GLYCOSYLATION OF BOVINE α-LACTALBUMIN

A thesis presented in partial fulfilment of the requirements
for the degree of
Master of Philosophy in Biochemistry,
at Massey University,
New Zealand.

Udumalagala Gamage Chandrika
1999

MASSEY UNIVERSITY LIBRARY

THESIS COPYRIGHT FORM

Title of Thesis: Glycosylation of α-Lactalbumin

Please delete section not applicable.

(1) (a) I give permission for my thesis to be made available to readers in Massey University Library under conditions determined by the Librarian.

(b) I do not wish my thesis to be made available to readers without my written consent for ___ months.

(2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.

(b) I do not wish my thesis, or a copy, to be sent to another institution without my written consent for ___ months.

(3) (a) I agree that my thesis may be copied for Library use.

(b) I do not wish my thesis to be copied for Library use for ___ months.

Signed: Chandrika Gamage

Date: 5/2/99

****

FOR READERS ONLY

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME and ADDRESS DATE

ABSTRACT

Bovine α-lactalbumin exists in four different forms. These are the F, M, S_1 and S_2 forms named after their positions on native gels. F, S_1 and S_2 are minor components representing 15% of the total α-lactalbumin fraction whereas M is the major component. S_1 and S_2 have been shown to be glycoforms of α-lactalbumin and although there are potentially three glycosylation sites in the protein, only asparagine 45 appears to be glycosylated. It has been suggested that F differs from M by the replacement of an amide group.

The glycoforms of α-lactalbumin (S_1 and S_2), and the non-glycosylated proteins (M and F) were isolated and purified using selective precipitation, affinity chromatography, size exclusion chromatography and preparative electrophoresis.

The potential glycosylation sites were investigated using selective proteolysis in conjunction with Edman sequencing and electrospray mass spectrometry (ES/MS).

It was found that although the main fraction (M) or non-glycosylated protein contained no covalently bound carbohydrate, selected ion monitoring experiments showed that there appeared to be a lactosamine sugar associated with the protein.

A number of methods were investigated for analysing and separating the glycoforms of α-lactalbumin. These included high pH anion exchange chromatography with pulsed amperometric detection (HPAEC/PAD), fluorophore assisted gel electrophoresis (FACE) and derivatisation of the oligosaccharides with 1-(p-methoxy)phenyl-3-methyl-5-pyralozone (PMPMP) and subsequent separation by RP-HPLC. Electrospray mass spectrometry was used to confirm the results of these various techniques.

Although it was not firmly established which of the three possible sites were glycosylated, refinement of the purification protocol resulted in several different glycans being identified on the basis of the ES/MS and FACE results. It would appear that there are up to 15 different glycoforms of α-lactalbumin, some of which are highly sialated. It is difficult to determine whether the simpler structures represent breakdown products of the more complex structures, or whether they are present naturally.

ACKNOWLEDGEMENTS

I wish to express my sincere thanks and heart felt gratitude to my supervisor Dr. Gill Norris, for her invaluable advice and guidance given to me throughout the project, suggestions on the manuscript preparation and ever willing help and co-operation over the last couple of years.

I would like to thank Professor Patrick Sullivan for giving me the opportunity to carry out this project and making the facilities available.

I would also like to thank Professor E. R. Jansz, from Sri Lanka, for his encouragement and kind co-operation for my postgraduate studies in New Zealand.

My sincere appreciation is also extended to...

The Dairy Research Institute for the opportunity to use the Dionex system and the electrospray mass spectrometer.
Jo Mudford for doing the Edman sequencing and the ES/MS for my samples.
Paul McJarrow for his invaluable assistance with the Dionex system (HPAEC/PAD).
Rochelle Ramsay for her assistance with the FACE gels.
Professor Dave Harding and Dick Poll for their help on the use of the FPLC.
Dr. Cristina Weinberg who helped me readily whenever I sought her help.
Debbie Frumau and Trever Loo for their co-operation as lab-mates.
Derek Body for his constant support and personal time invested in proof-reading.

I am forever grateful to my parents and husband who helped me in many ways.

ABBREVIATIONS

API	Atmospheric pressure ionisation
ANTS	8-amino-napthalene-1,3,6 trisulphonic acid
AUFS	Absorbance units at full scale
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
cAMP	Cyclic 3', 5'-adenosine monophosphate
CAP	3-(Cyclohexylamino)-propanesulphonic acid
CBR-250	Coomassie Brilliant Blue R-250
CID	Collision induced decomposition
Con A	Concanvalin A
3-D	Three dimensional
Da	Dalton
DDT	Dithiothretitol
DEAE	Diethylaminoethyl
Dp	Degree of poymerization
EDTA	Ethylenediamine tetra-acetic acid (di-sodium salt)
ENGase	Endo-N-acetyl-β-D-glucosaminidase or endoglycosidase
ER	Endoplasmic reticulum
ES/MS	Electrospray mass spectrometry
FACE	Fluorophore assisted carbohydrate electrophoresis
FPLC	Fast protein liquid chromatography
GPI	Glycosylphosphatidylinositol
HEWL	Hen egg-white lysozyme
H^1 NMR	Hydrogen nuclear magnetic resonance
HPAEC/PAD	High performance anion exchange chromatography with pulsed amperometric detection
IgGs	Immunoglobulins

IS	Ion spray voltage
IQ	Q exit lens
α-LA	α-lactalbumin
β-LG	β-lactoglobulin
M+H	Molecular ion (protonated)
PAGE	Polyacrylamide gel electrophoresis
PNGase	Peptide-N4-(N-acetyl-β-D-glucosaminyl) asparagine amidase
PMPMP	1-(p-methoxy)phenyl-3-methyl-5-pyrazolone
PVDF	Polyvinylidene difluoride
OR	Orifice
RER	Rough endoplasmic reticulum
RP-HPLC	Reverse phase-high performance liquid chromatography
RNG	Ring voltage
SDS	Sodium dodecyl sulphate
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
Tris	Tris(hydroxymethyl) aminomethane
UDP	Uridine diphosphate

Sugar abbreviations

Gal	Galactose
Man	Mannose
Fuc	Fucose
GlcNAc	N-acetylglucosamine
GalNAc	N-acetylgalactosamine
NeuNAc	N-acetylneuraminic acid (sialic acid)

The 20 common amino acids

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Outline

This section will provide a brief overview of the relationship between glycosylation and bovine α-lactalbumin. Firstly, protein glycosylation will be reviewed looking at the different glycans that are found covalently linked to proteins, the manner in which they are attached to the protein, and how effectively certain recognition sites are glycosylated. Secondly, the functional roles of these oligosaccharides will be discussed. Finally, there will be a detailed discussion about α-lactalbumin and glycosylated α-lactalbumin.

1.2 Glycoproteins

Glycoproteins are proteins which possess an oligosaccharide moiety that is covalently attached to either the amide side chain of asparagine via an N-glycosidic bond or the hydroxyl side chain of serine, threonine, hydroxy proline or hydroxy lysine via an O-glycosidic bond. The plethora of glycosyltransferases present in the endoplasmic reticulum (ER) and the Golgi apparatus results in the biosynthesis of a large heterogeneous population of structurally related oligosaccharides that can be attached to one or more sites on any one protein. These oligosaccharides can be different in their primary structure as well as their anomeric (α or β) and isomeric (links through the 2, 3, 4, or 6 position) configurations. Each sugar has an absolute configuration (D or L) and can have a different ring size (furanose or pyranose). Additional groups (usually phosphate or sulphate) can also be attached at various positions around the ring (Figure 1.1).

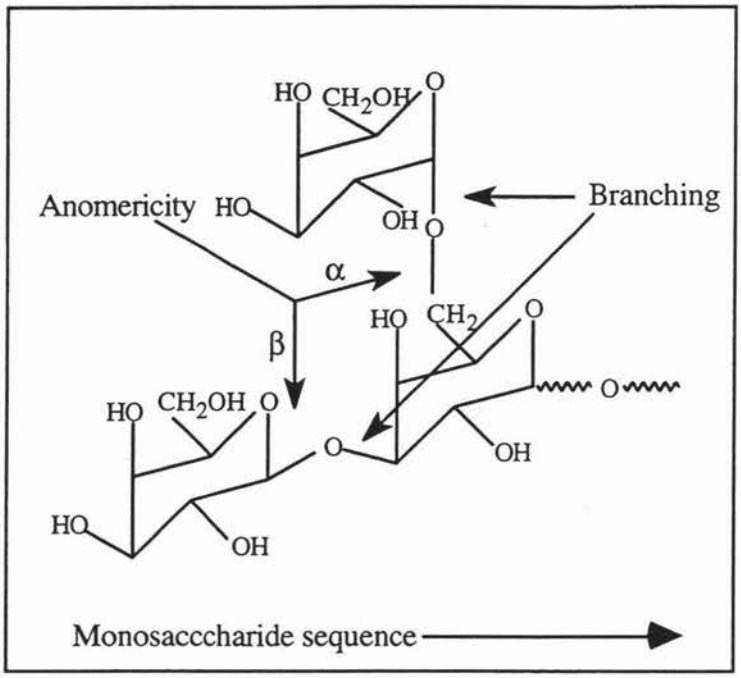
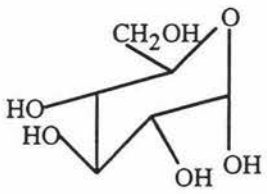


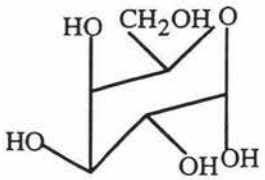
Figure 1.1 *Origins of structural diversity.*

Oligosaccharides are produced by the sequential addition of monosaccharide units to a core structure. Figure 1.2 illustrates the different types of monosaccharide units commonly found in glycoproteins.

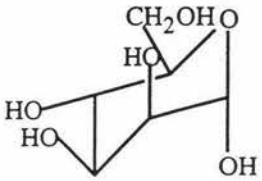
In general the oligosaccharides of glycoproteins can be classed into two groups by the way in which they are bound to the protein. N-linked oligosaccharides are linked via a chitobiose unit to the amide nitrogen of an asparagine residue. O-linked oligosaccharides are linked to a hydroxyl oxygen in the side chain of an amino acid, most frequently a serine or threonine residue. A third type of sugar chain is a component of the glycophosphatidylinositol (GPI) anchor, which links some cell surface proteins to the lipid bilayer.



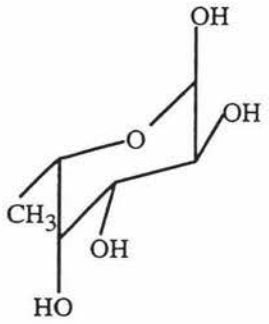
α-D-Glucose



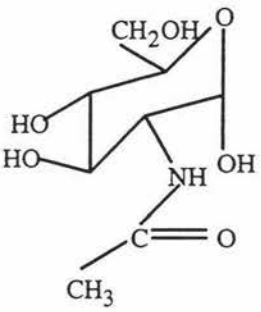
α-D-Galactose



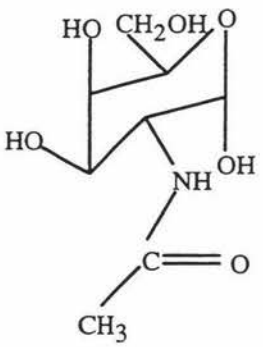
α-D-Mannose



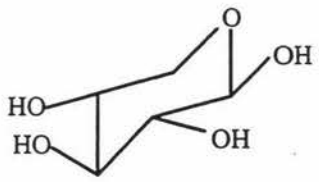
α-L-Fucose



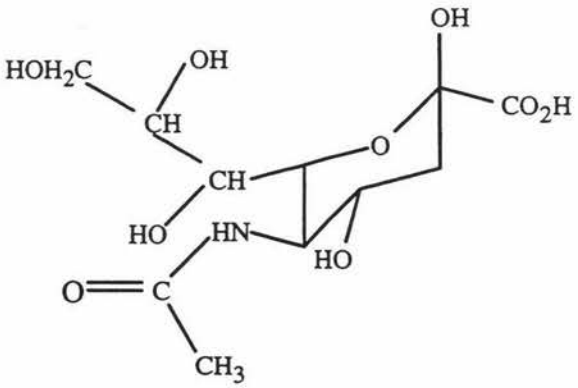
N-Acetyl-α-D-glucosamine



N-acetyl-α-D-galactosamine



β-D-Xylose



α-N-Acetylneuraminic acid

(Sialic acid)

Figure 1.2 *Structures of monosaccharides commonly found in eukaryotic glycoproteins.*

1.3 N-linked glycosylation

One of the most common types of eukaryotic protein modification is N-linked glycosylation. The biosynthesis of N-linked oligosaccharides begins with the synthesis of a lipid linked oligosaccharide moiety ($Glc_3Man_9GlcNAc_2PPDol$) that is transferred onto the nascent polypeptide chain in the rough endoplasmic reticulum (RER). The acceptor is the amino group of an asparagine side chain that is part of a tripeptide recognition sequence Asn-X-Ser/Thr. A series of trimming and elongation reactions is then catalysed by glucosidases, mannosidases and an array of glycosyltransferases (N-acetylglucosaminyl-, fucosyl-, galactosyl-, sialyltransferases) in the RER and the Golgi apparatus to produce a wide range of glycans that are specific to each cell type, tissue type and species. Glycoproteins which contain more than one glycan site can carry different N-glycan structures at each recognition sequence (Rademacher *et al.*, 1988).

1.3.1 Structure of the N-linked oligosaccharides

Structurally, the N-linked oligosaccharides can be classified into three sub groups:

(1) High mannose

(2) Complex

(3) Hybrid

All three classes contain an identical core pentasaccharide as shown below:

(1) High mannose type

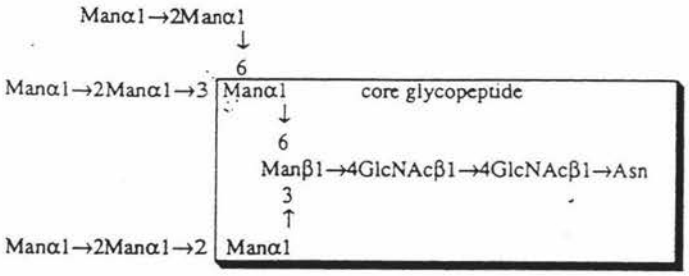


Figure 1.3 *High mannose type N-linked glycan (Mort and Pierce, 1995).*

Oligosaccharides that contain only mannose and N-acetylglucosamine residues are known as high mannose type sugar chains (Figure 1.3).

(2) Complex type

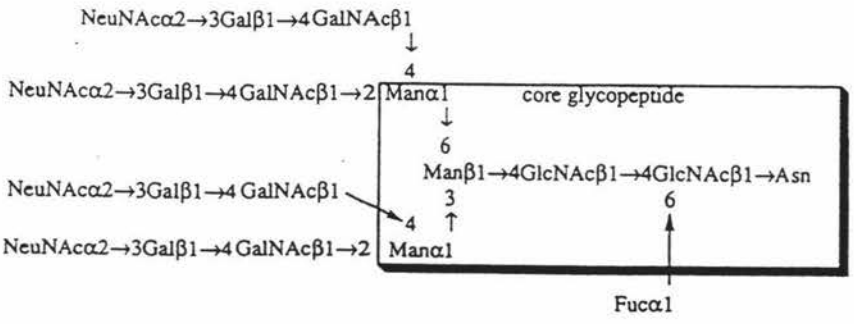


Figure 1.4 *Complex type N-linked glycan (Mort and Pierce, 1995).*

Complex-type sugar chains are formed when the oligosaccharide contains monosaccharides other than mannose and N-acetylglucosamine. Further structural variation is afforded by the presence or absence of fucose residues linked at the C-6 or C-3 positions of the proximal N-acetylglucosamine residue. Complex type oligosaccharides are known to be present in α-lactalbumin (Tilley *et al.*, 1991).

(3) Hybrid type

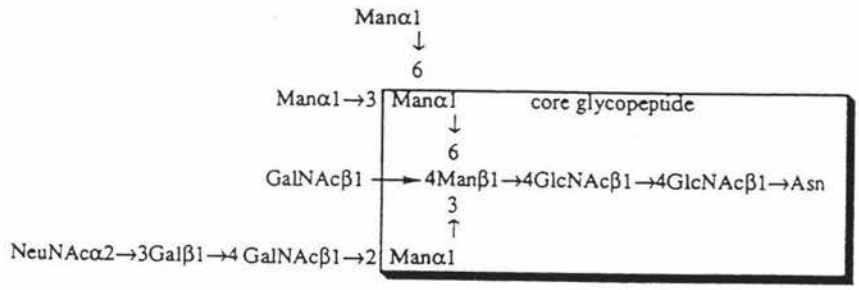


Figure 1.5 *Hybrid type N-linked glycan (Mort and Pierce, 1995).*

The hybrid type was found during a comprehensive study of the large sugar chains isolated from hen egg albumin (Kobata and Endo, 1993). As shown above, these sugar chains are classified as the hybrid type because they have structural features that are characteristic of both the high mannose type and complex type sugar chains.

1.3.2 Influence of local amino acids on core-glycosylation

N-linked glycosylation occurs only when the asparagine residue is part of the sequon Asn-X-Ser/Thr. There are some reports that amino acids near an asparagine residue are important determinants of core glycosylation efficiency. These include amino acids within the sequon itself (i.e: serine, threonine and X) as well as amino acids which flank the sequon.

(a) The influence of the hydroxy amino acids

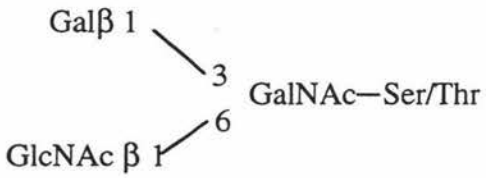
Studies using a variety of experimental approaches have shown that the hydroxy amino acid within the sequon can influence its potential glycosylation. Surveys of N-linked glycans have revealed that Asn-X-Thr sequons are 2-3 times more likely to be glycosylated than Asn-X-Ser sequons. These results were confirmed by experiments with a series of peptides that showed Asn-X-Thr sequons are more efficient oligosaccharide acceptors compared to Asn-X-Ser sequons (Kaplan *et al.*, 1987). A mechanism was proposed in which the hydroxy amino acid plays a direct, catalytic role in the glycosylation reaction through a series of hydrogen bond transfer reactions (Imperialli *et al.*, 1997). This involves an oligosaccharyltransferase, the hydroxy amino acid side chain, and the β amino group of asparagine. Interestingly it was also shown that the sulfhydril group of cysteine can support glycosylation at Asn-X-Cys sequons in peptides, although the oligosaccharide acceptor activity of such peptides is low (Shakin-Eshleman, 1996). Glycosylated Asn-X-Cys sequons have been observed in bovine and human protein C (Miletich and Broze, 1990), human von Willebrand factor (Titani *et al.*, 1986) and human α-lactalbumin (Giuffrida *et al.*, 1997). Only serine, glutamine, alanine and isoleucine have been observed at the X position in this sequon. The presence of proline at the X position seems to prevent glycosylation (Bause, 1983).

(b) The influence of the X amino acid

While there are instances of all the common amino acids apart from proline in position X of glycosylated sequons, some appear to be less favourable than others for glycosylation to occur. These amino acids include cysteine, tryptophan, aspartic acid and glutamic acid. Large hydrophobic amino acids at position X (e.g: tryptophan) also impair glycosylation, as do negatively charged amino acids (e.g: glutamic acid and aspartic acid). In contrast, small amino acids (e.g: glycine) or positively charged amino acids (e.g: lysine, arginine, histidine) in the X position seem to favour efficient glycosylation (Shakin-Eshleman, 1996).

1.4 O-linked glycosylation

Sugars may also be linked to proteins through a serine or threonine residue. In mammals, this link is commonly to an α-N-acetylgalactosamine (αGalNAc-Ser/Thr) residue and is known as a mucin type linkage. Mammalian O-linked glycoproteins have either linear chains formed by an extension of the oligosaccharide chain through a β1,3-linked galactose (Gal β1-3 GalNAcα-Ser/Thr) or branched structures by additional extension through β1,6-N-acetylglucosamine linked to the N-acetylgalactosamine as follows:



1.5 Functions of the oligosaccharide moiety

Many different functions have been attributed to the carbohydrate side chains that are covalently attached to the majority of eukaryotic membrane and secreted proteins (Varki, 1993, Lis and Sharon, 1993), all of which are not yet fully understood. The effect of carbohydrates on the physiochemical properties of proteins such as viscosity, isoelectric pH, the degree of hydration, solubility, thermal stability, and resistance to proteolysis, have been known for some time. Similarly, it has also been known that the oligosaccharide structures of glycoproteins are frequently different in various pathological states, including malignancy. Recently it has been shown that glycosylation has an important effect on protein folding and is involved in targeting within the cell. The first clear-cut demonstration of the functional significance of the carbohydrate moiety of a glycoprotein was in the blood group substances, where immunological specificity was found to be dependent on the structure of the oligosaccharide chains linked to the glycoproteins of the erythrocyte cell membrane. A major breakthrough in the search for functional significance of carbohydrates linked to proteins occurred when it was fortuitously discovered that the removal of sialic acid from the glycan

of proteins resulted in the rapid clearance of that glycoprotein from circulation. It was found that removal of sialic acid exposes the penultimate galactose residues at the non-reducing termini of the carbohydrate chains. These specifically bind to a Gal/GalNAc-binding receptor on hepatocytes and the glycoprotein becomes internalised by receptor-mediated endocytosis. Other specific carbohydrate-binding receptors or lectins for GlcNAc/Man and L-fucose were subsequently found to be located in the reticuloendothelial system. Carbohydrate-lectin interactions are also involved in lymphocyte migration, since treatment of lymphocytes with exoglysosidases or inhibitors of oligosaccharide-processing enzymes has been shown to alter their migratory or so called homing properties. Mammalian sperm and egg interaction has been shown to involve the binding of a sperm protein with a zona pellucida protein, ZP-3. In this case O-linked and not N-linked oligosaccharides on ZP-3 have been shown to be involved in binding (Paulson and Colley, 1989).

The variety of biological phenomena, such as the alteration of the carbohydrate structure of glycoproteins in cellular growth, development, and differentiation, and in various disease states, allude to the importance of the role that glycans play in biology. For some proteins these roles have been elucidated by studying the function of the protein in its glycosylated and deglycosylated states. For example when the glycan of the hormones human chorionic gonadotropin (hCG), lutropin (LH), and thyrotropin (TSH), is removed enzymatically, the receptor binding activity of the hormones is not affected but the ability of the deglycosylated hormones to stimulate cAMP and steroidogenesis is drastically reduced.

1. 6 Deglycosylation

1.6.1 PNGases

A number of endoglycosidases have been isolated that recognise and hydrolyse the N-linked glycans found on glycoproteins. Peptide-N-(N-acetyl-β-glucosaminyl)-asparagine amidase, or PNGase, selectively hydrolyses the β-asparatylglucosylamine bond of most known types of N-linked oligosaccharides as shown in Figure 1.6 (Tarentino *et al.*, 1985). The hydrolysis of this bond occurs in two steps, as proven by H^1 NMR and kinetic studies (Risley and Van Etten, 1985). In the first step, the carbohydrate is cleaved from the protein/peptide and the asparagine residue is converted to aspartic acid. The intermediate oligosaccharide retains the amino group. This intermediate then slowly, non-enzymatically degrades to an intact oligosaccharide and free ammonia (Risley and Van Etten, 1985, Tarentino *et al.*, 1982). Substrates for these enzymes include N-linked, high mannose, hybrid, and bi-, tri- and tetraantennary complex glycans (Tarentino *et al.*, 1982). It has been shown however, that PNGase F will not act on an oligosaccharide made up of less than two GlcNAc residues, or one that has a fucose linked 1-3 to the proximal GlcNAc (Chu, 1986). The asparagine has to be part of a peptide of at least five amino acids. Optimum deglycosylation occurs when the asparagine is near the middle of this peptide. Table 1.1 shows PNGases that have been isolated to date.

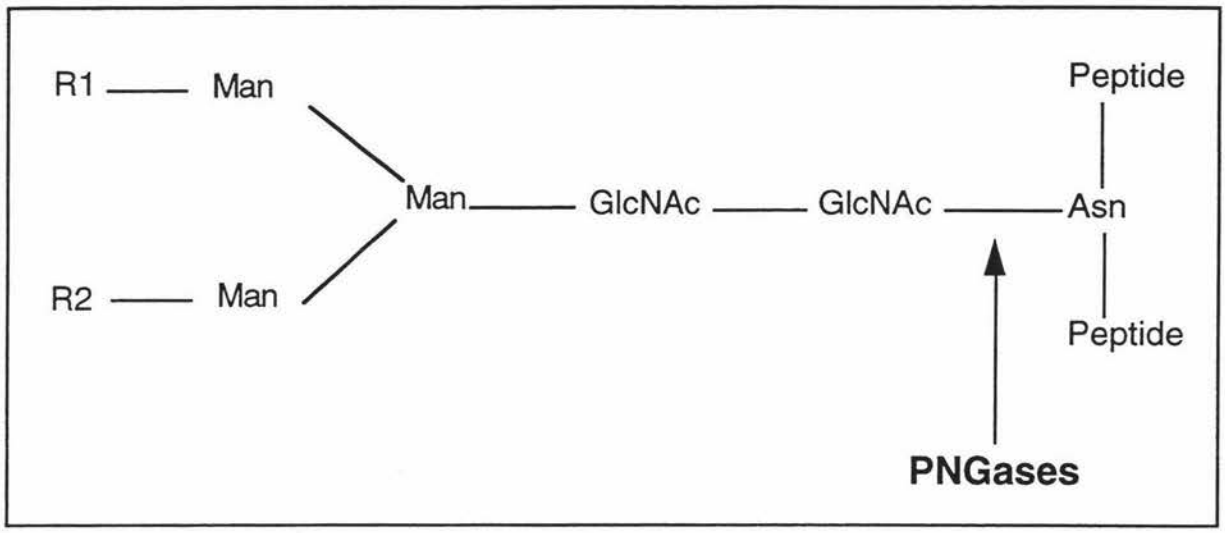


Figure 1.6 *The β-asparartylglucosylamine bond cleaved by PNGases.*

Table 1.1 *PNGases and their source.*

Enzyme	Source	Reference
PNGase A	Almond	Takahashi, 1977
PNGase F	Bacterium	Plummer *et al.*, 1984
PNGase L-929	Mouse fibroblast	Suzuki *et al.*, 1994
PNGase Se	White campion	Lhernould *et al.*, 1992
PNGase	Oryzias latipes	Seko *et al.*, 1991
PNGase J	Jack beans	Sugiyama *et al.*, 1983
PNGase P	Pea	Plummer *et al.*, 1987
PNGase R	Radish	Beger *et al.*, 1995
PNGase	Various mouse organs	Kitajima *et al.*, 1995
PNGase	Humans, chickens	Suzuki *et al.*, 1995
PNGase HO	Hen oviduct	Suzuki *et al.*, 1997
PNGase GM	Soybean	Kimura and Ohno, 1998

1.6.2 ENGases

A more specific enzyme that recognises only high mannose structures is endo-β-N-acetylglucosaminidase H (Endo H) which hydrolyses the linkage between the two N-acetylglucosamine (GlcNAc) residues in the chitobiose core (Tai *et al.*, 1977). This reaction cleaves the glucosidic bond in the chitobiose core, leaving a GlcNAc residue on the glycoprotein and liberating the oligosaccharide with one less reducing end GlcNAc. The site of cleavage catalysed by this enzyme is shown in Figure 1.7. A number of ENGases have been isolated from various sources as shown in Table 1.2.

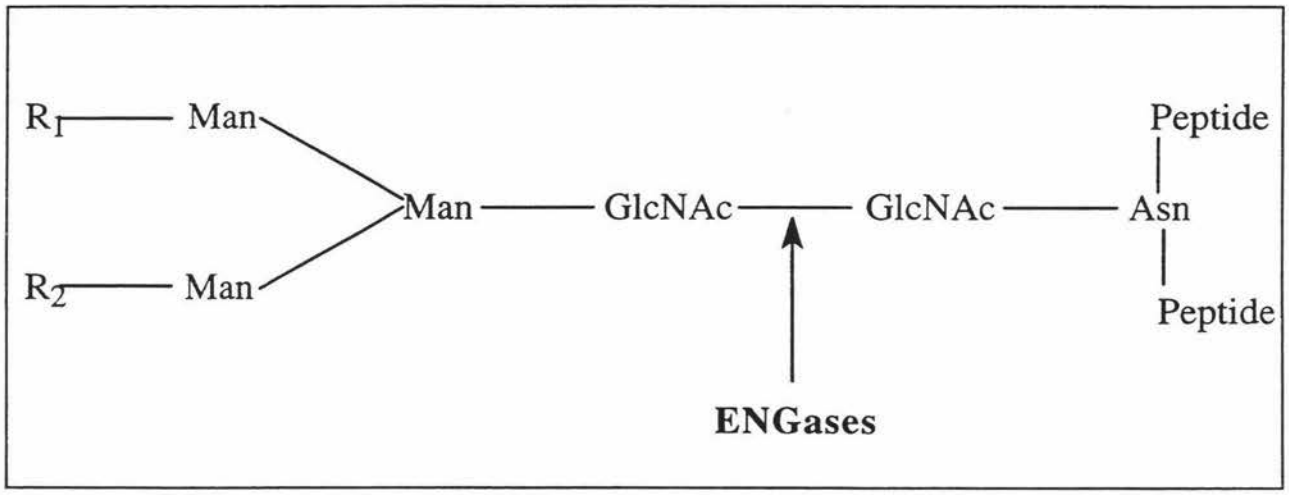


Figure 1.7 *Site cleaved by ENGases.*

Table 1.2 *Some ENGases and their source.*

Enzyme	Source	Reference
Endo B	Fungi	Bouquelet *et al.*, 1980
Endo C	Bacterium	Ito *et al.*, 1975
Endo D	Bacterium	Muramastu, 1971
Endo F1, F2, F3	Bacterium	Plummer *et al.*, 1984
Endo H	Bacterium	Tarentino and Maley, 1974
Endo L	Bacterium	Tarentino and Maley, 1974
Endo S	Fungi	Freeze and Etchinson, 1984

1.7 α-Lactalbumin

α-Lactalbumin, a globular whey protein consisting of 123 amino acid residues with a molecular weight of around 14 kDa, is produced in the lactating mammary gland. Three genetic variants (A, B, and C) are known to exist, although B is the predominant variant in bovine milk. The B variant differs from the A variant by a substitution of arginine for glutamic acid at position 10 (Gordon *et al.*, 1968). The results of investigations into α-lactalbumin have shown that it has an interesting role as a specificity regulator of lactose synthase (Brodbeck *et al.*, 1967). It has also been shown to have unexpected homology to the c-type lysozymes (Brew *et al.*, 1967). Although their functions are different both contain 51 homologous residues (including four disulphide bonds), and 24 additional residues with similar properties. It has been suggested that α-lactalbumin and lysozyme may have evolved from the same ancestral genes by a process of gene duplication and divergence. More recently, α-lactalbumin was shown to bind Ca^{2+} (Hiroaka *et al.*, 1980) and Zn^{2+} (Musci and Berliner, 1985), and is denatured at pH 3.5 due to the dissociation of these calcium ions from the protein (Desmet *et al.*, 1987). Later, it was shown that the α-lactalbumin molecule contained one tight binding site for Ca^{2+} and one weaker binding site for Zn^{2+} (Fitzgerald and Swaisgood, 1989).

1.7.1 Biological role of α-lactalbumin

The elucidation of the role played by α-lactalbumin in the synthesis of lactose represents a significant biochemical discovery. While studying the enzyme lactose synthase, Brodbeck and Ebner (1966) determined that it consisted of two components that were designated as "A-protein" and "B-protein". One year later, it was determined that α-lactalbumin was indeed the B protein and that it functioned in concert with the A protein as a "specifier" protein in the catalytic function of lactose synthase, possessing no catalytic function itself (Brodbeck and Ebner, 1966). Brew *et al.*, (1968) identified the A-protein as a galactosyltransferase.

In the absence of α-lactalbumin, galactosyltransferase transfers galactose from uridine diphosphate-galactose (UDP-galactose) to N-acetylglucosamine. In lactating mammary tissue, in the presence α-lactalbumin, the transfer of UDP-galactose is to glucose rather than to N-acetylglucosamine forming the milk sugar, lactose as shown in Figure 1.8. α-Lactalbumin has been found in milk of all species in which lactose is the principal sugar, and occurs in glycosylated and non-glycosylated forms.

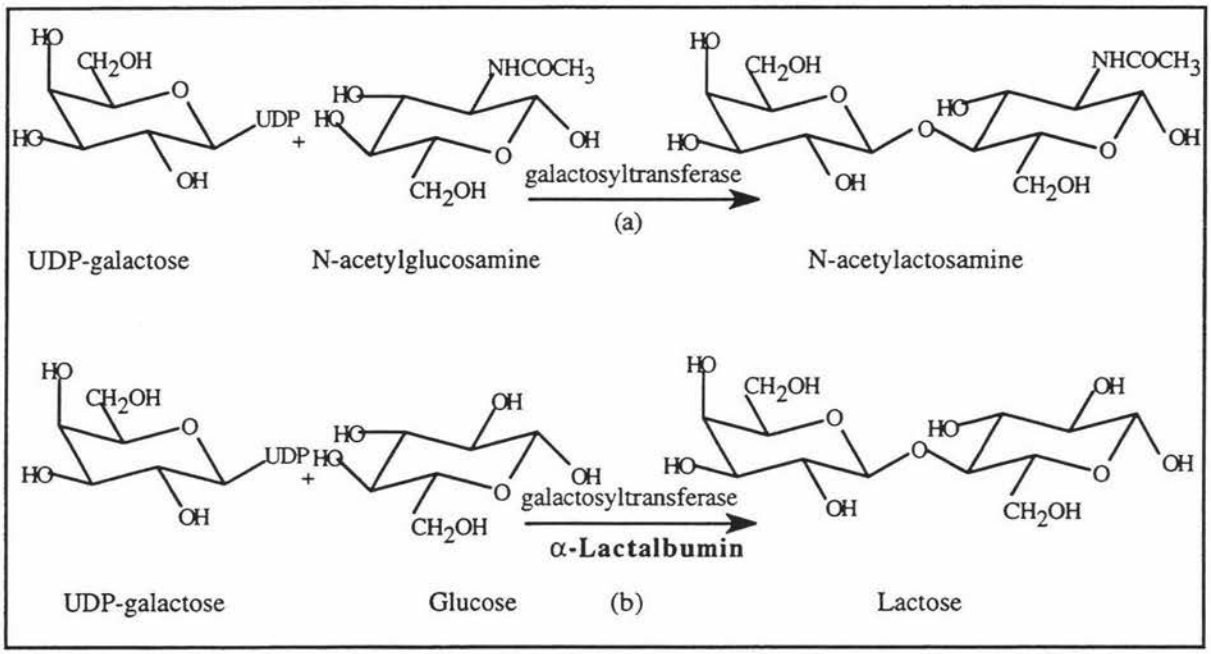


Figure 1.8 *Reactions catalyzed by galactosyltransferase. (a) In the absence of α-lactalbumin. (b) In the presence of α-lactalbumin.*

1.7.2 The importance of α-lactalbumin in infant nutrition

The complete primary structures of both the bovine and human forms of α-lactalbumin have been reported (Brew *et al.*, 1970, Findlay and Brew, 1972). Both consist of 123 amino acids that show 72% sequence homology. An additional 6% of the amino acid sequence are similar. Because of its amino acid composition α-lactalbumin is a remarkable protein. It has a very high chemical score and is rich in essential amino acids. The high concentration of α-

lactalbumin in human milk is largely responsible for its distinctive amino acid composition.

In the food industry the "humanising" of bovine milk for infant formulae has become an important process. Bovine milk has a different composition from human milk in that it has a lower content of α-lactalbumin, lactoferrin, lysozyme, and immunoglobulins (IgGs) (Hambraeus, 1977). In humans, α-lactalbumin comprises 28% of the total protein, whereas in

bovine milk α-lactalbumin represents only 4% of the total protein (Heine *et al.*, 1991). β-lactoglobulin is the major whey protein in bovine milk but it is present at only low levels in human milk. It has been claimed that β-lactoglobulin is one of the main causes of the allergy to bovine milk exhibited by some infants (Kuwata *et al.*, 1985). The aim of "humanising" milk is to achieve an adequate supply of essential and semi-essential amino acids but to limit the oversupply of non-essential amino acids. In order to produce infant formulae with a more similar composition to that of human milk, a proportion of β-lactoglobulin needs to be removed and α-lactalbumin needs to enriched.

1.7.3 The structure of α-lactalbumin

The primary structure of α-lactalbumin was reported by Brew *et al.*, (1970) (Figure 1.9), who drew scientists' attention to aspects of its structure that were similar to hen's egg-white lysozyme (HEWL). Direct X-ray crystallographic studies of the three dimensional structure of α-lactalbumin have been difficult for two reasons. First, it was difficult to find suitable crystals for data collection (Aschaffenburg *et al.*, 1972a,b, Fenna, 1982) and secondly, it was difficult to prepare heavy-atom derivatives necessary to solve the structure. Baboon (*Papio cynocephalus*) α-lactalbumin was found to give diffraction quality crystals and low resolution structures (6.0 Å and 4.5 Å) were reported (Smith *et al.*, 1987). However to obtain a useful isomorphous derivative, it proved necessary to chemically modify the protein by breaking the disulphide bond between cysteines 6 and 120 to provide a specific binding site for a mercury ion. X-ray crystallographic studies of baboon α-lactalbumin at high resolution revealed a

novel calcium binding loop and confirmed that in the native state, α-lactalbumin and c-type lysozyme have highly similar three dimensional structures (Stuart *et al.*, 1986., Acharya *et al.*, 1989). The Ca^{2+} seems to be important, as its removal results in conformational changes in the tertiary structure and may therefore affect the catalytic properties of the lactose synthase complex (Acharya *et al.*, 1991). It is proposed that Ca^{2+} binding may be important in controlling the release of α-lactalbumin from the Golgi membrane, a process necessary for the induction of lactation (Stuart *et al.*, 1986).

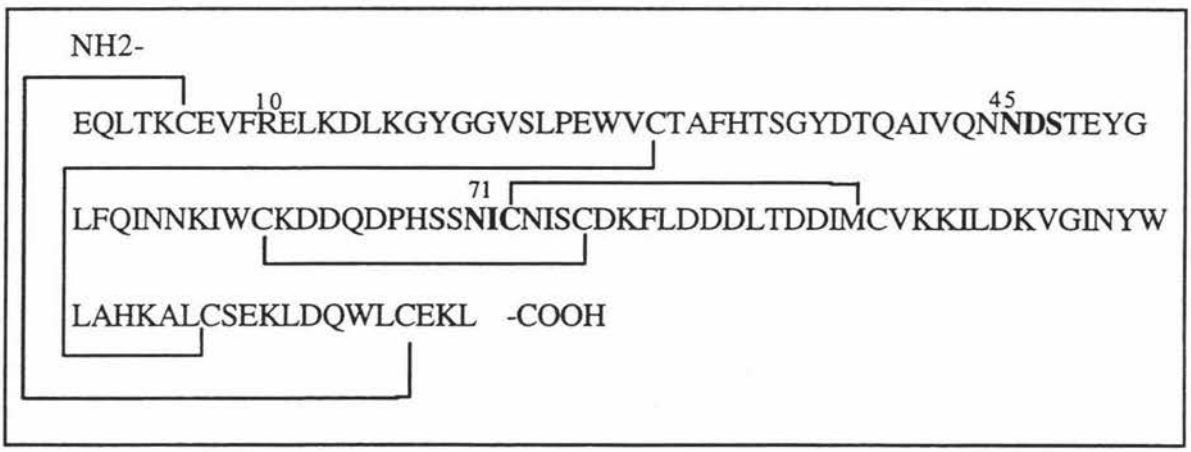


Figure 1.9 *Primary sequence of bovine α-lactalbmin. Intramolecular disulphide bridges are indicated between residues 6 - 120, 28 - 111, 61 - 77 and 73 - 91. Amino acids in bold represent the possible glycosylation sites.*

A comparison of the three dimensional crystal structures of human α-lactalbumin refined at 1.7 Å resolution with the baboon α-lactalbumin also refined at 1.7 Å (Acharya *et al.*, 1991, Acharya *et al.*, 1989) showed them to be very similar. The two proteins are known to have more than 90% amino acid sequence identity and crystallise in the same orthorhombic space group, $P2_12_12$.

In order to investigate the molecular basis of α-lactalbumin's function in lactose synthase, the three dimensional structures of guinea-pig, goat and a recombinant bovine α-lactalbumin were determined (Pike *et al.*, 1996). These structures were solved by molecular replacement techniques. Overall, the structures are very similar reflecting their high degree of amino acid sequence identity (66-94%). There is however, a part of the molecule, (residues 105-110) that does show a number of distinctly different conformations in the various structures. This region has been shown to be important in the function of the molecule and lies adjacent to two residues (Phe 31 and His 32) that are thought to be involved in stabilising the binding of the monosaccharide glucose to lactose synthase.

A comparative analysis of the structures showed that these different conformations significantly affected the enviroment of Phe 31 and His 32. It has been hypothesised that changes in the enviroment of the functional groups of these residues could affect the way in which they interact with the monosaccharide-lactose synthase complex. Interestingly, both residues are located in a region of the molecule that has relatively high thermal parameters. While this is indicative of disorder, it is also indicative of flexibility. So that each different conformation observed in each structure may represent a "snapshot" of one of many conformations adopted by the α-lactalbumin at any one time. Such flexibility may be important for the regulation of lactose synthase activity, an idea that is supported by the observation that metal ions have an enormous effect on the conformation of α-lactalbumin in solution (Pike *et al.*, 1996).

1.7.4 Glycosylation of α-lactalbumin

Minor amounts of bovine α-lactalbumin in milk have been isolated which are electrophoretically distinct from each other and have different carbohydrate contents. Some investigators obtained fractions that contained only hexosamine (Gordon *et al.*, 1968) whereas Barman, (1970) isolated a so-called glyco-α-lactalbumin from bovine milk using ion-exchange chromatography that contained mannose, fucose, galactose, galactosamine, and

sialic acid. Proctor and Wheelock, (1974) obtained at least two different carbohydrate containing fractions with varying amounts of the aforementioned monosaccharides. These species appear to have both the same specifier protein activity and the same amino acid sequence as the non-glycosylated form. However, the primary structures of the carbohydrate moieties have not yet been established. It has been suggested that all α-lactalbumins are glycosylated on secretion but are deglycosylated after leaving the secretary cells (Barman, 1970), but this suggestion has not been confirmed. The point of attachment of the carbohydrate group is probably at asparagine 45 (Tilly *et al.*, 1991) but this also remains to be confirmed. The sequence in this region of bovine α-lactalbumin is Asn-Gln-Ser, a glycosylation sequon (Marshall, 1972).

Studies on the glycosylation of the protein using *in vitro* systems are under active investigation. Although bovine α-lactalbumin contains two tripeptide sequences of -Asn_{45}-Asp_{46}-Ser_{47}- and Asn_{74}-Ile_{75}-Ser_{76}- glycosylation seems to only occur at aspargine 45 under *in vitro* conditions (Struck *et al.*, 1978). There is also a Asn_{71}-Ile_{72}-Cys_{73} sequence, although there have been no reports confirming the presence of glycans at this site.

Bovine, sheep, goat, rabbit, human, rat and wallaby α-lactalbumin all have a consensus N-glycosylation sequence (Asn-X-Ser) at residues 45-47. In bovine (and probably in other ruminants) a small proportion (around 15%) of molecules are glycosylated, whereas rabbit and rat α-lactalbumins appear to be uniformly glycosylated at this site (Prasad *et al.*, 1980). This difference has been attributed to the Asn-Gly-Ser sequence at this site in the rabbit and rat being a more effective substrate for N-glycosylation than the corresponding ruminant sequence Asn-Asp-Ser (Prasad *et al.*, 1982). Recently it was reported that asparagine 71 was glycosylated in human α-lactalbumin. This is an Asn-Ile-Cys sequon and only a relatively small proportion (about 1%) of the protein is glycosylated at this site (Guiffrida *et al.*, 1997).

Bovine α-lactalbumin possesses two main genetic variants, A and B, differentiated by one amino acid. The variant normally present in European dairy herds is variant B (Lyster, 1972). When these variants were studied using ion-exchange chromatography both variants behaved in a similar way except that the α-lactalbumin from variant A eluted at a higher salt concentration than variant B (Hopper and McKenzie, 1973). Homozygous genetic variants of bovine α-lactalbumin (i.e: AA or BB) were shown to contain a major component (component M) and three minor components (components F, S_1 and S_2) when analysed by gel electrophoresis. AB heterozygotes consist of two major and six minor components. F, M, S_1 and S_2 are named because of their position on starch gels after electrophoresis at pH 7.7 as shown in Figure 1.10 and Table 1.3. Only the S_1 and S_2 components have been shown to be glycosylated (Hopper and McKenzie, 1973). Analysis showed that S_1 contained 6 to 7 hexosamines, 6 hexoses, 1 deoxyhexose, and sialic acid. The S_2 protein had an identical carbohydrate composition except that it contained no sialic acids.

Hopper and McKenzie (1973), concluded that component F does not contain any carbohydrate residues, but suggested that it differs from component M because of an additional amide group, which accounts for its greater mobility on PAGE gels. Component F makes up approximately 5% of the total α-lactalbumin obtained from whole milk while S_1 and S_2 make up 3 to 5% of the total α-lactalbumin isolated from whey with component M making up the rest.

Starch gel electrophoresis

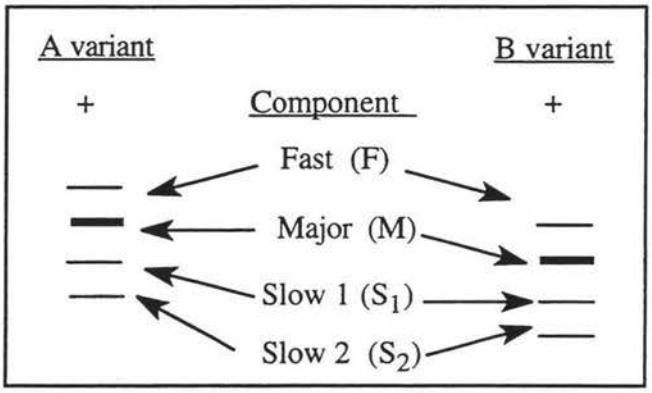


Figure 1.10 *Summary of the electrophoretic behaviour and designation of the components of bovine α-lactalbumins from A and B variants (adapted from Hopper and McKenzie, 1973).*

A recent abstract gave the structure of one of these glycoforms but did not detail the conditions used or indicate which particular glycoform (S_1 or S_2) was sequenced, it was shown to have the following structure (Tilly *et al.*, 1991) (Figure 1.11).

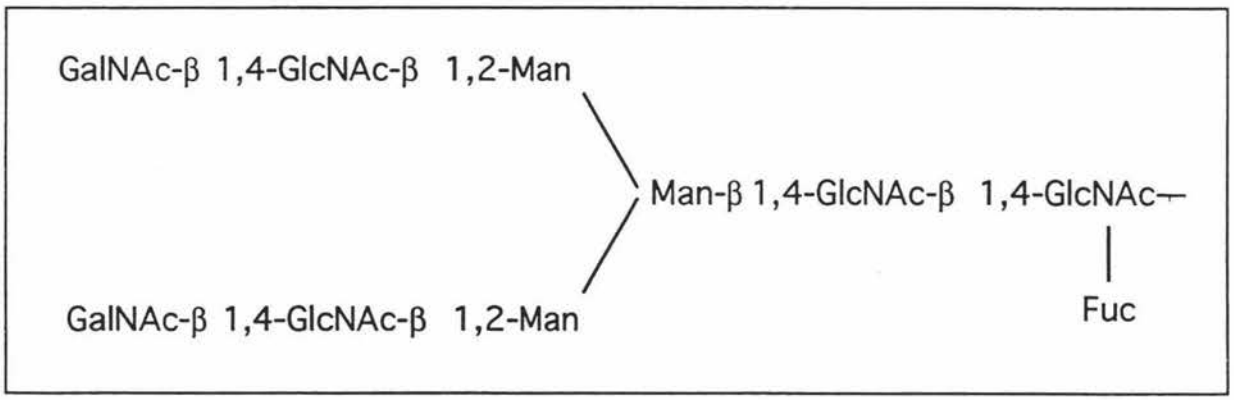


Figure 1.11 *Structure of one of the bovine α-lactalbumin glycans.*

There is no report as yet of the structure of a second glycoform that was found. Because it runs faster on native polyacrylamide gels, it is believed to contain a sialic acid residue. Glycoforms containing sialic acid should run faste, as sialic acid contains a negative charge.

Recent developments in the field of glycobiology suggest that the glycan of α-lactalbumin could have a variety of functions. Because of the current concern in improving human milk substitutes, the determination of the functional role of the glycan on α-lactalbumin is of interest. However the elucidation of the oligosaccharide structures must first be more thoroughly investigated to see whether or not the different glycan structures have different effects on protein function.

1.7.5 Incomplete glycosylation of bovine α-lactalbumin

It is unclear why all of the bovine α-lactalbumin in milk is not glycosylated at any of these possible glycosylation sequons since it has been proposed (Marshall, 1972) that glycosylation is a requirement for secretion of glycoproteins. Several possible reasons have been proposed:

1. All α-lactalbumins are glycosylated on secretion but are deglycosylated by endogenous ENGases or PNGases. Deglycosylation by ENGases would leave one GlcNAc residue attached to the glycosylated Asn.

2. The rate of synthesis, folding and secretion of α-lactalbumin could be sufficiently rapid to prevent glycosylation. Indeed, some other secreted proteins are also found in both glycosylated and non-glycosylated forms, for example, ribonuclease (Catley, 1973). Interestingly glycosylation has been shown to regulate the activity of ribonuclease.

3. The sequence is poorly glycosylated because of physical reasons due to the sequence near the vicinity of the sequon, such as:
   (a) The presence of aspartic acid at the X position.
   (b) The proximity of disulphide bonds.

1.8 Aims of current research

1.8.1 Objective 1

Deglycosylation of S_1 and S_2 using crude PNGase F (prepared in house) produced two bands in native gels. The major fraction ran to the same position as the F component of α-lactalbumin, while a faint band ran the same as the M component. This suggested that in the F protein, asparagine was replaced by aspartic acid as a consequence of the action of the deglycosylating enzyme PNGase. An ENGase, on the other hand, would leave a single N-acetylglucosamine on the protein. The presence of this single sugar inhibits further PNGase F digestion which requires a minimum of two monosaccharide units attached to a peptide for cleavage to occur. There is therefore the interesting possibility that the M component of α-lactalbumin may contain a single hexosamine attached to asparagine 45. Asparagine 71 and 74 sites are not known to be glycosylated under *in vitro* conditions (Struck *et al.*, 1978). Therefore the first objective of this current research was to analyse the M component of α-lactalbumin to investigate whether the main fraction contains polypeptide chains with one hexosamine per protein molecule attached to asparagine 45 or not.

1.8.2 Objective 2

Recent work carried out at Massey University, indicated that the S_1 and S_2 fractions actually represented families of glycoforms, not just two different glycoforms. Mass spectrometry and gel electrophoresis showed that these may be families of glycoforms with and without sialic acid (Chilcott, 1996). The second objective was designed to take the work one step further and analyse the glycosylated α-lactalbumin:

1. To confirm the number of glycans linked to glycosylated α-lactalbumin.

2. To separate and analyse the different oligosaccharide chains bound to α-lactalbumin.

1.8.3 Objective 3

The F fraction is to be sequenced to see whether it contains an aspartic acid residue at the asparagine 45 glycosylation site. This is to investigate whether it has been deglycosylated by endogenous PNGases. This requires a modified methodology due to the low concentration of the F fraction present in bovine milk.