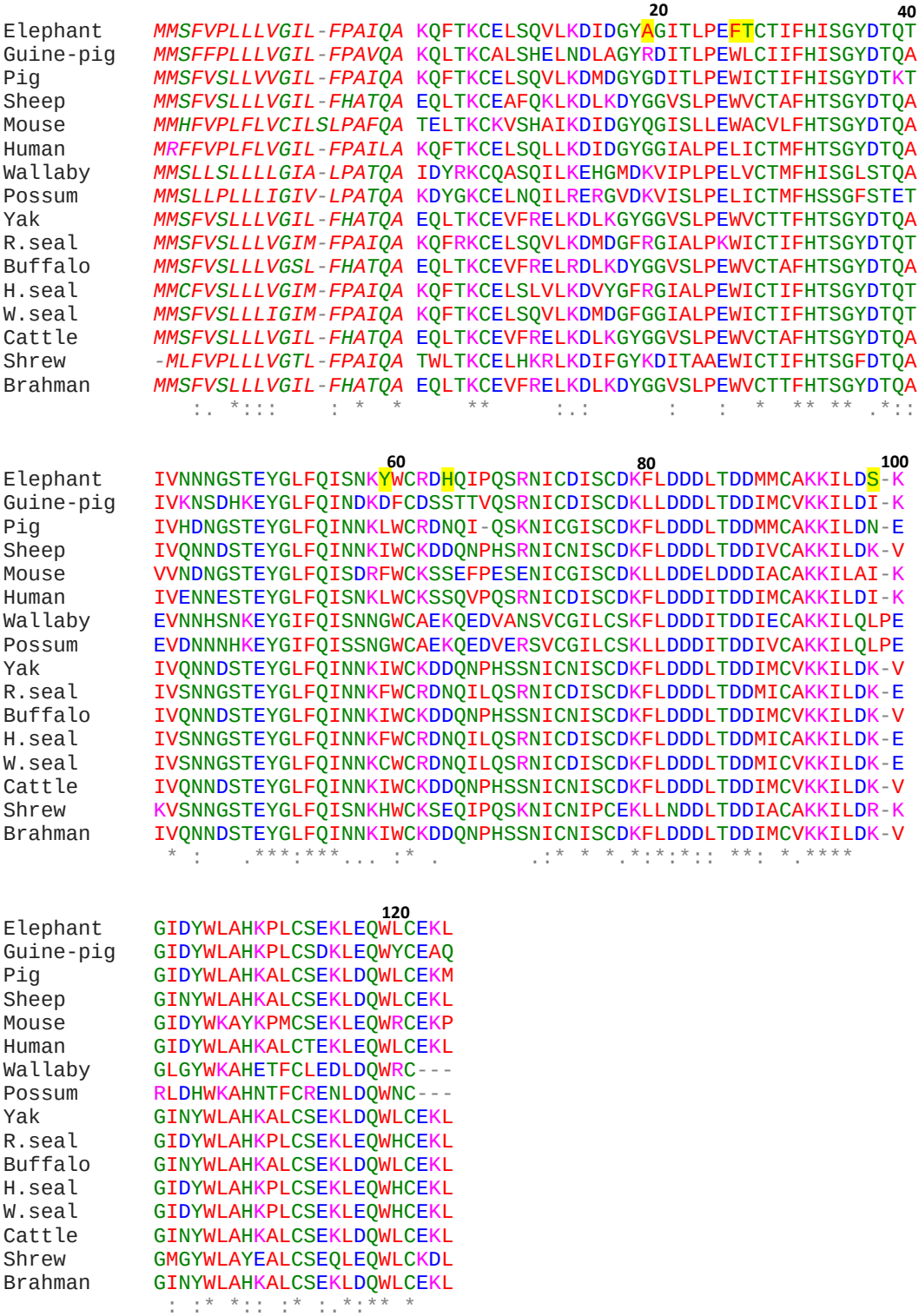

Figure 3.6. Multiple sequence alignment of α-LA from 16 mammalian species. The asterisks, colons and full stops beneath the alignment indicate highly conserved amino acid residues (>12/16) between species. The signal sequence is split from the rest of the sequence and is presented in italics. Amino acid residues are color coded according to their different properties, red (small and hydrophilic); blue (acidic); magenta (basic); green (hydroxyl, sulfhydryl, amine and Glycine).

```
                                                                  20
Elephant    MMKGFLLVVNILLLPLPFLAA EVQNQEESRCLEKDERWFCQKAVKYIPNDYVLKSYYRYE
Mouse       MMRNFIVVVNILALTLPFLAA EIQNPDSNCRGEKNDIVYDEQRVLYTPVRSVLNF-NQYE
Rat         MMRNFIVVMNILALTLPFLAA EVQNPDSNCRE-KNEVVYDVQRVLYTPVSSVLNR-NHYE
Rabbit      MMKHFLLVVNILAVTLPFLAA DIQNQEQTTCRENEERLFHQVTAPYIPVHYVMNRYPQYE
Goat        MMKSFFLVVTILALTLPFLGA QEQNQEQPICCEKDERFFDDKIAKYIPIQYVLSRYPSYG
Sheep       MMKSFFLVVTILALTLPFLGA QEQNQEQRICCEKDERFFDDKIAKYIPIQYVLSRYPSYG
Cow         MMKSFFLVVTILALTLPFLGA QEQNQEQPIRCEKDERFFSDKIAKYIPIQYVLSRYPSYG
W.Buffalo   MMKSFFLVVTILALTLPFLGA QEQNQEQPIRCEKEERFFNDKIAKYIPIQYVLSRYPSYG
Pig         MMKSSFLIVPILALTLPFLGA EEQNQEKLTRCESDKRLFNEEKVKYIPIYYMLNRFPSYG
Human       -MKSFLLVVNALALTLPFLAV EVQNQKQPACHENDERPFYQKTAPYVPMYYVPNSYPYYG
Horse       -MKSFFLVVNILALTLPFLGA EVQNQEQPTCHKNDERFFDLKTVKYIPIYYVLNSSPRYE
             *:  ::::  * : ****.. : ** ..     .:.  :    . * *   : .    *

            40                  60                  80                 100
Elephant    PNYNQFRAAVPIN-NPYLIYLYPAKQVAVRPHTQIPQWQVPSNIY--------PSPSVPH
Mouse       PNYYHYRPSLPATASPYMYYPLVVRLLLLRSPAPISKWQSMPNFP--------QSAGVPY
Rat         PIYYHYRTSVPV--SPYAYFPVGLKLLLLRSPAQILKWQPMPNFP--------QPVGVPH
Rabbit      PSYYLRRQAVPTL-NPFMLNPYYVKPIVFKPNVQVPHWQILPNIH--------QPKVGRH
Goat        LNYYQQRPVALIN-NQFLPYPYYAKPVAVRSPAQTLQWQVLPNTVPAKSCQDQPTTLARH
Sheep       LNYYQQRPVALIN-NQFLPYPYYAKPVAVRSPAQTLQWQVLPNAVPAKSCQDQPTAMARH
Cow         LNYYQQKPVALIN-NQFLPYPYYAKPAAVRSPAQILQWQVLSNTVPAKSCQAQPTTMARH
W.Buffalo   LNYYQQKPVALIN-NQFLPYPYYAKPAAVRSPAQILQWQVLPNTVPAKSCQAQPTTMTRH
Pig         F-FYQHRSAVSPN-RQFIPYPYYARPVVAGPHAQKPQWQDQPNVY--------PPTVARR
Human       TNLYQRRPAIAIN-NPYVPRTYYANPAVVRPHAQIPQRQYLPNSH--------PPTVVRR
Horse       PIYYQHRLALLIN-NQHMPYQYYARPAAVRPHVQIPQWQVLPNIY--------PSTVVRH
                  :         .       .       .   : *   *

                                120                 140                160
Elephant    TYLKPPFIVIPPKKTQDKPIIPPTGTVASIEATV----------EPKVNTVVNAEASSEF
Mouse       AIPNPSFLAMPTNENQDNTAIPTIDPITPIVSTPVPTMES------IVNTVANPEASTV-
Rat         PIPNPSFLAIPTNEKHDNTAIPASNTIAPIVSTPVSTTES------VVNTVANTEASTV-
Rabbit      --SHPFFMAILPNKMQDKAVTPTTNTIAAVEPTPIPTTEPV------VSTEVIAEASPEL
Goat        PHPHLSFMAIPPKKDQDKTEVPAINTIASAEPTVHSTPTT----EAIVNTVDNPEASSE-
Sheep       PHPHLSFMAIPPKKDQDKTEIPAINTIASAEPTVHSTPTT----EAVVNAVDNPEASSE-
Cow         PHPHLSFMAIPPKKNQDKTEIPTINTIASGEPTS--TPTT----EAVESTVATLEDSPE-
W.Buffalo   PHPHLSFMAIPPKKNQDKTEIPTINTIVSVEPTS--TPIT----EAIENTVATLEASSE-
Pig         PRPHASFIAIPPKKNQDKTAIPAINSIATVEPTIVPATEPIVNAEPIVNAVVTPEASSEF
Human       PNLHPSFIAIPPKKIQDKIIIPTINTIATVEPTPAPATEP------TVDSVVTPEAFSES
Horse       PCPHPSFIAIPPKKLQEITVIPKINTIATVEPTPIPTPEP------TVNNAVIPDASSEF
               .   *:.:  :: ::    *    :.    *              .     :

                           180
Elephant    IATNTPEATTVPVISPQI---
Mouse       -SINTPETTTVPVSSTAA---
Rat         -PISTPETATVPVTSPAA---
Rabbit      IISPETTTEATAA-SAAA---
Goat        SIASASETNTAQVTSTEV---
Sheep       SIASAPETNTAQVTSTEV---
Cow         VIESPPEINTVQVTSTAV---
W.Buffalo   VIESVPETNTAQVTSTVV---
Pig         LITSAPETTTVQVTSPVV---
Human       IITSTPETTTVAVTPPTA---
Horse       IIASTPETTTVPVTSPVVQKL
                    :. .
```

Figure 3.7. Multiple sequence alignment of κ-casein from 11 mammalian species. The asterisks, colons and full stops beneath the alignment indicate highly conserved amino acid residues (>7/11) between species. The signal sequence is split from the rest of the sequence and is presented in italics. Amino acid residues forming the chymosin sensitive bond are indicated by a red rectangular block.

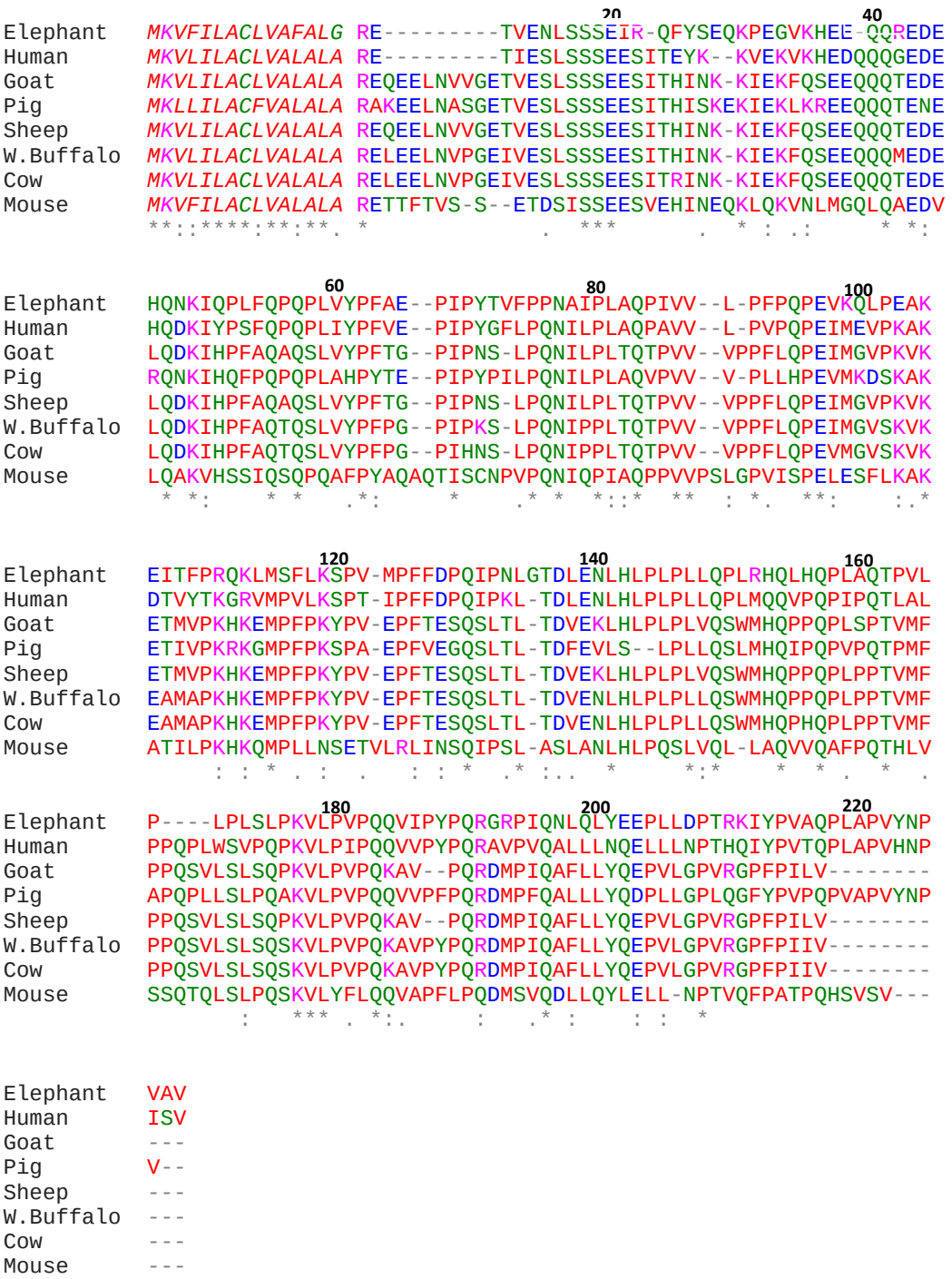

Figure 3.8. Multiple sequence alignment of β-casein from 8 mammalian species. The asterisks, colons and full stops beneath the alignment indicate highly conserved amino acid residues (>7/11) between species. The signal sequence is split from the rest of the sequence and is presented in italics. Amino acid residues are color coded according to their different properties, red (small and hydrophilic); blue (acidic); magenta (basic); green (hydroxyl, sulfhydryl, amine and Glycine).

3.4. Discussion

The 2D PAGE gels shown in Figure 3.1 have striking differences in terms of number and intensity of protein spots on the gel. Some protein spots seem to either increase in concentration (β-casein) or decrease (Immunoglobulins) significantly as the lactation period progresses. It appears that whey proteins are decreasing as lactation progresses and as a result the caseins become more prominent, specifically the β-caseins as identified on Figure 3.4. This data supports evidence observed in chapter two, where SDS PAGE was used to separate protein samples at different stages of lactation and adjacent protein bands showed significant differences in their intensities. This change in milk proteome at different stages of lactation can be attributed to changes in nutritional requirements of the neonate for growth, immunological protection and development (Kunz & Lonnerdal, 1992; Powe *et al*., 2010). Comparative studies have shown marked differences in terms of milk composition of mammalian species during the entire lactation period, after the expression of the initial colostrum, eutherian species produce milk of relatively constant composition, whereas milk of marsupials on the other hand changes progressively in composition during lactation (Murakami *et al*., 1998; Sharp *et al*., 2007).

The 2D PAGE gels in Figure 3.2 compare the distribution of spots on gels in which equal protein concentrations of fractionated milk proteins was loaded. The protein samples were obtained from milk at early and mid-lactation. Caseins normally occur in higher amounts than whey proteins, with human milk being an exception. However isoelectric precipitation of caseins (fractionation), allows concentration or enrichment of low abundance whey proteins (Martin *et al*., 2003; Yamada *et al*., 2002). Enrichment of whey proteins enables visualization of protein spots that are not initially visible as indicated by circled areas of Figure 3.2. Additionally this allows more options in terms of the number of spots that can be investigated.

Figure 3.3 shows a comparison of 2D PAGE gels for cow's milk, the two gels generally look similar in terms of the distribution of proteins, notably position of the major milk proteins. The cow's 2D PAGE gel gave a general idea concerning

probable positions from which certain proteins on the African elephant milk 2D PAGE gels, can be located as well as validate the methods. Figure 3.3a was adapted from (Chevalier, 2004), in this experimental work, a narrow pH range (4-7) was used compared to our own gel (Figure 3.3b) in which we used a wider pH range (3-10). Use of a narrow pH range results in increased resolution among the protein spots whereas a wider pH range has the advantage of displaying the most proteins in a single gel. However, for this study, both narrow (5-8) and wider (3-10) pH ranges were used. A wider pH range gave better coverage of protein spots that are displayed on the gel regarding the elephant milk proteins whereas a narrow pH range gave better resolution between the proteins. This extended protein spots coverage on 2D PAGE gels was of paramount importance during comparative studies of the 2D PAGE gels at different lactation stages.

Identification of individual spots on 2D PAGE gels shown in Figure 3.1 using LC-MS/MS only yielded a few positive results. The α-LA and α_{s1}-casein were the only two milk proteins of interest identified. At this stage, the database used did not have elephant milk proteins so these proteins were correlated to other species' proteins such as bovine and pig. Since this was not satisfactory, orbitrap MS was employed and an elephant protein database was used for the identification of peptides.

For orbitrap MS, a gel of a narrower pH range (5-8) was utilized to improve resolution between the larger spots (Figure 4.3a and 4.3b). Upon orbitrap MS analysis, more proteins were identified (Table 3.1). In addition to α-LA, lactoferrin, β- and κ-casein were identified. However, α_{s1} casein was not identified. To confirm the protein identities, the molecular weight and isoelectric points of the identified proteins, which were predicted by the MS, also correlated well with the molecular weight and isoelectric points on the 2D PAGE gel.

Beta caseins seem to dominate the total content of milk proteins of African elephant milk. Spots 5, 6, 7 and 8 on gel A (Figure 3.4) were all shown to contain β-casein. A

closer inspection of the spots showed that they are distinct from each other and are of approximately the same molecular weight. The different pI's of the proteins can be explained by the proteins being different isomers due to post translational modification, i.e. phosphorylation of serine residues (Horne, 2008). It is possible that differences in the numbers of the phosphoserines may result in the same group of proteins assuming different positions on a 2D gel since the basis of 2D separation is charge and size.

The amino acid sequences of α-LA, β- and κ-casein are depicted in Figures 3.5a, 3.5b and 3.5c. Sequences of α-LA, β- and κ-casein were aligned with corresponding proteins sequences from other mammalian species for comparison, the multiple sequence alignments are shown in Figures 3.6, 3.7 and 3.8.

The sequence alignment of α-LA is shown in figure 3.6, 16 amino acid sequences were aligned. From the alignment, several amino acids are shown to be unique to the African elephant α-LA sequence. On position 19, 26, 27, 59, 65 and 98, the elephant sequence has alanine, phenylalanine, threonine, lysine, histidine and serine respectively. These amino acids are unique to the African elephant α-LA sequence. Such substitutions may have some implications in the way the protein is folded, aswell as how it functions or interacts with other proteins, especially considering the role of α-LA in lactose synthesis. For example, histidine on position 65 is the only amino acid residue on that position that is not polar or charged, compared to the same protein of other species. The rest of the sequences on the latter position are either charged (basic or acidic) or hydrophilic. The same can be described for serine on amino acid position 98, which is polar but not charged whereas in other species the amino acids at that position are either polar and charged or hydrophobic.

Figure 3.7 shows the multiple sequence alignment of κ-casein. This protein is structurally amphipathic; it interacts hydrophobically with other caseins and provides a negatively charged and hydrophilic surface on the micelle. The signal peptide of

African elephant κ-casein comprise of 21 amino acids, one amino acid longer than some few mammalian species such as human and horse. The C-terminal of African elephant κ-casein comprises of mostly hydrophilic amino acids with serine residues that can potentially be phosphorylated. Unlike the cow κ-casein chymosin cleavage site, which constitutes of phenylalanine and methionine, the chymosin cleavage site of African elephant κ-casein comprises of phenylalanine and isoleucine. This property is also shared in pig, human and horse κ-casein (Nakhasi *et al*., 1984). The *N*-terminus of African elephant κ-casein constitutes of mostly hydrophobic amino acid residues. This structural arrangement of residues in African elephant κ-casein sequence allows the protein to function in stabilization and maintaining the colloidal nature of milk.

The multiple sequence alignment of β-casein is shown in Figure 3.8. In comparison to the other β-casein sequences in the alignment, African elephant β-casein also does not contain a cysteine residue in its sequence. There are several other properties that the African elephant β-casein shares with other sequences in the alignment. These include: presence of a single *N*-terminus located multiple phosphorylation site (residues 15-19), richness in proline residues (39 proline residues) as well as a highly negatively charged *N*-terminus with an increased number of glutamic acid residues. Additionally, the signal peptide of African elephant β-casein shares extensive homology with α_{s1}-casein signal sequence thus hinting some form of evolutionary relationship (Lefèvre *et al*., 2009).

The β-casein sequences from the alignment also show high homology between them, with most of the substitutions being conservative. However, some of the substitutions are not, for example on position 172, where a proline in African elephant β-casein sequence is replaced with a serine in the β-caseins of other species. Furthermore, it is important to note the exonic skipping that occurs in African elephant β-casein (position 3-11). This property is also shared with human β-casein and horse β-casein (not shown on the alignment). Exonic skipping may influence the properties of the peptide including self aggregation and interaction with other proteins.

The aim of this current chapter was to identify protein spots on 2D PAGE gels and obtain the sequences by orbitrap MS analysis. Furthermore, this chapter aimed to compare protein sequences of κ-, β- casein and α-LA with other mammalian species proteins by multiple sequence alignment. In the following chapter, the dynamics of these identified African elephant milk protein spots, at different stages of lactation will be evaluated as well as the protein ratios on the 2D PAGE gels, using PD QUEST (Bioinformatics program).

3.5. References

Aslam, M., Jimenez-Flores, R., Kim, H. Y. & Hurley, W. L. (1994). Two-dimensional electrophoresis analysis of proteins of bovine mammary gland secretions collected during the dry period. *J Dairy Sci* **77**, 1529-1536.

Chevalier, F., Rofidal, V., Vanova, P., Bergoin, A. & Rossignol, M. (2004). Proteomic capacity of recent fluorescent dyes for protein staining. *Phytochemistry* **65**, 1499-1506.

Denisov, E., Damoc, E., Lange, O. & Makarov, A. (2012). Orbitrap mass spectrometry with resolving powers above 1 000 000 *Int J Mass Spectrom* **325-327**, 80-85.

Goldfarb, M. (1999). Two-dimensional electrophoresis and computer imaging: quantitational of human milk casein. *Electrophoresis* **20**, 870-874.

Horne, D. S. (2008). Casein micelle structure and stability. In *Milk proteins: from expression to food,* pp. 133-162. Edited by A. Thomson, M. Boland. & H. Singh. New York: Elsevier.

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M. & Graham-Cooks, R. (2005). The Orbitrap: a new mass spectrometer, *J Mass Spectrom* **40**, 430-443.

Iametti, S., Tedeschi, G., Oungre, E. & Bonomi, F. (2001). Primary structure of κ-casein isolated from mare's milk. *J Dairy Res* **68**, 53-61.

Kunz, C. & Lonnerdal, B. (1992). Re-evaluation of the whey protein/casein ratio of human milk. *Acta Paediatr* **81**, 107-112.

Lefe`vre, C. M., Sharp, J. A. & Nicholas, K. R. (2009). Characterisation of monotreme caseins reveals lineage specific expansion of an ancestral casein locus in mammals. *Reprod Fertil Dev* **21**, 1015-1027.

Makarov, A., Denisov, E. & Lange, O. (2009). Performance evaluation of a high-field Orbitrap mass analyzer. *J Am Soc Mass Spectrom* **20**, 1391-1396.

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

Manso, M. A., Leonil, J., Jan, G. & Gagnaire, V. (2005). Application of proteomics to the characterization of milk and dairy products. *Int Dairy J* **15**, 845-855.

Martin, P., Ferranti, P., Leroux, C., & Addeo, F. (2003). Non-bovine caseins: Quantitative variability and molecular diversity. In *Advanced Dairy Chemistry,* pp. 387-419. Edited by P. F. Fox & P. L. H. McSweeney. New York: Plenum Publishers.

Murakami, K., Lagarde, M. & Yuki, Y. (1998). Identification of minor proteins of human colostrums and mature milk by two dimensional electrophoresis. *Electrophoresis* **19**, 2521-2527.

Nakhasi, H. L., Grantham, F. H. & Gullino, P. M. (1984). Expression of κ-casein in normal and neoplastic rat mammary gland is under the control of prolactin. *J Biol Chem* **259**, 14894-14898.

O'Donnell, R., Holland, J. W., Deeth, H. C. & Alewood, P. (2004). Milk proteomics. *Int Dairy J* **14**, 1013-1023.

Patterson, S. D. & Aebersold, R. A. (1995). Mass spectrometric approaches for the identification of gel separated proteins. *Electrophoresis* **16**, 1791-1814.

Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence database using mass spectrometry. *Electrophoresis* **20**, 3551-3567.

Powe, C. E., Knott, C. D. & Conklin-Brittain, N. (2010). Infant sex predicts breast milk energy content. *Am J Hum Biol* **22**, 50-54.

Sharp, J. A., Lefèvre, C. & Nicholas, K. R (2007). Molecular evolution of monotreme and marsupial whey acidic protein genes. *Evol Dev* **9**, 378-392.

Sievers, F., Wilm A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins D. G (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol* **7**, 1-6.

Rabilloud, T., Chevallet, M., Luche, S. & Lelong, C. (2010). Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J Proteomics* **73**, 2064-2077.

Roncada, P., Piras, C., Soggiu. A., Turk, R., Urbani, A. & Bonizzi, L. (2012). Milk proteins: Farm animal milk proteomics. *J Proteomics* **75**, 4259-4274.

Swissprot database. version 2.2.04: http://cbio.ufs.ac.za/mascot

Yamada, M., Murakami, K., Wallingford. J. C. & Yuki, Y. (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.

Chapter 4

Dynamic changes of African elephant milk proteins over lactation.

4.1. Introduction

Lactation is a defining attribute of mammals and, with evolution being a driving force, lactation has become an extremely efficient, effective and adaptable means of providing nutrient feed for the neonate (Sharp *et al*., 2007). The development of nutritionally enhanced milks has through evolution led to diversity in milk composition and function, quantity of milk output, length of lactation period, length of intervals between nursing and contribution of lactation to the nutrition of offspring. Comparative studies have shown marked differences between different mammalian groups (Murakami *et al*., 1998). The gestation period of eutherian species is longer, compared to their lactation stage, and the milk composition remains relatively constant throughout the whole lactation phase (Oftedal & Iverson 1995).

Two-dimensional polyacrylamide gel electrophoresis remains a popular, versatile and powerful tool for identifying proteins that are differentially expressed across treatment conditions. Additionally it has been instrumental and preceded the birth and development of proteomics (Rabilloud & Lelong, 2011). Although no longer exclusively used, 2D PAGE remains at the fore front of experimental schemes with advantages that include; economic efficiency, robustness, excellent reproducibility and finally, ease of quantitative analysis. Due to the overwhelming number of proteins and the tremendous variation shown in gel images, the differential analysis of 2D gel images is challenging (Li & Seillier-Moiseiwitsch, 2001). The 2D gel analysis often involves multiple comparisons of several gel images hence positional reproducibility is of paramount importance (Rabilloud *et al*., 2010). Commercial software packages

are available for such analysis and often require considerable human intervention for spot detection and matching (Li & Seillier-Moiseiwitsch, 2001).

In this chapter, the dynamics of African elephant milk proteins over lactation time will be evaluated. PD QUEST, a 2D gel image analysis software will be used to detect any differences in terms of distribution of protein spots on 2 D gels, at early (4 days), mid (12 months) and late (18 months) lactation stages. Ultimately, the aim is to detect any proteins that are only expressed at specific stages of lactation. Additionally, protein ratios of the main caseins and whey proteins will be calculated relative to each other by making use of the spots intensities. 2D PAGE gel from cow's milk (mid lactation) will also be incorporated in this analysis. This will allow comparison of the distribution of protein spots of African elephant milk and cow's milk at mid lactation. This latter analysis will also be done on PD QUEST. The spots analysis will allow further characterization of the African elephant proteins that have been identified in the previous chapter and possibly give further insight into their expression throughout lactation.

4.2. Materials and Methods

4.2.1. Sample preparation

The collection of African elephant milk, precipitation and fractionation of proteins was described in section 2.2.1.

4.2.2. Determination of protein concentration

The determination of protein concentration of the African elephant milk samples, using the Bradford assay, was described in section 2.2.2.

4.2.3. Two-Dimensional polyacrylamide gel electrophoresis (2D PAGE)

Separation of protein samples by two-dimensional electrophoresis was described in section 3.2.1.

4.2.4. 2D PAGE gels analysis

Gel images from two-dimensional electrophoresis analysis were analyzed with PD Quest software version 8.0 (Bio-Rad, South Africa). All African elephant milk samples were analyzed in duplicate. The averaged gels (average quantities of protein spots on duplicate gels) were compared to evaluate differences between samples belonging to the three groups (early, mid and late lactation). The same gel regions were selected by square markings and compared. The spot volume was determined as a percentage of total volume of all spots on respective gels. The complete pattern (including all spots) was used for qualitative and quantitative analysis. Qualitative analysis identified specific spots present only in one group, while quantitative analysis determined spots with over- or under-expression. Protein quantities of each identified spot on 2D PAGE gels were also measured on PD QUEST. This was done by measuring the intensities of each spot on the gel; the total intensity of a defined spot in a gel image corresponds to the amount of protein in the actual spot on the gel.

4.3. Results

A striking feature that was observed on the 2D PAGE gel is the presence of numerous protein spots of differing abundances as portrayed by the size and colour intensities of the observed spots. The high abundance proteins are represented by large dark spots, whereas the low abundance proteins are represented by small light spots. Figure 4.1 shows the comparison of spots on the African elephant and cow's milk protein 2D PAGE gels against the master gel using PD QUEST. The gels marked (2) are duplicate gels of African elephant milk at early (4 days) lactation, (3) and (4) are replica gels representing African elephant milk proteins at mid (12 months) and late (18 months) lactation respectively as well as cow's milk at mid lactation (5). The yellow squares show a specific protein spot under analysis (0203) and its position on the gel, while yellow circles show absence of that particular spot on any of the gels under comparison. The graph shows presence of the protein spot 0203 only during early lactation and its absence throughout the rest of the lactation stage as well as its absence in cow's milk at mid lactation.

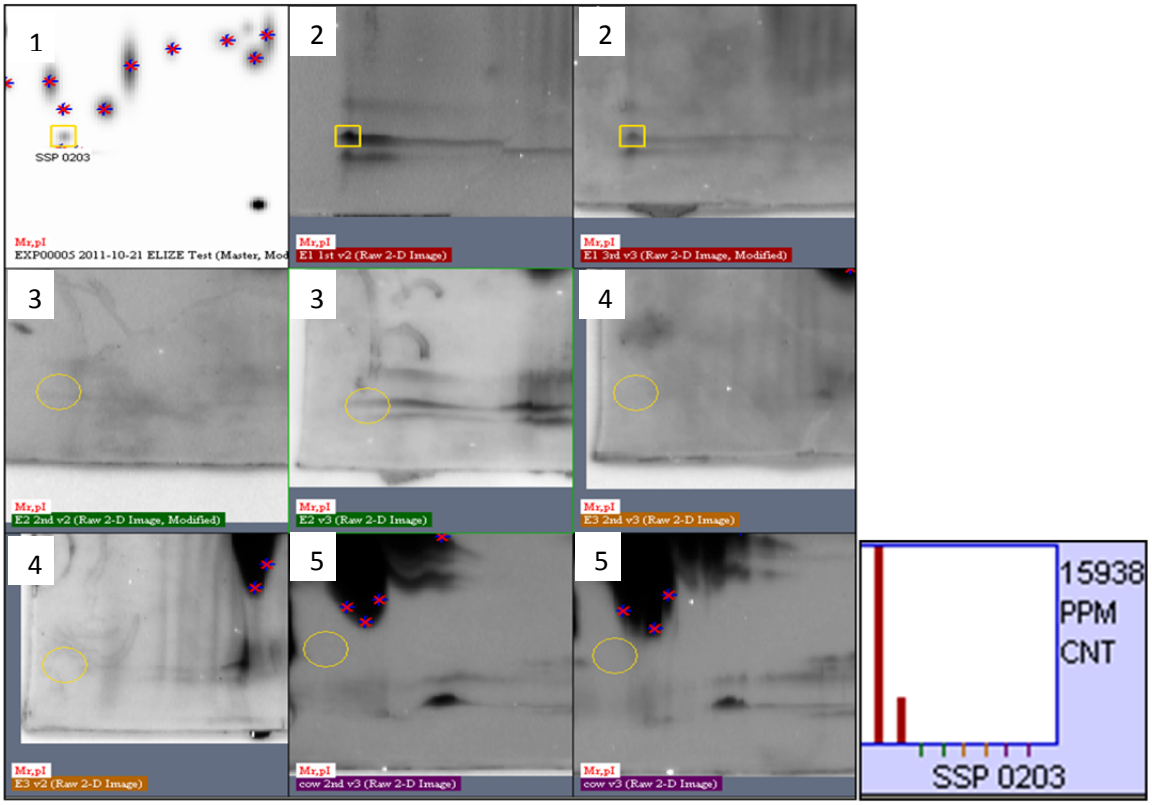

Figure 4.1. PD QUEST image of 2D PAGE gel analysis, three duplicates of African elephant milk proteins at early (4 days) (2), mid (12 months) (3) and late (18 months) lactation (4), as well as cow's milk proteins at mid lactation (5). The master gel is represented by gel (1). The histogram on the right shows a protein spot that is only observed on the duplicate gels at early lactation (red) but no longer present on the duplicate gels at mid (green) and late (orange) lactation. The same protein spot is also absent on cow's duplicate gels (purple).

PD Quest can show the relative quantities in terms of intensities which in this case, is interpreted as the protein concentration of the individual spots. A number of differences in terms of concentration were also observed showing presence of low and high abundance proteins. Additionally, Figure 4.2 shows the presence of a certain spot SSP 2602 only at mid and late lactation but absent at early lactation. This gives an indication of the difference in proteins that are coded for, and required at, certain stages during lactation. However, most of the spots on the gels shown including spots 0203 and 2602 are still unidentified proteins at this stage.

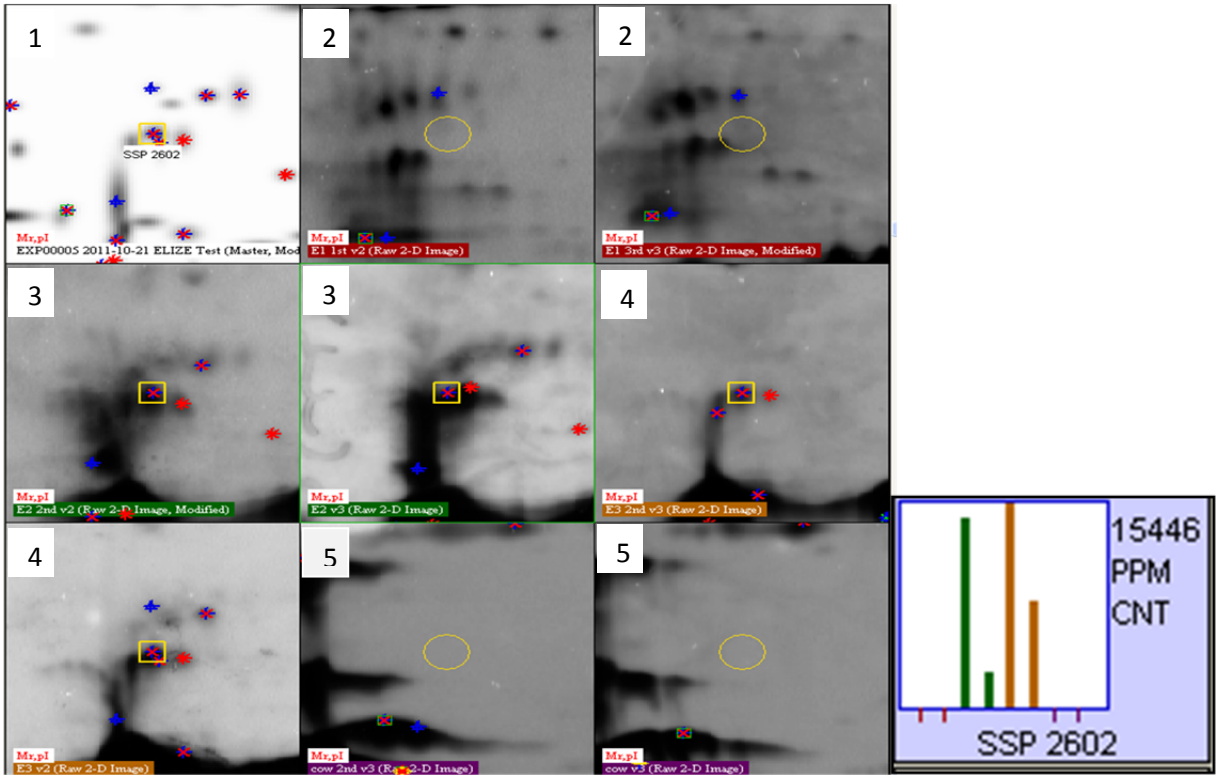

Figure 4.2. PD QUEST image of 2D PAGE gel analysis, the yellow squares shows presence of protein spot labelled SSP 2602 and its position on the gel, oval shapes shows absence of the spot. Three duplicates of African elephant milk proteins at early (4 days) (2), mid (12 months) (3) and late (18 months) (4) lactation stages; as well as cow's milk proteins at mid lactation (5) are shown. The master gel is represented by gel (1). The histogram on the right shows a protein spot that is only present on the duplicate gels at mid (green) and late (orange) lactation but no longer present on the duplicate gels at early (red) lactation. The same protein spot is also absent on cow's duplicate gels (purple).

The next step involved calculating the percentage ratios of protein spots on 2D PAGE gels that were successfully identified. For this analysis, the 2D PAGE in Figure 3.4a was used, because protein spots on this gel were successfully identified unlike the protein spots in Figures 4.1 and 4.2. The specific protein spots whose intensities were measured are shown in figure 4.3. All the visible spots on the gel were assigned a specific SSP number by the program. The larger spots with more peaks were assigned more than one SSP number per spot.

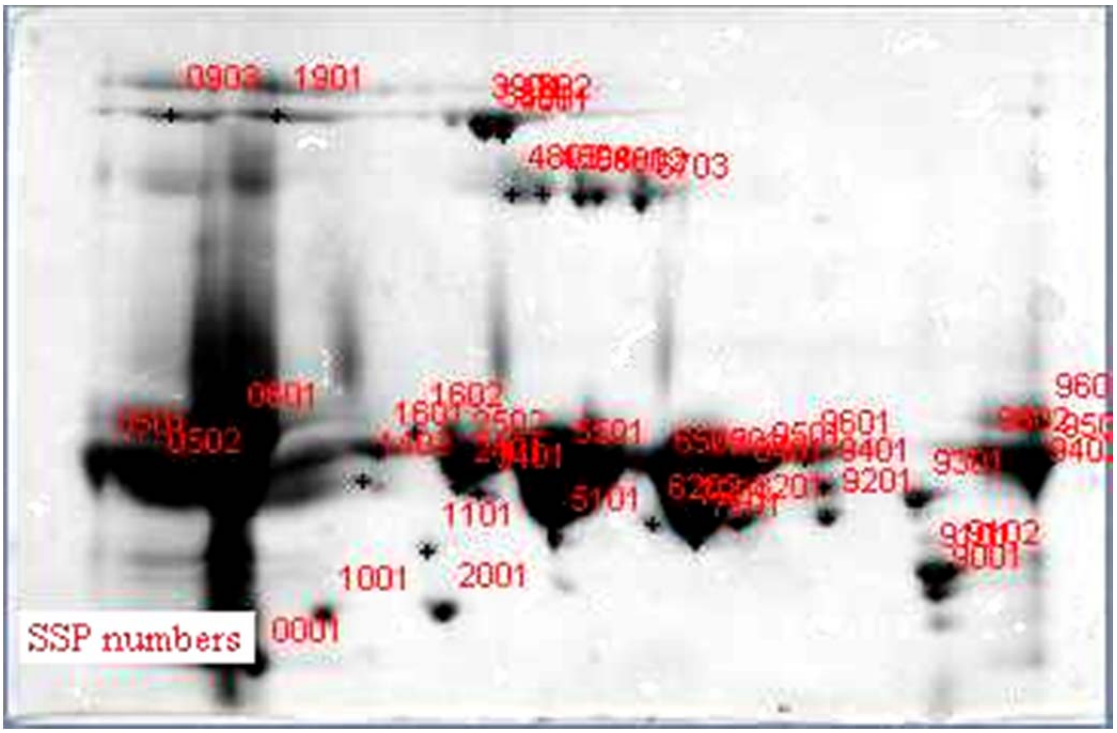

Figure 4.3. PD QUEST image of 2D PAGE gel analysis, the image shows the SSP numbers of the specific spots whose intensity was measured.

Table 4.1 represents the intensities of all the spots on the gel (Figure 4.3) at mid lactation. The sum of the intensities gives the total amount of protein in the gel. Protein spots identified as κ-, β-casein and α-LA were expressed as percentages of the total amount of proteins, as well as ratios of each other.

Table 4.1. The table gives a summary of the intensity values of all the protein spots on the 2D PAGE gel and the spot numbers. The identified proteins of interest are shown in bold. Intensity values refer to the concentration of proteins in a particular spot on a 2D PAGE gel.

SSP number	Intensity value
1 **(α-lactalbumin)**	185705.8
501	70373.7
502	752.8
601	194.6
903	242264
1001	195764.8
1101	89652.2
1402	51918.5
1601	271752.5
1602	64727.7
1901	238943.1
2001	311621.3
2401 **(β-casein)**	119345.6
2502 **(β-casein)**	11638.6
3401 **(β-casein)**	105977
3801	91930.2
3901	156319.7
4801	24725.9
4802	375150
4803	83613.1
4804	127059.4
5101 **(β-casein)**	300235.3
5501 **(β-casein)**	341.6
5801	132546.1
5802	110444.8
6202 **(β-casein)**	73446.5
6502 **(β-casein)**	199583.1
6703	153382.7
7101 **(β-casein)**	102177.5
7201 **(β-casein)**	201119.8
7501 **(β-casein)**	188.2
8201	127289.6
8401 **(β-casein)**	128501.6
9001	202668
9101 **(κ-casein)**	205347.1
9102	255330.4
9201	179375.7
9301	310797.1
9401	118469.5
9402	275739.1
9501 **(β-casein)**	220905.2
9503	301803.9
9601 **(β-casein)**	267333.3
9602	213064.8
9603	190232.8

4.4. Discussion

The analysis confirmed that the milk proteome is not constant throughout the entire lactation period. This was observed in both Figures 4.1 and 4.2 whereby in Figure 4.1, a protein spot of yet unknown identity was only observed during early lactation. The same observation is made in Figure 4.2 where a protein spot of yet unknown identity is present in mid and late lactation but absent in early lactation. These results also match the findings of (Osthoff *et al*., 2007), where the total protein content, casein and whey protein content of African elephant milk at different stages of lactation was compared. African elephant milk protein content increases steadily from early to mid lactation, before assuming a constant level into and during late lactation. The differences in milk protein composition throughout lactation are directly correlated to the neonatal demands which vary due to growth demands of the neonate (Nicholas, 1988; Nicholas *et al*., 1997). This may result in some proteins being expressed at only certain stages of the lactation period. Additionally, proteins in milk will change over time due to initiation of lactation where the first milk to be produced is colostrums, the proteome may also change due to expression of immunoglobulins that protect the mammary gland from infection during lactation and also during aptosis of mammary gland cells as a result of weaning (Roncada *et al*., 2013).

Table 4.1 gives a summary of the spots intensities which directly relates to the concentration of proteins in that particular spot. Some proteins such as β-casein were identified from more than one spot, possibly isomers that resulted from different levels of phosphorylation of serine residues during posttranslational modification. To account for this, the total intensity of such spots was calculated by adding together their individual intensities. The total intensity for all the spots on the gel, at mid lactation (12 months), was calculated as 6899521.4. The percentage values of κ- , β-casein and α-LA relative to the total volume or intensities of all the spots on the 2D gel were 2.97 %, 25.09 % and 2.69 % respectively. The ratio of African elephant milk κ-casein to β-casein is therefore approximately 1: 8.5, which is in the same order as reported for camel and rat milk (Martin *et al*., 2003). However, the method used to calculate the intensities has limitations, especially for large dark spots and hence the ratio of African elephant milk κ-casein to β-casein could be much larger than 1: 8.5.

This will then require loading lower concentrations of proteins on 2D PAGE gels, but also taking care not to exclude low abundance proteins, which may not be visible on the gels when low protein concentrations are loaded. The relative abundance of African elephant β-casein compared to κ-casein, and the distribution of hydrophobic and hydrophilic properties may be an indication of the role they may assume in the formation of casein micelles. It is possible that African elephant β-casein may stabilize the casein micelle in addition to provision of means for chain elongation and micelle growth (Horne, 1998).

The aim of this chapter was to evaluate the dynamics of elephant milk proteins over lactation time. It appears that some proteins are only expressed at specific stages of lactation as evident from the gels. In this study, the major casein and whey proteins were selected as the proteins of interest, the quantity experiment showed that β-casein dominates milk proteins, accounting for over a quarter of the total volume of proteins, at mid lactation. Having evaluated the dynamics aspect of African elephant milk proteins over lactation, the following chapter aims to evaluate the structural aspect of the lactose synthase complex (α-LA and β-1,4GT1 complex) in elephant milk. The lactose synthase system is involved in the biosynthesis of lactose and indirectly in oligosaccharide synthesis. The structure of α-LA and β-1,4GT1 may be key in explaining why there are high levels of both oligosaccharides and lactose in elephant milk. Homology modeling will be used to predict the structure of African elephant α-LA which has not been solved by X-ray crystallography as yet.

4.5. References

Horne, D. S. (1998). Casein interactions: casting light on the black boxes, the structure in dairy products. *Int Dairy J* **8**, 171-177.

Li, F. & Seillier-Moseiwitsch, F. (2001). Differential analysis of 2D gel images. *Method Enzym* **487**, 595-609.

Martin, P., Ferranti, P., Leroux, C., & Addeo, F. (2003). Non-bovine caseins: Quantitative variability and molecular diversity. In *Advanced Dairy Chemistry,* pp. 387-419. Edited by P. F. Fox & P. L. H. McSweeney. New York: Plenum Publishers.

Murakami, K., Lagarde, M. & Yuki, Y. (1998). Identification of minor proteins of human colostrums and mature milk by two dimensional electrophoresis. *Electrophoresis* **19**, 2521-2527.

Nicholas, K. R. (1988). Control of milk protein synthesis in the tammar wallaby: A model system to study prolactin-dependent development. In *The Developing Marsupial: Models for Biomedical Research,* pp. 68-85. Edited by C. H. Tyndale-Biscoe & P. A. Janssens. Heidelberg: Springer.

Nicholas, K., Simpson, K., Wilson, M., Trott, J. & Shaw, D. (1997). The tammar wallaby: A model to study putative autocrine-induced changes in milk composition. *J Mammary Gland Biol Neoplasia* **2**, 299-310.

Oftedal, O. T & Iverson, S. J. (1995). Comparative analysis of non-human milks: Phylogenetic variation in the gross composition of milks. In *Hand book of Milk Composition,* pp. 749-788. Edited by R. G. Jensen. New York: Academic Press.

Osthoff, G., de Wit, M., Hugo, A. & Kamara, B. I. (2007). Milk composition of three free-ranging African elephant (*Loxodonta africana africana*) cows during mid lactation. *Comp Biochem Physiol B* **148**, 1-5.

Rabilloud, T., Chevallet, M., Luche, S. & Lelong, C. (2010). Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J Proteomics* **73**, 2064-2077.

Rabilloud, T. & Lelong, C. (2011). Two-dimensional gel electrophoresis in proteomics: A tutorial. *J Proteomics* **74**, 1829-1841.

Sharp, J. A., Lefèvre, C. & Nicholas, K. R (2007). Molecular evolution of monotreme and marsupial whey acidic protein genes. *Evol Dev* **9**, 378-392.

Roncada, P., Stipetic, L. H., Bonizzi, L., Burchmore, R. J. S. & Kennedy, M. W. (2013). Proteomics as a tool to explore human milk in health and disease. *J Proteomics* **88**, 47-57.

Chapter 5

Structure modeling of African elephant alpha lactalbumin and beta-1,4-galactosyltransferase.

5.1. Introduction

α-LA is one of the major and extensively characterised proteins. The relative abundance and easy purification has led to α-LA being targeted as a model for investigating the chemical and biophysical properties of proteins (Brew, 2003). α-LA is present in milk of most mammals and plays a key role in the synthesis of lactose as part of the lactose synthase complex. The lactose synthase complex is composed of α-LA as the regulatory component and beta-1,4-galactosyltransferase 1 (β-1,4GT1) as the catalytic component (Ramakrishnan *et al*., 2001). α-LA, a protein that is structurally homologous to lysozyme, binds to β-1,4GT1 and subsequently alters its sugar acceptor specificity from *N*-acetylglucoseamine to glucose, consequently the disaccharide lactose is synthesized which in turn is a major carbohydrate in milk of most eutherian species. In the absence of α-LA, β-1,4GT1 transfers galactose from UDP-galactose to *N*-acetylglucoseamine residues forming poly-*N*-acetyllactoseamine, a primary building block of oligosaccharides.

The levels of lactose in milk are dependent on the amounts of α-LA in the mammary gland (Stacy *et al.*, 1995; Ramakrishnan & Qasba, 2007). High amounts of α-LA are observed in most eutherians whereas low amounts are observed in monotremes and marsupials. Furthermore, milks of monotremes, marsupials and a few eutherian species contain oligosaccharides as the major carbohydrate rather than lactose. The ratio of free lactose to oligosaccharides in milk is not constant among mammalian species. The milk of monotremes and marsupials contains minute quantities of lactose compared to large amounts of oligosaccharides (Urashima *et al*., 2009). Glucosyltransferases catalyzes the biosynthesis of oligosaccharides. The African

elephant is unique in that it contains relatively high levels of both oligosaccharides and lactose in milk, 1.5-2.7 % and 0.7-5.3 % respectively; however, the amounts are reduced during the course of lactation (Osthoff *et al*., 2005).

In order to shed more light onto this unique observation of high levels of both lactose and oligosaccharides in African elephant milk, a closer look at the African elephant lactose synthase complex was necessary. The close relationship between protein structure and function is of paramount importance with regards to biologically active compounds (Deschamps, 2009; Elbegdorj *et al*., 2013). Elucidation of the individual protein structures and comparison with known crystal structures could provide answers to the unique observation of the levels of lactose and oligosaccharides in African elephant milk.

In chapter 3, the sequence of African elephant α-LA was elucidated. In this current chapter, the aim is to predict the 3D structure of African elephant α-LA, and comparison thereof with α-LA crystal structures of other mammalian species that have been solved. The same will be done for African elephant β-1,4GT1, but in this case the gene sequence will be used. Consequently the structure of African elephant lactose synthase complex will be modelled. Ultimately, this may answer the question why elephant milk contains high levels of both lactose and oligosaccharides.

5.2. Materials and Methods

5.2.1. Sample preparation

The collection of African elephant milk, precipitation and fractionation of proteins was described in section 2.2.1.

5.2.2. Determination of protein concentration

The determination of protein concentration of our samples, using the Bradford assay, was described in section 2.2.2.

5.2.3. Two-Dimensional polyacrylamide gel electrophoresis (2D PAGE)

Separation of protein samples by electrophoresis was described in section 3.2.1.

5.2.4. Mass spectrometric analysis

Orbitrap MS analysis and sequence determination was described in section 3.2.4.

5.2.5. Homology modeling: Alpha lactalbumin

In this homology modeling experiment, YASARA-structure version 12.7.6 was used. The three-dimensional structure of the target sequence of African elephant α-LA (Figure 5.1) which contains 123 amino acid residues in one molecule was predicted by YASARA's homology modeling experiment. The sequence was obtained from the National Centre for Biotechnology Information protein database after a BLAST search of peptides from orbitrap mass spectrometry (Accession: XP_003405795.1). The default parameters were used for structure modeling of α-LA sequence.

```
>α-LA
KQFTKCELSQVLKDIDGYAGITLPEFTCTIFHISGYDTQTIVNNNGSTEYGLFQISNKYWCRDHQIPQS
RNICDISCDKFLDDDLTDDMMCAKKILDSKGIDYWLAHKPLCSEKLEQWLCEKL
```

Figure 5.1. African elephant α-LA sequence, the sequence was used for 3D structure prediction of African elephant α-LA in the homology modeling experiment using YASARA. The sequence was obtained from orbitrap MS analysis.

A total of 253 hits for the homology modeling templates were identified by searching through the PDB using the target sequence of African elephant α-LA provided. The total score was used in selection of the template. This ensures that good template structures are used even if the alignment score is low (Hooft *et al*., 1996). A secondary structure prediction was done by running a PSI-BLAST to create a target sequence profile and feeding it to the PSI-Pred secondary structure prediction algorithm (Jones, 1999). A secondary structure prediction is important for loop modeling and alignment correction. A target sequence profile was then created from a multiple sequence alignment built from related UniRef90 sequences. A model was then built for all the sequences in the multiple sequence alignment.

For all the alignments which were certain, a single model was built and for those alignments which were ambiguous, a number of alternative models were built. No loop modeling was done, since the sequence was trivial. After the side-chains had been built, optimized and fine-tuned, all newly modeled parts were subjected to a combined steepest descent and simulated annealing minimization. Then a full unrestrained simulated annealing minimization was run for the entire model. The overall quality Z-score improved to -0.139 during the minimization and this fully refined model has been accepted as the final one for this template and alignment. Finally, YASARA tried to combine the best parts of the 5 models to obtain a hybrid model, hoping to increase the accuracy beyond each of the contributors. Since the score of the hybrid model was lower than the best scoring initial model (1HML-A with score -0.139), the hybrid model was discarded.

5.2.6. Homology modeling: Beta galactosyltransferase

The three-dimensional structure of the target sequence (Figure 5.2) of African elephant β-1,4GT1 which contains 396 amino acid residues in one molecule was predicted by YASARA's homology modeling experiment. The sequence was obtained from the National Centre for Biotechnology Information protein database (Accession: XP_003407363.1).

```
>β-GT1
MRFREPLLGGSAAMPGASLQRACRLLVAVCALHLGVTLVYYLSGRDLSRLPRLIGVPSPL
QGGSNGTATLGQPSGELRPGGASPPAAFTLSSEPRPSRSPSPSAGAGAASNLTTAPVPLT
TALSLPACPEESPLLVGPMVIEFNVAVDLDRVAKKNPEVKVGGRYAPKNCVSPHKVAIII
PFRNRQEHLKYWLYYLHPILQRQQLDYGVYVINQDGDSMFNRAKLLNIGFQEALKDYDYN
CFVFSDVDLIPMDDHNTYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLTING
FPNNYWGWGGEDDDIFNRLVFKGMSISRPNAAVGKCRMIRHSRDKKNEPNPQRFDRIAHT
KETMRLDGLNTLTYKVLDTQRYPLYTKITVDVGKPS
```

Figure 5.2. Gene sequence of African elephant β-1,4GT1, the sequence was used for 3D structure prediction of African elephant β-1,4GT1 in the homology modeling experiment using YASARA. The gene sequence was obtained from (http://www.ncbi.nlm.nih.gov).

The homology modeling parameters were the same as those for α-LA 3D structure prediction, 15 possible templates were identified for structure prediction of the β-1,4GT1. To aid alignment correction and loop modeling, a secondary structure prediction for the target sequence was obtained by running PSI-BLAST to create a

target sequence profile and feeding it to the PSI-Pred secondary structure prediction algorithm (Jones, 1999). A target sequence profile was created to help align target and templates, models were then built for the templates. A total of nine models were built. Only one loop was modeled and after side-chains had been built, optimized and fine-tuned, all newly modeled parts were subjected to a combined steepest descent and simulated annealing minimization. After a full unrestrained simulated annealing minimization was run for the entire model, the overall quality Z-score improved to 0.149 during the minimization. This fully refined model has been accepted as the final one and the overall best of the nine models.

5.3. Results

Five initial homology models were set up for African elephant α-LA. Among the five models, one was accepted as the best model. The models were generated based on the template of human alpha lactalbumin-Zn^{2+} complex X-ray crystal structure at a resolution of 1.7 Å (Ren *et al*., 1993).The best model had the following parameters, in the alignment: 123 out of 123 target residues were aligned to template residues and the sequence identity and similarity (BLOSUM62 score is greater than zero) among the residues was 83.7 % and 89.4 % respectively. The alignment is shown in Figure 5.3. Figure 5.4 shows the final 3D structure model of the African elephant α-LA. This structure was then superimposed on goat, pig, human and bovine α-LA X-ray structures for comparison as shown in Figure 5.5.

Figure 5.6 shows a model of the 3D structure of African elephant β-1,4-GT1as built by YASARA. Initially nine models were built for β-1,4-GT1 but only one was selected as the best model, it had the following parameters, 272 of 396 target residues (68.7 %) were aligned to template residues. Among these aligned residues, the sequence identity is 86.0 % and the sequence similarity is 94.1 % ('similar' means that the BLOSUM62 score is > 0). African elephant β-1,4-GT1 was superimposed on cow β-1,4-GT1 (Figure 5.7) to show the structure similarity of the two proteins.

```
                                                                  20                  40
Human       MRFFVPLFLVGILFPAILAKQFTKCELSQLLKDIDGYGGIALPELICTMFHTSGYDTQAI
Pig         MMSFVSLLVVGILFPAIQAKQFTKCELSQVLKDMDGYGDITLPEWICTIFHISGYDTKTI
Goat        MMSFVSLLLVGILFHATQAEQLTKCEVFQKLKDLKDYGGVSLPEWVCTAFHTSGYDTQAI
Cattle      MMSFVSLLLVGILFHATQAEQLTKCEVFRELKDLKGYGGVSLPEWVCTAFHTSGYDTQAI
Elephant    MMSFVPLLLVGILFPAIQAKQFTKCELSQVLKDIDGYAGITLPEFTCTIFHISGYDTQTI
            *  ** *::***** *  *:*:****: : ***:. *. ::***  ** ** *****::*

                                  60                  80                  100
Human       VENNESTEYGLFQISNKLWCKSSQVPQSRNICDISCDKFLDDDITDDIMCAKKILDIKGI
Pig         VHDNGSTEYGLFQINNKLWCRDNQI-QSKNICGISCDKFLDDDLTDDMMCAKKILDNEGI
Goat        VQNNDSTEYGLFQINNKIWCKDDQNPHSRNICNISCDKFLDDDLTDDIVCAKKILDKVGI
Cattle      VQNNDSTEYGLFQINNKIWCKDDQNPHSSNICNISCDKFLDDDLTDDIMCVKKILDKVGI
Elephant    VNNNGSTEYGLFQISNKYWCRDHQIPQSRNICDISCDKFLDDDLTDDMMCAKKILDSKGI
            *.:* *********.** **:. *  :* *** **********:***::*.*****  **

                                  120
Human       DYWLAHKALCTEKLEQWLCEKL
Pig         DYWLAHKALCSEKLDQWLCEKM
Goat        NYWLAHKALCSEKLDQWLCEKL
Cattle      NYWLAHKALCSEKLDQWLCEKL
Elephant    DYWLAHKPLCSEKLEQWLCEKL
            :****** **:***:******:
```

Figure 5.3. Multiple sequence alignment of α-LA sequences from human, pig, goat, cattle and African elephant. The amino acids highlighted in rectangular borders are involved in interaction with β-1,4-GT1 and binding of glucose, these sequences are well conserved amongst the five sequences aligned. The alignment was produced using clustal omega. Alanine106, Histidine107, Leucine110, Glutamine117 and Trp118 of α-LA are important for binding to β-1,4GT1, whereas, Phenylalanine31 and Histidine32 are important for Glucose binding.

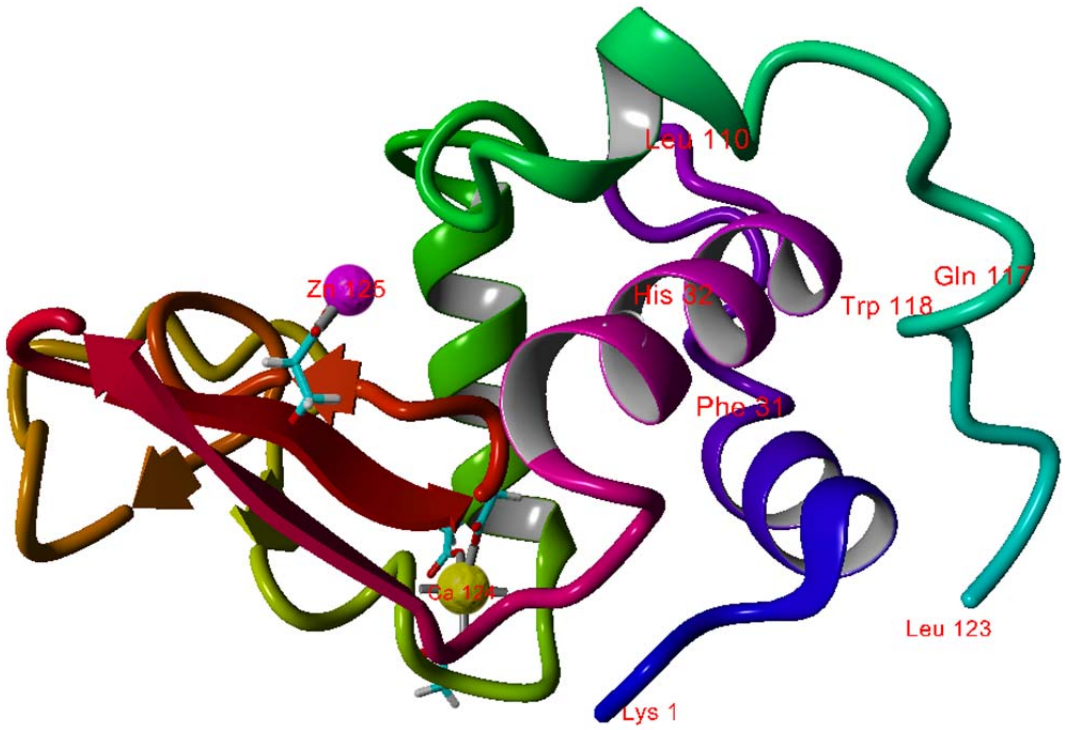

Figure 5.4. A ribbon diagram representation of the African elephant α-LA homology model, the first and last amino acids in the structure are labelled, as well those important for the catalytic mechanism of lactose biosynthesis. The location of the calcium and zinc ions is indicated.

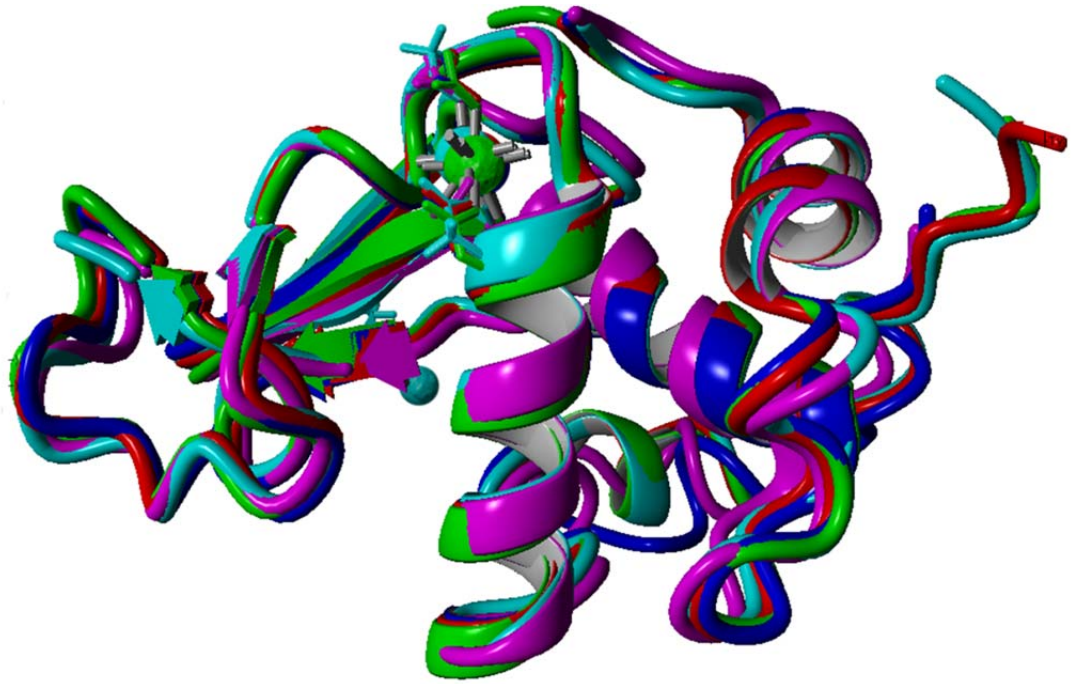

Figure 5.5. Homology model of African elephant α-LA (cyan) superimposed on human (magenta), bovine (green), pig (red), goat (blue) α-LA crystal structures in ribbon representation.

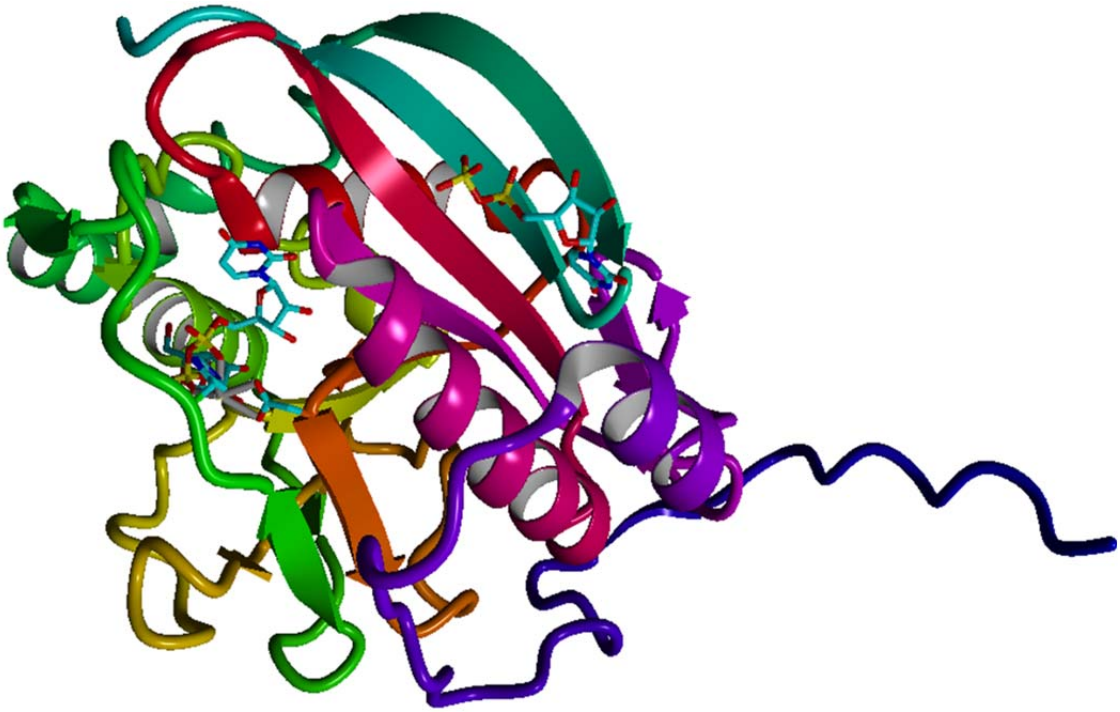

Figure 5.6. A ribbon structure representation of African elephant β-1,4-GT1 structure model.

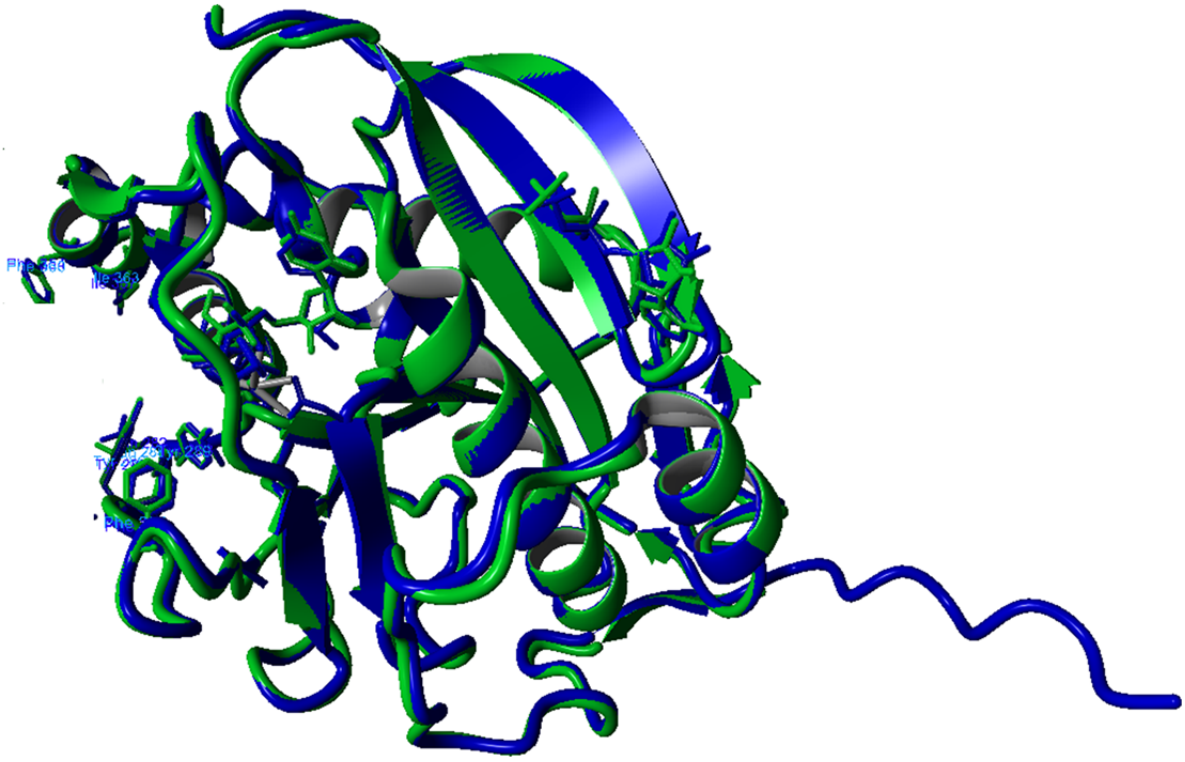

Figure 5.7. Superposition of modelled African elephant β-1,4-GT1(blue) and bovine β-1,4-GT1 X-ray structure (green). Amino acid side chains important for the catalytic mechanism of lactose biosynthesis are labelled.

The lactose synthase is composed of two enzymes working together, where α-LA modulates the catalytic action of the β-1,4-GT1. The interaction of these two proteins is key in lactose biosynthesis (Ramakrishnan *et al*., 2001). During the homology modeling studies, a model of the interaction of the two proteins in African elephant was modeled; Figures 5.8 and 5.9 show the interaction of the two enzymes after energy minimization on YASARA. Furthermore a structural alignment was done between the modeled African elephant lactose synthase complex with the X-ray crystallography structures of the cow lactose synthase complex. The superimposition was done in order to see how the two complexes compare to each other structurally (Figure 5.10).

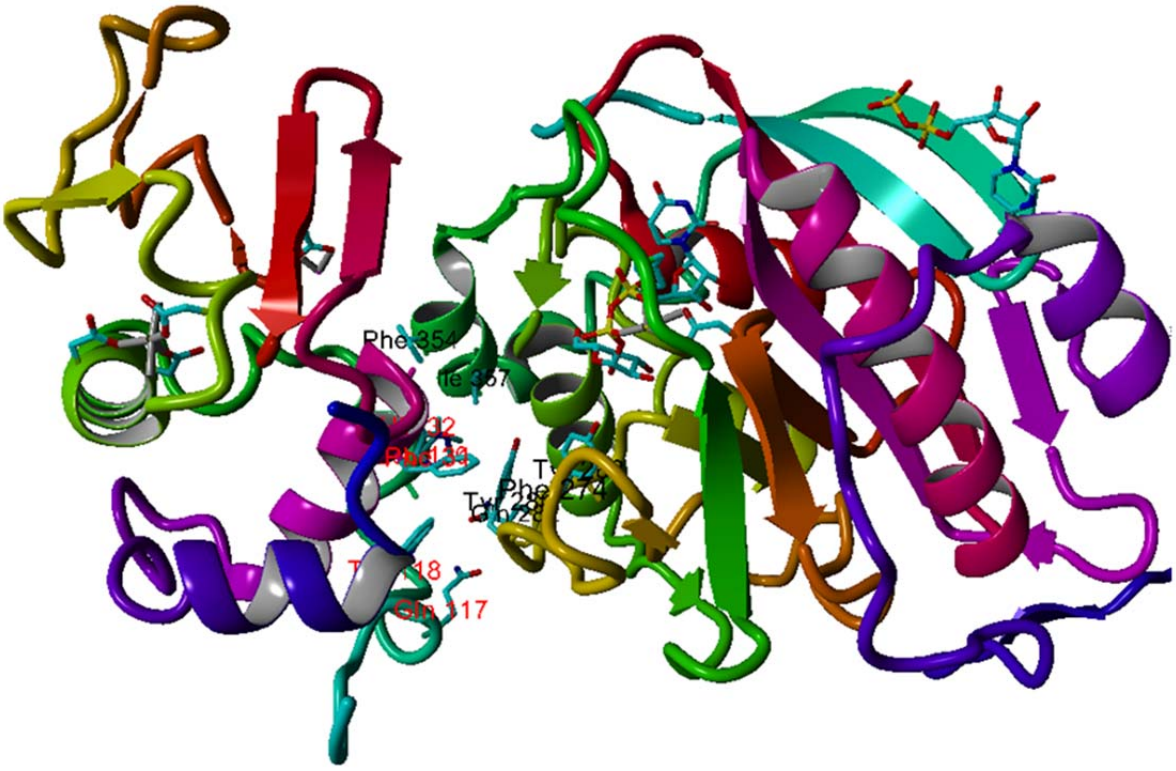

Figure 5.8. African elephant lactose synthase complex model. Amino acid side chains involved in the interaction of the two proteins during lactose biosynthesis are labelled.

Figure 5.9. A close up representation of the African elephant lactose synthase complex showing residues that are involved in the interaction between α-LA(left) and β-1,4-GT1 (right).

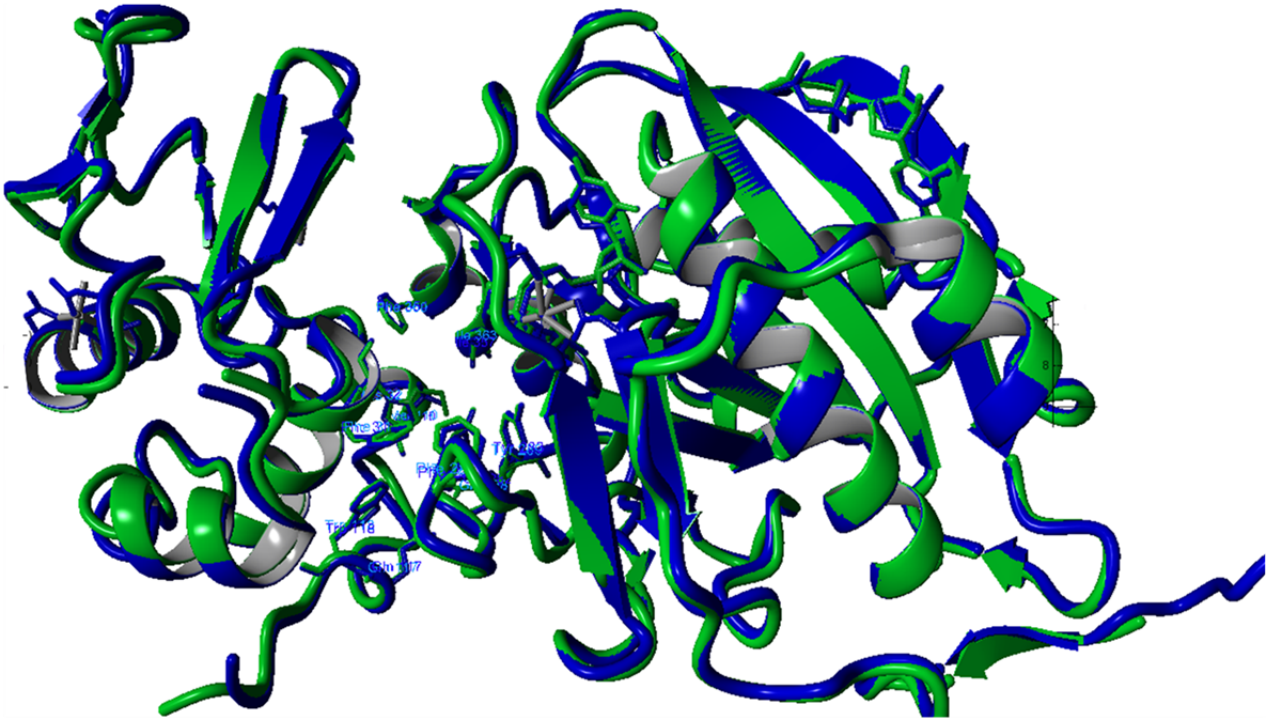

Figure 5.10. Homology model of African elephant lactose synthase complex (blue) aligned with the template structure of bovine lactose synthase complex (green).

5.4. Discussion

The African elephant α-LA structure shown in Figure 4.5 shows a very close structural similarity to α-LA X-ray crystallography structures reported for other species such as cow and pig. The structure is divided into a large α-domain and a much smaller β-domain with the two domains separated by a deep cleft. Figure 5.5 compares five α-LA structures including the African elephant α-LA. The structural alignment between the query (African elephant model) and bovine crystal structure had an RSMD of 0.818 Å with a sequence identity of 71.9 %, (African elephant model) and goat crystal structure has an RSMD of 0.991 Å with a sequence identity of 73.04 %, (African elephant model) and pig crystal structure had an RSMD of 0.836 Å with a sequence identity of 66.12 %, (African elephant model) and human crystal structure had an RSMD of 1.122 Å with a sequence identity of 82.61 %. The structure alignments show that there are no significant differences between these five structures, and it may be assumed that the structure of African elephant α-LA as a modulator should not have an effect on the catalytic mechanism of lactose biosynthesis. It is therefore unlikely that the structure of African elephant α-LA could

be the key factor in explaining why African elephant milk contains relatively high levels of both lactose and oligosaccharides.

While, the overall amino acid homology among the five α-LA sequences was approximately 70 %, the important amino acid residues that are involved in lactose biosynthesis, those involved in binding to β-1,4-GT1 (Ala106, His107, Leu110, Gln117 and Trp118) and interaction with glucose (Phe31 and His32), are also highly conserved, as is evident from the multiple sequence alignment in Figure 5.3.

The 3D structure of African elephant β-1,4-GT1, based on homology modeling is shown in Figure 6.5. As was observed for α-LA, there were very few differences when compared to the x-ray crystallography structure of the cow β-1,4-GT1 (Figure 5.7). It may therefore also be assumed that it would be highly unlikely that β-1,4-GT1 of African elephant has an influence on the high levels of oligosaccharides observed in African elephant milk.

Finally, the African elephant lactose synthase complex was modeled (Figures 5.8 and 5.9) by combination of the β-1,4-GT1 and α-LA. No differences in orientation or alignments of amino acid were noted between the modeled lactose synthase from African elephant and X-ray crystallography structure of bovine lactose synthase complex (Figure 5.10). It may therefore be concluded that the differences in amino acid composition in α-LA, and also in β-1,4-GT1, of African elephant do not affect the 3D structure of the proteins, and also not their interaction and catalytic sites. This in turn may implicate that it is not the small structural differences of these proteins that are responsible for a saccharide synthesis resulting in high levels of both oligosaccharides and lactose in African elephant milk, compared to other mammals. The reason for this unique saccharide composition should be searched elsewhere in the saccharide synthesis pathway, perhaps the other galactosyltransferases can be a target of such investigation.

There are at least two other human mammary β-galactosyltransferases which catalyze the synthesis of Gal (β1-3) GlcNAc-R and Gal(β1-4)GlcNAc-R structures (Urashima *et al*., 2009). The catalytic function of these β-galactosyltransferases is independent of α-LA. Additional glycosyltransferases are also found in lactating mammary glands of nonhuman species. These synthesize other branched saccharide types. The α-galactosyltransferase that synthesizes α-3′-galactosyllactose [Gal(α1-4)Gal(β1-4)Glc, isoglobotriose], a trisaccharide that is present in bovine, ovine, and caprine colostrums and the milk or colostrums of several other species but not in human milk or colostrums. The synthesis of α-4′-galactosyllactose [Gal(α1-4)Gal(β1-4)Glc, globotriose], which has been found in the colostrums of the bottle-nosed dolphin, is presumably catalyzed by a different α-galactosyltransferase. Lactating mammary glands of the tammar wallaby contain a very active α-galactosyltransferase that is involved in the synthesis of a series of β(1-3)-linked galactosyllactoses that are unique to the milk of marsupials. Derived from the different branch types of the oligosaccharides found in African elephant milk, it is suspected that eight different galactocyl transferases might be involved (Urashima *et al*., 2009).

Since isoglobotriose [Gal(αl-3)Gal(β1–4)Glc] is the most abundant of the oligosaccharides in Asian and African elephant milk (Uemura *et al*., 2006; Osthoff *et al*., 2008), it is most probably the β-galactosyltransferase that links galactose via an α l-3 bond to the galactosyl moiety of the lactose, that is in competition with α-LA in its affinity for lactose, which may result in disproportionate synthesis of oligosaccharides or lactose.

While the structural aspects of α-LA were considered in this chapter, the next chapter is devoted to structural aspects of caseins, specifically β- and κ-casein with regards to their role in casein micelle composition in African elephant milk.

5.5. References

Brew, K. (2003). α-Lactalbumin. In *Advanced Dairy Chemistry,* pp. 387-419. Edited by P. F. Fox & P. L. H. McSweeney. New York: Plenum Publishers.

Deschamps, J. R. (2009). X-ray crystallography of chemical compounds. *Life Sci* **86**, 585-589.

Elbegdorj, O., Westkaemper, R. B. & Zhang, Y. (2013). A homology modeling study toward the understanding of three-dimensional structure and putative pharmacological profile of G-protein coupled receptor GPR55. *J Mol Graph Model* **39**, 50-60.

Hooft, R. W. W., Vriend, G. & Abola, E. E. (1996). Errors in protein structures. *Nature* **381**, 272-272.

Jones, D. T. (1999). GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* **287**, 797-815.

Osthoff, G., De Waal, H. O., Hugo, A., De Wit, M. & Botes, P. (2005). Milk composition of a free-ranging African elephant (*Loxodonta africana*) cow during early lactation. *Comp Biochem Physiol A* **141**, 223-229.

Osthoff, G., Dickens, L., Urashima, T., Bonnet, S. L., Uemura, Y. & van der Westhuizen, J. H. (2008). Structural characterization of the oligosaccharides in the milk of an African elephant (*Loxodonta africana africana*). *Comp Biochem Physiol B* **150**, 74-84.

Ramakrishnan, B., Shah, P. S. & Qasba, P. K. (2001). α-Lactalbumin (LA) stimulates milk β-1,4-Galactosyltransferase I (β4Gal-T1) to transfer glucose from UDP-glucose to N-Acetylglucosamine. *J Biol Chem* **276**, 37665-37671.

Ren, J., Stuart, D. I. & Acharya, K. R. (1993). Alpha lactalbumin possesses a distinct zinc binding site. *J Biol Chem* **268**, 19292-19298.

Stacy, A., Schnieke, A., Kerr, M., Scott, A., McKee, C., Cottingham, I., Binas, B., Wilde, C. & Colman, A. (1995). Lactation is disrupted by α-lactalbumin deficiency and can be restored by human α-lactalbumin gene replacement in mice. *Proc Natl Acad Sci U S A* **92**, 2835-2839.

Uemura, Y., Asakuma, S., Yon, L., Saito, T., Fakuda, K., Arai, I. & Urashima, T. (2006). Structural determination of the oligosaccharides in the milk of an Asian elephant (*Elephas maximus*). *Comp Biochem Physiol A Mol Integr Physiol* **145**, 468-478.

Urashima, T., Kitaoka, M., Asakuma, S. & Messer, M. (2009). Milk oligosaccharides. In *Advanced Dairy Chemistry,* pp. 295-349. Edited by P. McSweeney & P. F. Fox. New York: Springer.

Chapter 6

Beta and Kappa caseins of African elephant milk and the casein micelle model

6.1. Introduction

Milk is secreted by mammalian species and serves as a complete and balanced food for the nutritional requirements of the neonate. It composes of nutrients such as lipids, proteins, carbohydrates and minerals (Fox & McSweeney, 1998). The composition of milk varies extensively between species, mostly due to specific nutritional demands of the neonate for growth and development (Kulski & Hartmann, 1981; Horne, 2008). Bovine milk contains approximately 3.5 % protein which provides necessary amino acids for muscular tissue development (Phadungath, 2005). Caseins in bovine milk possess functional as well as nutritional properties; they also carry mineral elements in the form of calcium, magnesium and phosphorous (Farrell *et al*., 2004). Milk of most mammalian species contain α_{s1}-, α_{s2}-, κ- and β-caseins, which, exist as colloidal aggregates of caseins with calcium and phosphate, known as the casein micelle. The high heat stability property of caseins enables high heat treatment of milk and milk products, and as such, the importance of caseins and casein micelles for the functional behavior of dairy products cannot be overemphasized (Phadungath, 2005). A great deal of research has gone into unravelling the structure of caseins and the casein micelle, however the biggest challenge remains that caseins cannot be crystallized for X-ray crystallography nor can they be accurately studied by Nuclear Magnetic Resonance spectroscopy, and as such, research could only come up with models to describe the casein micelle structure using bovine milk (Farrell *et al*., 2004). A variety of models have been put forward to describe the casein micelle structure, these models are based on experimental data which covers several micellar properties (Phadungath, 2005). The models proposed fall in mainly three categories, the coat-core, the subunit and the internal structure models.

There seem to be a variety of models to explain the casein micelle. Some models are widely accepted and favored over the others. The underlining factor which is consistent throughout these models is the important role of κ-caseins in chain termination and maintenance thus putting κ-caseins at the center of the casein micelle models. The amount of κ-caseins present is also equally important because of the role they are assigned to and this knowledge has been exploited by the dairy industry targeting κ-casein hydrolysis using the enzyme rennin. Bovine milk generally has high levels of κ-casein, but there are other mammalian species, such as horse, whose milk contains low amounts of κ-casein and high levels of β-casein (Horne, 2008). Such species have casein micelles, but the difference in the amounts of casein types challenges the existing micelle models. Further investigations that focus on β-caseins of such species, in this case African elephant milk β-caseins might provide answers. This may shed more light in the otherwise overlooked role that β-caseins may play in the formation of casein micelles.

In this current chapter, we unravel the casein micelle model, by analyzing the hydrophilic and hydrophobic properties of African elephant κ- and β-casein sequences, which were discussed in chapter 3. The hydropathy plots of African elephant κ- and β-casein will give an indication of how these proteins self-associate or interact with other proteins, during casein micelle formation and maintenance.

6.2. Materials and Methods

6.2.1. MS analysis and sequence determination

Orbitrap MS analysis and sequence determination was described in section 3.2.4.

6.2.2. Sequence alignment and Hydropathy plots

A multiple sequence alignment of β-caseins was done using the latest edition of clustal omega (Sievers *et al*., 2011), followed by hydrophobicity studies (Kyte & Doolittle, 1982). A window size of nine was used for plotting all the hydropathy plots, which meant that nine amino acids were averaged to obtain the hydropathy score at each position or data point on the hydropathy plots.

6.3. Results

A multiple sequence alignment of eight β-casein sequences, with the 15 amino acid long signal sequence included, was carried out as depicted in Figure 6.1 using clustal omega. The sequences had approximately 20 % amino acid identity between them. Additionally, hydropathy plots for the eight β-casein sequences were plotted. Hydropathy plots show parts of the sequences that are hydrophilic or hydrophobic and hence give a possible indication of their role and how they interact with each other or with other proteins (Kyte & Doolittle, 1982). Peaks above zero indicate hydrophobic regions of the protein and peaks below zero indicate hydrophilic regions. The Kyte and Doolittle plots were used determine the hydropathy of β-casein from cow, sheep, goat, African elephant, human, water buffalo, pig and mouse. The combined plots are depicted in the figures below. The ruminants (cow, sheep and goat) plots were combined in Figure 6.2 for comparison whereas Figures 6.3 and 6.4 depicts (cow, mouse, human and pig) and (cow and African elephant) plots combined respectively. The ruminant hydropathy plots show a similar trend in terms of the distribution of hydrophobic and hydrophilic amino acids. This aspect is also evident in the sequence alignment of β-casein shown in figure 6.1, where ruminant sequences have over 90 % amino acid homology and thus a similar pattern on the hydropathy plot would be expected. It was also interesting to the note the amino acid sequence homology between African elephant and human β-casein sequences since they both lack exon 3. The sequence homology was calculated to be over 63 %.

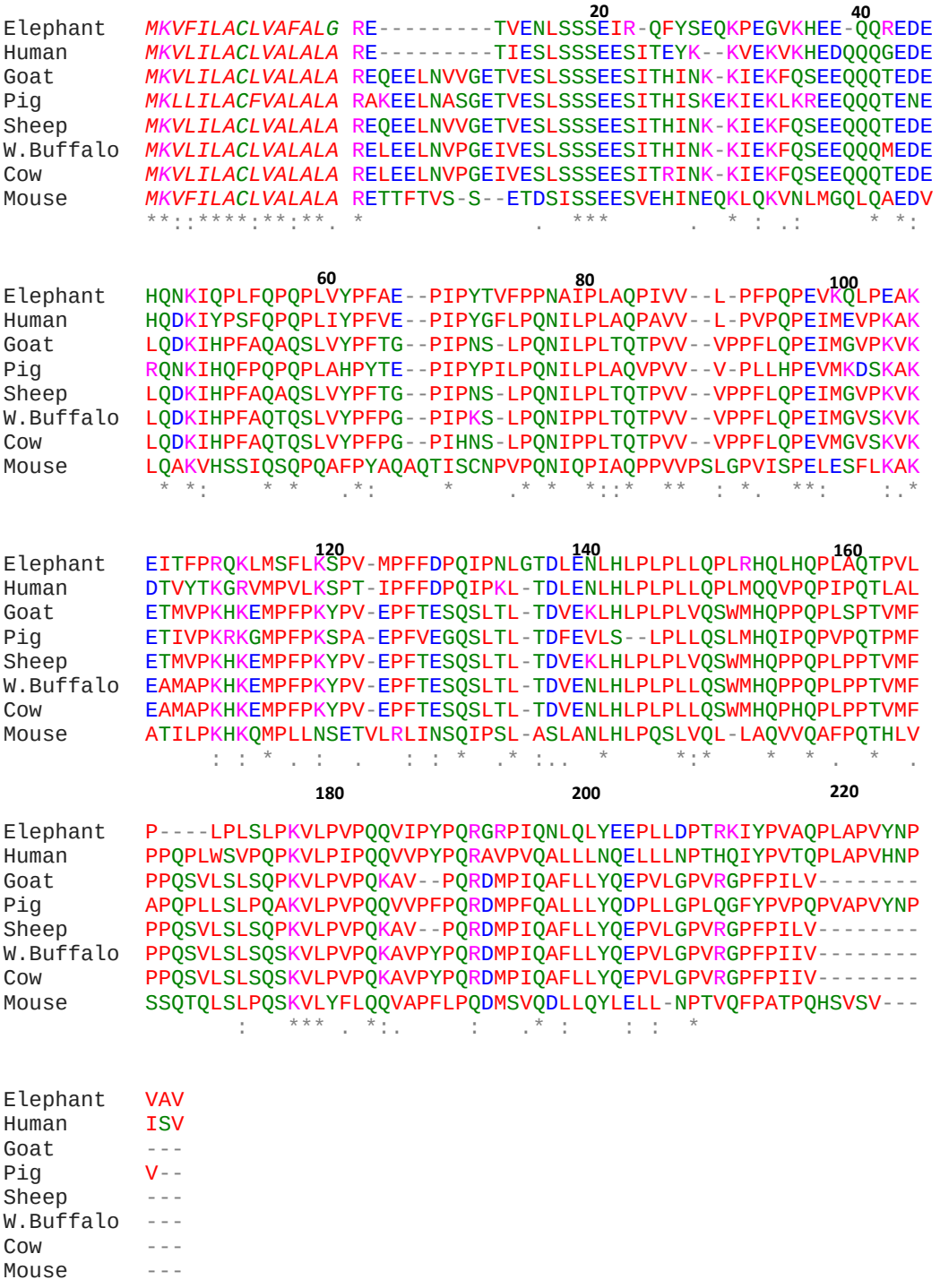

Figure 6.1. Multiple sequence alignment of β-casein from 8 mammalian species. Amino acid residues are color coded according to their different properties, red (small and hydrophilic); blue (acidic); magenta (basic); green (hydroxyl, sulfhydryl, amine and Glycine). The 15 amino acid long signal sequence is included and is shown in italics.

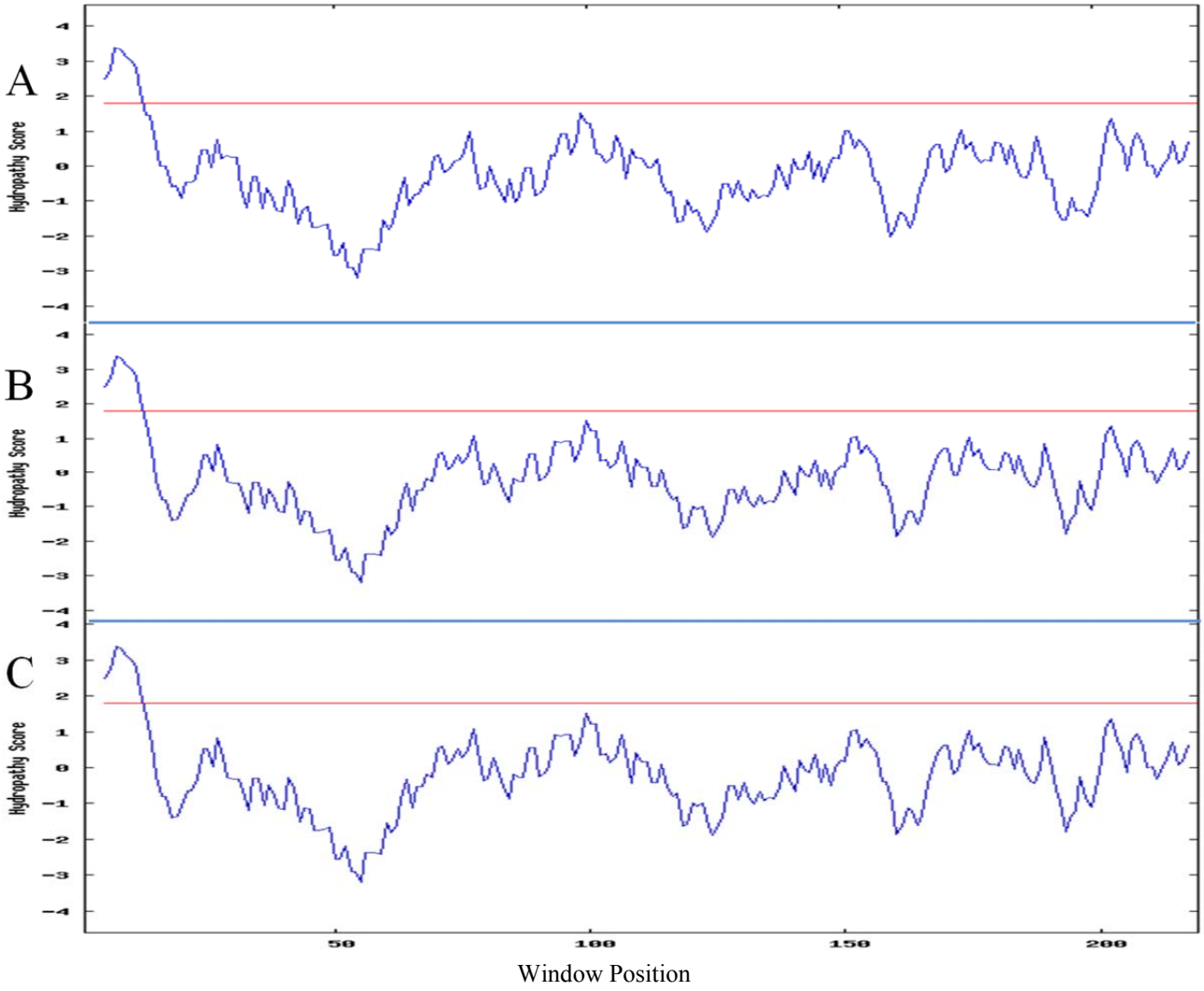

Figure 6.2. Kyte and Doolittle hydropathy plots for ruminant β-casein, (a) cow, (b) goat and (c) sheep. Points above zero indicate hydrophobic residues while points below zero represent hydrophilic regions.

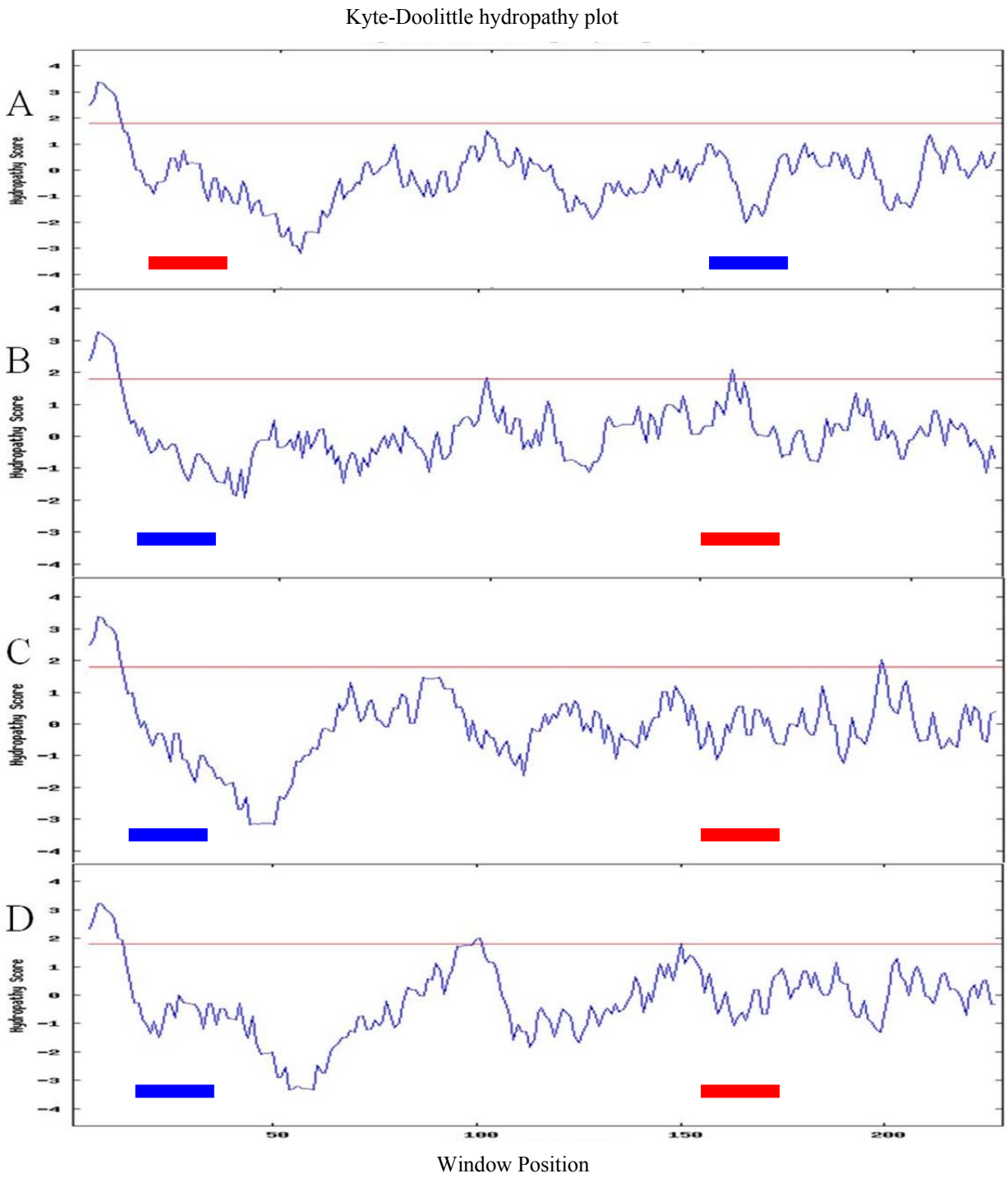

Figure 6.3. Kyte and Doolittle hydropathy plots for β-casein from cow (A), mouse (B), human (C) and pig (D). Points above zero indicate hydrophobic residues while points below zero represent hydrophilic regions. The bars show differences in the distribution of hydrophobic (red) and hydrophilic (blue) regions among the four proteins.

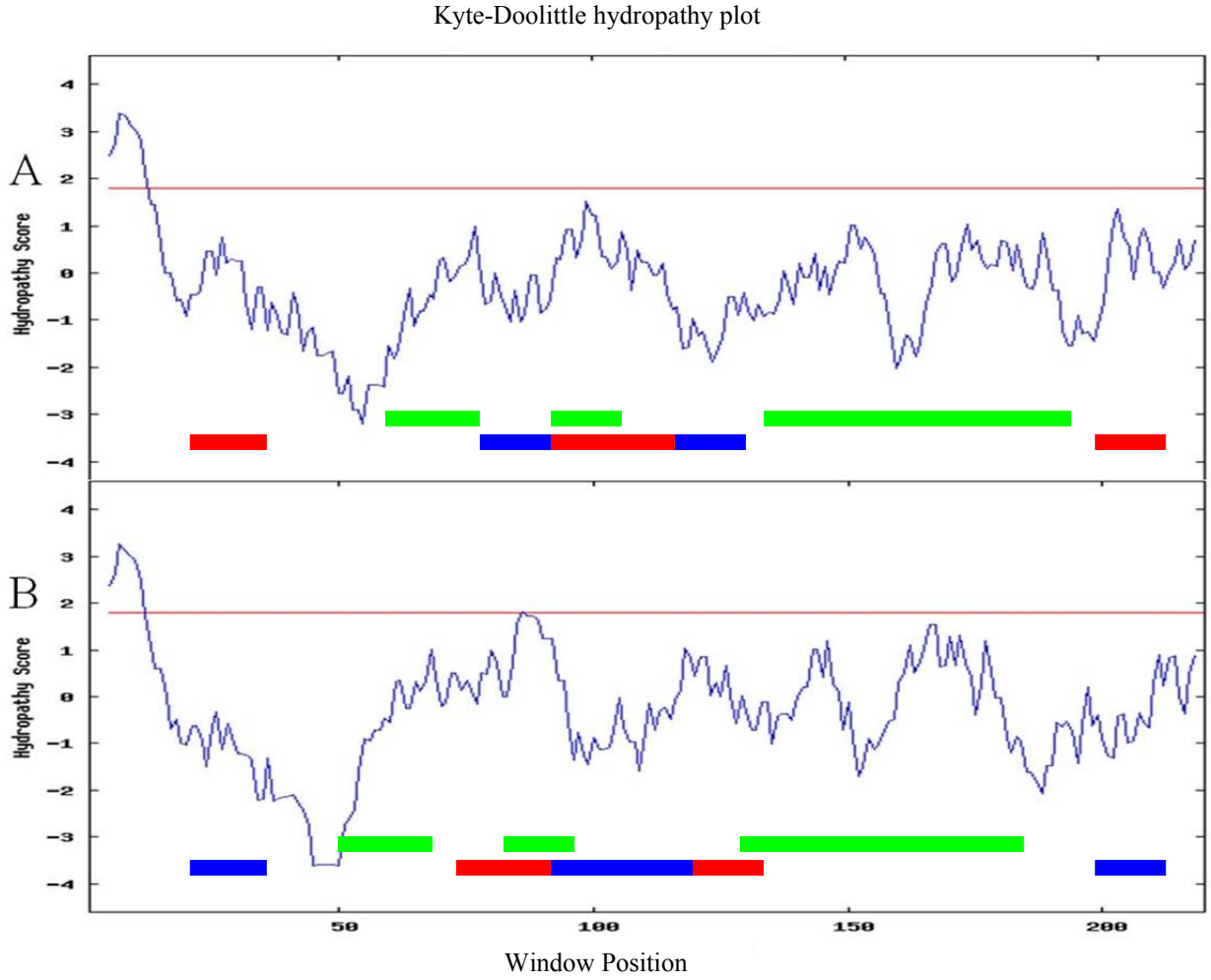

Figure 6.4. Kyte and Doolittle hydropathy plots for cow (A) and African elephant (B) β-caseins. The bars show differences in the distribution of hydrophobic (red) and hydrophilic (blue) regions. In African elephant β-casein, regions between amino acids 25-30, 95-115 and 200-210 are hydrophilic, and 75-90 and 120-130 are hydrophobic, whereas in the cow's β-casein the opposite is observed. Regions in green indicate the regions of high homology in amino acid sequences, which are shifted to the left in African elephant β-casein, due to its extended C-terminal, and which display the same hydropathy.

In order to investigate the properties of African elephant κ-casein, which are crucial in electrostatically stabilizing the casein micelle according to many theories which are universally accepted (Horne, 1998), hydropathy plots were used. These plots, depicted in Figures 6.5 and 6.7 show the hydrophobic and hydrophilic character of all the amino acids along κ-casein sequences, including the 21 amino acid signal sequences.

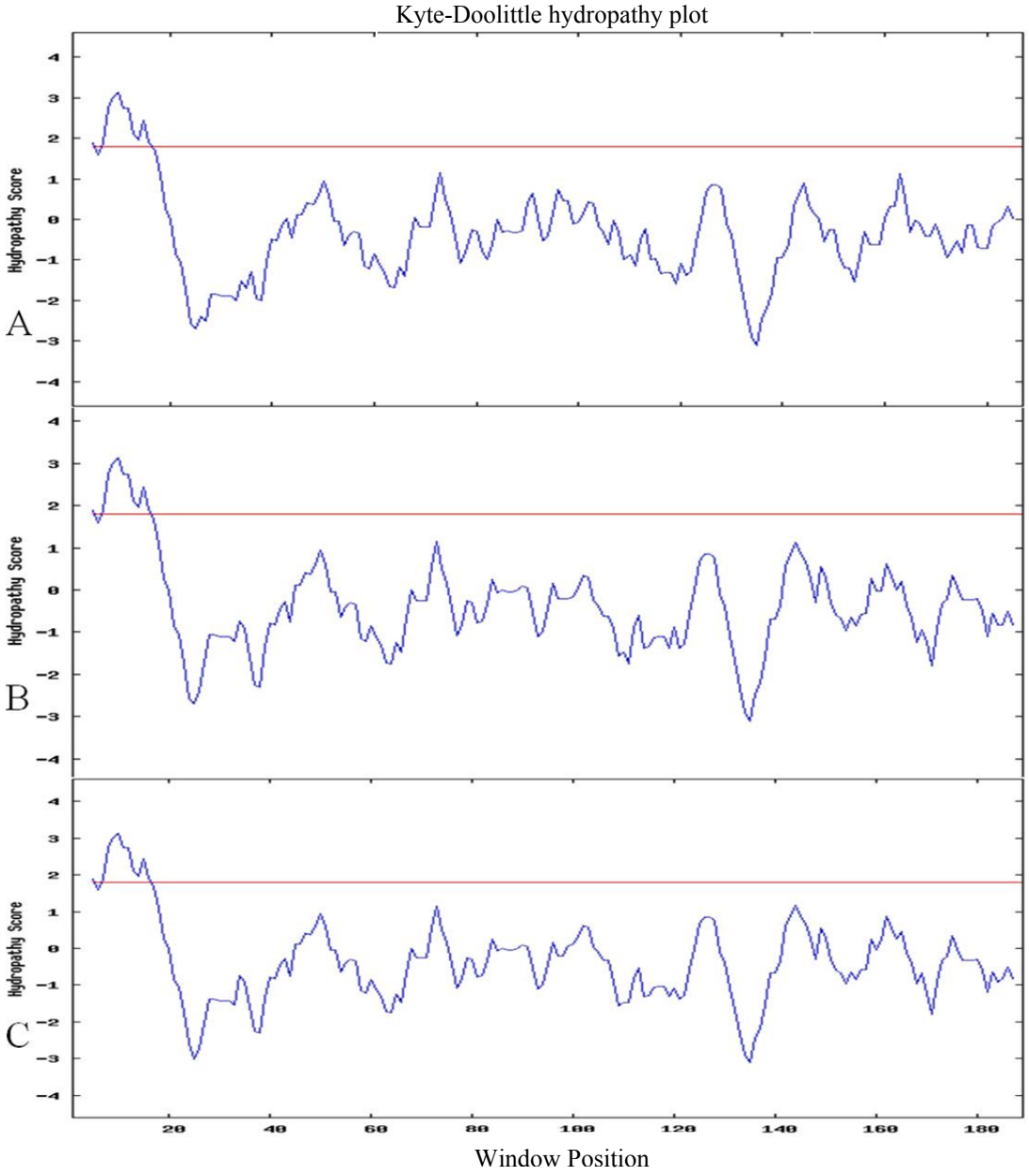

Figure 6.5. Kyte and Doolittle hydropathy plots for ruminant κ-casein, (a) cow, (b) goat and (c) sheep. Points above zero indicate hydrophobic residues while points below zero represent hydrophilic regions.

Figure 6.6 shows the alignment of cow and African elephant κ-casein. Since most of the experimental work on the casein micelle models was based on bovine milk and hence bovine caseins, it was imperative to deduce how these two sequences compare.

```
Elephant  MMKGFLLVVNILLLPLPFLAA EVQNQEESRCLEKDERWFCQKAVKYIPNDYVLKSYYRYE60
Cow       MMKSFFLVVTILALTLPFLGA QEQNQEQPIRCEKDERFFSDKIAKYIPIQYVLSRYPSYG
          ***.*:***.** * ****.*  :****:    *****:*.:* .**** :***. *  *
Elephant  PNYNQFRAAVPINNPYLIYLYPAKQVAVRPHTQIPQWQVPSNIYPSPSVP--------HT120
Cow       LNYYQQKPVALINNQFLPYPYYAKPAAVRSPAQILQWQVLSNTVPAKSCQAQPTTMARHP
           ** * : .. *** :* * * ** .***  :** **** **  *: *          *
Elephant  YLKPPFIVIPPKKTQDKPIIPPTGTVASIEATVEPKVNTVVNA----EASSEFIATNTPE180
Cow       HPHLSFMAIPPKKNQDKTEIPTINTIASGEPTSTPTTEAVESTVATLEDSPEVI-ESPPE
          : :  *:.*****.***  **   *:** * *  *.::* .:    * * *.*  . **
Elephant  ATTVPVISPQI
Cow       INTVQVTSTAV
           .** * *  :
```

Figure 6.6. Sequence alignment of elephant and cow κ-casein. The asterisks beneath the alignment indicate highly conserved amino acid residues between the two species. Colons and full stops indicate conservative and semi conservative substitutions respectively. Amino acid residues forming the chymosin sensitive bond are indicated by a red rectangular block.

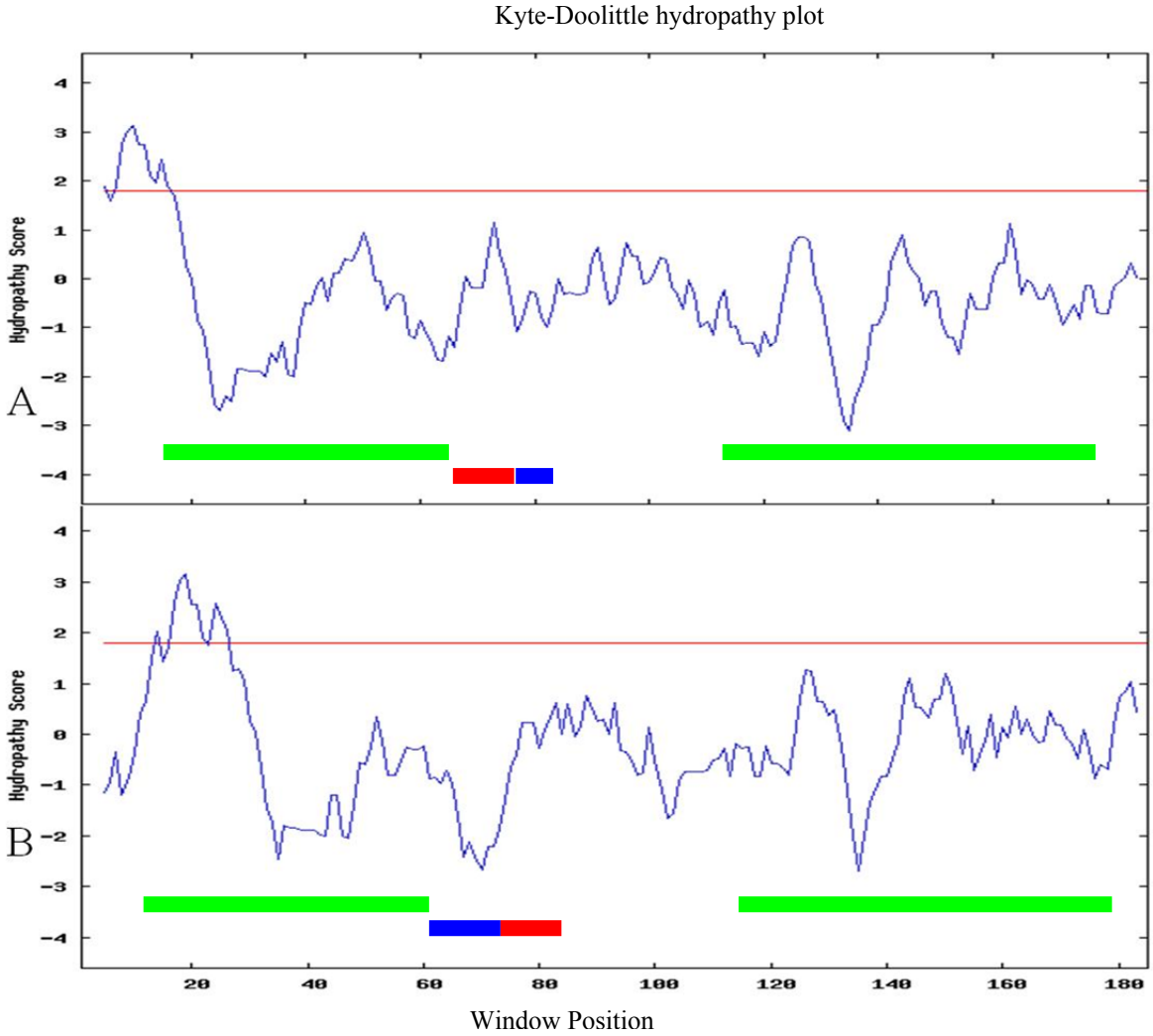

Figure 6.7. Kyte and Doolittle hydropathy plots for African elephant (A) and cow (B) κ-caseins. The bars show differences in the distribution of hydrophobic (red) and hydrophilic (blue) regions. In cow's κ-casein, a region between amino acids 79-81 is hydrophilic, and a region between amino acids 82-85 is hydrophobic, whereas in the African elephant κ-casein the opposite is observed on the corresponding regions. Regions in green indicate the regions of high homology in amino acid sequences and which display the same hydropathy.

6.4. Discussion

There is approximately 20 % amino acid homology between the amino acid sequences of the β-casein molecules in Figure 6.1. The homologous positions are mainly represented by single, fully conserved residues across the alignment of the β-casein sequences. This is low, compared to the approximately 70 % observed for α-LA discussed in chapter 5. While α-LA is a biologically active protein with high amino acid homology amongst mammalian species, β-caseins are storage proteins that provide essential amino acids for the nutritional requirements of the neonate, amino acid substitutions seem to be more tolerated (Farrell *et al*., 2004). The African elephant and human β-casein sequences are unique compared to the rest of the sequence, especially at the *N*-terminal, where a sequence constituting of charged amino acid residues is absent (17-26) which also constitute exon 3 of β-caseins (Martin *et al*., 2003; Ginger & Grigor, 1999). These sequences also have an additional 8 amino acids at their C-terminal which forms exon 7. This may suggest that African elephant β-caseins may interact differently as part of a casein micelle with reference to the casein micelle model. Hydropathy plots indicate potential surface or transmembrane proteins, each of the 20 amino acids is assigned a hydropathy score, either being hydrophobic (peaks above zero) or hydrophilic (peaks below zero) (Kyte & Doolittle, 1982). A comparison of the hydropathy plots of the β-caseins shows that the plots of cow, sheep and goat, depicted in Figure 6.2 are very similar to each other, this is expected since the three species are all ruminants and therefore much more closely related.

Figure 6.3 shows a comparison of hydropathy plots of β-caseins from cow, mouse, pig and human. Mouse and human hydropathy plots show more or less neutral proteins, the sequence regions between amino acids 25-30 acids is hydrophobic in cow β-casein, whereas the corresponding regions in mouse, human and pig β-casein are hydrophilic. The same contrasting feature is also observed between amino acids 155-165, where cow β-casein is hydrophilic while mouse, human and pig β-casein are hydrophobic. However, exon 3 in human β-casein is spliced out which results in shifting of amino acid sequences and hence changes in the trend of hydropathy. The African elephant β-casein hydropathy plot exhibits a rather different plot compared to

ruminant plots as shown in Figure 6.4. These differences are mainly attributed to the missing exon 3 and an extended exon 7 in African elephant β-casein, consequently the sequence of African elephant β-casein that is homologous to cow β-casein sequence is shifted to the left as depicted in Figure 6.4 by the green blocks. Be that as it may, the African elephant β-casein constitutes more hydrophilic stretches compared to cow β-casein, which may suggest that African elephant β-caseins may interact differently as part of the casein micelle.

The hydropathy plots of ruminant κ-casein, shown in Figure 6.5, are very similar to each other. This supports the notion that the κ-casein gene is identically organised in ruminants (Martin *et al*., 2003). Since most of the casein micelle models were drawn from experimental work with bovine κ-casein, it is more likely that goat and sheep κ-casein will behave the same way in the same way in the formation of micelles. The alignment of cow and African elephant κ-casein sequences, depicted in Figure 6.6, shows some interesting aspects. It appears that cow and African elephant κ-casein sequences have approximately 50 % amino acid homology. Most of the substitutions are conservative, meaning substitutions are mainly of amino acids of similar properties regarding charge or hydrophobicity.

The comparison of the hydropathy plots of cow and African elephant κ-casein, depicted in Figure 6.7, shows very few differences between the two. A section of the plots between amino acids 65-83 have contrasting features due to a number of charged amino acids occurring in the cow's protein, which have been exchanged with hydrophobic ones in the elephant's. It is possible that this difference may have an effect on the interaction with other caseins within the casein micelle. Of importance is that the C terminal sections after the chymosin sensitive bond at position 126-127 are highly conserved and also display almost identical hydropathy profiles. This implies that the African elephant κ-casein should functionally behave in the same manner as all κ-caseins by presenting the charged C-terminal on the surface of the casein micelle to avoid coagulation.

The casein micelle models were developed after experimental procedures involving bovine caseins. Since methods such as NMR and x-ray crystallography proved improbable to use in determining the structure of casein micelles, only models could be used to explain the structure of casein micelles. These models explain how the characteristic behavior of bovine caseins allows them to form the casein micelle and also the influence of hydrophobic and hydrophilic residues they possess. Now, after the characterization of African elephant β-casein, a number of observations were made which makes it unique compared to other β-caseins referred to in this study.

Firstly, looking at the sequence alignment shown in Figure 6.1, it is observed that African elephant β-casein has an extended C-terminal with charged amino acids so that it is more hydrophilic than the bovine β-casein sequence. The missing amino acid sequence between amino acids 18-26, which constitutes exon 3 near the N-terminal, is a loss of charged (QEE-N) as well as hydrophobic (L-VVG) amino acids. This character is also shared with human β-casein. Additionally, the presence of a rather unique sequence IR-QFYSEQKPEGVKH (between amino acids 36-51) of African elephant β-casein (of which the PEGVKH sequence shares some similarities with the protein from human), which contains a mixture of charged and hydrophobic amino acids, was also observed. In general the African elephant protein has large hydrophilic stretches between amino acids 10-55, 90-125 and 190-210. The bovine protein is more amphipathic in these regions. It could therefore be possible that these regions would protrude from the surface of the micelle causing steric and electrostatic stabilization of the micelles in milk (De Kruif & Roefs, 1997). The hydrophilic part between amino acids 90-110, therefore causes the middle section to be less hydrophobic, so that the hydrophobic interaction proposed by the casein models may not be possible with this protein. Additionally, as discussed in chapter 4, the analysis of 2D PAGE of African elephant milk proteins using PD QUEST (Bio-Rad) has revealed that the ratio of kappa casein to β-casein is approximately, 1: 8.5 compared to the same ratio in cow's milk, which is approximately 1:5 (Martin *et al*., 2003), it is possible that there is not enough κ-casein to promote repulsion between casein micelles, and because of this, African elephant β-casein could contribute to this repulsion by contributing two loops and a C-terminal on the surface of the micelle.

6.5. References

De Kruif, C. G. & Roefs, S. P. F. M. (1997). Skim milk acidification at low temperature. *Neth Milk Dairy J* **50**, 113-120.

Farrell, H. M., Jiminez-Flores, R., Bleck, G. T., Brown, E. M., Butler, J. E., Creamer, L. K., Hicks, C. L., Hollar, C. M., Ng-Kwai-Hang, K. F. & Swaisgood, H. E. (2004). Nomenclature of the proteins of cow's milk-sixth- revision. *J Dairy Sci* **87**, 1641-1674.

Fox, P. F. & McSweeney, P. L. H. (1998). *Dairy Chemistry and Biochemistry*. London: Chapman & Hall.

Ginger, M. R. & Grigor, M. R. (1999). Comparative aspects of milk caseins. *Comp Biochem Physiol B* **124**, 133-145.

Horne, D. S. (2008). Casein micelle structure and stability. In *Milk proteins: from expression to food,* pp. 133-162. Edited by A. Thomson, M. Boland. & H. Singh. New York: Elsevier.

Kulski, J. K. & Hartmann, P. E. (1981). Changes in human-milk composition during the initiation of lactation. *Aust J Exp Biol Med Sci* **59**, 101-114.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132.

Martin, P., Ferranti, P., Leroux, C., & Addeo, F. (2003). Non-bovine caseins: Quantitative variability and molecular diversity. In *Advanced Dairy Chemistry,* pp. 387-419. Edited by P. F. Fox & P. L. H. McSweeney. New York: Plenum Publishers.

Phadungath, C. (2005). Casein micelle structure: A concise review. *Songklanakarin J Sci Technol* **27**, 201-212.

Sievers, F., Wilm A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins D. G (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 1-6.

Chapter 7

General Discussion and Conclusions

Milk has co-evolved with mammals and serves as a complete, complex and key element suited for specific nutritional requirements of the neonate (Martin *et al*., 2003; Casado *et al*., 2009). Comprehensive studies on milk, especially milk proteins, goes no further than human milk and milk from a few other mammalian sources (Oftedal & Iverson, 1995). The availability of accurate and reliable data on milk of non-dairy mammals is limited by factors such as: inaccurate determination of lactation stage, flaws in analytical procedures and availability of small sample numbers and sample volumes among other factors. A deep understanding of milk proteomics of other mammalian species could be an answer to the increased demand of the food industry for functional proteins as ingredients and therefore their inclusion in functional foods. Moreover, the study of milk proteomics can have a massive contribution to human as well as animal welfare, thus providing important knowledge for improving milk for nutrition. Therefore, the main objective of this study was to investigate the milk proteome of species that differs substantially from human and cow milk proteomes. In this case, the milk proteome of African elephant milk was investigated using the proteomics approach.

The first comprehensive study of African elephant milk was conducted by (McCullagh & Widdowson, 1970). However, the samples were collected post mortem. In another study by (Osthoff *et al*., 2005), milk samples were collected ante mortem. In the latter study, details of protein and sugar content of the African elephant milk were also provided.

In this study, milk samples collected from a live African elephant, were prepared by a precipitation method so that only proteins were retained. One dimensional (SDS

PAGE) separation of African elephant milk proteins was then conducted; this was done in order to obtain a basic picture of the array of proteins that are present in African elephant milk. Milk samples were from African elephant at early (4 days), mid (12 months) and late (18 months) lactation. It was evident from the SDS PAGE gel, specifically the intensity of the bands on the gels that some proteins occur in higher amounts than others. Furthermore, it was also concluded that whey proteins occur in higher amounts at early lactation, whereas at later stages of lactation, caseins seem to occur in higher amounts. From LC MS/MS analysis, proteins such as α_{s1}-, α_{s2}-, β-casein, α-LA, β-LG and serum albumin were identified, although some were matched to proteins from other mammalian species. This suggested that the African elephant milk proteome is more or less similar to that of other mammalian species, such as cow's milk, which have been extensively characterized. Precipitation of caseins was also undertaken in an effort to investigate the less abundant whey proteins. Following LC MS/MS, whey proteins such as α-LA could now be identified with high confidence values. Of interest is that κ-casein, one of the well-studied casein (Hidalgo *et al*., 2010) could not be identified at this stage, thus it was necessary to conduct further investigations.

The next experimental work involved using 2D PAGE as a means of protein separation (early, mid and late lactation samples) and consequent MS analysis (tandem and orbitrap MS). Comparison of 2D PAGE gels at different lactation stages showed that whey proteins decrease as lactation progresses and as a result the caseins become more prominent. After LC MS/MS analysis of the proteins spots on 2D PAGE gels, α-LA and α_{s1}-casein were the only proteins of interest identified. This prompted further analysis of protein spots using orbitrap MS, which is a more sensitive analytical tool, with higher resolution, mass accuracy and faster cycling times, compared to tandem MS and in addition, an elephant protein database was used for identification. This approach was successful, more protein spots were identified and their sequences were also obtained. The major proteins identified included α-LA, lactoferrin, β- and κ-casein. Moreover, several protein spots on the 2D PAGE gels were identified as β-casein. The latter could be a result of posttranslational modifications; in this case, serine residues of caseins which are widely and variably

phosphorylated. Multiple sequence alignments for comparison of α-LA, β- and κ-casein with corresponding protein sequences from other mammalian species were conducted. From the α-LA alignment which consisted of 16 sequences, several amino acids were shown to be unique to the African elephant α-LA sequence at certain positions on the sequences. Some of the substitutions were not conservative; such substitutions may have functional implications depending on where they are situated, especially considering the role of α-LA in lactose synthesis (Ramakrishnan *et al*., 2001).

The African elephant κ-casein chymosin cleavage site constitutes of phenylalanine and isoleucine rather than phenylalanine and methionine, of which the same property is shared with pig, human and horse κ-casein. The sequence alignment of β-casein showed some interesting results. Like other caseins in the alignment, African elephant β-casein does not have any cysteine residues in its sequence, is rich in proline and has a single *N*-terminal located multiple phosphorylation site. These characteristics may allow β-casein to self-aggregate and interact via hydrophobic interactions with other caseins and there-by provide means for chain elongation and micelle growth (Horne, 1998). However, the exonic skipping that occurs in African elephant β-casein between (residues 3-11), may influence the properties of the peptide including self aggregation and interaction with other proteins.

2D PAGE gels of African elephant proteins were analyzed using computer software (PD QUEST). The major findings of this experiment were that the milk proteome of the African elephant is not constant throughout lactation. Furthermore, some proteins of unknown identity were only visible on the gels at specific stages of lactation. This could be a consequence of changes in the neonatal nutritional demands (Nicholas *et al*., 1997; Nicholas, 1988). The ratio of the main proteins of interest was also calculated by making use of the spot intensities. The ratio of African elephant milk κ-casein to β-casein is approximately 1: 8.5, which is in the same order compared to camel and rat milk (Martin *et al*., 2003). However, as mentioned in chapter 4, the method used in intensity calculation has some limitations especially at very high spot

intensities, thus the ratio of African elephant κ-casein to β-casein maybe far greater than the figure reported here, it is imperative that further investigations are conducted.

YASARA (Modeling program) was used to model the structure of African elephant lactose synthase complex, this was done in order to evaluate the structure of African elephant α-LA and β-1,4GT1, and their possible structural influence in the presence of high levels of both lactose and oligosaccharides in African elephant milk (Osthoff *et al*., 2008). After analysis of the models, which were structurally homologous to X-ray crystallography structures of other mammalian sources, it was concluded that the differences in amino acid composition in α-LA, and also in β-1,4-GT1, of African elephant neither affect the 3D structure of the proteins, nor their interaction and catalytic sites. This in turn may implicate that it is not the small structural differences of these proteins that are responsible for the eventual high levels of both oligosaccharides and lactose in African elephant milk, compared to other mammals. This observation could be an effect of other additional or over expression of glycosyltransferases relative to α-LA, which may be present in the mammary gland of African elephant as it is the case in some nonhuman species (Messer & Urashima, 2002; Urashima *et al.*, 2004; Urashima *et al*., 2009). Since isoglobotriose [Gal(αl-3)Gal(β1–4)Glc] is the most abundant of the oligosaccharides in Asian and African elephant milk (Uemura *et al*., 2006; Osthoff *et al*., 2008), it is most probably the β-galactosyltransferase that links galactose via an α l-3 bond to the galactosyl moiety of the lactose, that is in competition with α-LA in its affinity for lactose, which may result in disproportionate synthesis of oligosaccharides or lactose

Lastly, the role of caseins in micelle formation was investigated. Amino acid sequence comparison and hydropathy profiles of African elephant κ- and β-casein were employed. The hydropathy plot of African elephant β-casein differs from that of ruminants. This feature is mainly attributed to the missing exon 3 in African elephant β-casein sequence, consequently resulting in shifting of the plot to the left. The African elephant β-casein constitutes more hydrophilic regions compared to cow β-caseins, which may suggest that African elephant β-caseins may interact differently as

part of the casein micelle, possibly by assuming a role that is similar to that of κ-casein, i.e. with parts of the hydrophilic, charged regions protruding from the surface of the micelle, causing steric hindrance and electrostatic repulsion in order to prevent micelle aggregation.

The peak picking shows that elephant milk contains a low ratio of κ-casein. The amino acid sequence and hydropathy plots predict that African elephant κ-casein would function in the same way as in other species. This means that there would be very little κ-casein on the surface to effect repulsion of the micelles. There should therefore be other protein regions protruding from the micelle surface to aid in this function. It is suggested that the hydrophilic regions of β-casein plays this role.

It may be asked what the usefulness of the knowledge gained from this research is on milk nutrition and technology in general. Many plant foods are produced by genetic modification or manipulation of plants. Whether the nutritional composition of milk might be improved by genetic manipulation of the cow has come up in several debates. This would involve stimulation of the production of enzymes. If α-La could be manipulated to produce less lactose, the milk content of the probiotic oligosaccharides would increase. The results of this study show that the predicted tertiary structure of the α-La most probably does not differ, and would therefore not function differently compared to its equivalent in other species. This structure-function activity of this protein is therefore not the reason for the high content of oligosaccharides in African elephant milk, and probably also not in other species with this property.

The results obtained for the structural aspects of the β- and κ-caseins show that micelle formation and prevention of their aggregation might not solely be the responsibility of the κ-caseins, but that the β-caseins may contribute to a large extent. It might therefore be possible that site specific hydrolysis of the β-caseins should be investigated, with the consequential coagulation of the micelles. This might be very

useful for cheese production from other species such as horse milk. New cheese products could then be developed. Perhaps this approach might also be investigated on cow's milk in order to increase cheese recovery and also texture and mouth feel of cheeses.

7.1. Future research

Future experimental work will involve searching for alpha caseins of African elephant milk and consequently their characterisation provided they are present. Peak picking experiments should also be repeated to improve the current result regarding the casein ratios. Finally, investigation of other functional proteins such as β-Lactoglobulin and lactoferrin can also be undertaken.

7.2. References

Casado, B., Affolter, M. & Kussmann, M. (2009). OMICS-Rooted studies of milk proteins, oligosaccharides and lipids. *J Proteomics* **73**, 196-208.

Hidalgo, M. E., Pires, M. S. & Risso, P. H. (2010). A study of bovine kappa-casein aggregation after the enzymatic action of chymosin. *Colloids Surf B Biointerfaces* **76**, 556-563.

Horne, D. S. (1998). Casein interactions: casting light on the black boxes, the structure in dairy products. *Int Dairy J* **8**, 171-177.

Martin, P., Ferranti, P., Leroux, C., & Addeo, F. (2003). Non-bovine caseins: Quantitative variability and molecular diversity. In *Advanced Dairy Chemistry,* pp. 387-419. Edited by P. F. Fox & P. L. H. McSweeney. New York: Plenum Publishers.

McCullagh, K. G. & Widdowson, E. M. (1970). The milk of the African elephant. *Br J Nutr* **24**, 109-117.

Messer, M. & Urashima, T. (2002). Evolution of milk oligosaccharides and lactose. *Trends Glycosci Glycotechnol* **77**, 153-176.

Nicholas, K. R. (1988). Control of milk protein synthesis in the tammar wallaby: A model system to study prolactin-dependent development. In *The Developing Marsupial: Models for Biomedical Research,* pp. 68-85. Edited by C. H. Tyndale-Biscoe & P. A. Janssens. Heidelberg: Springer.

Nicholas, K., Simpson, K., Wilson, M., Trott, J. & Shaw, D. (1997). The tammar wallaby: A model to study putative autocrine-induced changes in milk composition. *J Mammary Gland Biol Neoplasia* **2**, 299-310.

Oftedal, O. T & Iverson, S. J. (1995). Comparative analysis of non-human milks: Phylogenetic variation in the gross composition of milks. In *Hand book of Milk Composition,* pp. 749-788. Edited by R. G. Jensen. New York: Academic Press.

Osthoff, G., De Waal, H. O., Hugo, A., De Wit, M. & Botes, P. (2005). Milk composition of a free-ranging African elephant (*Loxodonta africana*) cow during early lactation. *Comp Biochem Physiol A* **141**, 223-229.

Osthoff, G., Dickens, L., Urashima, T., Bonnet, S. L., Uemura, Y. & van der Westhuizen, J. H. (2008). Structural characterization of the oligosaccharides in the milk of an African elephant (*Loxodonta africana africana*). *Comp Biochem Physio B* **150**, 74-84.

Ramakrishnan, B., Shah, P. S. & Qasba, P. K. (2001). α-Lactalbumin (LA) stimulates milk β-1,4-Galactosyltransferase I (β4Gal-T1) to transfer glucose from UDP-glucose to N-Acetylglucosamine. *J Biol Chem* **276**, 37665-37671.

Urashima, T., Kitaoka, M., Asakuma, S. & Messer, M. (2009). Milk oligosaccharides. In *Advanced Dairy Chemistry,* pp. 295-349. Edited by P. McSweeney & P. F. Fox. New York: Springer.

Summary

Milk is a complex and complete food for the specific nutritional requirements of the neonate. For the dairy industry, milk is a suitable raw material for the production of other high value products. Although extensive research has been carried out on milk of economically exploited dairy animals such as cow, goat, sheep, buffalo and camel, there are properties which are not explicit in the milk of these, that are not completely understood. Research of the milk from non-dairy animals, in which these properties are explicit, may provide answers.

One of the unique properties is the content of oligosaccharides, which is low in the dairy animals, but high in some species. This property points to a specialized saccharide synthesis in the latter, where the whey protein α-lactalbumin may play a role. Another unique property is the structure of casein micelles, which in the dairy animals is stabilized by the presence of a specific ratio of four casein types, as well as their specific structural properties. The most important is the κ-casein with its amphipathic nature. In some non-dairy species, stable casein micelles are formed in spite of the absence, or low content, of some of the casein types, specifically the κ-casein.

The milk of the African elephant (*Loxodonta africana*) displays several unique properties. In this research the proteome of its milk was investigated, with a focus on α-lactalbumin and the caseins, in order to shed light on the mentioned unique properties. The proteomics approach was used, which includes gel electrophoresis and mass spectrometry. Computer modeling was also employed. The major proteins α_{s1}-, α_{s2}-, β-casein, α-lactalbumin, β-lactoglobulin and serum albumin of African elephant milk were identified with one-dimensional electrophoresis and mass spectrometry. Better results were obtained with two-dimensional electrophoresis and orbitrap mass spectrometry, with which α-lactalbumin, lactoferrin, β- and κ-casein were identified.

The multiple sequence alignment of α-lactalbumins showed that there are six amino acid positions that are unique to that of African elephant. Most of the amino acid substitutions in this protein were found to be conserved, and the structure model of African elephant α-lactalbumin was found to be homologous to the X-ray crystallography structures of several species. Consequently the structure models of β-1,4-galactosyltransferase 1 and the lactose synthase complex were built, again showing homology to crystallographic data from other species. It may therefore be concluded that structures of α-lactalbumin and β-1,4-galactosyltransferase 1 are highly conserved amongst species. The saccharide synthesis in the African elephant milk would probably not differ from that of other mammals, and may therefore not be the reason for high levels of oligosaccharides in its milk.

The comparison of amino acid sequences and hydropathy plots of African elephant β-casein with that of other species showed that it would self-aggregate and interact via hydrophobic interactions with other caseins to form casein micelles, similar to the model proposed for cow's milk micelles. However, the African elephant β-casein displays several more hydrophilic regions, compared to the cow's protein.

The amino acid sequence and hydropathy plots predicted that African elephant κ-casein would function in the same way as the equivalent of other species. The ratio of African elephant milk κ-casein to β-casein was calculated to be approximately 1:8.5, which is in the same order as camel and rat milk, compared to the 3:8 of cow's milk. This means that there would be very little κ-casein on the surface to effect repulsion of the micelles. There should therefore be other protein regions protruding from the micelle surface to aid in this function. It is suggested that in African elephant milk the hydrophilic regions of β-casein carry out this role.

Keywords: African elephant; Electrophoresis; Milk proteomics; Hydropathy analysis; Casein micelle; Structure modeling.

Opsomming

Melk is ‘n komplekse en voedingsgewys volmaakte voedsel vir die pasgeborene. Vir die suiwelindustrie is melk ‘n geskikte grondstof vir die vervaardiging van ander produkte van hoë waarde. Alhoewel uitgebreide navorsing uitgevoer is op melk van ekonomies bruikbare suiwelvee soos beeste, bokke, skape, waterbuffels en kamele, is daar eienskappe wat nie eksplisiet in hierdies is nie, wat nie heeltemal verstaan word nie. Navorsing op die melk van nie-suiweldiere, waarin hierdie eienskappe eksplisiet is, mag antwoorde verskaf.

Een van die unieke eienskappe is die inhoud van oligosakkariede, wat laag is in die suiweldiere, maar hoog in sommige spesies. Hierdie eienskap dui op ‘n gespesialiseerde sakkariedsintese in laasgenoemdes, waar die weiproteïen, α-laktalbumien, ‘n rol mag speel. ‘n Verdere unieke eienskap is die struktuur van kaseïenmiselle, wat in die suiweldiere gestabiliseer word deur die teenwoordigheid van ‘n spesifieke verhouding van vier kaseïentipes sowel as hul spesifieke strukturele eienskappe. Die belangrikste is die κ-kaseïen met sy amfipatiese einskappe. In sommige nie-suiwelspesies word stabiele kaseïenmiselle gevorm ten spyte van die afwesigheid, of lae inhoud, van die kaseïentipes, spesifiek die κ-kaseïn.

Die melk van die Afrika olifant (*Loxodonta africana*) besit verskeie unieke eienskappe. In hierdie navorsing is die proteoom van sy melk bestudeer, met ‘n fokus op α-laktalbumien en die kaseïene, met die doel om lig te werp op die genoemde unieke eienskappe. Die proteomiese benadering is gevolg, wat insluit jelelektroforese en massaspektrometrie. Rekenaarmodelering is ook ingespan.

Die hoofproteïene α_{s1}-, α_{s2}-, β-kaseïen, α-laktalbumien, β-laktoglobulien en serumalbumien van Afrika olifantmelk is geïdentifiseer met eendimensionele elektroforese en massaspektrometrie. Beter resultate is behaal met tweedimensionele elektroforese en orbitrap massaspektrometrie, waarmee α-laktalbumien, laktoferrien, β- and κ-kaseïene identifiseer is.

Veelvoudige vergelyking van aminosuurvolgordes van α-laktalbumiene het gewys dat ses aminosuurposisies uniek is in die olifantproteïen. Dit is vasgestel dat die meeste van die aminosuurvervangings konserwatief is, en struktuurmodelering van die Afrika olifant α-laktalbumien het gewys dat dit homoloog is met X-straalkristallografiese strukture van verskeie spesies. Gevolglik is die struktuurmodel van β-1,4-galaktosieltransferase 1 en die laktosesintasekompleks gebou, wat weereens homologie met die kristallografiese data van ander spesies getoon het. Daar mag dus aanvaar word dat die strukture van α-laktalbumien en β-1,4-galaktoesieltransferase 1 hoogs gekonserveerd is tussen spesies. Die sakkariedsintese in die Afrika olifantmelk behoort dus nie te verskil van die van ander soogdiere nie, en behoort dus nie verantwoordelik te wees vir die hoë vlakke van oligosakkariede in sy melk nie.

Die vergelyking van aminosuurvolgordes en hidropatisiteitsprofiele van Afrika olifantmelk β-kaseïen met die van ander spesies het aangedui dit mag selfaggregeer en interreageer met ander kaseïene deur middle van hidrofobe interaksies om kaseïenmiselle te vorm, soortgelyk as die model wat vir beesmelkkaseïene voorgestel is. Maar die Afrika olifantmelk β-kaseïen vertoon meer hidrofieliese gebiede, vergeleke met die beesproteïen.

Die aminosuurvolgorde en hidropatisieteitsprofiele voorspel dat die Afrika olifant κ-kaseïen soortgelyk mag funksioneer as die eweknie in ander spesies. Die verhouding van die κ-kaseïene tot β- kaseïene in Afrika olifantmelk is bereken as ongeveer 1:8.5, wat in dieslefde orde is as kameel- en rotmelk, vergeleke met die 3:8 in beesmelk. Dit beteken dat daar baie min κ-kaseïene op die oppervlak mag wees om afstoting tussen miselle te bewerkstellig. Daar moet dus ander proteïengebiede uit die misel uitsteek om hierdie funksie te vervul. Daar word voorgestel dat in Afrika olifantmelk die hidrofieliese gebiede van β-kaseïen hierdie rol mag vervul.

Sleutelwoorde: Afrika olifant; Melk proteomika; Hidropatisiteitsanalise; Kaseïen; Alfalaktalbumien; Structuurmodelering.