How sweet are our gut beneficial microbes:

Protein glycosylation in *Lactobacillus reuteri*

Dimitrios Latousakis

A thesis submitted for the degree of Doctor of Philosophy (PhD)

To the University of East Anglia

Quadram Institute Bioscience

Gut Health and Food Safety

Norwich Research Park

Colney, Norwich

NR4 7UA

September 2017

Abstract

Protein glycosylation is a well-established post translational modification occurring in all forms of life. In the past two decades, protein glycosylation has been extensively studied in bacterial pathogens underscoring its importance in virulence and colonisation. However, despite the wealth of information regarding protein glycosylation in bacterial pathogens, little is known about this process in gut commensal bacteria. The gut microbiota has co-evolved with and is largely adapted to its host, leading to mutually beneficial interactions. These interactions often involve the adhesion of the bacteria to the gastrointestinal tract using specialised adhesins. *Lactobacillus reuteri* is a common gut symbiont found in a wide range of vertebrate hosts. As such, it is used as a model organism to study the co-evolution between gut commensal bacteria and their hosts. Here, we used *L. reuteri* 100-23 (rodent isolate), ATCC 53608 (pig isolate) and MM4-1a (human isolate) to study the glycosylation of proteins in gut commensal bacteria. An initial bioinformatics approach to identify putative glycosylation clusters suggested the presence of two putative O-glycosylation systems (*gtf1/gtf2* and $secA_2/Y_2$) that could be involved in post translational modification of proteins in *L. reuteri* strains. Further genetic and biochemical analyses suggested that Gtf1 is involved in a general glycosylation system targeting multiple proteins in *L. reuteri*, whereas proteins encoded form the $secA_2Y_2$ cluster are dedicated to the glycosylation of a serine rich repeat protein (SRRP). Lectin screening of the secreted proteins from *L. reuteri* combined with mass spectrometry analysis identified the mucus binding protein MUB_{53608}, $SRRP_{53608}$ and $SRRP_{100\text{-}23}$, the major adhesins of *L. reuteri* ATCC 53608 and 100-23, respectively, as glycoproteins. MUB_{53608} is a putative target of Gtf1-mediated glycosylation. The glycans present on this adhesin were biochemically characterised and found to carry α-galactose and galactofuranose epitopes that could be involved in interactions with the host immune system. The glycosylation of the SRRPs was studied by mass spectrometry and biochemical assays. $SRRP_{53608}$ harboured GlaNAcα-GlcNAc moieties, whereas $SRRP_{100\text{-}23}$ was found to be glycosylated with Glc-Glc-GlcNAc trisaccharides. The data produced in this work provided novel insights into the *L. reuteri* glycosylation systems, the nature of glycoproteins and the structure of their glycans, furthering our understanding of the underpinning mechanisms behind their beneficial interactions with the host.

Acknowledgements

This work was funded by IFR Extra and none of this work would have been possible without their contribution.

Starting my PhD, I knew I was stepping very far outside my comfort zone – finishing my PhD, and looking back, I am really excited I made that choice. The journey through this period was not easy, but fortunately there were people around me willing to help, every step of the way. First and foremost, I would like to thank Dr. Nathalie Juge, my supervisor, for the constant support, help and guidance throughout my PhD. I am really grateful for all the time and effort Nathalie put on my project and, most importantly, on me. A big "thank you" also goes to Professor Rob Field, for all the advice, ideas and useful input he offered. Special thanks are due to Dr. Donald Mackenzie; Don trained me, guided me and advised me on an every-day basis when I started, and his help was invaluable for the rest of my PhD. Many thanks also to Dr. Norihito Kawasaki for his discussions and suggestions on my work.

Many thanks are due to Professor Arjan Narbad and Professor Sabine Flitsch, who agreed to discuss and examine my work.

I also want to thank the members of the Juge group with whom I shared this experience. They were always helpful, willing to discuss and make suggestions and they made me feel happy being in the lab; especially Dr. Devon Kavanaugh, with whom we had endless discussions on my project. Many thanks are also due to the scientists in the QIB, who trained me or contributed to this project with their expertise and especially to Dr. Patrick Gunning for the AFM experiments and Dr. Alfonsina D'Amato and Dr. Francis Mulholland for the proteomics work.

I would also like to thank my friends around the NRP and back home for their support and encouragement throughout this period of my life.

My deepest gratitude and appreciation goes to my family and especially to my parents, for their tireless support and faith in me, throughout my PhD, and life in general. Last, but definitely not least, I owe sincere gratitude to Eleftheria for putting up with me through all the difficult times and sharing with me all the happy ones.

Table of Contents

Table of figures

List of Tables

Table 1 Table of abbreviations used in this thesis.

Abbreviation	Full name
2'FL	2'-fucosyl-lactose
3'FL	3'-fucosyl-lactose
6'SL	6'-sialyl-lactose
Ab	Antibody
ACN	Acetonitrile
AFM	Atomic force microscopy
AIM	Auto induction medium
Amp	Ampicillin
AP	Alkaline phosphatase
Ara	Arabinose
Ara_f	arabinofuranose
aSecA	Accessory secretion system
Asp	Aspartic acid
Asp1-5	Accessory secretion protein 1-5
asp2	Gene encoding Asp2
AST	Alanine-serine-threonine rich domain
BR	Binding (or basic region)
CAZyme	Carbohydrate active enzyme
Cb	Carbenicillin
CHAPS	(3-[(3-cholamidopropyl) dimethylammonio]1-propanesulfonate)
CM	Cell membrane
Cm	Chloramphenicol
ConA	Concanavalin A
CPS	Capsular polysaccharide
Cys	Cysteine
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
DHB	2,5-dihydroxy-benzoic acid
DMSO	Dimethylsulfoxide
DSF	Differential scanning fluorimetry
DSS	Dextran sodium sulfate
dTDP	Deoxythymidine diphosphate
DTT	Dithiothreitol
DUF	Domain of unknown function
ECM	Extracellular matrix
EF-Tu	Elongation factor Tu
EPEC	Enteropathogenic *E. coli*
EPS	Exopolysaccharide
ER	Endoplasmic reticulum
Ery	Erythromycin
FbpA	Fibrinogen binding protein A
FITC	Fluorescein isothiocyanate
FITC- (*f*-) lectin	Fluorescein isothiocyanate-labelled lectin
Ftf	Fructosyltransferase
ftf	Gene encoding Ftf
Fuc	Glucose
Gal	Galactose

Gal_f	Galactofuranose
GalNAc	N-acetyl-galactosamine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC-MS	Gas chromatography – mass spectrometry
gDNA	Genomic DNA
GF	Germ-free
GH	Glycoside hydrolase
GI	Gastrointestinal
Glc	Glucose
GlcNAc	N-acetyl-glucosamine
Gln	Glutamine
GluC	Endoproteinase C
GPC	Gel permeation chromatography
GSL-I B4	*Griffonia simplicifolia* Lectin I isolectin B4
GT	Glycosyltransferase
Gtf1	Glycosyltransferase 1
gtf1	Gene encoding Gtf1
GtfA	Glycosyltransferase A
gtfA	Gene encoding GtfA
GtfB	Glycosyltransferase B
gtfB	Gene encoding GtfB
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hex	Hexose
HexNAc	N-acetyl-hexosamine
h-Int1	Human intelectin 1
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
IEC	Intestinal epithelial cells
IMAC	Immobilised metal ion affinity chromatography
IMO	Iso-maltooligosaccharides
Kan	Kanamycin
Lac	Lactose
LB	Luria Bertani
LCR	Low complexity region
LDM II	Lactobacillus defined medium II
LDS	Lithium dodecyl sylphate
Leg	Legionaminic acid
LPS	Lipopolysaccharide
*Lr*100-23	*Lactobacillus reuteri* 100-23
*Lr*ATCC 53608	*Lactobacillus reuteri* ATCC 53608
*Lr*MM4-1a	*Lactobacillus reuteri* MM4-1a
LTA	Lipoteichoic acid
LTL	*Lotus tetragonolobus* lectin
MAL I	*Maackia amurensis* lectin I
MALDI-ToF	Matrix assisted laser desorption/ionisation – time of flight
Man	Mannose
ManNAc	N-acetyl-mannosamine
MPL	*Maclura pomifera* lectin
MRS	deMan-Rogosa-Sharpe
MUB	Mucus binding protein

MucBP	Mucus binding protein domain
MurNAc	N-acetyl-muramic acid
MW	Molecular weight
Neu5Ac	N-acetyl-neuraminic acid
Neu5Gc	N-glycolyl-neuraminic acid
OD_{600}	Optical density at 600 nm
O-Ac	O-acetyl
O-OTase	O-oligosaccharyl-transferase
OST	Oligosaccharyl-transferase
PAS	Periodic acid/Schiff stain
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline – Tween-20
PCR	Polymerase chain reaction
PG	Peptidoglycan
PNA	Peanut agglutinin
Pse	Pseudaminic acid
PTM	Post translational modification
PTS	Phosphostransferase system
PUL	Polysaccharide utilization *loci*
RCA	*Ricinus communis* agglutinin
Rha	Rhamnose
SCFA	Short chain fatty acids
SecA2	Accessory secretion system protein A
secA2	Gene encoding SecA2
Ser	Serine
SlpA	Surface layer protein A
SM	Spent media
SNA	*Sambucus nigra* agglutinin
S.O.C.	Super optimal broth with catabolite repression
SPase	Signal peptidase
srr	Gene encoding SRRP
SRRP	Serine rich repeat protein
Suc	Sucrose
TFA	Trifluoroacetic acid
Thr	Threonine
Tm	Melting temperature
TMB	3,3',5,5'-Tetramethylbenzidine
$tMUB_{1063N}$	Truncated MUB_{53608} from *L. reuteri* 1063N
T_{reg}	Regulatory T-cells
Trp	Tryptophan
TTP	Triple phase partitioning
Tyr	Tyrosine
UDP	Uridine diphosphate
UEA	*Ulex europaeus* agglutinin
WB	Western blot
WT	Wild type
WTA	Wall teichoic acid
Xyl	Xylose

Chapter 1

Introduction

1.1. Gut Microbiota

1.1.1. Occurrence in the gastrointestinal tract

Microbes can be found on any surface of the human body, such as the skin, the oral cavity and the respiratory tract, as well as the gastrointestinal (GI) tract and the urogenital system. The largest population of microorganisms, however, is found in the GI tract, which harbours hundreds of trillions of microbes per gram of tissue, collectively referred to as the *gut microbiota*. An uneven distribution of the microbial population is observed along the GI tract; while a few thousands cells per gram of tissue survive the highly acidic conditions of the stomach, more than 10^{11} cells per gram of tissue reside in the colon (Figure 1) (1).

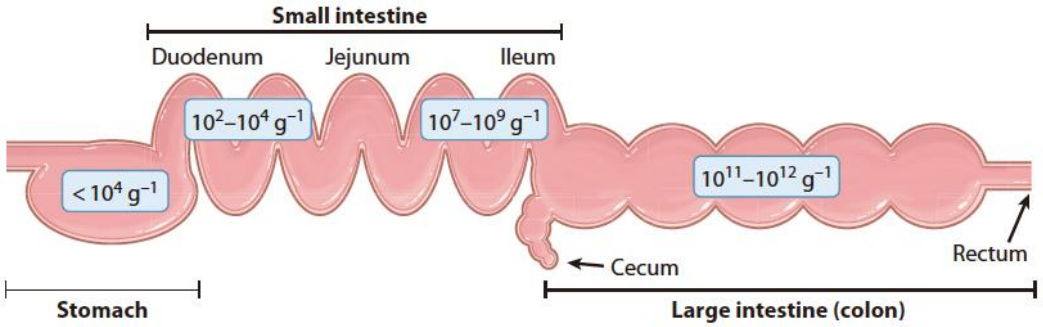

Figure 1. Distribution of the gut microbiota throughout the GI tract (adapted from (1)).

The gut microbiota is comprised mainly of bacteria, but archaea, Eukaryotes and viruses are also part of the microbial community (2). The composition of the gut microbiota is highly dynamic and depends on many environmental and genetic factors, thus it differs greatly between individuals, as well as within a single host.

The colonisation of the GI tract begins at birth - some studies suggest even earlier (3) - and changes throughout life, with different genera prevailing at different stages [reviewed in (4)]. The great majority of bacterial gut symbionts found in healthy humans fall into the phyla of *Firmicutes* and *Bacteroidetes*, and a minor proportion of the microbiota consists of *Actinobacteria*, *Proteobacteria* and other phyla (2,5).

While related individuals share some similarity in terms of microbiota composition (6), a core microbiota, *i.e.* a group of species that are common across the population, is considered unlikely to exist. In contrast, common functions between different individuals' microbiota have been identified and this has led to the hypothesis of a core microbiome, which includes genes that perform conserved functions across all microbiota (7). This also implies that functional redundancies exist between different sets of species (8), favouring the phylogenetic diversity of the microbiota between and within individuals. Phenotypic diversity is considered to be an advantage in evolution, as it provides organisms with various traits which may benefit them under environmental challenges, such as in the gut environment.

1.1.2. Factors affecting the microbiota composition

1.1.2.1. Mode of delivery

The initial colonisation of the GI tract plays a pivotal role in shaping the future composition of the microbiota (5). The method of delivery greatly affects the initial microbiota composition; children born pre-termly or by cesarean section are exposed mainly to nosocomial microbes, such as *Clostridia, Staphylococci* and *Streptococci*, as well as skin related species. In addition, Jacobsson et al. (2014) showed that children delivered by C-section had a decreased microbiota diversity and that colonization of the gut by members of the *Bacteroidetes* phylum, one of the predominant phyla in adult microbiota, was significantly delayed (9). In contrast, naturally born babies, as well as those that did not require the administration of antibiotics showed high numbers of Lactobacilli, a species that dominates the vaginal flora, as well as increased numbers of *Bifidobacteria* and *Bacteroides* (5,10).

1.1.2.2. Diet

Diet is one of the primary driving forces that shapes the microbiota composition. After birth, the environment of the gut favours the colonisation of facultative anaerobic bacteria, and bacteria that carry genes for lactose utilization, which agrees with a diet enriched in milk (4,5). In addition

to lactose, breast milk contains short oligosaccharides that act as prebiotics to promote colonization by specific species, such as *Bifidobacteria* and *Bacteroidetes*, which preferentially utilize these carbohydrates (11–14). In contrast, formula-fed children are colonized by *Bacteroides fragilis, Clostridium coccoides* and Lactobacilli (12) and show greater microbial diversity (10,15). In addition to providing nutrients for early colonisers, breast milk also carries a plethora of bacterial species that babies get exposed to during lactation; the composition of this group of bacteria changes throughout the lactation period (16,17). Once the oxygen is all consumed, strictly anaerobic bacteria take the opportunity to colonise the large intestine (4) and by the age of one year old, and after the introduction of solid food, bacteria that can digest dietary carbohydrates emerge and the microbiota diversity increases (5,18). At the age of 2-3 years old, the gut microbiota matures and resembles that of an adult (18,19).

In an adult-like microbiota, the spatial distribution of bacteria differs, not only in the number of microbial cells found in each part of the GI tract, but also in the type of bacteria. For instance, the small intestine harbours facultatively anaerobic bacteria that can adhere onto the host tissue and feed on simple or short sugars (20,21). In contrast, strictly anaerobic bacteria that exploit complex and -otherwise- indigestible carbohydrates live in the large intestine (20).

Even though an adult-like microbiota is considered relatively stable, it is very dynamic and responds to changes in environmental factors such as the diet. A common trend is that a diet rich in dietary fibres favours the colonisation and growth of *Bacteroidetes*, whereas a diet rich in fat promotes the growth of *Firmicutes* (22,23). In addition, high-fat diets promote the growth of bacteria that are bile-acid tolerant, like *Alistipes sp.* and *Bilophila sp*, as bile acids are bactericidal for many species (24,25).

However, it is difficult to establish a clear connection between diet and specific bacterial species, as most studies report changes at the phylum level. For example, a study by Wu et al. (2011) showed that *Bacteroidetes* and *Actinobacteria* positively correlated with a high-fat diet, whereas

Firmicutes and *Proteobacteria* benefited from a fibre-rich diet (26). Similarly, murine microbiota from mice on a low-fat diet showed higher abundance of *Firmicutes*, in contrast to those on high-fat diet that showed increased numbers of *Bacteroidetes* (27). Such differences were also reported using specific oligosaccharides, such as iso-maltooligosaccharides (IMOs) which have been shown to either benefit *Bifidobacteria* in rats (28,29), or lead to a decrease in their population (29). In addition, different dietary fibres seem to benefit different species. For instance, inulin promotes growth of *Bifidobacteria* (28,30) whereas IMOs favour the growth of Lactobacilli (28). Interestingly, *Lactobacillus* species that are adapted to the GI tract can utilize a wider range of carbohydrates compared to those isolated from food sources (31).

It is apparent that the way various bacterial species respond to changes in diet is inconsistent between individuals. As there is functional redundancy between different microbiota and even different species, it is likely that a specific diet will select for a microbiome rather than a microbiota. For example, a fibre-rich diet will select for genes that can digest complex carbohydrates rather than specific bacterial species. Such genes can be found in members of the *Bacteroidetes* or the *Firmicutes* phyla (19). This could account for the differences observed between studies, where identification of the different bacteria reaches down to the species level, but does not inform about the genetic repertoire of the strains.

1.1.2.3. Host genetics

It has been proposed that the genetic profile of a host plays a role in shaping the gut microbiota. This arose from studies showing that related individuals have more similar microbiota than unrelated ones (7,32,33). Host genes encoding proteins involved in bile acid metabolism (34), antimicrobial peptide production (35), metabolism and mucin glycosylation have been shown to affect the microbiota composition (Figure 2). Most of our understanding on how specific genes affect the microbiota comes from studies investigating the role of single genes. A few studies

have linked genes associated with metabolism (e.g. Apo-E1 and leptin) with differences in the microbial community (36–38).

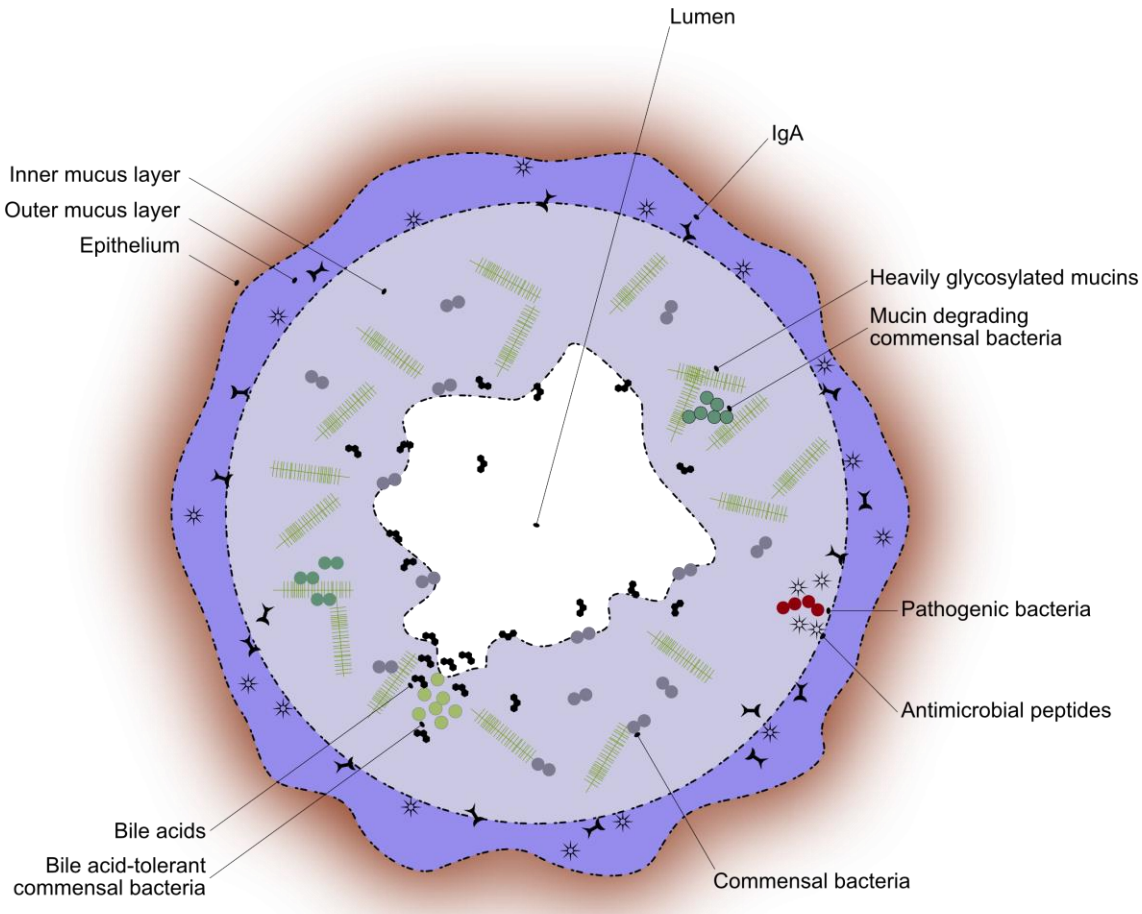

Figure 2 Host genetics affect the microbiota composition by mucin production, to provide binding sites and nutrients to the commensal bacteria, by regulating the immune system to tolerate or attack commensal or pathogenic bacteria, respectively, or by secretion of bile acids, which promote the survival of bile-acid tolerant bacteria (adapted from (39)).

The mucus layer is a gelatinous layer that covers the GI tract and is comprised mainly of mucins, large glycoproteins in which the glycans account for more than 80% of their mass. Mucin O-glycosylation (see section 1.4.1.1) occurs in the Golgi apparatus where mucins are modified by carbohydrate chains that consist of N-acetyl-galactosamine (GalNAc), N-acetyl-glucosamine (GlcNAc) and galactose (Gal), and are often capped by sialic acid and/or fucose (Fuc) (40). In early life, the profile of the capping monosaccharides is the same in the mucus across the GI tract, but later in life an increasing gradient of sialic acid and decreasing gradient of Fuc is

observed from the small intestine to the colon in humans (29). Mucin glycosylation can affect bacterial colonization by either selectively providing adhesion sites to bacteria expressing the right repertoire of adhesins (40) or by providing carbohydrates as nutrients that only certain bacteria can utilize (41). It is well documented that mouse models with altered or deficient mucin glycosylation show changes in the microbiota composition (42). For example, deletion of FUT2, a fucosyl-transferase responsible for capping mucin glycans with Fuc, results in lower microbial diversity and altered microbiota composition in mice (43). An earlier study with Finnish individuals showed a link between FUT2, the presence of ABO antigens in the mucus and the microbiota composition, with differences in the populations of *Bifidobacteria*, *Bacteroides* and *Clostridia* between individuals with different antigens (44).

The increased sialic acid content found in the large intestine also contributes in shaping the microbiota. As sialic acid protects the mucin carbohydrates, it requires specific bacteria expressing sialidases to release it and expose the underlying glycans. Upon release, other bacteria can either use free sialic acid or the rest of the glycan to feed (45,46). Recently, a new type of sialidase was discovered in strains of the gut commensal *Ruminococcus gnavus*, which released a modified sialic acid that the bacteria can preferentially metabolise, providing an ecological advantage to this species (41). A similar synergy between bacterial species is observed in cases where bacteria rely on degradation or metabolic products of other species to survive, as for example with species that utilize short chain fatty acids (SCFAs) as carbon source (47).

Genes encoding proteins that are involved in the regulation of the immune system also contribute to the selection of the gut microbiota. These include, antimicrobial peptides, proteins, such as defensins, various types of lectins, cell-wall degrading enzymes and immune modulators, all of which are part of the innate immune system and directly interact with microbes. Some of these, such as RegIIIγ, can lead to disruption of the cell wall or membrane and eventually lysis of the bacteria (48–50). In addition, secretory IgAs that are found in large

amount in the mucus interacts with bacteria and contribute to biofilm formation, minimizing exposure of the epithelium to the bacteria (51).

Bile acids, although their primary role is the emulsification of fats, are also part of the innate immune system, owing to their antimicrobial properties (52). As such, they play a direct role in shaping the microbiota community, by stressing the bacteria, damaging their membranes or leading to protein misfolding (52,53), thus selecting for strains that are bile acid tolerant. Bacterial species that feed on bile acids can also be selected. For example, mice fed a high milk-derived-fat diet led to production of bile acids conjugated with taurine to more effectively emulsify the fats. This, in turn, increased the availability of organic sulfur and sulfite-reducing bacteria, such as *Bilophila wadsworthia,* causing colitis in mice (27).

1.1.2.4. Antibiotics

Antibiotics, which are often used to treat infectious diseases, target a wide range of bacteria, regardless of their pathogenicity; thus, gut symbionts are also affected. Depending on the mechanism of action of a given antibiotic, different bacterial classes are affected. For example, a recent study showed a strong increase in damaged *Firmicutes* when cell wall synthesis inhibitors were used, such as ampicillin or vancomycin (54). Interestingly, while vancomycin specifically targets Gram-positive bacteria, it was found that the *Bacteroides* population was substantially reduced (55,56), suggesting a complex interplay between different species. In the same study, members of the *Enterobacteriaceae* family, which is normally under-represented in the gut microbiota, dominated the population following antibiotic treatment (55). Also, antibiotics can disturb the balance of commensal bacteria and reduce their diversity (57). This causes a disruption to the normal gut microbiota also known as “dysbiosis”, but also exposes the host to pathogenic species.

1.1.3. The role of the gut microbiota in human physiology

It is now clear from analysis of metagenomics studies that there is a significant correlation between the composition of the gut microbiota and the health or disease state of individuals (58). Human gut bacteria play a crucial role in our health and well-being, by affecting various levels of the human physiology, as discussed below.

1.1.3.1. Food digestion

The gut microbiota is essential in processing otherwise indigestible polysaccharides from our diet. For example, several *Ruminococcus* species are able to digest insoluble carbohydrates like cellulose and resistant starch (59) whereas *Bacteroidetes* species are characterized by a number of polysaccharide utilization *loci* (PUL) clusters (60). These species, along with others from the *Bacteroidales* and the *Clostridiales* orders, can convert these polysaccharides into SCFAs, which can be utilized by the host (60,61). These SCFAs provide up to 10% of the daily energy requirement (62). When the diet is rich in simple sugars and depleted in dietary fibres, SCFA producers, such as *Prevotella,* are diminished (26). Additionally, bacteria that feed on dietary fibres will consume host sugars when fibres are not available from the diet (63), leading to a thin mucus layer and eliciting an immune response from the host (64). Therefore, the gut microbiota can enhance the host's ability to extract energy and nutrients from the food. As a consequence, an imbalance in microbiota composition has been associated with nutritional disorders, such as diabetes (61,65) and obesity (36). Also, conventionalization of germ-free (GF) mice leads to an increase in body mass and insulin resistance (65) and the transfer of microbiota from obese mice into lean donors can result in a phenotypic change of these mice (66). In addition, the gut microbiota is able to increase the host's monosaccharide uptake from the gut lumen, and, at the same time, increase the *de novo* hepatic lipogenesis (65).

1.1.3.2. Vitamin production

Humans only carry limited biosynthetic pathways for vitamins. To cover our needs in all essential vitamins, we rely on vitamins available in the food or provided by members of the gut microbiota that can synthesise them. It has been shown, for example, that members of the *Bifidobacteria* genus can *de novo* synthesise folate, while *Lactobacillus reuteri* strains can produce vitamin B_{12}, the only vitamin exclusively produced by bacteria, and closely related analogues (67,68). Gut bacteria have also been implicated in the *de novo* synthesis of riboflavin (69), vitamin K (70), biotin (68,71).

1.1.3.3. Short-chain fatty acids production

The most important SCFAs are acetate, propionate and butyrate. These are small molecules that act as signals in the GI tract and regulate various processes. They are end products of the anaerobic fermentation of dietary fibres by the microbiota, which takes place mainly in the proximal colon (72). *Bacteroidetes* produce mainly acetate and propionate, whereas butyrate mainly results from *Firmicutes* activity (73). In addition to their role in energy provision, SCFAs are involved in regulation of appetite (74) as well as fatty acid, glucose and cholesterol metabolism (72). In addition, anti-inflammatory and anti-cancerous properties have been attributed to butyrate (75) and the gradient of butyrate that is formed between the lumen of the gut and the bottom of the crypts has been proposed to mediate cell apoptosis at the tip of the *villi* and cell proliferation at the bottom of the crypt, thus contributing in the regeneration of the brush border epithelium (76). Butyrate was also shown to enhance the tight-junction assembly (77), probably by increasing the production of IL-18, which is responsible for maintaining epithelial integrity (78) and mucin production (79).

1.1.3.4. Maturation of the immune system

In addition to contributing to the food digestion capacity of the host, the gut microbiota is also required for the development and maturation of the immune system. Many studies in GF mice

revealed various malfunctions, reflecting an immature immune system. For instance, GF mice show defective development of lymphoid tissue, and administration of gut bacteria restored its proper function (80). Insufficient microbial exposure in early life can result in an alteration of the gut microbiota in adulthood and a defective or allergy-susceptible immune system (81). Anti-inflammatory immune responses establish tolerance against non-pathogenic microbes, and the gut microbes play a pivotal role in maintaining this tolerance, by regulating responses of the innate immune system (82). For instance, it has been shown that microbes producing butyric acid, one of the main SCFAs, can induce colonic regulatory T cells (T_{reg}), which are a key component in inflammation and allergy suppression (83). In another study, administration of a cocktail of three commensal bacteria in mice resulted in higher counts of T_{reg} cells, which led to increased IL-10 and TGF_{β} production, which are known for their anti-inflammatory activity (84). Similarly, polysaccharide A from *Bacteroides fragilis* showed a protective activity against colitis in mice, by inducing T_{reg} differentiation, while reducing the level of pro-inflammatory cytokines (85). It was recently shown that *Lactobacillus reuteri* strains was necessary for the differentiation of $CD4^{+}$ $CD8\alpha\alpha^{+}$ double-positive intraepithelial lymphocytes (86), cells that are involved in the suppression of inflammation in the gut (87).

Interestingly, host-specific microbiota composition is required for the proper development of a mature immune system. It was recently shown that only GF mice colonised by mouse-derived microbiota could develop a fully mature immune system, in contrast to mice colonised with human- or rat-derived microbes (88), supporting the concept of co-evolution between the gut microbiota and its host.

Members of the gut microbiota can also promote restoration of damaged epithelial tissue in an *ex vivo* model of intestinal epithelial cells (89). In addition, it is well established that certain commensal bacteria, such as *Akkermansia muciniphila*, can induce mucin production leading to

thicker mucus layer (90), or change the glycosylation profile of mucins (41,91), thus protecting the host against infections (92).

1.1.3.5. Competition against pathogens

The protective effects of the gut microbiota also derive from the ability of gut symbionts to compete with pathogens for available nutrients and binding sites. For example, commensal *E. coli* can outcompete *Citrobacter rodentium* in nutrient utilisation in mice, as they both utilise similar carbon sources. The same effect was also observed with *B. thetaiotaomicron* (93). In addition, the imbalance in the gut microbiota composition caused by antibiotics administration or gut inflammation in mice can provide enteric pathogens, such as *Salmonella typhimurium* or *Clostridium difficile* a greater opportunity to infect the host (45,94). Commensal bacteria can also prevent adherence of pathogens to the host tissue, as shown by reduced the adherence of enteropathogenic *E. coli* (EPEC) to epithelium following pre-treatment with *L. reuteri* strains, due to the occupation of available binding sites (95).

To assist the host fight against pathogens, commensal bacteria can also produce antimicrobial compounds, known as bacteriocins, which can have potent inhibitory effects on harmful microbes (96). It has also been postulated that commensal bacteria may exclude pathogens by forming aggregates and trapping them into it, thus preventing invasion in the host (97), They can also protect the host from viral infections (for a review see (98)).

Another protective property of gut commensal bacteria comes from their ability to metabolise unabsorbed bile acid conjugates, which act as germinants in spore germination of *C. difficile*, using bile acid hydrolases. These enzymes remove the amino acid conjugate and release primary bile acids that competitively inhibit germination (99).

1.1.3.6. Probiotic effects

To confer the beneficial effects listed above, the gut microbiota is in a constant interaction/communication with the host. This occurs either directly, where the microbial cells

come in contact with the host tissue, mainly *via* surface proteins or carbohydrates, or indirectly, *via* the action of secreted proteins or other molecules. These interactions often trigger a response cascade which either exerts an immune response (pro-inflammatory effect) or attenuates the immune system (anti-inflammatory effect). Gut-derived microbes can be used as probiotics (*i.e.* live organisms that, when administered in adequate amounts, confer a health benefit on the host, according to the World Health Organization (100)) in order to exploit these immunomodulatory effects. Probiotics can either suppress immune responses to mitigate chronic inflammation or elicit a stronger immune response against pathogens. They are also actively involved in metabolism, directly compete with pathogens and maintain the epithelium integrity (Figure 3).

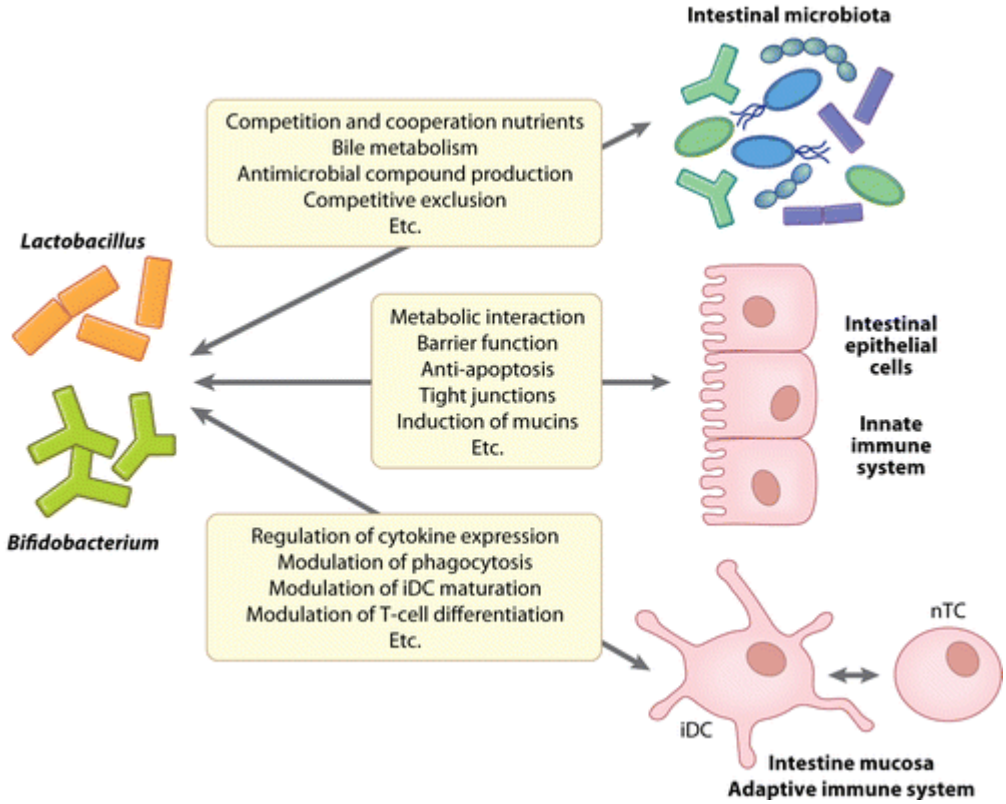

Figure 3 Schematic of the mechanisms underpinning probiotics health benefits (1).

1.2. *Lactobacillus reuteri*: Occurrence in the GI tract and its impact on host health

The Lactobacillus genus is one of the most commonly represented genus in the gut microbiota across vertebrates and one of the first to colonise the mammalian GI tract (4) as it is often found in breast milk (101). It includes facultative anaerobic, Gram-positive, rod-shaped bacteria. *L. reuteri,* in particular, is a wide-spread bacterial species commonly found in the gut of humans (101,102), as well as other vertebrates, including rodents, pigs and chicken (103).

L. reuteri strains from different hosts show a great level of host-specificity (104). For example, murine isolates were found to be the most successful colonisers of the murine GI tract, compared to strains isolated from different hosts (103,105,106). Similar results were observed for poultry strains, when colonizing the chicken GI tract (107). This evolutionary adaptation of different *L. reuteri* strains to their respective hosts was first reported by Oh *et al.* (2010) (108) and many genes and gene clusters specific to strains from the same host have since been identified (108,109). An important feature of host adaptation is the difference in the nature and the number of large cell-surface adhesins between human, porcine, poultry and rodent isolates. The specialization of different strains towards specific hosts has been shaped evolutionally by various events of horizontal gene transfer and deletions of genes (103).

Lactobacilli have been used in a range of human intervention studies as probiotics, to investigate their health promoting effects (110). For instance, *L. reuteri* DSM 17938 was used in clinical trials against infant colic and showed important improvement in the treated groups compared to the placebo group (111,112). In addition, administration of *L. reuteri* ATCC 55730 resulted in milder cystic fibrosis symptoms by reducing pulmonary exacerbations caused by infection or inflammation, and also by reducing the infections of the upper respiratory tract (113). *L. reuteri* RC-14*,* along with *L. rhamnosus* GR-1, was used in a study of HIV-infected patients in sub-Saharan Africa. Despite having no effect on the white blood cell numbers, the patients who received the

probiotics had a 15% increase in $CD4^+$ cells and also showed a significant improvement in symptoms like clinical diarrhea and nausea (114). In addition, many *L. reuteri* strains have demonstrated anti-inflammatory properties in animal studies. For example, administration of *L. reuteri* DSM17938 increased $Foxp3^+$ T_{reg} cells in an induced necrotizing enterocolitis (NEC) mouse model (115) and reduced lipopolysaccharide (LPS)-induced gut inflammation (116). *L. reuteri* anti-inflammatory properties were also shown in a *C. rodentium* infection mouse model (117). The results showed that infected mice that were also exposed to stress exerted a stronger immune response compared to infected-only mice. But infected/stressed mice that were given *L. reuteri* as a probiotic showed normal immune responses, *i.e.* higher levels of β-defensin 3 and reduced expression of CCL2 (which recruits inflammatory monocytes to the colon) and TNF-α (117). While *L. reuteri* was not able to outcompete *C. rodentium,* there was a decreased translocation of the pathogen to the spleen, probably due reduced loosening of the tight epithelial junctions (117). *L. reuteri* was also tested for its ability to reduce inflammation in the GI tract. A mixture of four *L. reuteri* strains was found to protect mice against dextran sodium sulfate (DSS)-induced colitis (118). More specifically, administration of *L. reuteri* R2LC, JCM 5869, ATC PTA 4659 and ATCC 55730 prior to and during the DSS treatment showed a protective effect, by significantly reducing mucosal damage in the Lactobacilli-treated mice. In addition, p-selectin, a protein associated with colitis, was found to be expressed in "healthy" levels in Lactobacilli-treated mice, whereas it was significantly elevated in DSS-treated mice (119).

Other strains of *L. reuteri* have shown potent activity against infectious diseases in preliminary experiments conducted *in vitro* or in animals. For example, the human strain *L. reuteri* L22 produces bacteriocins that kill and inhibit the growth of methicillin-resistant *Staphylococcus aureus in vitro* (120). Reuterin, one of the *L. reuteri* L22 bacteriocins produced, can reduce the growth of *Salmonella pullorum* ATCC 9120 *in vitro*, while *L. reuteri* ATCC 55730, another reuterin producer, exhibits a protective role in a Salmonella-induced pullorum disease model chicken. However, *L. reuteri* L22 did not improve the survival of chicken in the same study, suggesting

that the protective role of *L. reuteri* ATCC 55730 could not be attributed to reuterin exclusively (121). Table 2 summarises some of the studies testing *L. reuteri* strains as a probiotic in various diseases, ranging from infectious diseases to genetic disorders. The results of these studies suggest that *L. reuteri* strains have anti-inflammatory properties, as in most cases they attenuate immune responses.

Table 2 Summary of studies on the *in vivo* probiotics effects of *L. reuteri* strains

Strain	Host	Condition	Outcome	Ref.
***L. reuteri* DSM 17938**	Human	Infant colitis	Reduction in crying time	(111)
***L. reuteri* ATCC 55730**	Human	Cystic fibrosis	Reduction of pulmonary exacerbations, reduction of upper respiratory tract infections	(113)
***L. reuteri* RC-14,** ***L. rhamnosus* GR-1**	Human	HIV infection	No effect in white blood cells, increased $CD4^+$ cells, reduction of clinical diarrhea and nausea	(114)
***L. reuteri* DSM 17938**	Human	T2Diabetes	No change in glycated haemoglobin, improved insulin sensitivity in a subset of participants	(122)
***L. reuteri* DSM 17938,** ***L. reuteri* ATCC PTA 6475**	Human	*Helicobacter pylori infection*	Reduced side effects, reduced gastrin-17, increased rate of eradication	(123)
***L. reuteri* DSM 17938**	Mouse	Necrotising enterocolitis	Increased $Foxp3^+$ T_{reg} cells, reduced LPS-induced gut inflammation	(115)
***L. reuteri* ATCC 23272**	Mouse	*C. rodentium* infection/stress	Decreased translocation of *C. rodentium* to the spleen,	(117)
***L. reuteri* R2LC,** ***L. reuteri* JCM 5869,** ***L. reuteri* ATC PTA 4659,** ***L. reuteri* ATCC 55730**	Mouse	DSS-induced colitis	Reduction of mucosal damage, normal expression of p-selectin	(118)
***L. reuteri* TB-B11**	Mouse	Ovalbumin-induced food allergy	Reduction of diarrhea, reduction of ovalbumin-specific IgE titre	(124)
***L. reuteri* ATCC 55730**	Chicken	*Salmonella*-induced pyllorum disease	Increase in survival rate of the model organism	(121)
***L. reuteri* L22**	Chicken	*Salmonella*-induced pyllorum disease	No significant protective effect *in vivo*.	(121)

Although the exact mechanisms underpinning the health promoting effects of *L. reuteri* remain undefined, immunomodulatory effects and the ability to colonise the human intestine are considered as key elements of these properties.

1.3. *Lactobacillus* cell surface: role in adhesion to the host

The ability of gut commensal bacteria and probiotics to colonise the host is one of the key steps in symbiosis and health promotion. One of the mechanisms mediating a specific interaction with the host is *via* specific surface structures, such as exopolysaccharides (EPS) and cell wall or surface proteins (Figure 4), which can bind to components of the epithelium, the mucus or the extracellular matrix (ECM), depending on the niche the bacteria colonise in the GI tract. In Gram-positive bacteria, these proteins are anchored on the peptidoglycan (PG) layer, *via* specific cell-wall anchoring domains, or the cell membrane (CM), through lipid modifications (Figure 4).

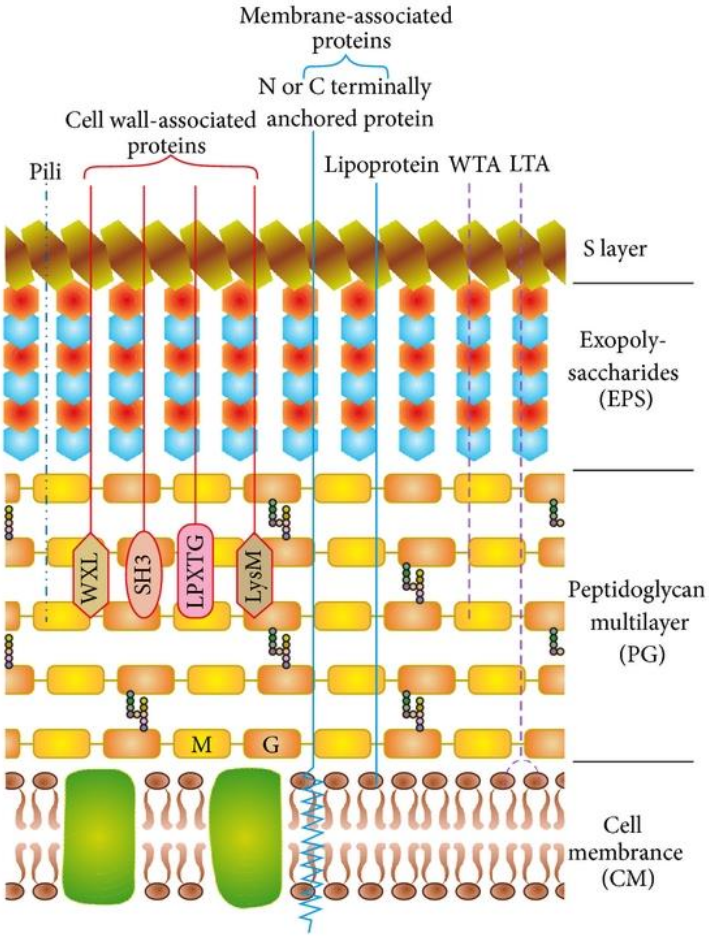

Figure 4 Architecture of the surface of Lactobacilli species. The cell membrane (CM) is protected by a thick layer of peptidoglycan (PG) covered with exopolysaccharides (EPS) and surface layer proteins. Adhesins and other surface proteins are anchored onto the PG layer by various domains such as SH3 and LysM, as well as LPxTG motifs. Wall teichoic (WTA)- and lipoteichoic acids (LTA) are bound onto the PG and the CM, respectively (125).

1.3.1. Exopolysaccharides

EPSs are high molecular weight carbohydrate polymers that are produced by bacteria and secreted into their microenvironment to form the EPS layer. They have a diverse composition, even within the same species, and range in size from 10 to 1000 kDa (126). The EPS layer is a multifunctional structure, involved in bacterial aggregation, as well as adhesion to the host, among other biological roles. EPS is believed to play either a direct role in adhesion by acting as a ligand that mediates adhesion and co-aggregation or indirectly by holding surface proteins in place (127). Bacterial aggregation can also enhance adhesion to host tissue, by increasing hydrophobicity (128). However, in some cases, EPS removal can lead to enhanced adhesion,

which may result from exposure of proteins involved in adhesion to the microenvironment (129). For instance, mutants of *Lactobacillus johnsonii* that do not produce EPS showed greater biofilm formation and adhesion to chicken gut explants, as compared to the wild-type strain, probably due to exposure of hydrophobic components on their surface (130). Similar results have been observed with the pathogen *Streptococcus pneumoniae*, where removal of the capsular polysaccharide led to the exposure of lectins that were responsible for agglutination of human red blood cells (131). These seemingly contradicting results suggest that there are different mechanisms in bacterial adhesion and aggregation, where EPS may favour or impede cell-cell interactions.

1.3.2. Teichoic acids

Wall-teichoic acid (WTA) and lipoteichoic acid (LTA) are cell wall glycopolymers characteristic of Gram-positive bacteria (132). WTA is attached to the PG layer whereas LTA is anchored *via* a lipid moiety to the cell membrane (133). Both WTA and LTA have been implicated in adhesion of pathogenic bacteria to host tissue, and specifically to scavenger receptors and C-type lectins expressed on the surface of macrophages, dendritic cells and epithelial tissue of the host (132). An early study by Granato et al. (1999) showed that treatment of Caco-2 cells with pure LTA decreased binding of *L. johnsonii* to the cells (134), suggesting a direct interaction between LTA and the cells. This interaction between LTA and Caco-2 cells may be mediated by the carbohydrate component of LTA (which could also be found on other glycoconjugates), or by the lipid moiety, which could also be involved in other functions (e.g. protein-protein interactions), by occupying hydrophobic pockets. However, a more recent study also suggested that LTA may be directly involved in adhesion, as *Listeria monocytogenes* that lacked the D-alanine modification of LTA adhered inadequately to Caco-2 cells, compared to the wild-type strain (135). Despite the original view that LTA only had an indirect role in adhesion, these data suggest that it may be an important contributor of bacteria adhesion to host tissue.

1.3.3. Cell surface adhesins

Surface proteins, such as adhesins or lectins, are an important class of cell surface molecules that are involved in adhesion of gut bacteria to the host tissue, especially the gut epithelium and mucus layer, by interacting with host proteins or glycoconjugates.

The mucus layer is a thick mixture of glycoproteins (mainly mucins), immunoglobulins, antimicrobial peptides and other intestinal proteins, lipids and electrolytes (40), covering the GI tract. Many bacterial adhesins have been experimentally shown to bind mucus components such as mucins, mucin glycans or immunoglobulins (40). These include moonlighting proteins, surface appendages like pili and flagella, as well as specialized surface adhesins that bind to host tissue components (summarized in Table 3).

Table 3 Summary of the main characterised adhesins in *Lactobacillus* species.

Adhesin	***Organism***	**Ref.**
32-Mmubp	*L. fermentum* BCS87	(136)
CmbA	*L. reuteri* JCM1112	(137)
DnaK	*L. agilis*	(138)
EF-Tu	*L. johnsonii* La1, *L. reuteri* JCM1081,	(139–141)
Enolase	*L. johnsonii, L. agilis*	(138)
Formyl-CoA transferase	*L. johnsonii*	(138)
GAPDH	*L*. acidophilus, *L. plantarum*	(142–144)
GroEL	*L. johnsonii* La1	(145)
Lam29	*L. mucosae* ME-340	(146)
Lar_0958	*L. reuteri* MM4-1a	(147)
MapA	*L. reuteri* 104R	(148)
MBF	*L. rhamnosus* GG	(149)
Msa	*L. plantarum* WCFS-1	(150)
Mub	*L. reuteri, L. plantarum, L. acidophilus*	(151–157)
Peptidase C1	*L. johnsonii*	(138)
phosphofructokinase	*L. johnsonii*	(138)
Phosphoglycerate kinase	*L. agilis*	(138)
Phosphoglycerolmutase	*L. johnsonii*	(138)
Pili	*L. rhamnosus* GG	(158)
SRR	*L. reuteri* (pig and rodent isolates)	(109,159)
Triosephosphate isomerase	*L. johnsonii*	(138)

1.3.3.1. Moonlighting proteins

Moonlighting proteins are proteins with more than one functions. Many moonlighting proteins that have been experimentally shown to bind mucus have no apparent mucus binding domains or a signal peptide necessary for secretion. Elongation factor Tu (EF-Tu) is a well-studied example of moonlighting proteins involved in bacterial adhesion to the host tissue. Its primary role is believed to be in the translation of mRNA, where it carries the amino acyl-tRNA to the ribosome (160). However, studies have shown its involvement in the formation of bacterial cytoskeleton (161), as well as to adhesion to host tissue (139,140,162). In particular, EF-Tu was found to be upregulated when Lactobacilli strains were grown in presence of mucin and soluble EF-Tu inhibited binding of Lactobacilli as well as *E. coli* and *Salmonella enterica* strains (163). In some cases though, EF-Tu may be masked by EPS, thus its contribution to adhesion is limited (129). EF-Tu from *L. reuteri* JCM1081 was found to bind to mucin sulphated glycans (141). Other similar examples of mucin binding, moonlighting proteins include the heat shock protein GroEL (145), the enzyme GAPDH (143), enolase and DnaK, all of which can be found both intracellularly and on the bacterial surface (125,138).

1.3.3.2. Surface appendages

Pili and flagella are large polymeric proteins that form long surface structures which are involved in bacterial adhesion. Although rare in Gram-positive bacteria, pili have been identified in *Lactobacillus rhamnosus* GG, where they confer binding to mucus (164) and are predicted to exist in other Lactobacilli species, based on genomics analyses (125). Flagellar proteins have been extensively studied in pathogens but not in commensal bacteria. For example, flagellar proteins of the enteropathogenic and enterohaemorrhagic *E. coli* as well as those of *Campylobacter sp.* have been shown to be important components of adhesion to host tissue (165,166). A few Lactobacilli species have the potential to produce flagella which can induce pro-inflammatory responses by the host (167), but their role in adhesion remains to be elucidated.

1.3.3.3. Cell-surface adhesins

This group of adhesins contains proteins that have been primarily characterized for their ability to bind components of the host tissue, such as collagen, fibronectin and mucus. The Collagen Binding Protein from *L. reuteri* NCIB 11951, (168) and *L. crispatus* JCM 5810 (169,170), as well as Fibronectin binding protein A (FbpA) from *L. acidophilus* NCFM (171) and Surface Layer Protein (SlpA) from *L. brevis* ATCC 8287 (172) are examples of this group. The latter comprises the majority of the bacterial surface protein load and plays key roles in aggregation and binding to mucus or ECM (125,173).

Many adhesins are cell surface-bound; they include a signal peptide at the N-terminus that directs their secretion by specialized secretion systems, as well as a C-terminal cell-wall anchoring motif through which they are covalently bound onto the cell wall by specific enzymes called sortases (174). Lar_0958 from *L. reuteri* MM4-1a is such a protein, containing a YSIRK signal motif at the N-terminus, as well as a LPxTG anchoring motif at the C-terminus. This protein family was recently identified as a major adhesin of human *L. reuteri* isolates (147). In addition to the signal and anchoring motifs, these proteins were shown to contain a varying number of repeating domains, the presence of which correlates with the capacity of the microorganism to adhere to mucus. The crystal structure of this repeat showed structural similarities with internalins from *L. monocytogenes*, proteins that are suggested to mediate adhesion to the GI mucus (147). Other examples of major cell-surface adhesins from Lactobacilli are detailed below.

1.3.3.3.1. Mucus binding protein

Mucus Binding proteins (MUBs) containing Mub repeats have been identified primarily in lactic acid bacteria (125) and are more common in those colonizing the GI tract (155); no mucus binding protein has been identified in food-derived Lactobacilli (31). MUB_{53608} from *L. reuteri* ATCC 53608 is one of the most studied examples of mucus adhesins (154). It is a high molecular weight protein that consists of six type 1 Mub (Mub1) repeats and eight type 2 Mub repeats

(Mub2), based on sequence homology. Each repeat is further divided into two domains, a mucin binding (MucBP) domain and an immunoglobulin binding (Ig-binding protein) domain (155,156) (Figure 5). The Mub repeats mediate binding to mucin glycans, through interactions with terminal sialic acid (152,156), and immunoglobulins in a non-antigenic manner (155). In addition, MUB_{53608} presents a C-terminal LPxTG anchoring motif, and an N-terminal secretion signal peptide. MUB_{53608} has the shape of a long, fibre-like structure, of around 180 nm in length, and forms appendices similar to pili found in pathogenic and, more rarely, other commensal bacterial species. This elongated structure allows for the exposure of all 14 repeats, each of which has the capacity to bind onto mucus components. This is proposed to allow the bacteria to firmly adhere to their host, in contrast to many pathogenic adhesins which only show adhesion capacity at the N-terminal tip (156). In addition, MUB_{53608} was recently shown to interact with elements of the immune system *in vitro,* as it elicited immune responses from monocyte-derived dendritic cells, leading to increased levels of pro-inflammatory cytokines and CD83 (175). Proteins containing one or more copies of MucBPs have been identified across most species of the *Lactobacillus* genus, as well as proteins containing Mub repeats (40).

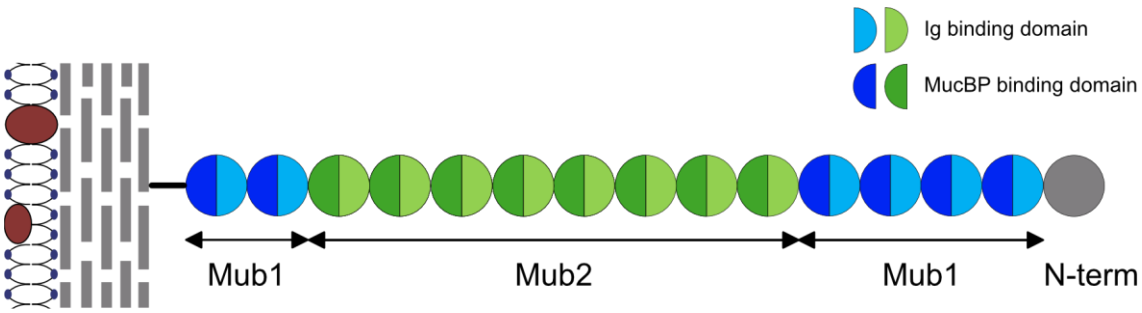

Figure 5 Graphic representation of MUB_{53608} from *L. reuteri* ATCC 53608. Six Mub1 and eight Mub2 repeats comprise MUB_{53608}. Each repeat has two domains, an Ig binding domain at the N-terminus and a mucin binding domain (MucBP) at the C-terminus. Adapted from (156)

1.3.3.3.2. Serine rich repeat proteins

Another recently identified type of adhesin in Lactobacilli belongs to the serine-rich-repeat protein (SRRP) family. This protein family was originally identified in Gram-positive oral pathogenic bacteria, such as streptococci and staphylococci (176–179), where it was shown to

be important for adhesion to host tissue and virulence (180). As a result, most of our knowledge on these adhesins is derived from studies conducted with SRRPs from pathogenic species. These include SraP from *Staphylococcus aureus* COL, Srr1 and Srr2 from *Streptococcus agalactiae* NCTC 10/84 and COH1, respectively, Hsa from *Streptococcus gordonii* DL1, GspB from *Streptococcus gordonii* M99, Fap1 from *Streptococcus parasanguinis* FW213 and PsrP from *Streptococcus pneumoniae* TIGR4 (summarised in Table 4).

Table 4 SRRPs from different microorganisms and their binding substrates.

Origin	SRRP name	Substrate	Reference
***Lactobacillus reuteri* 100-23**	SRR	Forestomach epithelium	(103)
***Staphylococcus aureus* COL**	SraP	Surface molecules of blood platelets	(181)
***Streptcoccus gordonii* M99**	GspB	Human salivary proteins, glycoprotein Ibα	(182,183)
***Streptococcus agalactiae* NCTC 10/84**	Srr1	Platelet fibrinogen	(177,184)
***Streptococcus agalactiae* COH1**	Srr2	Fibrinogen	(184)
***Streptococcus gordonii* DL1**	Hsa	Sialic acid containing MUC7, human salivary proteins	(131,182)
***Streptococcus parasanguinis* FW213**	Fap1	Saliva coated-tooth model	(185)
***Streptococcus pneumoniae* TIGR4**	PsrP	Keratin-10, BR_{PsrP}, extracellular DNA	(186–188)
***Streptococcus salivarius* JIM8777**	SrpA, SrpB, SrpC	Epithelial cells (SrpB), ECM proteins (SrpC)	(189)

The s*rr* gene encoding SRRP is located in a genomic island that encodes for proteins that form an accessory secretion system, dedicated to the glycosylation and secretion of SRRP. The secretion system typically consists of two translocases, $SecA_2/SecY_2$, three to five accessory sec system proteins (Asp1-5), and a variable number of glycosyltransferases, ranging between two to eight (190,191).

The SRRPs are composed of distinct subdomains: a cleavable and unusually long signal peptide which, in some cases, is followed by an alanine-serine-threonine rich (AST) motif, a short serine rich repeat region, a binding region (also known as “basic region” due to its unusual composition

of basic amino acids), a second and much larger serine rich repeat region, and a cell wall anchoring motif (Figure 6) (192).

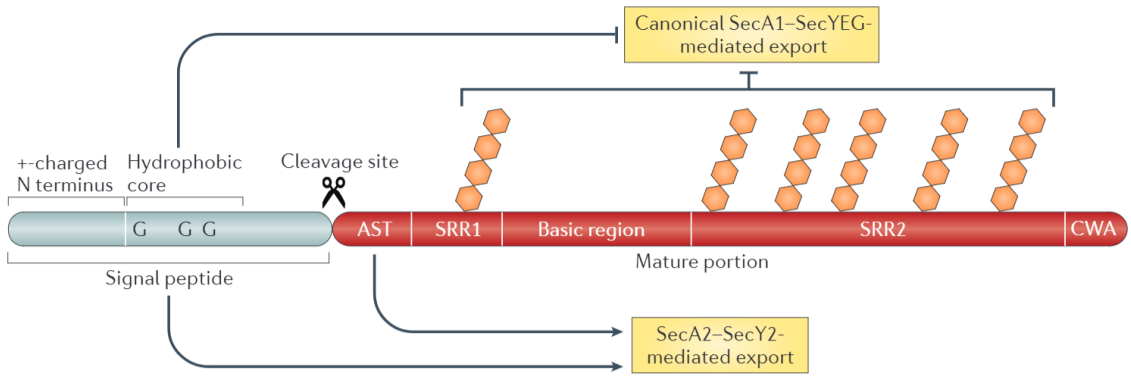

Figure 6 Schematic representation of the SRRP domain organisation. There is an unusually long signal peptide that blocks secretion from the canonical SecA system, followed by an alanine-serine-threonine (AST) domain and a short serine-rich repeat region (SRR1). The binding region (or basic region; BR) is followed by a second, larger serine-rich repeat domain (SRR2) and the cell wall anchoring motif (CWA) LPxTG is found at the C-terminus (193).

Substrates for the various SRRPs characterised to date include a tooth surface model for streptococcal Fap1 and staphylococcal SraP (185,194), which implies its importance in dental diseases such as caries and periodontitis, the platelet surface for the staphylococcal SraP, and the streptococcal Hsa and GspB (181,183,195), suggesting a role in transport to heart valves and bacterial endocarditis, as well as various keratins that can be found in the lungs for the pneumonococcal Srr-1 and PsrP (177,186,188). In addition, SrpB and SrpC from the oral commensal *S. salivarius* JIM8777 were found to specifically bind to epithelial cells and ECM, respectively, suggesting a role in colonisation of the oral cavity (189). The binding or the SRRPs to their substrates is mediated by the binding region (BR). However, different substrate specificities can be observed within the same BR domain. For instance, PsrP was shown to bind keratin 10 in mice lungs, as well as other PsrP molecules to promote aggregation, but different regions in the BR domain were responsible for each interaction (188). While binding of PsrP to keratin-10 is not glycan mediated, it was found to bind various Lewis blood group related glycans in a sialic acid independent manner (131), further suggesting different binding specificities for a single BR region. Strikingly, in addition to protein and carbohydrate substrates, PsrP was found

to bind extracellular DNA during biofilm formation, which greatly expands the nature of substrates this protein binds onto, using a single BR domain (187). Pyburn et al. (2011) also observed distinct binding domains within the BR region of GspB by X-ray crystallography (196), supporting the idea that there may be different binding sites in a BR domain. In addition, defective glycosylation of SRRPs also leads, in many cases, to impaired binding of these adhesins onto their respective substrates. This suggests that the glycans have a role in adhesion, either by mediating glycan-protein interactions, or by contributing to proper folding and stability of the adhesin.

The mechanism of secretion of SRRPs is not yet fully understood, and seems to vary slightly from one organism to another. The two translocases along with the Asps are important for the export, as $SecA_2$ and $SecY_2$ form the channel through which SRRP is secreted (Figure 7). Deletion of either *$secA_2$* or *$secY_2$* leads to accumulation of GspB in *Streptococcus gordonii* (180), but the deletion of *$secY_2$* in *Streptococcus parasanguinis* did not fully abolish Fap1 export (197). $SecA_2$ was shown to interact with GspB *via* three glycine residues found in a hydrophobic region of the N-terminal signal peptide to mediate secretion, and when these residues are mutated export occurs through the canonical SecA system (198). In addition, $SecA_2$ was found to readily interact with the N-terminal of the AST motif, but interaction with the C-terminal of the same motif occurs only when the Asps are present (199). Deletion of any of the *asp* genes results in the defective export of SRRPs (200), showing the importance of these proteins in the secretion process. Asp3 appears to have a central role in the coordination of the export machinery, as it was found to interact with the other Asps, as well as with $SecA_2$ and SRRP (201,202). The interaction with SRRP occurs in the SRR regions, prior to complete glycosylation, so mutations in *asp2* or *asp3* also lead in altered glycosylation of the SRRP (197,202,203). It was recently shown that in addition to its role in export of GspB, Asp2 is also an acyltransferase that O-acetylates GlcNAc residues found on the adhesin (204). However, the role of Asp3 in glycosylation remains elusive. Interestingly,

secretion of SRRP through the $SecA_2/Y_2$ system does not depend on the glycosylation of SRR, but this modification blocks export through the canonical SecA system (205).

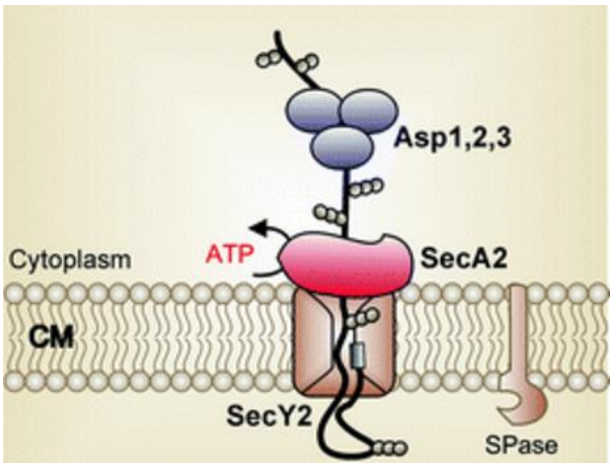

Figure 7 Predicted model of the $SecA_2/Y_2$ export mechanism. Asp1-3 (purple) target SRRPs (black) to the accessory $SecA_2$ protein (red). During the translocation process through the $SecY_2$ channel across the cell membrane (CM), Asp2 and 3 modify the glycan composition and complete the glycosylation of SRRPs (brown dots). A signal peptidase (SPase) cleaves the signal peptide adapted from (206).

To date, *$secA_2/Y_2$* clusters have been identified in the genomes of various *Lactobacillus species* (159). In *L. reuteri*, the cluster has only be found in isolates of murine or porcine origins, and it appears to be absent from isolates of human origin (103,159). The cluster in the murine isolate *L. reuteri* 100-23 is crucial for adhesion of the bacteria to the forestomach epithelium of the murine GI tract, as shown by colonisation experiments in mice with *L. reuteri* 100-23 and mutants lacking putative adhesins. Mutants lacking the *$secA_2$* gene showed defective adhesion, whereas mutants lacking *srr* showed the most reduced colonisation, compared to other putative adhesins (109) tested.

1.4. Protein glycosylation

Protein glycosylation, *i.e.* the covalent attachment of a carbohydrate moiety onto a protein, is a highly ubiquitous protein modification in nature, and considered to target the most diverse group of proteins, as compared to other post-translational modifications (PTM) (207). Although

it was originally believed to be restricted to eukaryotic systems and later to archaea, it has become apparent nowadays that protein glycosylation is a common feature in all three domains of life. In fact, it is now believed that at least 70% of eukaryotic and 50% of prokaryotic proteins are glycosylated by post-translational modification (208).

1.4.1. Mechanisms of protein glycosylation

Protein glycans are secondary gene products, as they are synthesized in a no-template manner by the sequential action of multiple enzymes, called glycosyltransferases (GTs) (209). These enzymes transfer monosaccharides from activated sugar nucleotides (donor) onto a lipid carrier, or the target protein (acceptors). The nature of the primary acceptor defines the glycosylation mechanism that each species has developed. Two distinct glycosylation mechanisms have been identified to date. In the first one, the glycan is first synthesized onto a lipid carrier (dolichol in Eukaryotes and undecaprenol in Prokaryotes), and then transferred as a whole onto the acceptor protein. In the second mechanism, the glycans are built directly onto the acceptor protein. In Eukaryotes, in most cases, an initial, *en bloc* glycosylation occurs, followed by further extension of the glycans directly onto the protein.

Of the 20 amino acids found in proteins, only six of these have been found to accommodate glycans, with asparagine (Asp, N), serine (Ser, S), threonine (Thr, T) being the most common ones, whereas tyrosine (Tyr, Y), cysteine (Cys, C) and tryptophan (Trp, W) and glutamine (Gln, Q) have only been identified as glycosylated in very few and specific cases. On these amino acids, the glycans can be found on the nitrogen of the Asp or Gln amido group (N-linked glycans), the oxygen of the Ser, Thr, or Tyr hydroxyl group (O-linked glycans), the sulfur of the Cys thiol (S-linked glycosylation) or the C2 of the Trp indol ring.

1.4.1.1. Protein glycosylation in Eukaryotes

Protein glycosylation in eukaryotic systems is a process that occurs co-translationally, as well as post-translationally and can be either N- or O-linked. N-glycosylation begins in the endoplasmic

reticulum (ER), where six GTs catalyse the synthesis of the core $Mannose_5$N-acetyl-$glucosamine_2$ ($Man_5GlcNAc_2$) on dolichyl-pyrophosphate on the outer membrane of the ER. $Man_5GlcNAc_2$ is then flipped towards the ER lumen and expanded by four Man and three glucose (Glc) molecules. The complete $Glc_3Man_9GlcNAc_2$ glycan is then transferred by the Stt3b active subunit of the oligosaccharyl-transferase (OST) complex onto the asparagine of the highly conserved consensus sequon Asp-Xxx-Ser/Thr, (where Xxx can be any amino acid, other than proline) (210). The glycoproteins are then subjected to further modifications by two ER α-glucosidases that remove the Glc_3, and sometimes by an α-mannosidase that removes one or two mannose units. Up to this point, the glycans serve as quality control markers; misfolded proteins are tagged for ER-degradation, whereas folded proteins are transferred to the Golgi apparatus. There, the glycans are further trimmed by mannosidases down to $Man_3GlcNAc_2$ and then extended with galactose (Gal), N-acetyl-galactosamine (GalNAc), or GlcNAc and capped with N-acetyl-neuraminic acid (Neu5Ac) and/or fucose (Fuc) to yield complex N-glycans (211) (Figure 8). Mammalian systems other than humans can also cap the glycans using N-glycolyl-neuraminic acid (Neu5Gc) whereas yeasts generally extend the glycans with mannose only, yielding high mannose structures (212).

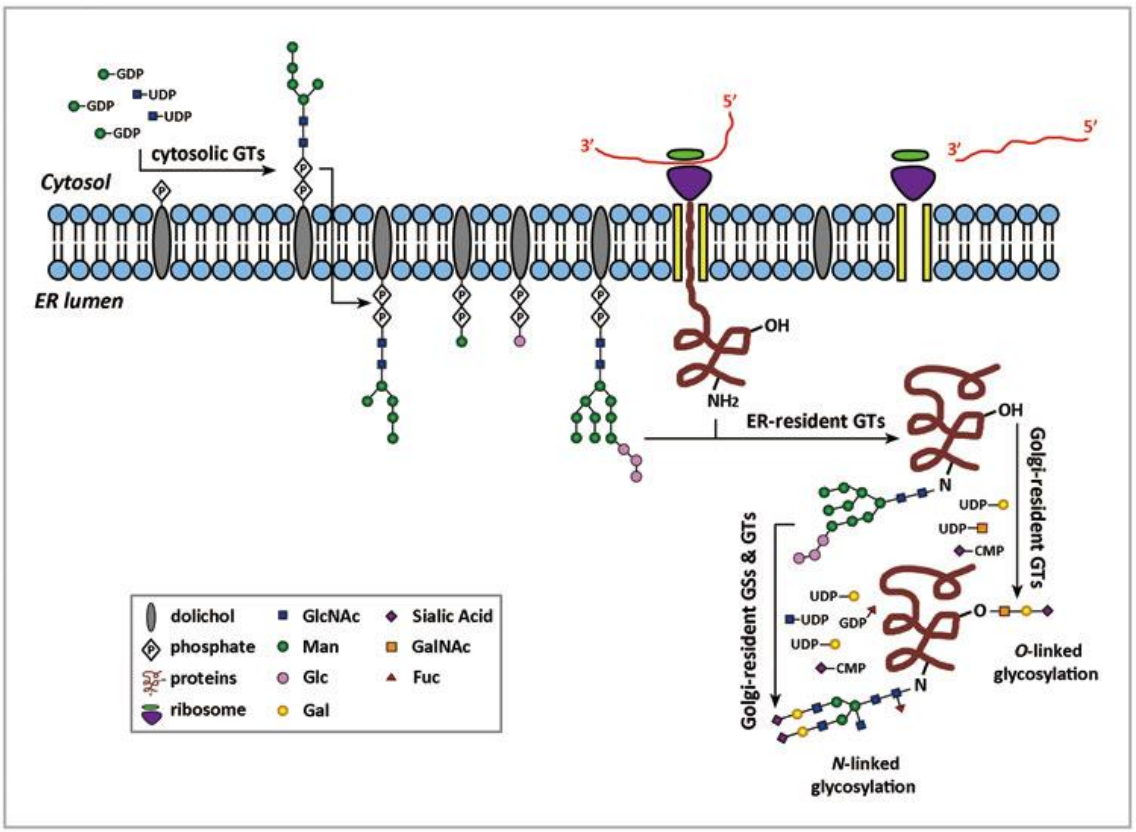

Figure 8 *En bloc* N-glycosylation in Eukaryotes. The core glycan is synthesised on the cytosolic side of the endoplasmic reticulum (ER), flipped over to the ER lumen and transferred onto the acceptor protein before it is translocated to the Golgi apparatus where the glycan is further matured (213).

O-glycosylation (often referred to as "mucin-type O-glycosylation") takes place in the Golgi apparatus (214,215) and is characterized by the synthesis of the glycan chain directly onto the acceptor protein by the sequential action of multiple GTs (Figure 9). The process is initiated by the addition of an α-GalNAc to a Ser or Thr residue (Tn-antigen), which, under normal conditions, is extended by a β-1,3-Gal (T-antigen). As an exception, the Tn-antigen can be extended by β-1,3-GlcNAc (core 3), or β-1,3 GalNAc (core 5) in intestinal epithelial cells (216). The T-antigen and core 3 (GlcNAcβ-1-3-GalNAcα) structures are further elongated or branched with Gal and GlcNAc, and capped with Fuc or Neu5Ac, similarly to N-glycans. Core 5 structures can be capped by an α-2,6-Neu5Ac (214). In disease states, such as cancer, the Tn- and T-antigens may not be extended, or immediately capped with sialic acid, generating sialyl-Tn- or sialyl-T-antigens (216). While no consensus sequon has been identified for O-glycosylation (in contrast to the N-X-S/T sequon in N-glycosylation), there is evidence that the priming N-actetyl-α-galactosaminyl-

transferase shows a preference for Thr over Ser residues, while glycosylation is favoured by the presence of a proline close to the glycosylation site and may be inhibited by charged amino acids flanking the glycosylation site (214).

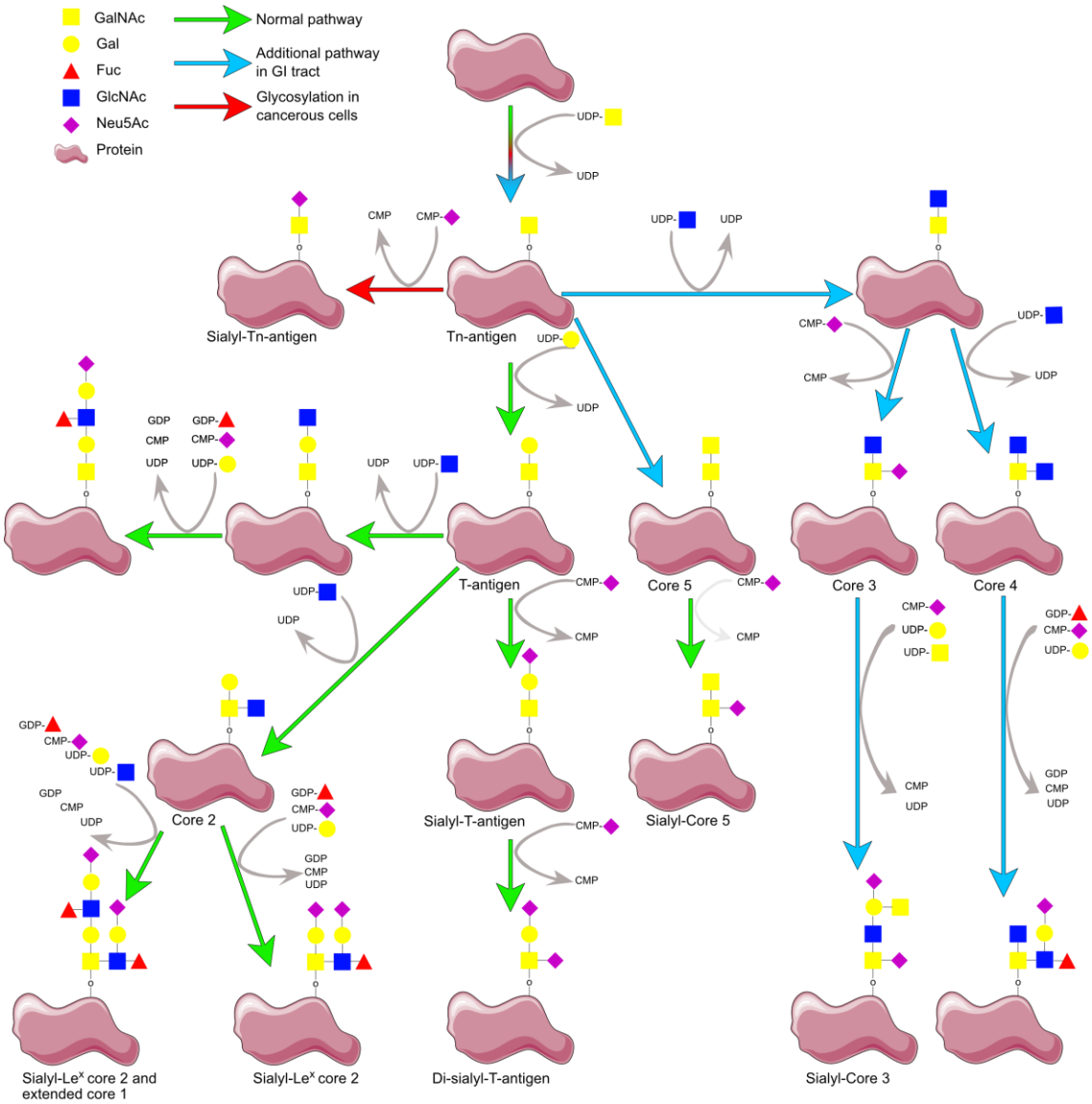

Figure 9 Mucin-type protein O-glycosylation in Eykaryotes. Each GT extends the glycan directly ont the acceptor protein to generate diverse carbohydrate epitopes (adapted from (216)). Core structures are shown.

To summarise, protein N-glycosylation in Eukaryotes starts during protein synthesis in the ER by the addition of the pre-assembled core N-glycans on the asparagine residue of the sequon N-X-S/T. The glycans mature in the Golgi apparatus, where O-glycosylation also takes place, by sequential addition of monosaccharides to the extending glycan, directly onto the acceptor protein. A limited number of monosaccharides are employed in eukaryotic protein glycosylation,

including GlcNAc, Man, Glc (only in the premature N-glycans), Gal, GalNAc, Fuc, and Neu5Ac or Neu5Gc. Additional modifications, such as addition of sulfates may occur under certain conditions (217,218).

1.4.1.2. Protein glycosylation in bacteria

Protein glycosylation in bacteria shares several similarities with that of Eukaryotes, but also shows some significant differences. Similar to eukaryotic glycosylation, bacterial glycoproteins can be modified primarily on Asp (N-glycosylation) or Ser/Thr (O-glycosylation). However, in contrast to eukaryotic glycosylation, where N-glycans are pre-assembled onto a lipid carrier before being transferred onto the acceptor protein and O-glycans are synthesized directly onto the acceptor protein, bacterial glycosylation is more diverse, both in terms of mechanisms employed, as well as in the nature of the carbohydrates used. In addition, while glycosylation in Eukaryotes occurs co-, as well as post- translationally, glycosylation in bacteria is believed to occur post-translationally.

1.4.1.2.1. N-Glycosylation: The *Campylobacter jejuni* paradigm

The first complete glycosylation system ever identified in bacteria, and indeed the most well studied, is that from *C. jejuni*. *C. jejuni* 81-176 harbours a protein glycosylation cluster (pgl) of 13 genes (219) that are responsible for the glycosylation of various proteins (220). These genes encode i) enzymes that synthesise bacilosamine (2,4-diacetamido-2,4,6-trideoxyglucose) found at the reducing end of the glycan, ii) GTs that are involved in the production of the glycan {α-GalNAc-(1,4)- α-GalNAc-(1,4)- [β-Glc-(1,3)-]- α-GalNAc-(1,4)- α-GalNAc-(1,4)- α-GalNAc-(1,3)-α-diNAcBac} on undecaprenol-phosphate, iii) a transporter that flips the glycan to the periplasm and iv) an oligosaccharyl-transferase (N-OSTase) that glycosylates the target protein (219) (see Figure 10A).

This system resembles the eukaryotic N-glycosylation pathway, as the glycan is synthesized onto a lipid carrier, before being flipped across a membrane and transferred *en bloc* onto the acceptor

protein. In addition, PglB, the N-OSTase, is homologous to the Stt3b subunit that glycosylates the newly synthesised proteins in the ER of Eukaryotes and, as such, it recognizes a similar consensus sequon, although extended to Asp/Glu-Tyr-Asp-Xxx-Ser/Thr (221). This may reflect the fact that the glycosylation site of unfolded proteins in eukaryotic systems are more flexible and, thus, readily accessible by the N-OSTase, whereas fully folded bacterial proteins are more rigid and require this extended sequon to better expose the glycosylation site to PglB. This was shown in *in vitro* glycosylation reactions were PglB was used to glycosylate native *C. jejuni* glycoproteins or bovine ribonuclease A. PglB had greater affinity for the native substrates, compared to folded RNase A, but could glycosylate unfolded RNase A as efficiently as the native proteins (222). Homologues of PglB from other organisms, however, have been shown to possess a more relaxed specificity towards the acceptor peptide, yet retaining the requirement for the Asp-Xxx-Ser/Thr motif (223). PglB has been shown to have a relaxed specificity towards the oligosaccharide it can transfer (224,225). The nature of the monosaccharides does not seem to restrict transfer, as heterologous expression and glycosylation systems have been used to modify proteins even with eukaryotic-like glycans (225). The length of the glycan is not strigent either; transfer of longer or shorter glycans has been achieved in cases were O-antigen-derived glycans were used for protein glycosylation (224).

Homologues of PglB, the key enzyme of this N-glycosylation system in *C. jejuni*, are present in most δ- and ε- proteobacteria analysed to date. However, the organisation of the cluster varies between species and strains in terms of the number of genes and in particular the GTs expressed. For example, some *Campylobacter* and *Helicobacter* species contain two putative copies of *pglB*, whereas the *pgl* cluster in *Helicobacter canadensis* MIT 98-5491 is split between multiple loci (226). *H. pylori* seems to be an exception, as it lacks all homologues from the *pgl* cluster and, to date, no N-glycosylation has been identified in this species (227).

In addition, as this type of glycosylation takes place in the periplasm, and requires flipping of the lipid-linked glycan across the inner membrane, this N-glycosylation system has not been identified or predicted to exist in any Gram-positive species.

1.4.1.2.2. Alternative N-glycosylation in β- and γ-proteobacteria

In contrast to the "typical" *en bloc* N-glycosylation system found in *Campylobacter* and other δ- and ε- proteobacteria (as described above), a different N-glycosylation system has been reported in γ-proteobacteria. In particular, HMW1, an adhesin in *Haemophilus influenza* (*Hi*), was found to undergo glycosylation in the cytoplasm by HMW1C with one or two hexose (Hex) molecules at over 30 sites (228). In contrast to other glycosylation systems, where one enzyme is responsible for adding the initial monosaccharide and other enzymes act to elongate the glycan, HMW1C can perform both tasks, *i.e.* form an N-bond between the first Hex and the acceptor protein, and extend the glycan by forming an O-glycosidic bond with a second Hex (229) (see Figure 10B). Based on its amino acid sequence, the enzyme is classified as a GT family 41 of the Carbohydrate Active Enzymes (CAZy) database (www.cazy.org) (230), which contains almost exclusively O-glycosyltransferases (231). HMW1C recognises the same N-X-S/T consensus sequence required for the typical eukaryotic or prokaryotic N-glycosylation process, with a single exception in the native substrate HMW1, in which the consensus sequence appears to be reversed (T-F-N-V-E) (229). It is still unknown, however, if HMW1C is involved in the glycosylation of other proteins, or if it is solely dedicated to the modification of HMW1. Regarding its sugar specificity, the enzyme was shown to initiate glycosylation with either Glc or Gal, but the glycan was further extended with Glc or Gal only when Glc was the initiating monosaccharide (228). This is in contrast to other N-glycosylation systems were the reducing sugar is a N-acetylhexosamine (HexNAc).

HMW1C-like proteins are predicted to exist in many families of β- and γ- proteobacteria (231). Some of them are clustered closely with predicted adhesins similar to HMW1 and translocases,

suggesting that the adhesins may be the target proteins, but others have no apparent glycosylation targets (231). Of these, HMW1C-like protein from *Actinobacillus pleuropneumoniae* (*Ap*), *Kingella kingae* and *Aggregatibacter aphrophilus* were experimentally shown to possess a function similar to *Hi*HMW1C. Most importantly, *Ap*HMW1C-like protein was shown to have a relaxed specificity towards the glycosylation site, which also included Ser or Gln residues (232).

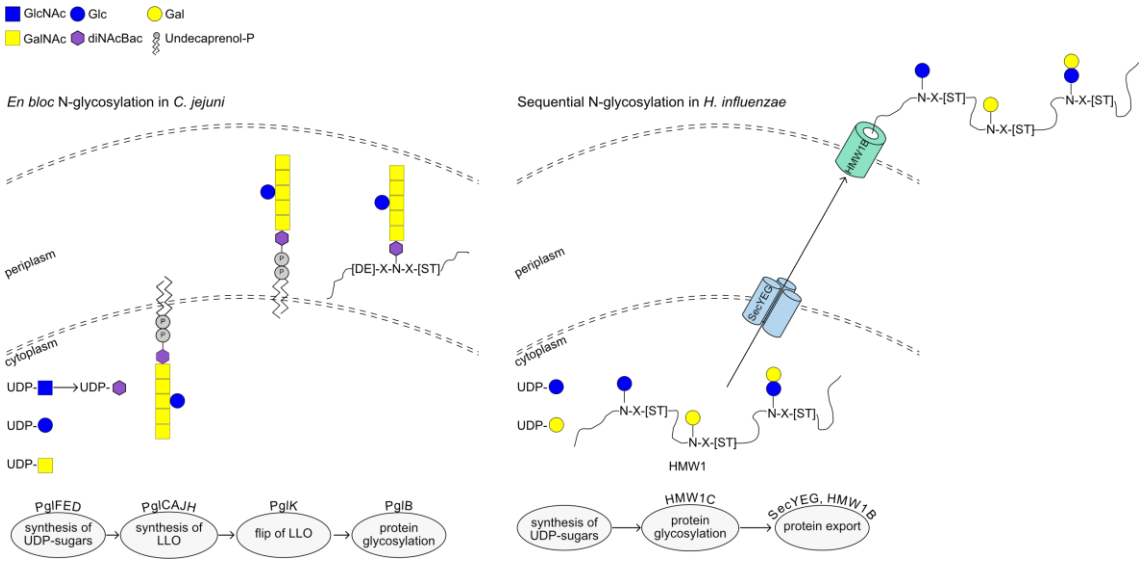

Figure 10 N-glycosylation pathway in Prokaryotes. A) The *en bloc* glycosylation system in *C. jejuni*. Synthesis of diNAcBac is mediated by the enzymes PglFED. The heptasaccharide is assembled onto udecaprenol-phosphate by the GTs PglCAJH, before flipped to the periplasmic space by PglK and transferred onto the target protein by PglB. B) The sequential N-glycosylation system in *Haemophilus influenzae.* HMW1 is glycosylated in the cytoplasm by HMW1C, and is then secreted through the SecYEG and HMW1B channels (233).

1.4.1.2.3. N-glycosylation in mycoplasmas

While the family of enzymes involved in cytoplasmic N-glycosylation appears to be restricted to limited classes of Gram-negative proteobacteria, similar glycosylation mechanisms cannot be excluded from Gram-positive bacteria. In fact, evidence for N-glycosylation in mycoplasmas species has emerged. Mycoplasmas are Gram-positive bacteria with a unique surface structure. In a recent study, Asn and Gln residues outside of the N-glycosylation consensus sequence were found to carry a single Hex in *Mycoplasma pulmonis and Mycoplasma arthritis*, which could suggest a similar glycosylation mechanism to that found in *H. infuenzae (234).* However, no intracellular glycoproteins could be identified, suggesting that this process takes place

extracellularly. In addition, it was shown that the bacteria could use free oligosaccharides from the growth media, without the need to synthesise glycans internally to use for protein glycosylation (234).

1.4.1.2.4. O-glycosylation in bacteria

Similarly to Eukaryotes, bacteria also have mechanisms to perform O-glycosylation by modifying protein targets with glycans on Ser or Thr residues, and, like prokaryotic N-glycosylation, two mechanisms have been identified: either *en bloc* transfer of a pre-assembled lipid-linked oligosaccharide, or modification of the acceptor protein directly, by the sequential action of GTs (235). In the first case, the mechanism is identical to that of the N-glycosylation system, *i.e.* the glycan is synthesised on undecaprenol, the lipid carrier found in bacteria, flipped over to the periplasmic space and transferred onto the acceptor protein by the action of an O-oligosaccharyltransferase (O-OTase). In the second case, multiple GTs act directly on the acceptor protein to extend the glycan, using sugar nucleotides as donors.

En bloc O-glycosylation

Several Gram-negative species have been identified to harbour genes encoding O-oligosaccharyltransferase (O-OTase), including *Neisseria, Pseudomonas*, *Aeromonas* and *Burkholderia spp* (209). The most well studied example of *en bloc* O-glycosylation is that of *Neisseria gonorhoeae*, where PglO, the active O-OTase, glycosylates multiple proteins with a O-acetylated (OAc)-Gal-Gal-diNAcBac (236–238). Often, the O-OTase utilises lipid-linked glycans used in O-antigen biosynthesis (239), as in the case of PilO from *Pseudomonas aeruginosa* (240,241) and *Francisella tularensis* (242). In this case, the O-antigen subunit is built on an undecaprenyl phosphate lipid carrier on the cytosolic side of the inner membrane, as in the case of N-glycosylation, by the sequential action of a varying number of GTs. Then it is flipped across the membrane to the periplasmic space by the transmembrane flippase Wzx. At this stage, either the O-antigen polymerase Wzy forms a glycosidic bond to link two subunits together

(243), or the O-OTase uses the synthesised subunit to transfer the glycan onto a glycoprotein (Figure 11A) (239).

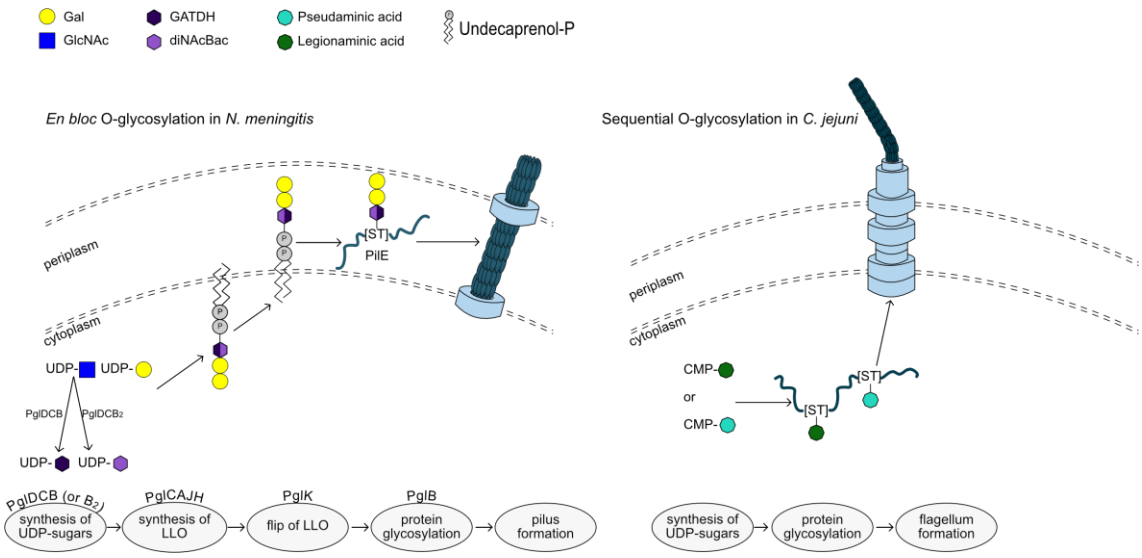

Figure 11 O-glycosylation in Gram-negative bacteria. A) *En bloc* O-glycosylation of *Neisseria meningitis* pili; after the synthesis of the necessary nucleotide-activated monosaccharides, the glycan is assembled onto undecaprenol phosphate, before flipped in the periplasmic space, where it is transferred onto the target protein. B) Sequential O-glycosylation of the *C. jejuni* flagella; the glycan is synthesised directly onto the target protein in the cytoplasm, before the glycosylated protein is secreted.

As *Neisseria* species lack O-antigens (244), their glycosylation clusters may be dedicated to protein glycosylation. *Burkholderia cenocepasia* also has an O-glycosylation system that is independent of the O-antigen biosynthesis and can transfer HecNAc-HexNAc-Hex moieties onto acceptor proteins (245). Most of these systems have been studied in the context of flagellar or pili glycosylation in their respective organisms, and it is not clear whether flagellar subunits are the only targets of glycosylation. Recent studies have shown that the O-glycosylation systems in *B. cenocepasia* and *N. gonorrhoeae* can target multiple proteins (237,245), whereas PilO from *P. aeruginosa* was found to target proteins other than pilins, when heterologously expressed in *E. coli*, suggesting that it could potentially glycosylate other periplasmic proteins (246). The *N. gonorrhoeae* glycoproteins were found to be modified in low-complexity regions (LCRs), rich in Ala, Ser and Pro residues (237), suggesting there may be some structural features that are identified by the O-OTases, since no consensus sequence has been identified. Interestingly, two

O-OTases were identified in *Acinetobacter baylyi* ADP1, one being specific to pilin glycosylation, whereas the other one could target several proteins (247).

While the specificity of the enzymes towards the acceptor protein is not yet fully understood, it is clear that O-OTases from different organisms can transfer a wide range of sugars. For instance, PglL from *N. meningitidis* and PilO from *P. aeruginosa* can transfer, in addition to their native glycans, the *C. jejuni* glycan synthesised during N-glycosylation, as well as the *E. coli* O7 antigen. It is worth mentioning that while PglL could facilitate O-glycosylation using glycans comprised of more than three repeating units of the *E. coli* O7 antigen, PilO was not active when more than two repeats were used (244). In addition, PglL was also able to glycosylate a pilin subunit using UDP-diNAcBac, but not UDP-GlcNAc or UDP-GalNAc, as donor *in vitro*. However, this donor is not present in the periplasmic space, where this reaction naturally takes place, so it is considered unlikely for this reaction to occur *in vivo* (248). Putative O-OTases from *Burkholderia thailandensis* and *Vibrio cholerae* identified and expressed heterologously were shown to be active, when provided with the right combination of glycan donor and protein acceptor (249).

O-glycosylation by sequential action of GTs

In addition to the *en bloc* O-glycosylation systems, many bacteria encode enzymes that mediate mucin-type O-glycosylation, where the acceptor protein is modified intracellularly by the direct action of a priming GT, followed by extension of the glycan by the action of additional GTs. In its simplest form, the acceptor protein is modified by a single monosaccharide, with no further elongation of the glycan, as in the case flagellar proteins from *C. jejuni* and *C. coli* which are modified by a single pseudaminic acid (Pse) or legionaminic acid (Leg) or their derivatives, respectively (Figure 11B). In both cases, the genes coding for enzymes involved in the biosynthesis and subsequent transfer of the sugar onto the protein are all located downstream of the *flaA* flagellin gene (250,251).

Examples of this O-glycosylation mechanism have also been described in a range of Gram-positive species. For example, strains of *C. difficile* glycosylate their flagella with two types of glycans: Type A is composed of an O-GlcNAc modified with a thr *via* a phosphodiester bond (Thr-P) and Type B of an O-GlcNAc, extended with two rhamnose (Rha) molecules, occasionally methylated, and capped with a unique sulfonated peptidyl fucosamine (252,253). Glycosylation with the Type A glycan is mediated by the action of a GT encoded by *CD0240* (254), followed by the addition of the pre-formed Thr-P at C-3 of the GlcNAc moiety by CD242, before CD243 methylates the amine on the thr (255). The synthesis of Type B glycans requires the action of five gene products: GT1 initiates glycosylation by adding the reducing GlcNAc, GT2 extends the glycan by adding two Rha residues and methylating either or both Rha, whereas GT3 adds the non-reducing terminal peptidyl-modified fucosamine, which is synthesised by the action of CDR20291_0245 and 0246 (252). In contrast to *C. difficile, C. botulinum* glycosylates its flagella with a single hexuronic acid or Leg derivative per glycosylation site (256). *Bacillus antharacis* and the closely related *Bacillus cereus* glycosylate their spore protein with 3-O-Me-Rha-α-1,2-Rha-α-1,3-GalNAc, capped either with anthrose or cereose, respectively, sugars characteristic for each strain (257). A distinct glycosylation system is found in *L. monocytogenes*, where the flagella is modified on several amino acids by a single β-O-linked GlcNAc (258). This is unexpected as it resembles the cytosolic O-GlcNAcylation mechanism that is involved in signalling pathways (259), rather than the typical O-glycosylation pathways where the reducing sugar is most often in α-configuration (see section 1.4.1.1).

The accessory secretion system $SecA_2/Y_2$ glycosylation pathway

As described above (1.3.3), several pathogenic Gram-positive bacteria possess an auxiliary secretion system (aSecA), in addition to the canonical SecA system. This system contains the necessary genes encoding proteins that facilitate the expression, glycosylation and subsequent secretion of serine-rich-repeat containing proteins (SRRPs) (for details see 1.3.3). The cluster also contains a variable number of GTs, depending on the microorganisms. The most well

studied $SecA_2$-mediated glycosylation system is that of *Streptococcus parasanguinis* (191,205,260–265) which shows some unique features, not found in other glycosylation systems. In total this cluster contains six GTs. First, glycosylation of Fap1, the SRRP in *S. parasanguinis* FW213, is initiated by the combined action of two GTs, GtfA and GtfB. These enzymes interact with the acceptor SRRP and with each other through a conserved domain DUF1975 and mediate the addition of the reducing GlcNAc. GtfA acts as the active GT, whereas GtfB interacts with the acceptor protein, acting as a chaperone (266,267). An enzyme previously annotated as a sugar nucleotide synthase like protein (NSS) extends the glycan by adding a glucose unit (thus this enzyme has been renamed GtfC). dGT1 contains two distinct GT domains (DUF1792 in N-terminus, which is a recently described GT-D type glycosyltransferase fold, and a GT-A type GT fold in C-terminus (262)) and creates a branching point by adding a Glc and a GlcNAc residue on Glc. GalT2 adds a Rha residue onto the second glucose and the glycosylation is completed by the addition of a Glc residue onto GlcNAc by Gly (Figure 12) (191).

Recently, the $SecA_2/Y_2$ glycosylation system from *S. pneumoniae* TIGR4 was elucidated (268). This cluster contains ten GTs. Glycosylation of PsrP, the respective SRRP in this organism, is initiated by the addition of GlcNAc by GtfA and GtfB. GtfA was recently shown to belong to GT family 4, suggesting the addition of an α-O-GlcNAc (267). Gtf3 catalyses the second step in glycosylation by adding a Glc unit, similarly to the Fap1 glycosylation. The third step is catalysed by GlyD, GlyG or GlyE. Similarly to dGT1, GlyD is also composed of two GT domains: a DUF1792 at the N-terminus and a GT8 domain at the C-terminus. $GlyD_{DUF1792}$ primarily mediates the third step in PsrP glycosylation by adding either a Glc or a Gal residue. When GlyG, a GT family 2 glycosyltransferase, mediates the third step, the glycan is extended with Glc, whereas GlyE, a GT8 enzyme, adds a Gal residue. These three enzymes, in addition to GlyA can also mediate the fourth glycosylation step; in particular, GlyG can add an additional Glc to the extending glycan, whereas $GlyD_{GT8}$, GlyE and GlyA can extend the glycan with a Gal residue. $GlyD_{DUF1792}$ can contribute by adding either a Glc or a Gal, during the third glycosylation step (Figure 12). So far,

no activity has been attributed to GlyB, GlyC and GlyF, although GlyB was shown to possess hydrolytic activity against both UDP-Glc and UDP-Gal, whereas GlyF was more active against UDP-Gal (268).

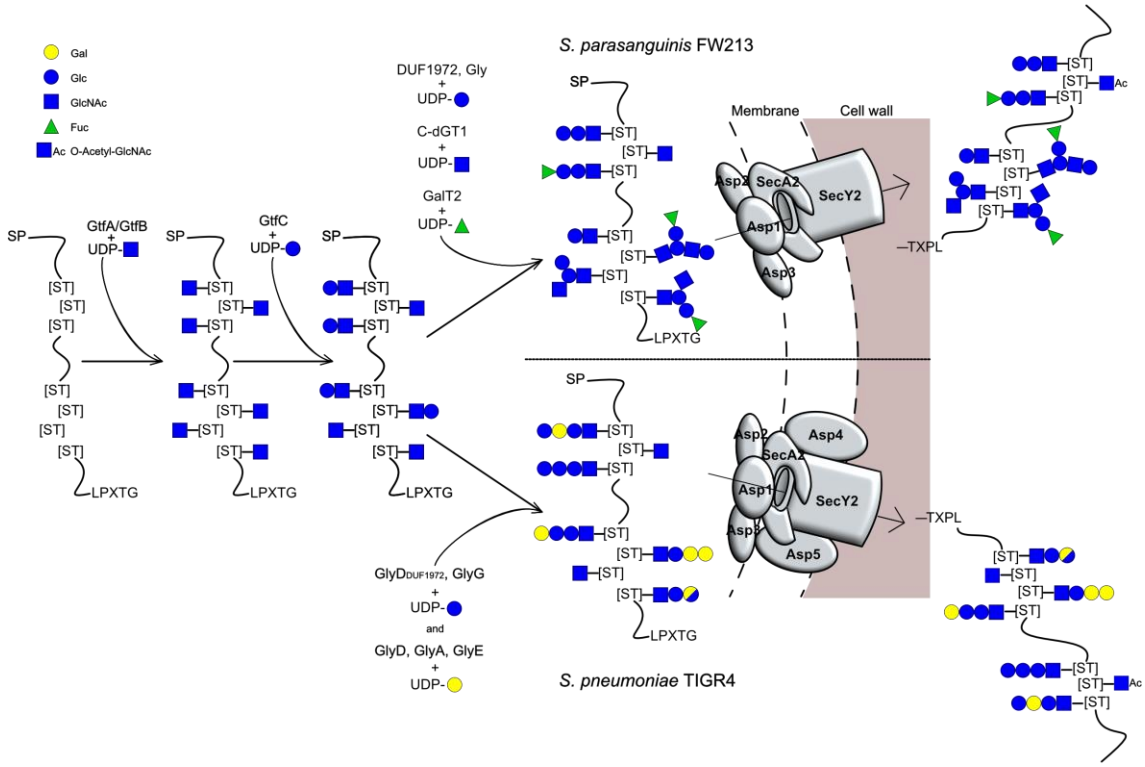

Figure 12 SRRP Glycosylation mechanism of Fap1 in *Streptococcus parasanguinis* FW213 (top) and PsrP in *S. pneumoniae* TIGR4 (bottom). The adhesin is glycosylated intracellularly, by the sequential action of a number of GTs. The glycosylated product is secreted through the $SecA_2/Y_2$ export machinery. During secretion the signal peptide (SP) is cleaved and Asp2 acetylates GlcNAc residues on the SRRPs.

Glycosylation of the *S. agalactiae* Srr-1 is mediated by the GTs found in the $SecA_2/Y_2$ cluster. However, it shows great glycan heterogeneity and the role of each GT is not yet understood. Importantly, a novel modification was identified in the Srr-1 glycan: some HexNAc residues also carry an additional O-acetyl group, probably on the O-6 position (269). This modification had previously been identified on GlcNAc in the peptidoglycan layer of various Lactobacilli species, where it protects the PG from enzymatic hydrolysis (270). However, while OatB was responsible for generating the O-Acetylated GlcNAc (O-AcGlcNAc) used in the peptidoglycan, Asp2 was found to modify GlcNAc residues on SRRPs (204).

Recently, a functional $SecA_2/Y_2$ cluster has been identified in the oral commensal *Streptococcus salivarius* JIM8777, which harbours genes encoding three SRRPs and four GTs (189). In contrast to other studied $SecA_2/Y_2$ systems, GtfA and GtfB in the cluster were not involved in glycosylation of the SRRPs. Instead, two genetically linked GTs found outside of the cluster were shown to act on the SRRPs, as well as other surface and secreted proteins (189). SRRPs in this cluster were found to be glycosylated with O-acetyl-HexNAc, HexNAc and HexNAc-Hex residues, only in the serine rich repeat regions (189), in contrast to Srr-1 from *S. agalactiae* in which glycopeptides from the N-terminal domain were identified (269).

Table 5 SPPR glycosylation in bacteria.

SRRP	**Organism**	**Glycans**
PsrP	*Streptococcus pneumoniae* TIGR4	Glc-Glc-Glc-GlcNAc Glc-Gal-Glc-GlcNAc Gal-Glc-Glc-GlcNAc Gal-Gal-Glc-GlcNAc
Fap1	*Streptococcus parasanguinis* FW213	Glc-GlcNAc-(Rha-Glc)-Glc-GlcNAc
Srr1	*Streptococcus agalactiae* H36B	N.D.* Contain HexNAc, O-AcHexNAc, Hex
GspB	*Streptococcus gordonii* M99	N.D.* Contain HexNAc, O-AcHexNAc

*N.D: Not determined

1.4.1.3. Protein glycosylation in gut commensal bacteria

While protein glycosylation has attracted much interest in pathogenic bacteria underscoring the role of glycans in virulence and pathogenicity in many clinically important bacterial species, the nature and function of glycoproteins in commensal bacteria remains largely unexplored.

Various glycoproteins have been identified in *B. fragilis*, one of the most studied gut commensal bacteria. Most proteins were shown to be fucosylated as shown using the fucose-specific *Aleuria aurantia* lectin (AAL). The bacteria could both synthesise GDP-fucose from GDP-mannose, and acquire Fuc from the growth media and activate it with GDP, after phosphorylation (both phosphorylation and subsequent activation are catalysed by a single enzyme) (271). Affinity chromatography with AAL followed by MS analysis identified glycoproteins of various functions, including peptidases, chaperones and proteins predicted to be involved in protein-protein interactions. All the identified proteins were predicted to be periplasmic or associated with the bacterial outer membrane (272). It was also found that glycosylation took place in the periplasm, which suggested an *en bloc* glycosylation mechanism. Indeed, a gene cluster resembling a capsular polysaccharide (CPS) biosynthesis cluster was identified, which lacked a polymerase gene. After its deletion, the protein recognition by AAL was lost, suggesting that this cluster plays a critical role in a general O-glycosylation system and that this system is independent of the CPS biosynthesis pathway. Using an antibody specific for the *B. fragilis* glycan against protein extracts of various *Bacteroides* species, it was suggested that most of them, including *B. thetaiotaomicron* and *B. ovatus*, produce similar glycans. No glycosylation was observed in *B. vulgatus,* suggesting either a lack of glycosylation, or, more likely, a different glycan structure, as this bacterial species contains a homologous glycosylation system (272). Interestingly, even though there is no consensus sequence identified for O-glycosylation, the *B. fragilis* O-OTase seems to be specific for the three-amino-acid long sequon D-(S/T)-(A/I/L/M/T/V). Mutation of the first Asp led to a loss of glycosylation in the proteins tested, and there was a clear requirement for an amino acid with at least one methy-group in its side chain in the position following the glycosylation site (273). Based on this sequence, more than 1000 putative glycoproteins were identified in *B. fragilis*, and by introducing this sequence to a putative α-fucosidase from *B. fragilis*, which does not carry a glycosylation motif and was shown to lack

glycosylation, site-specific glycosylation was achieved *in vivo* (273). However, the structure of the glycan has not yet been determined.

Lactobacillus species have been extensively studied owing to the importance of certain *Lactobacillus* strains as probiotics, evidence for protein glycosylation has recently emerged.

In *L. plantarum* WCFS1*,* a general glycosylation system has been described (274). This system is very similar to that initiating glycosylation of SRRPs, *i.e.* two GT, named as GtfA and GtfB in *L. plantarum,* but here will be referred to as Gtf1 and Gtf2 (so as to avoid confusion with the GtfA and GtfB from the $SecA_2/Y_2$ cluster) are involved in the addition of a single HexNAc molecule onto each glycosylation site of the acceptor proteins (274); deletion of either Gtf1 or Gtf2 leads to a loss of protein recognition by the GlcNAc-specific lectin wheat germ agglutinin (WGA). This suggests that both GTs are required for protein glycosylation and that the added sugar is most likely GlcNAc. These two enzymes contain a DUF1975 in N-terminus which probably mediates the interaction between the two GTs and the target proteins and a GT domain in C-terminus, as described for the SRRP-specific GtfA and GtfB, suggesting a similar mode of action. In contrast to the *L. monocytogenes* flagellar glycosylation and the O-GlcNAcylation systems that add a single β-O-GlcNAc, this system is likely to add α-linked O-GlcNAc, as Gtf1 is highly similar to GtfA from *S. pneumoniae* (267). The muramidase Acm2 from *L. plantarum* WCFS1 was identified as a glycoprotein targeted by this glycosylation system. Analysis of the glycosylation of the muramidase Acm2 by MS showed that it is glycosylated by single HexNAc residues at more than 20 sites, all found within the AST domain (275). This is in accordance to previous studies showing that O-glycosylation occurs in low complexity regions. By deleting the secretion signal peptide of Acm2, Fredrikesn et al. (2012) showed that glycosylation occurs intracellularly and therefore precedes secretion (275). It was also shown that glycosylation of Acm2 inhibited its enzymatic activity, by causing conformational changes. These were proposed to occur by interaction of GlcNAc moieties either with the active site, or with the SH3b motifs, which are responsible for

binding the GlcNAc-rich peptidoglycan layer (276). Glycosylation also increased the resistance of the AST domain against trypsin (276). MS analysis of surface proteins in *L. plantarum* WCFS1 revealed numerous additional proteins carrying O-GlcNAc residues, suggesting that this is a general glycosylation system. These proteins include among others DnaK, a chaperone involved in protein folding, PdhC, which is involved in the anaerobic metabolism of Lactobacilli, as well as a mucus binding protein, similar to MUB_{53608} from *L. reuteri* ATCC53608, probably involved in adhesion of the bacteria to the host surface (277). This analysis also identified a HexNAc-Hex moiety on gamma-D-glutamate-meso-diaminopimelate muropeptidase, which suggests that either the GlcNAc residues can be further extended by the action of other GTs, or that there is an additional glycosylation system in *L. plantarum* (277). In addition to WGA, *Dolichos biflorus agglutinin*, a lectin specific for α-GalNAc, and *Lens culinaris* lectin, which is specific for α-Man, were also shown to interact with *L. plantarum* proteins (274). This would also suggest the presence of additional glycosylation system(s). However, deletion of four other putative GTs (including one with a DUF1975), similar to GtfA from *S. parasanguinis* in *L. plantarum* WCFS1 did not lead to any changes in the recognition of the proteins by these lectins (274).

The Major secreted protein 1 (Msp1) is a muramidase, similar to Acm2, found to be glycosylated in *L. rhamnosus* GG (278). It has a predicted molecular weight (MW) of 48 kDa, but was found to migrate at 75 kDa on SDS-PAGE and interact with Concanavalin A (ConA), a lectin specific for Man and Glc residues. Msp1 shows low complexity, as it consists of 23% ala residues. Monosaccharide composition analysis of Msp1 confirmed the presence of Man, in agreement with ConA affinity to Msp1 (278). As reported for Acm2, the glycosylation of Msp1 protected the protein against proteases. However, in contrast to Acm2, glycosylation of Msp1 did not affect the hydrolytic activity of the enzyme or its ability to activate the Akt signalling pathway in Caco-2 cells (278). *L. rhamnosus* GG also expresses proteins that are responsible for the formation of pili, a rare feature in the *Lactobacillus* genus. These are composed of the three-protein complex spaCBA, which is assembled by a pilin-specific sortase (158) The spaCBA proteins have been

involved in adhesion to intestinal epithelial cells (IEC) and in the attenuation of proinflammatory responses from these cells (279). Atomic force spectroscopy (AFM) of the pili using functionalised tips with lectins specific for Man and Fuc suggested the presence of these two monosaccharides, in contrast to the glycosylation analysis of *L. rhamnosus* GG Msp1, which only showed Man residues and no Fuc. Furthermore, the glycosylated pili were shown to interact with dendritic cells *via* the DC-SIGN lectin, an important receptor of the immune system that recognises primarily high-mannose structures, and induce the expression of the anti-inflammatory cytokine IL-10, as well as IL-6 and IL-12p35 (280). While *L. rhamnosus* GG contains a pair of putative GTs containing a DUF1975, these have not been experimentally assessed for their involvement in glycosylation of either Msp1 or the spaCBA pilin.

Similar to pili, flagella are also rare structures in the *Lactobacillaceae* family. However, they were recently discovered in a motile strain of *Lactobacillus agilis* and were shown to be glycosylated by means of periodic acid/Schiff (PAS) stain (281), but the nature of glycosylation was not further investigated.

S-layer proteins (Slps) are expressed by many bacterial species and form a two-dimensional layer that surrounds the bacterial cells (282). In Gram-positive bacteria, Slps are found attached onto components of the PG layer, such as (lipo)teichoic acids or neutral polysaccharides (282). In *Lactobacillus* species, these proteins usually consist of a C-terminal carbohydrate-binding domain, used for attachment of the protein onto the cell wall, and a self-assembly N-terminal domain that forms the 2D layer (283). Although there are older reports of glycosylated Slps in *Lactobacillus helveticus* and *L. plantarum*, based on PAS stain (284), it is believed that *Lactobacillus* Slps are generally non-glycosylated (173). However, recent detailed studies of *L. kefir* strains (173), *L. acidophilus* (285) and *L. buchneri* (284) revealed glycosylated S-layer proteins. Screening of various *L. kefir* strains showed that Slp glycosylation is conserved within this species (173), but the nature of the glycan or the glycosylation mechanism remains

unexplored. SlpA in *L. acidophilus* NCFM was found to be glycosylated with glycans containing Man and Fuc, as shown by AFM experiments with specific lectins (285). Similarly to spaCBA pili from *L. rhamnosus,* SlpA from *L. acidophilus* induced the production of IL-10 from DC, by interacting with DC-SIGN (285). MS analysis of SlpA from *L. buchneri* CD034 showed that the protein is glycosylated on Ser residues in the sequon SSASSASSA, consistent with previous reports for O-glycosylation in low complexity and AST-rich regions. The glycans found on each glycosylation site consisted of on average 7 residues of α1-6 linked Glc. It was also suggested that glycosylation occurs extracellularly, as no glycosylated SlpA was found in the cytosolic fraction (284). In the same study a glycoside hydrolase (GH), belonging to GH family 25 according to the CAZy database classification, was also found to carry a modification corresponding to the mass of eight glucose units, suggesting that the system that is responsible for SlpA glycosylation may target other proteins as well (284).

Table 6 A summary of identified glycoproteins from *Lactobacillus* species.

Protein	Organism	Glycan	Method	Reference
Msp1	*L. rhamnosus* GG	Man-containing	Pro-Q Emerald stain, Lectin affinity (WB, AFM) MS	(278)
spaCBA	*L. rhamnosus* GG	Man and Fuc-containing	Lectin affinity (AFM, WB, ELLA)	(280)
$FliC_1$/$FliC_2$	*L. agilis*	uncharacterised	PAS-stain	(281)
SlpB/N	*L. buchneri* CD034, *L. buchneri* NRRL B-30929	Glc_1-Glc_7	MS	(284)
LbGH25B/N Putative glycosyl-hydrolase	*L. buchneri* CD034, *L. buchneri* NRRL B-30929	Glc_8	MS	(284)
Slp	*L. kefir*	uncharacterised	PAS stain	(284)
Acm2	*L. plantarum* WCFS1	GlcNAc	MS, lectin affinity (WB)	(275)
DnaK	*L. plantarum* WCFS1	$GlcNAc_1$, $GlcNAc_1Hex_1$	MS	(277)
Lp_2162 (muropeptidase)	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)
Lp_2260	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)

Lp_1643 (mucus binding protein)	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)
PdhC	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)
FtsY	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)
Lp_2793	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)
FtsK1	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)
Lp_3421 (muropeptidase)	*L. plantarum* WCFS1	$GlcNAc_1$, $GlcNAc_1Hex_1$	MS	(277)
FtsZ	*L. plantarum* WCFS1	$GlcNAc_1$	MS	(277)

In addition to the characterised glycosylation systems and glycoproteins described above (Table 6 provides a summary of the identified glycoproteins from *Lactobacillus sp.*), there is also indirect evidence for additional protein glycosylation systems in gut commensal bacteria. For instance, the presence of a functional $SecA_2/Y_2$ cluster in the murine isolate *L. reuteri* 100-23 (103) shows that this glycosylation system is not exclusive to pathogenic bacteria. In addition, MUB_{53608} from the porcine isolate *L. reuteri* ATCC 53608 was recently shown to interact with DC-SIGN *via* C-type lectins (175), suggesting the presence of a carbohydrate component. This, in addition to its aberrant electrophoretic mobility (154), suggests that that protein may be glycosylated. However, glycosylation of the major adhesins in *L. reuteri* strains has not been investigated.

1.5. Aims- objectives

The aim of this project is to analyse the secreted glycoproteome of *L. reuteri* strains isolated from different vertebrate hosts. Specific objectives include (i) identifying the cell-surface and secreted glycosylated proteins and (ii) determining the nature of the glycans.

Chapter 2

Materials and Methods

2.1. Materials

Chemical reagents were purchased from Sigma Aldrich (UK), and enzymes were purchased from New England Biolabs (UK), unless otherwise stated. Antibodies with their suppliers are listed in Table 7. Plant lectins were purchased from Vector Labs (UK) (Table 8). Human intelectin-1 (h-Int1) was provided by Dr. Amanda Ducan and Prof. Laura Kiessling (University of Wisconsin-Madison, WI, USA).

Table 7 Antibodies used in this study.

Antibodies	**Specificity**	**Supplier/Reference**
Rabbit anti-MUB_{53608}-R5 polyclonal ab	MUB_{53608}-R5 repeat	(155)
Rabbit anti-MUB_{53608}-RI polyclonal ab	MUB_{53608}-RI repeat	(155)
Rabbit anti-$SRRP_{53608}$ polyclonal ab	Binding region of $SRRP_{53608}$	(286)
Mouse anti-β-O-GlcNAc monoclonal antibody	Serine O-linked β-GlcNAc	Millipore, UK
Mouse anti-α-gal monoclonal antibody	Galα1-3Galβ1-(3)4GlcNAc-R	Absolute antibody, UK
StrepMAB-Classic-HRP conjugate	Strep-tag II peptide	IBA lifesciences, Germany
Goat anti-rabbit antibody, AP conjugate	Rabbit antibodies	Sigma Aldrich, UK
Anti-mouse antibody, HRP conjugate	Mouse antibodies	Jackson laboratories, USA

Table 8 Fluorescein-labelled plant lectins used in this study.

Lectin	**Abbreviation**	**Specificity**	**Fluorescein/ protein ratio**
Concanavalin A	ConA	α-mannose	5:1
Lotus tetragonolobus lectin	LTL	α-L-fucose	4:1
Peanut agglutinin	PNA	Gal-(β-1,3)-GalNAc (T-antigen)	4.2:1
Ricinus communis agglutinn	RCA	Gal or GalNAc	6:1
Sambucus nigra agglutinin	SNA	α-2,6-linked sialic acid (α-2,3-linked sialic acid with lower affinity)	7.7:1
Ulex europaeus agglutinin	UEA	α-L-fucose (differs from LTL)	2.6:1
Wheat germ agglutinin	WGA	GlcNAc or sialic acid	3.1:1
Griffonia simplicifolia Lectin I isolectin B4	GSL-I B4	α-Gal	2.5:1

Maackia amurensis lectin I	MAL I	Galβ-1,4-GlcNAc	8.9:1
Maclura pomifera lectin	MPL	Galβ-1,3-GalNAc	3.9:1

Table 9 Antibiotics used in this study

Antibiotic	**Concentration (μg/ml)**	
	E. coli	*L. reuteri*
Kanamycin (Kan)	50	N/A
Chloramphenicol (Cm)	34	10
Erythromycin (Ery)	250	10
Ampicillin (Amp)/ Carbenicillin (Cb)	100	N/A

2.2. Media and bacterial growth

All strains used in this study are listed in Table 10.

Table 10 Bacterial strains used in this study.

Strain	**Antibiotic required**	**Reference**
Lactobacillus reuteri 100-23 wild type (WT)	N/A	(109)
Lactobacillus reuteri 100-23 *Δasp2*	Ery	Prof. Jens Walter (University of Alberta, Edmonton, Canada), (109)
Lactobacillus reuteri 100-23 *ΔsecA2*	Ery	(109)
Lactobacillus reuteri 100-23 *ΔgtfB*	Ery	Prof. Jens Walter (University of Alberta, Edmonton, Canada) (109)
Lactobacillus reuteri 100-23 *Δsrr*	Ery	(109)
Lactobacillus reuteri 100-23 *Δgtf1*	N/A	Dr. Rebbeca Duar and Prof. Jens Walter (University of Alberta, Edmonton, Canada) – This study
Lactobacillus reuteri 100-23 *Δftf*	Ery	(287)
Lactobacillus reuteri ATCC 53608	N/A	(288)
Lactobacillus reuteri 1063N	N/A	(151)
Lactobacillus reuteri MM4-1a	N/A	(147)
Escherichia coli DH5a	N/A	
Escherichia coli DH5a_pOPINF	Amp/Cb	
Escherichia coli DH5a_pOPINF-gtf1	Amp/Cb	This study
Escherichia coli DH5a_pOPINF-gtf2	Amp/Cb	This study
Escherichia coli DH5a_pOPINF-gtfA	Amp/Cb	This study
Escherichia coli DH5a_pOPINF-gtfB	Amp/Cb	This study
Escherichia coli DH5a_pOPINF-gtfC	Amp/Cb	This study
Escherichia coli DH5a_pRSFDuet-1	Kan	This study
Escherichia coli DH5a_pRSFDuet-1-gtf1	Kan	This study
Escherichia coli DH5a_pRSFDuet-1-gtf1/2	Kan	This study
Escherichia coli BL21 (DE3)	N/A	

Escherichia coli BL21 (DE3)_pOPINF	Amp/Cb	
Escherichia coli BL21 (DE3)_pOPINF-gtf1	Amp/Cb	This study
Escherichia coli BL21 (DE3)_pOPINF-gtf2	Amp/Cb	This study
Escherichia coli BL21 (DE3)_pOPINF-gtfA	Amp/Cb	This study
Escherichia coli BL21 (DE3)_pOPINF-gtfB	Amp/Cb	This study
Escherichia coli BL21 (DE3)_pOPINF-gtfC	Amp/Cb	This study
Escherichia coli BL21 (DE3)_pRSFDuet-1	Kan	
Escherichia coli BL21 (DE3)_pRSFDuet-1-gtf1	Kan	This study
Escherichia coli BL21 (DE3)_pRSFDuet-1-gtf1/2	Kan	This study

Typically, de Man-Rogosa-Sharpe (MRS; Table 11; ThermoFisher Scientific, UK) medium or Lactobacillus defined medium-II (LDM-II; prepared in house, Table 11), supplemented with the appropriate antibiotics (Table 10), was used for the growth of *L. reuteri*.

E. coli strains were grown Luria-Bertani (LB) medium (Table 12), supplemented with the appropriate antibiotics (Table 10). Where plates were used, LB was supplemented with 1.5% w/v bacteriological agar. Super optimal broth with catabolite repression (S.O.C.) medium (Table 12) was used for the recovery of *E. coli* strains after heat-shock transformation. Auto-induction medium (AIM; Table 12; Formedium, UK) was used for the overexpression of recombinant proteins.

Table 11 Composition of media used for *L. reuteri* growth.

LDM II			
Ingredients	**Amount per L**	**Ingredients**	**Amount per L**
K_2HPO_4 (anhydrous)	1.5 g	Glutamic acid	50 mg
KH_2PO_4 (anhydrous)	1.5 g	Arginine	50 mg
Sodium acetate	15 g	Lysine	50 mg
Sodium citrate	0.22 g	Thiamine-HCl	0.2 mg
Tryptophan	50 mg	p-Aminobenzoic acid	0.04 mg
Asparagine	50 mg	Calcium pantothenic acid	0.4 mg
Cysteine	50 mg	Niacin	1.0 mg
Glycine	50 mg	Pyridoxine-HCl	0.5 mg
Serine	50 mg	Biotin	0.05 mg
Alanine	50 mg	Folic acid	0.1 mg
Phenylalanine	50 mg	Riboflavin	0.4 mg
Histidine	50 mg	Adenine sulphate	10 mg
Isoleucine	50 mg	Uracil	20 mg
Methionine	50 mg	Guanine-HCl	10 mg
Proline	50 mg	Cytidine (acid)	50 mg

Threonine	50 mg	Thymidine	1.6 mg
Valine	50 mg	Tween-80	1.0 ml
Tyrosine	50 mg	$MgSO_4$ H_2O	0.163 g
Leucine	50 mg	$MnSO_4$ $7H_2O$	23.4 mg
Glutamine	50 mg	$FeSO_4$ $7H_2O$	13 mg
Aspartic acid	50 mg	Sucrose	30 g
MRS			
Ingredients	**Amount per L**	**Ingredients**	**Amount per L**
Enzymatic digest of casein	10 g	$MgSO_4$ $7xH_2O$	0.2 g
Meat extract	10 g	$MnSo_4$ $4xH_2O$	0.05 g
Yeast extract	4 g	K_2HPO_4 (anhydrous)	2 g
Tri-ammonium citrate	2 g	Sorbitan mono-oleate	1.08 g
Sodium acetate	5 g	Glucose	20 g

Table 12 Composition of media used for *E. coli* growth.

LB		**S.O.C.**	
Ingredients	**Amount per L**	**Ingredients**	**Amount per L**
Tryptone	10 g	Tryptone	10 g
Yeast extract	5 g	Yeast extract	5 g
NaCl	10 g	NaCl	4 g
AIM			
Tryptone	12 g	Na_2HPO_4	7.1 g
Yeast extract	24 g	Glucose	0.5 g
$(NH_4)_2SO_4$	3.3 g	α-Lactose	2 g
KH_2PO_4	6.8 g	$MgSO_4$	0.15

2.2.1. *L. reuteri g*rowth assays

2.2.1.1. 96-well plate format assay

To assess the growth of *L. reuteri* strains on different disaccharides, LDM-II (200 µl), supplemented with 2% w/v sucrose, maltose or lactose was added to each well on a 96-well plate. The wells were inoculated with 1 µl from an overnight culture of *L. reuteri* in MRS. Wells with no carbon source, or no inoculum were included as controls. The OD_{600} was monitored for 24 h on an Infinite® F50 plate reader (Tecan, Switzerland) and plotted as a function of time.

2.2.1.2. End-point growth assay

To assess the ability of *L. reuteri* strains to utilise different carbohydrates, overnight cultures in MRS (50 µl) were used to inoculate 10 ml filter-sterilised LDM-II containing appropriate carbohydrates at 2% w/v concentration. The reactions were incubated under static conditions

at 37°C for 48 h. OD_{600} measurements were taken after 48 h. Cultures in LDM-II supplemented with 2% sucrose were used as control.

2.3. Sedimentation assay

L. reuteri 100-23 WT and *Δgtf1*, or *L. reuteri* ATCC 53608 and 1063N were grown from glycerol stocks in MRS overnight. The cells were centrifuged at 4000 g for 5 min and washed twice with phosphate buffered saline (PBS). The OD_{600} was normalised to 1 unit for each sample. 2 ml of each sample was loaded to a cuvette. The wild type cells were inserted into the reference cell of a U-3010 spectrophotometer (Hitachi, UK) and the mutant cells in the measuring cell. The change in OD_{600} was monitored for 18 h. At the end of the experiment, the samples were vortexed and the OD_{600} was measured.

2.4. Preparation of *E. coli* competent cells

E. coli DH5α or BL21 (DE3) were streaked from glycerol stock onto LB agar plates, containing appropriate antibiotics, if needed. The plates were incubated overnight at 37°C. A single colony was used to inoculate 25 ml LB broth and the culture was grown for 6 h, until the OD_{600} was approximately 0.8-1. The culture was placed on ice for 10 min and then centrifuged at 3000 g and 4°C for 3 min. The pellet was resuspended in ice-cold $CaCl_2$ 0.1M and incubated on ice for 20 min. The cells were centrifuged again and after removal of the supernatant were resuspended in 5 ml ice-cold $CaCl_2$ 0.1 M in 15% glycerol. The suspension was aliquoted and stored at -80°C.

2.5. Transformation of *E. coli*

An ice-thawed aliquot of *E. coli* DH5α or BL21 (DE3) competent cells (100 µl) was mixed with 1 µl of plasmid (Table 13), or all of the In-Fusion® reaction (see 2.7.2.2) and the cells were kept on ice for 15 min. The cells were then heat shocked at 42 °C for 45 s and then put back on ice for 2 min. S.O.C. medium (900 µl) was added to the cells and they were incubated at 37°C for 1 h. An

aliquot (typically 100 µl) was plated onto LB plates supplemented with appropriate antibiotics and the culture was incubated overnight at 37°C.

2.6. Heterologous expression in *E. coli*

Glycerol stock of *E. coli* recombinant clones (Table 10) was used to inoculate 10 ml LB broth supplemented with appropriate antibiotics (Table 9). The culture was incubated for 16 h at 37°C and 250 rpm. This culture was used to inoculate either LB or AIM, with the necessary antibiotics at 1% vol/vol. The cultures were incubated at 37°C for 2.5 h and 16°C for 3 days at 250 rpm. If LB was used, IPTG was added to the culture before temperature decrease at a final concentration of 1 mM.

2.7. DNA/RNA manipulation

2.7.1. Polymerase chain reaction (PCR)

PCR reactions were performed using Q5 DNA polymerase (New England Biolabs, UK), following the manufacturer's instructions. Typically, PCR reactions were set up at a final volume of 50 µl and contained 10 µl of reaction buffer, 1 µl of 10 mM dNTPs, 2.5 µl of each primer, 1 µl of template DNA, and 0.5 µl of Q5 polymerase. H_2O was used to adjust the volume to 50 µl. When the PCR product was only required for agarose electrophoresis (e.g. colony PCR or RT-PCR), the PCR reactions were scaled down proportionally to a final volume of 10 µl. PCR reactions were performed on using Veriti Thermal Cycler (ThermoFischer Scientific, UK). A typical protocol consisted of an initial DNA denaturation step at 98°C for 30 s, followed by 35 cycles of denaturation at 98°C for 10 s, annealing between 55 and 68°C for 20 s, depending on the properties of the primers used, and extension at 72°C for 30 s/kb. The reaction was completed with a final extension step for 5 min. For a list of primers, see Table 15. The PCR products were analysed on an agarose gel (1% w/v agarose in 1x Tris-acetic acid-EDTA (TAE; 40 mM, 20 mmM and 1 mM, respectively) buffer, at 100 V for 30 min. For RT-PCR products, 1.2% w/v agarose was used and the electrophoresis was performed at 100 V for 1 h.

2.7.2. Preparation of recombinant plasmids

2.7.2.1. Plasmid preparation

Plasmids were extracted from 1.5 ml of overnight *E. coli* cultures, using the Monarch® Plasmid Miniprep Kit (New England Biolabs, UK), following the manufacturer's instructions.

2.7.2.2. Cloning

Restriction digests of the purified plasmids were performed using High Fidelity® (HF) restriction enzymes (New England Biolabs, UK) for 15 min at 25-37^{o}C, depending on the enzyme used. Typically, a reaction contained a maximum of 1 µg DNA, 2 µl CutSmart® reaction buffer, and 0.5 µl of the required restriction enzyme in a final volume of 20 µl. The reaction products were analysed by agarose electrophoresis, and digested DNA was purified using the Monarch® PCR and DNA Cleanup Kit (New England Biolabs, UK). The digested vectors were mixed in 1:1 ratio with the desired insert before addition of the In-fusion buffer and enzyme (Clonetech laboratories, USA), as instructed by the manufacturer. The reaction was incubated at 50^{o}C for 15 min and then quenched on ice. The reaction was used to transform *E. coli* DH5α cells, as described in section 2.5. After plating and growth at 37 ^{o}C for 16 h, three colonies were picked and grown in LB medium with the appropriate antibiotic for 6 h. Plasmids were prepared from the three cultures and digested with appropriate enzymes to confirm insertion. The reaction products were analysed by agarose electrophoresis, as described above. Positive clones were sequenced to confirm correct orientation and fidelity. The plasmids were then used to transform *E. coli* BL21 (DE3), as described in section 2.5.

Table 13 Plasmids used in this study

Plasmid	Selection marker	Restriction enzymes used for cloning and plasmid analysis	Reference
pRSFDuet-1	Kan		Dr. Susan Schlimpert (John Innes Centre, UK)
pRSFDuet-1_gtf1 (first position)	Kan	PstI/HindIII	This study
pRSFDuet-1_gtf1/2 (second position)	Kan	EcorV/KpnI	This study

pOPIN-F	Amp/Cb		
pOPIN-F_gtf1	Amp/Cb	HindIII/KpnI	This study
pOPIN-F_gtf2	Amp/Cb	HindIII/KpnI	This study
pOPIN-F_gtfA	Amp/Cb	HindIII/KpnI	This study
pOPIN-F_gtfB	Amp/Cb	HindIII/KpnI	This study
pOPIN-F_gtfC	Amp/Cb	HindIII/KpnI	This study
pET28b_GH89 (α-GlcNAcase)	Kan		Dr. Lucy Crouch and Dr. David Bolam (University of Newcastle, UK)

2.7.3. DNA extraction from *L. reuteri*

Genomic DNA (gDNA) was extracted from *L. reuteri* grown in MRS for 18 h, using the GeneJET Genomic Purification Kit (ThermoFischer Scientific, UK), following the manufacturer's instructions for preparation of DNA from Gram-positive bacteria, with minor modifications, including the addition of 10 U/ml mutanolysin in the lysis buffer and extension of the incubation period to 1 h. DNA preparations were stored at -20^{o}C until needed.

2.7.4. RNA extraction from *L. reuteri*

L. reuteri 100-23 and ATCC 53608 cultures in MRS were used to inoculate fresh MRS and LDM-II, respectively, at 1 % vol/vol. The cultures were incubated under static conditions at 37^{o}C for 24 h. An aliquot (1.5 ml) was collected from each culture after 7 h incubation, when the OD_{600} reached 1.2. the cells were centrifuged at 5000 g for 10 min and after removal of the spent media (SM), the cell pellets were treated with RNAprotect, (Qiagen, USA), following the manufacturer's instructions. The treated cells were centrifuged at 5000 g for 5 min and the supernatant was removed. The cell pellet was stored at -20^{o}C overnight. A second aliquot was collected after 24 h of growth, and treated in the same way. RNA was extracted from the cell pellets using the RNeasy Mini Kit (Qiagen, USA). The cells were resuspended in lysis buffer (100 µl, Tris-HCl 30 mM, EDTA 1 mM, pH 8), containing 10 µl Qiagen Proteinase K, 20 mg/ml lysozyme and 5 U/ml mutanolysin) and the lysis reaction was incubated on an Eppendorf thermomixer C at 25^{o}C, 1000 rpm for 1 h. RLT buffer (700 µl) was added to the samples. These were transferred to autoclaved 2-ml safe-lock tubes with acid washed glass beads (150-600 µm diameter). The

cells were mechanically disrupted 3 times for 1 min at 6.5 m/s on a FastPrep®-24 Classic homogeniser (MP Biomedicals, USA). Cells were kept on ice for 2 min in-between runs. The samples were centrifuged at 17000 g for 10 s. Ethanol (80%, 590 µl) was added to the supernatant, the lysate was loaded onto an RNeasy Mini spin column and centrifuged for 15 s at 17000 g. The column was washed with 350 µl RW1 buffer. On column DNA digestion was carried out, using Qiagen DNase I, following the manufacturer's instructions. The column was then washed with 500 µl RPE buffer and RNA was eluted twice with 50 µl RNase-free H_2O. The concentration was measured on a NanoDrop 2000 (ThermoFisher Scientific, UK) and the recovered RNA was further treated with TURBO™ DNase (ThermoFisher Scientific, UK), following the manufacturer's instructions. The RNA samples were stored at -20^{o}C until needed.

2.7.5. RT-PCR analysis

RNA purified from *L. reuteri* 100-23 and ATCC 53608 was treated with TURBO DNase (ThermoFischer Scientific, UK), following the manufacturer's instructions. cDNA was synthesised from the recovered RNA, using the Quantitect Reverse Transcription kit (Qiagen, USA), following the manufacturer's instructions. Briefly, 12 µl of RNA was treated with 2 µl gDNA wipeout at 50^{o}C for 15 min, before addition of the reverse transcription buffer, the primers and the reverse transcriptase. Initially, the random hexamers provided with the Reverse Transcription kit was used as primers. When the RT-PCR did not yield any products, gene specific reverse primers (Table 14) were used for cDNA synthesis, at a final concentration of 50 nM each.

Table 14 Primers used for RT-PCR.

Primer name	Sequence (5'-3')	DNA template	Intergenic region targeted
10023gtfc-f-op	ACTTTGCACGCTTGATCC	*L. reuteri* 100-23	2500070887-2500070888 (*gtfC-secY*$_2$)
10023secy2-r-op	GCACAGGAACGGTACAACC		
10023secy2-f-op	TGATTGTAATGTTTATGCAGG		2500070888-2500070889 (*secY*$_2$*-asp1*)
10023asp1-r-op	AGTTAATTGGTTTAATTTAGTGG		
10023asp1-f-op	CATTAATCAAGGATTAGATTATTTCC		2500070889-2500070890 (*asp1-asp2*)
10023asp2-r-op	ATCCGCTAATTTGGTCC		
10023asp2-f-op	GAAAGCTAATAAGGCAATCC		2500070890-2500070891 (*asp2-asp3*)
10023asp3-r-op	GAGTTATGAAACTCAACACTAGC		
10023asp3-f-op	CCATTACTGCCAGCTACTACTTGG		2500070891-2500070892 (*asp3-SecA*$_2$)
10023secA2-r-op	GCCGCAAATGCCCGTGG		
10023secA2-f-op	ATAATTTGATGCTGAGTACC		2500070892-2500070893 (*secA*$_2$*-gtfA*)
10023gtfA-r-op	AAGTCAGCGATATTATTTCC		
10023gtfA-f-op	TGCTGACTTGACTAAATTCC		2500070893-2500070894 (*gtfA-gtfB*)
10023gtfB-r-op	AAAATTTAATGGGAGATTCC		
10023gtfB-f-op	TAGCTAGTGAAGTTAGTGTCC		2500070894-2500070895 (*gtfB-gtfB*$_{C\text{-}term}$)
70895-r-op	TGAGCGAAAACTTACAGG		
70895-f-op	TGCTCTTGATTTTGATATTGG		2500070895-2500070896 (*gtfB*$_{C\text{-}term}$*- gtfD*)
70896-r-op	TATCGTAATTTGCGCATAGG		
70896-f-op	ATTGGTAAGAATAAGCATACG		2500070896-2500070897 (*gtfD-gtfE*)
70897-r-op	CGTAGATTTTGACATCACG		
70897-f-op	TGGCGAATAAGCATACC		2500070897-2500070898 (*gtfE-gtfF*$_{N\text{-}term}$)
70898-r-op	AACGATTAATATTCACAAACC		
70898-f-op	TTACTTTGATTAATGGTCAGG		2500070898-2500070899 (*gtfF*$_{N\text{-}term}$*-Tnase*)
70899-r-op	GATTCACTTTTATAACGAACG		

70899-f-op	GCCCTTGCTTGTTGTAGAGAC		2500070899-2500070900 (*tnase- gtfF$_{C\text{-}term}$*)
70900-r-op	TTTGCTAAAATTTGTTGATTATGAC		
70900-f-op	GGTATGCTTATTCTTTAACTT		2500070900-2500070901 (*gtfF$_{C\text{-}term}$-RTase*)
70901-r-op	TTTACCTTCACCATAATCG		
70901-f-op	CTTATGAATATGTCCAAGACG		2500070901-2500070902 (*RTase-srr*)
70902-r-op	AAAAGAGTTTGAAGAAGCG		
70902_int-F	CCAGTGCATCTACCAGTACCTCG		2500070902 (*srr, internal region*)
70902_int_R	GAGGCTAGGGCAGCATTCG		
SRR-ts-F	TTTCGATGAGTGAAAGTCTCAGC	*L. reuteri* ATCC 53608	LRATCC53608_0906 (*srr, internal region*)
SRR-ts-R	TTGGTAGTCTTAAGACCATTCCC		
srr-F-op	TGGTAATGAAAAGCATTCAACGG		LRATCC53608_0906 - LRATCC53608_0907 (*srr – gtfC*)
nss-R-op	GGCAATTTGATTTTGAGCAATT		
op_nss_f	AATGAAATGGTCCAACGAG		LRATCC53608_0907 - LRATCC53608_0908 (*gtfC-secY$_2$*)
op_secY2_r	TGTCGAAATGTTAACTAATGGC		
op_secY2_f	GATTGTAATGTTTATGCAGGG		LRATCC53608_0908 - LRATCC53608_0909 (*secY$_2$-asp1*)
op_asp1_r	TTGCGTTGGTGAGATCG		
op_asp1_f	GCAAATTTCTGATATCAATAAGGG		LRATCC53608_0909 - LRATCC53608_0910 (*asp1-asp2*)
op_asp2_r	CTGGACGCCTTCTGTAATTATC		
op_asp2_f	CTATTTGTTAGTAAAGGATTTCC		LRATCC53608_0910 - LRATCC53608_0911 (*asp2-asp3*)
op_asp3_r	GGACCAATTTAGTTGCC		
op_asp3_f	CTTTCCACCATCAATGCAG		LRATCC53608_0911 - LRATCC53608_0912 (*asp3-SecA$_2$*)
op_secA2_r	CGTGGTAAGATATCATCAACTG		
op_secA2_f	AAACTATCAGTTATCGTTCAGC		LRATCC53608_0912 - LRATCC53608_0913 (*secA$_2$-gtfA*)
op_gtfA_r	AATCTGATCAGCATTAAAACC		
op_gtfA_f	ACTGAAGATTGGAGTAATTCAC		LRATCC53608_0913 - LRATCC53608_0914 (*gtfA-gtfB*)
op_gtfB_r	GCCAGTAACGTGGTAAGG		

Table 15 Primers used for gene expression. The underlined sequences indicate the overlapping region of the insert and the plasmid.

Primer name	Sequence (5'-3')	DNA template	Region targeted	Vector used
0089-F	AAGTTCTGTTTCAGGGCCCGATGCTTTTTTTCCTAAACGATAATA	*L. reuteri* ATCC 53608	*gtf1*	pOPINF
0089-R	ATGGTCTAGAAAGCTTTACTCTTCCCCCTTTTCCTTCATG			
0090-F	AAGTTCTGTTTCAGGGCCCGGTGTATTTTTTTGTAAACCAATACT		*gtf2*	
0090-R	ATGGTCTAGAAAGCTTTAAATAGCTGATAGTTGTTGCC			
0913-F	AAGTTCTGTTTCAGGGCCCGATGACTGTATATAATATTAATTTAGGAATT		*gtfA*	
0913-R	ATGGTCTAGAAAGCTTTAAGCATGCTGTAACTCTCCTATCAATT			
0914-F	AAGTTCTGTTTCAGGGCCCGATGCTTAATTTATTTGATAACTTTG		*gtfB*	
0914-R	ATGGTCTAGAAAGCTTTATTCACTTTGCAAGGCTCCAATC			
0907-F	AAGTTCTGTTTCAGGGCCCGTTGACAGTTCATATTACTAACCTTT		*gtfC*	
0907-R	ATGGTCTAGAAAGCTTTATTGATAATATAATTTAAATACTGCTTC			
EcoRI -Gtf1-F	GCCAGGATCCGAATTCGATGCTTTTTTTCCTAAACGATAATA		*gtf1*	pRSFDuet-1 Cloning site 1
HindIII -Gtf1-R	ATGCGGCCGCAAGCTTTACTCTTCCCCCTTTTCC			
EcoRV -Gtf2-F	AGATCTCAATTGGATAGTGTATTTTTTTGTAAACCAATAC		*gtf2*	pRSFDuet-1 Cloning site 1
KpnI -Gtf2-R	TTACCAGACTCGAGGTTAAATAGCTGATAGTTGTTG			
pRSFDuet-1-F1	GGATCTCGACGCTCTCCCT	pRSFDuet-1	Cloning site 1	N/A
pRSFDuet-1-R1	GATTATGCGGCCGTGTACAA			
pRSFDuet-1-F2	TGTACACGGCCGCATAATC		Cloning site 2	
pRSFDuet-1-R2	GCTAGTTATTGCTCAGCGG			

2.8. Protein purification

2.8.1. Spent media (SM) Protein preparation from *L. reuteri*

L. reuteri strains were grown from glycerol stocks in MRS, overnight at 37°C under static conditions. This culture was used to inoculate fresh MRS or LDM-II (289), supplemented with 2% sucrose, unless otherwise stated, at 0.2 % vol/vol. The fresh cultures were incubated under static conditions at 37°C overnight. Cells were centrifuged at 4000 g for 5 min. The SM fraction was concentrated 10-fold by spin filtration using Vivaspin protein concentrator 10 kDa MWCO (Sartorius, UK) and proteins were extracted by addition of 1.33 vol of chloroform and 2.67 vol of methanol. The supernatant was removed after centrifugation at 17,000 g for 10 min and proteins were resuspended in H_2O and stored at -20 °C.

2.8.2. Purification of MUB_{53608} and $SRRP_{53608}$ from *L. reuteri* ATCC 53608

Glycerol stock of *L. reuteri* ATCC 53608 was used to inoculate MRS medium. The culture was incubated under static conditions overnight and was used to inoculate 1 L of LDM-II at 1% vol/vol. The fresh culture was grown under anaerobic and static conditions at 37°C for 24 h. The cells were removed by centrifugation at 10000 g, 4°C and SM proteins were precipitated upon addition of ammonium sulfate, at a final concentration of 60%. The precipitate was dissolved in H_2O and the proteins were extracted in a step-wise manner, using a triple phase partitioning (TTP) system. Ammonium sulphate was again added to the sample to a final concentration of 20% and one volume of tert-butanol was added to the sample. After thorough mixing by vortexing, the sample was centrifuged at 5000 g for 10 min. The upper phase was removed, the protein pellet in the interphase was recovered and dissolved in H_2O and CHAPS (3-[(3-cholamidopropyl) dimethylammonio]1-propanesulfonate) was added to the sample at 0.5% w/v final concentration. MUB_{53608} was purified from this fraction by gel permeation chromatography (GPC), using a Superose 6 10/30 GPC column (GE Healthcare, UK). PBS supplemented with 0.5% CHAPS was used as the elution buffer, at a flow rate of 0.4 ml/min. The collected fractions were analysed by SDS-PAGE and fractions containing MUB_{53608} and $SRRP_{53608}$ were pooled together

and dialysed in a 100 kDa MWCO dialysis membrane against 50 mM ammonium bicarbonate for 48 h, with 4 changes of buffer. The dialysed sample was loaded onto an agarose-bound wheat germ agglutinin (aWGA) affinity column (Vector labs, UK), which was pre-conditioned with 10 vol of lectin washing buffer [10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, 150 mM NaCl]. The column was washed with 5 vol of lectin washing buffer. The flow-through and wash fractions containing MUB_{53608} were concentrated by spin-filtration using a 100 kDa MWCO Vivaspin 2 spin filter. Bound $SRRP_{53608}$ was eluted with 3 vol of wash buffer, supplemented with 0.5 M GlcNAc. The eluant and the concentrated wash fraction were dialysed using a 100 kDa MWCO dialysis membrane against 50 mM ammonium bicarbonate for 48 h, with four changes of buffer, and stored at 4°C.

For purification of truncated MUB_{1063N} ($tMUB_{1063N}$), a similar workflow was followed, with the following modifications. *L. reuteri* 1063N was grown in MRS. After GPC, the fractions containing $tMUB_{1063}$ were pooled and were not further purified by affinity chromatography.

2.8.3. Purification of $SRRP_{100\text{-}23}$ from *L. reuteri* 100-23

An overnignt culture of *L. reuteri* 100-23 in MRS was used to inoculate 1 L of LDM-II, supplemented with 2% maltose. The culture was incubated under static conditions overnight at 37°C and then the cells were removed by centrifugation at 10000 g for 10 min. $SRRP_{100\text{-}23}$ was purified from the supernatant by an aWGA affinity chromatography column, as described for $SRRP_{53608}$ (section 2.8.2). The eluted protein was dialysed extensively in ammonium bicarbonate, to remove the excess of GlcNAc and then stored at -20°C.

2.8.4. IMAC purification of recombinant proteins

After induction of the expression (see section 2.6), cells were pelleted by centrifugation at 5000 g and washed twice with Tris buffer (50 mM Tris, 150 mM NaCl, pH 7.5). Cells were either lysed using the BugBuster® reagent (EMD Millipore), following the manufacturer's instructions, or resuspended in Tris buffer and lysed by sonication (15 µm, 15 cycles of 10 s sonication, 15 s

break on ice), using a Soniprep 150 (MSE, UK). The cell debris and insoluble material were removed by centrifugation at 12000 g for 15 min and the soluble and insoluble fractions were analysed by SDS-PAGE. Where the target protein was found in the soluble fraction, this was loaded onto a His-bind® resin (Bio-Rad, UK) column, charged with Ni^{2+}, following the manufacturer's instructions. The unbound material was washed with 10 column vol of Tris buffer supplemented with 10 mM imidazole. The bound protein was eluted with 6 column vol of Tris buffer supplemented with either 250 mM imidazole or 100 mM EDTA. The eluted protein was concentrated by spin filtration, using a 10 kDa MWCO Vivaspin® Turbo 15 spin filter (Sartorious, UK) and buffer exchanged into Tris buffer using a PD-10 desalting columns (GE Healthcare Lifesciences, UK), following the manufacturer's instructions.

2.9. Protein analysis

2.9.1. SDS-PAGE

Typically, 6.5 µl of protein was mixed with 2.5 µl lithium dodecyl sulphate (LDS) buffer 4x (ThermoFischer Scientific, UK), and 1 µl dithiothreitol (DTT; 100 mM). The samples were heated at 70^{o}C for 7 min and were then loaded onto 4-12% Bis-Tris NuPAGE polyacrylamide gel (ThermoFischer Scientific, UK). The electrophoresis was typically performed at 200 V for 50 min. The gel was then washed in H_2O before staining with InstantBlue (Expedeon, UK), for 30 min, before being washed with H_2O. Alternatively, the gel was used for western blotting.

2.9.2. Western blot analysis

Proteins from SDS-NuPAGE gels were blotted onto an Amersham™ Hybond® P Western blotting PVDF membrane (GE Healthcare Lifesciences, UK), using the XCell II Blot module (ThermoFischer Scientific, UK), following the manufacturer's instructions for assembly and run with the following modifications: no methanol was used in the transfer buffer and transferring was performed for 2 h for large proteins (> 300 kDa). The membrane was blocked with Pierce protein-free PBS blocking buffer (ThermoFisher Scientific, UK) for 1 h and then immersed in 10 ml PBS blocking buffer supplemented with the primary antibody (Table 7) at appropriate concentration for 1 h.

The probed membrane was then washed three times with PBS, supplemented with 0.1% Tween®20 (PBST) and then probed with the appropriate secondary antibody (Table 7), before being washed again three times with PBST. When horseradish peroxidase (HRP)-conjugated antibody was used, the blots were visualised by the addition of the chromogenic 1-step 3,3',5,5'-Tetramethylbenzidine (TMB) blotting substrate solution (ThermoFisher Scientific, UK), until bands appeared. The blot was then washed extensively with H_2O. When alkaline phosphatase (AP)-conjugated antibody was used, the membrane was first washed once with 10 ml Tris-HCl 0.1 M, pH 9.6, before 10 ml of the visualisation solution (40 µM $MgCl_2$, 0.1 mM nitroblue tetrazolium, 0.1 mM 5-bromo-4-chloro-3indolyl-phosphate toluidine) was added until bands appeared. The blots were then washed extensively with H_2O. After colour development, the blots were air-dried and then scanned on a GS-800 calibrated densitometer (Bio-Rad, UK).

Alternatively, PBS blocking buffer supplemented with 5 µg/ml of an appropriate fluorescein isothiocyanate FITC- labelled lectin (*f*-lectin; Table 8) was used instead of primary antibodies for 1 h. The membrane was then washed with PBST three times and scanned using a Pharos-FX™ Plus molecular imager (Bio-Rad, UK), using excitation and emission wavelengths of 495 and 520 nm, respectively.

2.9.3. Slot blot

Slot blots were performed using a Hoefer™ PR648 blotting manifold (ThermoFischer Scientific, UK). The protein sample was loaded onto a prewet Amersham™ Hybond® P Western blotting PVDF membrane (GE Healthcare Lifesciences, UK), followed by two washes with PBS. The membrane was blocked and probed with *f*-lectins or antibodies, as described in section 2.9.2.

2.9.4. Trypsin digestion and mass spectrometry

Protein bands of interest were excised from SDS-NuPAGE gels and cut up to small cubic pieces. The gel plugs were washed twice with 200 µl of ABC buffer (200 mM aqueous ammonium bicarbonate in 50% acetonitrile; ACN) for 15 min and then twice with ACN for 10 min. The excess

of ACN was removed and the gel plugs were air-dried for 15 min. A DTT solution (200 µl, 10 mM in 50 mM ammonium bicarbonate) was added and the samples were incubated at 60^{o}C for 30 min. Then, the DTT solution was removed and, upon addition of 200 µl iodoacetamide solution (10 mM in 50 mM ammonium bicarbonate), the plugs were incubated at room temperature in the dark for 30 min. The iodoacetamide solution was removed and the washing steps were repeated. Trypsin Gold (10 µl; 10 ng/µl; Promega, UK) was added to the gel plugs along with equal amount of 10 mM ammonium bicarbonate. After incubation at 37^{o}C for 3 h, 20 µl of 1% formic acid was added and the samples were further incubated at room temperature for 10 min. The solution was then transferred to a clean tube and tryptic peptides were further extracted from the gel plugs by addition of 40 µl of 50% ACN and incubation for 10 min at room temperature. The two samples were pooled together and dried on a Vacufuge® Plus vacuum concentrator (Eppendorf, UK). The peptide mixtures were analysed by nLC MS/MS, using and Orbitrap Fusion trihybrid mass spectrometer coupled with a nano flow UHPLC system (ThermoFischer Scientific, UK). The peptides were separated, after trapped on a C_{18} pre-column, using a gradient of 3-40% ACN in 0.1% formic acid (vol/vol), over 50 min at a flow rate of 300 nL/min, at 40°C. The peptides were fragmented in the linear ion trap by a data-dependent acquisition method, selecting the 40 most intense ions. Mascot (Matrix Science, UK) was used to analyse the raw data against an in-house maintained database of the *L. reuteri* proteome. The tolerance on parent ions was 10 ppm and on fragments was 0.5 Da. Carboxymethylation of cysteine was selected as fixed modification and oxidation of methionine as variable modification. One miscleavage was allowed.

2.9.5. Endoproteinase C digest

Endopeptidase C (GluC; 5 µg) was added to 5 µl purified protein (MUB_{53608} or $SRRP_{53608}$; 5 µl) in 50 mM Tris-HCl, pH 8, containing 0.5 mM glutamate-glutamate. The reaction was incubated at 37^{o}C for 16 h and quenched by the addition of LDS loading buffer. The reaction products were analysed by western blot, using suitable probes.

2.9.6. Differential scanning fluorimetry (DSF)

DSF was used to assess glycosyltransferase – sugar donor interactions by measuring changes in the melting temperature (Tm) of the protein, upon interaction with the sugar nucleotides. The reactions were set up at a final volume of 20 µl in Tris-HCl 50 mM, pH 7.5. Proteins were used at a final concentration of 10 µM and SYPRO Orange (ThermoFischer Scientific, UK), the fluorescent dye used in the assay was used at 5x final concentration. Ligand and ion concentration ranged from 0-50 mM. The reactions were initially kept at 10°C for 10 min and then the temperature increased in a step-wise manner, with increments of 0.5°C every 15 s, up to 90°C. Measurements of the fluorescence were taken every 15 s on a Bio-Rad CFX96 Touch™ Real-Time PCR Detection System. The results were analysed using CFX Manager 3.5 (Bio-Rad, UK).

2.10. Enzymatic assays

2.10.1. N-acetyl-glucosaminidase assay

Purified $SRRP_{100-23}$ or $SRRP_{53608}$ (3 µl) was mixed with 2 µl reaction buffer (sodium acetate 200 mM, NaCl 100 mM, pH 5), 5 µl α-GlcNAcase (cloned and expressed in-house) at a final concentration of 0.17 U/µl and/or 1 µl of β-N-acetyl-hexosaminidase$_f$ (0.5 U/µl; New England Biolabs, UK). The reaction was incubated at 25°C, 37°C, or 42°C for 24 h. The reaction was quenched by addition of 3.5 µl LDS loading buffer supplemented with DTT and the samples were analysed by western blot, using *f*-WGA and anti-$SRRP_{53608}$ antibodies (see section 2.9.2, and Table 7).

2.10.2. Sialidase assay

Protein sample (5 µl) was mixed with 0.6 µl of GlycoBuffer 1 (New England Biolabs) and 0.4 µl neuraminidase A (~1.5 U/µl final concentration). The reaction was incubated at 37°C for 16 h, quenched by the addition of 2.5 µl LDS buffer and 1 µl DTT 100 mM and analysed by western blot (see section 2.9.2) using *f*-SNA.

2.10.3. MUB_{53608} enzymatic deglycosylation assay

Purified MUB_{53608} (125 µl) was mixed with 100 µl reaction buffer (sodium acetate 200 mM, NaCl 100 mM, pH 4.4) and 0.3 µl α-galactosidase and/or β-galactosidase at 1 mU/ml final concentration. The reaction was incubated at 25°C for 16 h and the products were analysed by western blot using *f*-RCA and anti-MUB_{53608} antibodies.

2.11. Glycan analysis

2.11.1. Monosaccharide composition analysis

2.11.1.1. Methanolysis of glycoproteins and trimethyl-silylation of monosaccharides

SM proteins (~100 µg;) were precipitated as described in section 2.8.1 and resuspended in H_2O. Myo-inositol (2.5 µg) was added to the suspension, which was then lyophilised. The dried sample was resuspended in 0.5 ml methanolic HCl 1N and incubated at 80°C for 16 h. The methanolysed sample was cooled down to 25°C and silver carbonate (~50 mg) was added, followed by 100 µl addition of acetic anhydride. The N-acetylation reaction was incubated at room temperature for 24 h. The sample was centrifuged at 3000 g for 5 min and the supernatant was transferred into a clean vial and dried under a gentle stream of nitrogen. Tri-Sil HTP reagent (200 µl; ThermoFischer Scientific, UK) was added to the sample and the derivatisation reaction was incubated at 80°C for 30 min. The sample was dried again under nitrogen, resuspended in 1 ml hexane and sonicated for 15 min. After centrifugation at 3000 g for 5 min, the supernatant was transferred into a clean vial and dried under nitrogen. The dried sample was dissolved into 100 µl dichloromethane and transferred to a GC-compatible vial.

2.11.1.2. GC-MS chromatography

The samples were analysed on an Agilent 7890B GC-MS system paired with an Agilent 5977A mass spectrometry detector (Agilent, UK), using an BPX70 column (SGE Analytical Science, Australia). Helium was used as the carrier gas. The inlet was maintained at 220°C, 12.9 psi, and 23 ml/min flow. The injection volume was 1 µl in split mode (1:20). The oven temperature

increased initially from 100°C to 120°C over 5 min, followed by a second increase from 120°C to 230°C over 40 min.

2.11.2. Sialic acid release, labelling and HPLC-based analysis

Protein sample (100 µl) was mixed with 100 µl acetic acid (4 M) and incubated at 80°C for 2.5 h. The reaction mixture was dried using a centrifugal evaporator and the sample was dissolved in 100 µl DMB labelling reagent (14 mM 1,2-diamino-4,5-methylenedioxybenzene, 3 mM acetic acid, 1.5 mM β-mercaptoethanol, 36 mM sodium hydrosulphite). An aliquot of the Glyko® sialic acid reference panel (Prozyme, USA) was used as a positive control. The reaction was incubated for 3 h at 55°C and the mixture was transferred in an HPLC vial. The products were analysed on a Luna® 5µm C18(2) 100 A column (250x4,6 mm; Phenomenex, UK), using a Shimadzu Prominence HPLC system (Shimadzu, UK). The solvents used were H_2O (A), ACN (B) and 50% methanol (C). Sialic acids were separated on a gradient from 82:4:14 A/B/C to 75:11:14 over 40 min, followed by a wash step with 30:70:0 A/B/C for 10 min and an equilibration step with the starting solvent for 10 min.

2.11.3. Glycan release, permethylation and mass spectrometry analysis

For MS analysis of protein glycosylation, purified protein (~100 µg) in ammonium bicarbonate was freeze-dried and dissolved in 200 µl 50 mM NaOH, containing 1 M $NaBH_4$. The β-elimination reaction was incubated at 45°C for 16 h and quenched by addition of acetic acid, until effervescence stopped. The sample was desalted in a DOWEX 50WX8 column (H^+ form, 100-200 mesh) and glycans were eluted with 3 ml acetic acid 5% (vol/vol). The recovered sample was then freeze-dried and reconstituted in H_2O containing 0.1% trifluoroacetic acid (TFA; vol/vol). The sample was further desalted using a graphitized carbon NuTip (Glygen, USA), prewashed with 80% ACN containing 0.1% TFA, and 0.1% TFA. Hydrophilic contaminants were washed with sequential washes with 0.1% TFA and 0.1% TFA in 10% ACN. Glycans were eluted with 0.1% TFA in 25%, 50% and 80% ACN. The elution fractions were pooled and dried on a Vacufuge® Plus vacuum concentrator. The dried sample was dissolved in 300 µl anhydrous dimethylsulfoxide

(DMSO), followed by the addition of ~20 mg NaOH and 400 µL iodomethane. The permethylation reaction was incubated at room temperature for 60 min under vigorous shaking and quenched by the addition of 1 ml of 5% (vol/vol) acetic acid. The permethylated glycans were purified using a Hydrophilic-Lipophilic Balanced (HLB) copolymer Oasis cartridge (Waters, UK), prewet with 4 ml of methanol and equilibrated with 5% (vol/vol) methanol in H_2O. Salts and other hydrophilic contaminants were washed with 5 ml of 5% methanol and permethylated glycans eluted with 3 ml of 100% methanol. The samples were dried under a gentle stream of nitrogen, dissolved in 10 µl of TA30 [30% (vol/vol) ACN, 0.1% (vol/vol) trifluoroacetic acid] and mixed with equal amount of 2,5-dihydroxybenzoic acid (DHB; Sigma-Aldrich, UK; 20 mg/ml in TA30), before being spotted onto an MTP 384 polished steel target plate (Bruker, UK). The samples were analysed by MALDI-ToF, using the Bruker Autoflex™ analyzer mass spectrometer (Bruker, UK) in the positive-ion and reflectron mode.

2.11.4. Force spectroscopy

The interactions between MUB_{53608} and h-Int1 or RCA were examined by covalently attaching lectin molecules to atomic force microscopy (AFM) tips and MUB_{53608} to the glass slides to enable binding interactions to be measured in a specific manner (290). Silicon nitride AFM tips (PNP-TR, Nanoworld AG, Switzerland) and freshly cleaned glass slides were functionalized using a four-step procedure (carried out at 21 ◦C): the first step involved incubation in a 2% solution of 3-mercaptopropyltrimethoxy silane (MTS, Sigma–Aldrich, UK) in toluene (dried over a 4Å molecular sieve) for 2 h, followed by washing with toluene and then chloroform. In the second step, the silanised tips were incubated for 1 h in a 1 mg/ml solution of a heterobifunctional linker: MAL-PEG-SCM, 2 kD (Creative PEGWorks, USA) in chloroform. The silanised glass slides were incubated in 5 mM N g -maleimidobutyryl-oxysuccinimide ester (GMBS) in ethanol (ThermoFisher Scientific, USA). The tips and slides were rinsed with chloroform/ethanol, respectively, and then dried with argon. The third step involved covalent attachment of the lectins by incubation of the tips/slides in 1 mg/ml solutions of the proteins in PBS at pH 7.4 for

1 h at 21 °C, followed by a PBS washing step. The fourth step involved incubation of the functionalized cantilevers/slides in a 10 mg/ml solution of glycine in PBS to 'amine'-cap any unreacted succinimide groups, followed by washing in PBS. Binding measurements were carried out in PBS using a MFP3D BIO AFM (Asylum Research Inc., USA). The experimental data were captured in 'force-volume' (FV) mode at a rate of 2 mm/s in the Z direction and at a scan rate of 1 Hz and a maximum load force of 150-300 pN (pixel density of 32x32). Adhesion in the force spectra was quantified using a bespoke Excel macro (291) which fits a straight line to the baseline of the retract portion of the force-distance data. In order to explore the specificity of the binding interactions, the force measurements were repeated after addition of 100 mM galactose (for RCA assays) or EGTA (for h-Int1 assays) to the liquid cell.

Chapter 3

Analysis of carbohydrate metabolism and protein glycosylation in Lactobacillus reuteri

3.1. Introduction:

L. reuteri is a common gut commensal bacterial species that colonises the gastrointestinal (GI) tract of a wide variety of vertebrate hosts, including humans, rodents, birds, and livestock (103,292). *L. reuteri* 100-23 (*Lr*100-23) is a rat isolate that naturally colonises the stratified squamous epithelium of the forestomach (103), whereas *L. reuteri* ATCC 53608 (*Lr*ATCC 53608) is a pig isolate that colonises the porcine small intestine mucosa (292). *L. reuteri* MM4-1a (*Lr*MM4-1a) was isolated from human breast milk (www.biogaia.com). The localisation of *L. reuteri* in the GI tract suggests that it preferentially feeds on simple sugars that are available from the diet of the host, rather than dietary polysaccharides which are normally fermented by the microbiota in the large intestine (60). As *L. reuteri* strains are often used for their probiotic properties, research is focused on understanding the colonisation process of these bacteria in the GI tract. A key step in bacterial colonisation is adhesion of the bacteria to the mucus or the epithelium *via* specific surface adhesins. It is often reported that surface adhesins are glycosylated, but information on protein glycosylation in beneficial microbes and, in particular, in *L. reuteri* is lacking.

The building blocks of the glycans found on glycoproteins are sugar nucleotides. These are monosaccharides conjugated onto a nucleotide, most often uridine diphosphate (UDP). The synthesis of sugar nucleotides requires the phosphorylation of a monosaccharide, its subsequent conjugation with the nucleotide and further modifications of the sugar moiety that lead to diversification of the nature of the sugar. Key enzymes in this process are kinases that phosphorylate monosaccharides, thus enabling their subsequent activation by nucleotides (243). Alternatively, in bacteria, carbohydrates can be phosphorylated by specialised, sugar specific systems like the phosphoenolpyruvate: sugar phosphotransferase system" (PTS) (293,294). The PTS system consists of three enzymes in most cases: Enzymes I and II (EI and EII) and a low molecular weight heat-stable protein HPr. Enzyme I and HPr are sugar independent

and are responsible for the phosphorylation of the sugars. In contrast, Enzyme II is comprised of three, or, rarely, four subunits, and forms a sugar-specific permease (Figure 13) (294).

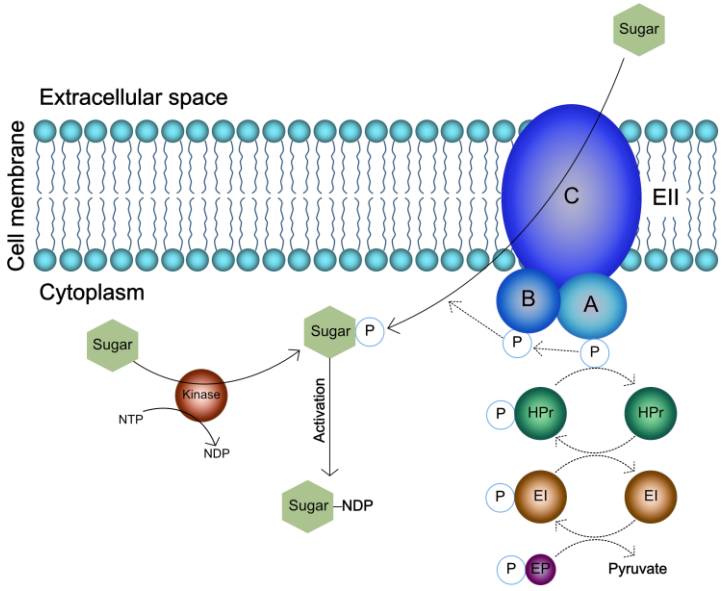

Figure 13 Phosphorylation of monosaccharides by cytosolic kinases or the phosphotransferase system (PTS) and subsequent activation with nucleotides. Adapted from (295).

The enzymes that mediate the synthesis of the protein glycan are called glycosyltransferases (GTs). GTs belong to a large group of carbohydrate active enzymes (CAZymes). CAZymes are divided into subgroups, based on their activity [e.g. GTs, glycoside hydrolases, (GH)]. Each subgroup is further divided into families, based on the amino acid sequence of the enzymes (296). Members of each family share significant similarities, which can often provide useful insights regarding the activity, mechanism, as well as sugar specificity of CAZymes. GTs are divided into two large categories, based on the sugar donor: GTs that utilise sugar nucleotides are called Leloir-GTs, whereas GTs that utilise other donors, such as lipid-linked oligosaccharides (e.g. PglB involved in N-glycosylation) are called non-Leloir GTs. Further information on each family of GTs can be found in the CAZy database (www.cazy.org). In addition, based on the mechanism they employ for glycosylation, they are divided into inverting and retaining GTs. Retaining GTs transfer the glycan onto the acceptor with the same conformation it had on the donor. In contrast, inverting GTs cause a switch in the conformation of the glycan during

glycosylation. For example, a retaining GT that uses a α-linked sugar nucleotide donor, will generate a glycoconjugate with an α-linked glycan (Figure 14).

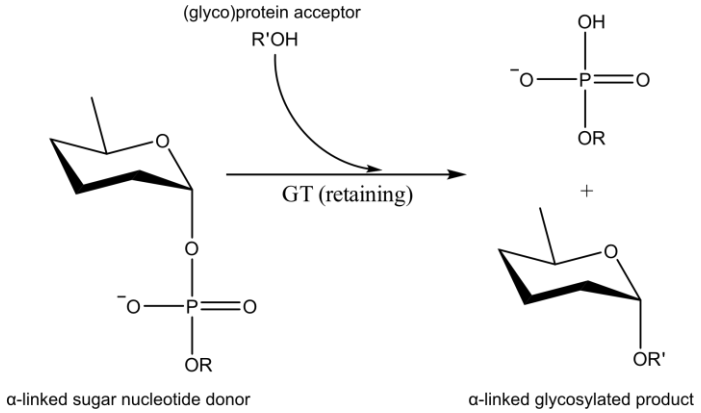

Figure 14 Schematic representation of the retaining glycosylation mechanism. Adapted from (230).

In bacteria, genes encoding for GTs are often grouped together in clusters encoding proteins involved in the biosynthesis of glycans. These include examples for protein glycosylation or exopolysaccharide (EPS) biosynthesis. However, it is often the case that GTs do not act on a single biosynthesis pathway, and can be involved in the formation of different glycoconjugates.

3.2. Results

3.2.1. *In silico* analysis of *L. reuteri* carbohydrate metabolism

In silico analysis of *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a genomes was carried out, in order to identify potential enzymes involved in sugar nucleotide biosynthesis and GTs involved in protein glycosylation.

The annotated genomes were first analysed in order to identify sugar kinases or sugar-specific PTS components. The analysis revealed the presence of predicted kinases for simple monosaccharides such as glucose (Glc; EC 2.7.1.2), galactose (Gal; EC 2.7.1.6) and fructose (Fru; EC 2.7.1.4) (Figure 15). *Lr*100-23 and *Lr*ATCC 53608, but not *Lr*MM4-1a, also harbour genes encoding a xylose (Xyl) isomerase (EC 5.3.1.5) and a xylulose kinase (EC 2.7.1.17), suggesting that these strains can also utilise this pentose (297), which may reflect an adaptation of the strains to the host diet. In contrast to the presence of hexose (Hex) kinases, no kinases specific to N-

acetyl-hexosamines (HexNAcs) were found. Components of Glc-, lactose (Lac)- and galactitol-specific PTS systems were identified in all three strains. While all three strains harbour genes encoding proteins involved in the utilisation of Glc and Lac, no genes involved in galactitol metabolism were identified. As galactitol can be converted to Gal by the reverse action of an aldehyde reductase, the *Lr*100-23 genome was also inspected for genes encoding proteins with similarity to known aldehyde reductases. The closest candidate was found to be a gene (gene ID: 2500069715) encoding a predicted aldo/keto reductase related to diketogulonate reductase, that shared ~45% identity (E-value < 1.0^{-100}), with an aldehyde reductase (EC 1.1.1.21) from the β-proteobacterium *Cupriavidus nantongensis*. This gene was also present in *Lr*ATCC 53608 and *Lr*MM4-1a and it is possible that its product may be involved in galactitol metabolism.

The lack of genes encoding HexNAc-specific kinases or N-acetyl-glucosamine (GlcNAc)-specific PTS was further confirmed by performing a BLAST search with the amino acid sequences of known GlcNAc kinases and GlcNAc-specific PTS, against the predicted proteome of *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a. The *in silico* analysis of these three *L. reuteri* strains revealed genes encoding enzymes that are predicted to be involved in the synthesis of UDP-GlcNAc from Glc, as well as its modification and conversion into UDP-N-acetyl-muramic acid (MurNAc), which is used in cell wall synthesis, or UDP-N-acetyl-mannosamine (ManNAc), but not UDP-N-acetyl-galactosamine (GalNAc), the monosaccharide found at the reducing end of mammalian O-glycans (Figure 15). Despite the predicted ability of these strains to produce UDP-ManNAc, they all lack the downstream enzymes required for the biosynthesis of sialic acid (Neu5Ac) or N-acetyl-mannosaminuronic acid, which raises questions about the role of UDP-ManNAc in the physiology of *L. reuteri* strains. Enzymes involved in the synthesis of guanidyl-mannose (GDP-Man) from fructose were also identified. However, the analysis did not identify genes involved in the synthesis of GDP-fucose (Fuc), a monosaccharide often employed by microbes to synthesise glycans resembling those of the host (271).

Furthermore, all three strains carried a gene encoding a predicted galactopyranose mutase (Figure 15), an enzyme that can convert UDP-Gal to UDP-galactofuranose (UDP-Gal_f), which is often found in fungal cell walls and glycoproteins (298), as well as carbohydrate structures in important human parasites and bacterial pathogens (299–301). In addition, a gene involved in the conversion of UDP-arabinose (UDP-Ara) to UDP-arabinofuranose (UDP-Ara_f) was identified in the three genomes analysed, but no genes encoding enzymes involved in UDP-Ara synthesis could be found. While most sugar nucleotide biosynthetic pathways were common between the three *L. reuteri* strains, *Lr*100-23 was also found to carry genes encoding proteins necessary for the synthesis of thymidyl-rhamnose (TDP-Rha), while *Lr*ATCC 53608 and *Lr*MM4-1a only harboured part of the biosynthetic gene cluster (Figure 15).

In addition to the kinases, and the other enzymes involved in sugar nucleotide metabolism, the three *L. reuteri* strains carry genes that encode glycoside hydrolases, and specifically α-glucosidases, α- and β-galactosidases and endo-galactanases (data not shown). Notably, no hydrolase specific for Neu5Ac or Fuc -monosaccharides naturally found in the host glycans of the small intestine- were identified in any of the three strains. Taken together, these observations suggest that *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a may rely on diet-derived, rather than host-derived, carbohydrates and can utilise free, short oligosaccharides, such as lactose, which is naturally found in excess in milk. In addition, the presence of genes involved in sugar nucleotide biosynthesis suggests that these strains can diversify the repertoire of monosaccharides they may use in glycoconjugates.

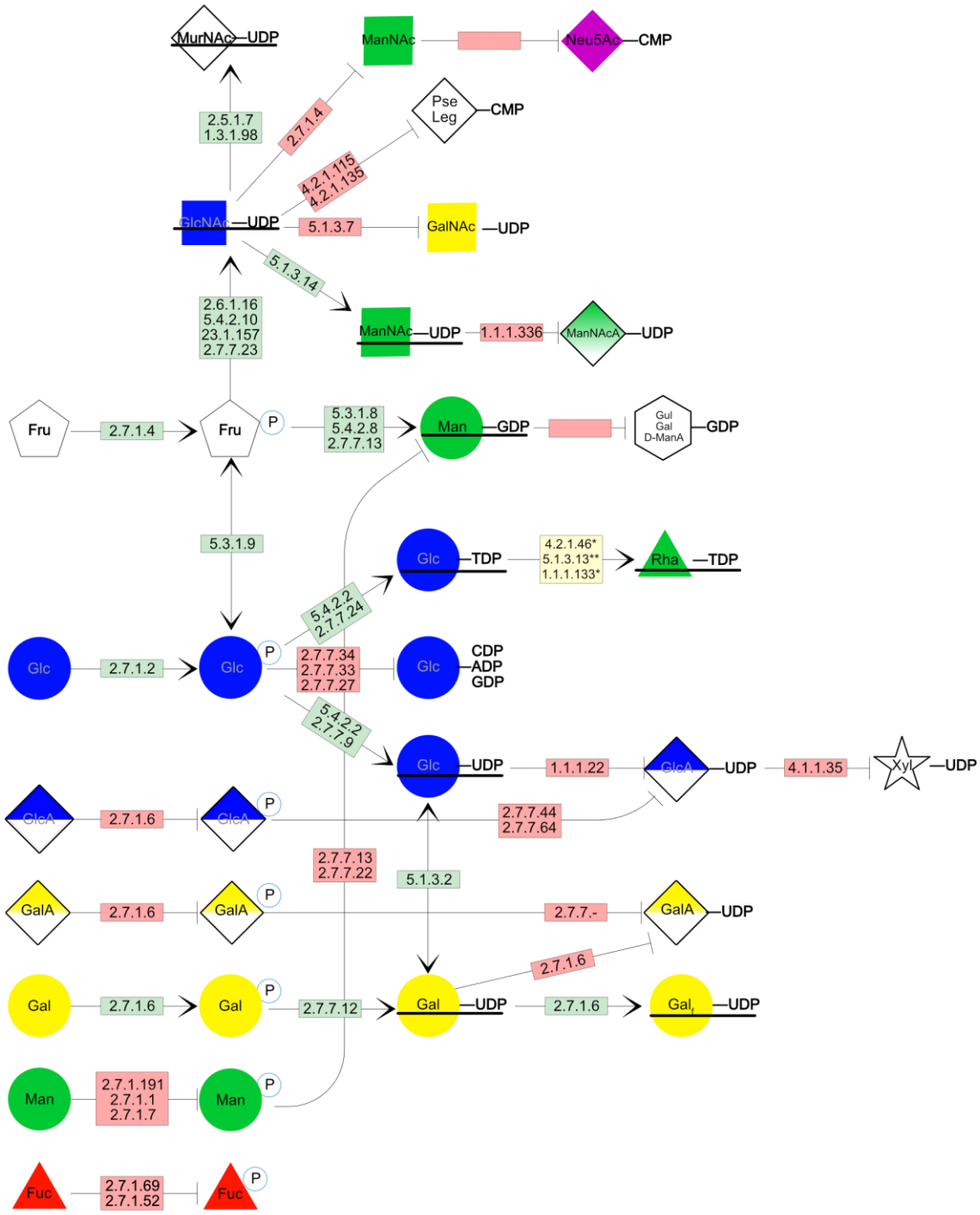

Figure 15 A simplified map of the "Amino sugar and nucleotide sugar metabolism" KEGG pathway for *L. reuteri*. The Enzyme Commission (EC) numbers of the enzymes required in each pathway are noted in the boxes. Green boxes indicate enzymes predicted to be produced from the genomes of *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a, whereas the yellow box indicates the TDP-Rha biosynthetic pathway predicted to be expressed only in *Lr*100-23. Red boxes indicate enzymes that are not encoded in the three *L. reuteri* genomes. Underlined are the sugar nucleotides predicted to be expressed from these *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a strains. *EC 4.2.1.46 and 1.1.1.133 are missing from the *Lr*ATCC 53608 genome. **EC 5.1.3.13 is missing from the *Lr*MM4-1a genome.

3.2.2. *In vitro* analysis of *L. reuteri* growth on carbohydrates

To complement the results from the bioinformatics analysis, and as there have been reports of GlcNAc utilisation, without the requirement of a functional GlcNAc kinase in *E. coli* (302), the growth of each *L. reuteri* strain was tested in the presence of lactose, sucrose (Suc), maltose, Xyl, 2-fucosyl-lactose (2'FL), 3-fucosyl-lactose (3'FL), 6-sialyl-lactose (6'SL), GlcNAc or GalNAc as sole carbon source. Overnight cultures in deMan-Rogosa-Sharpe (MRS) medium were used to inoculate Lactobacillus-defined-media-II (LDM-II) (154) supplemented with 2% (w/v) of the carbon source. The cultures were grown for 24 h and OD_{600} measurements were taken every 30 min, for the disaccharide-containing growth media, or at 24 h for the HexNAc, Xyl or trisaccharide-containing media.

All three strains grew in the disaccharide-containing media, however, the growth rate and the preference for a carbohydrate differed between strains (Figure 16). In particular, *Lr*100-23 grew best in the presence of maltose, in terms of the highest OD_{600} value, as well as the rate of growth, as it reached stationary phase after 6 h. In contrast, there was an extended lag phase when *Lr*100-23 was grown in lactose, reaching stationary phase after 13 h, and the growth was impaired in sucrose, showing an OD_{600} of 0.05 units at stationary phase, compared to 0.3 units when grown in maltose. This may suggest that *Lr*100-23 cannot metabolise Fru as efficiently as Glc. It is also possible that sucrose is hydrolysed by a fructosyltransferase (ftf, gene no. 2500071010, EC 2.4.1.10), generating Glc, which becomes available to the bacteria for growth, and using fructose to synthesise levan (287), making it unavailable to use in metabolism. This is also supported by the significantly impaired growth of a *Lr*100-23 *Δftf* mutant in the presence of sucrose as a carbon source (data not shown).

*Lr*ATCC53608 grew optimally in the sucrose-containing medium, reaching stationary phase (OD_{600} = 0.8) after 8.5 h. Its growth was mediocre in maltose, as it reached stationary phase (OD_{600} =0.6) after 11 h and significantly impaired in lactose, as it only reached an OD_{600} of 0.2 after 15 h growth. Even though *Lr*ATCC 53608 carries an *ftf* homologue, it displayed optimal

growth in presence of sucrose, which raises questions about the different sucrose metabolism pathways employed by *Lr*100-23 and *Lr*ATCC 53608. *Lr*MM4-1a showed comparable growth in all three disaccharides. However, growth in lactose was slower, as it reached stationary phase after 12 h, compared to 6 h in maltose and sucrose. *Lr*MM4-1a does not have a gene encoding a ftf, so it probably relies on other metabolic enzymes for the utilisation of sucrose.

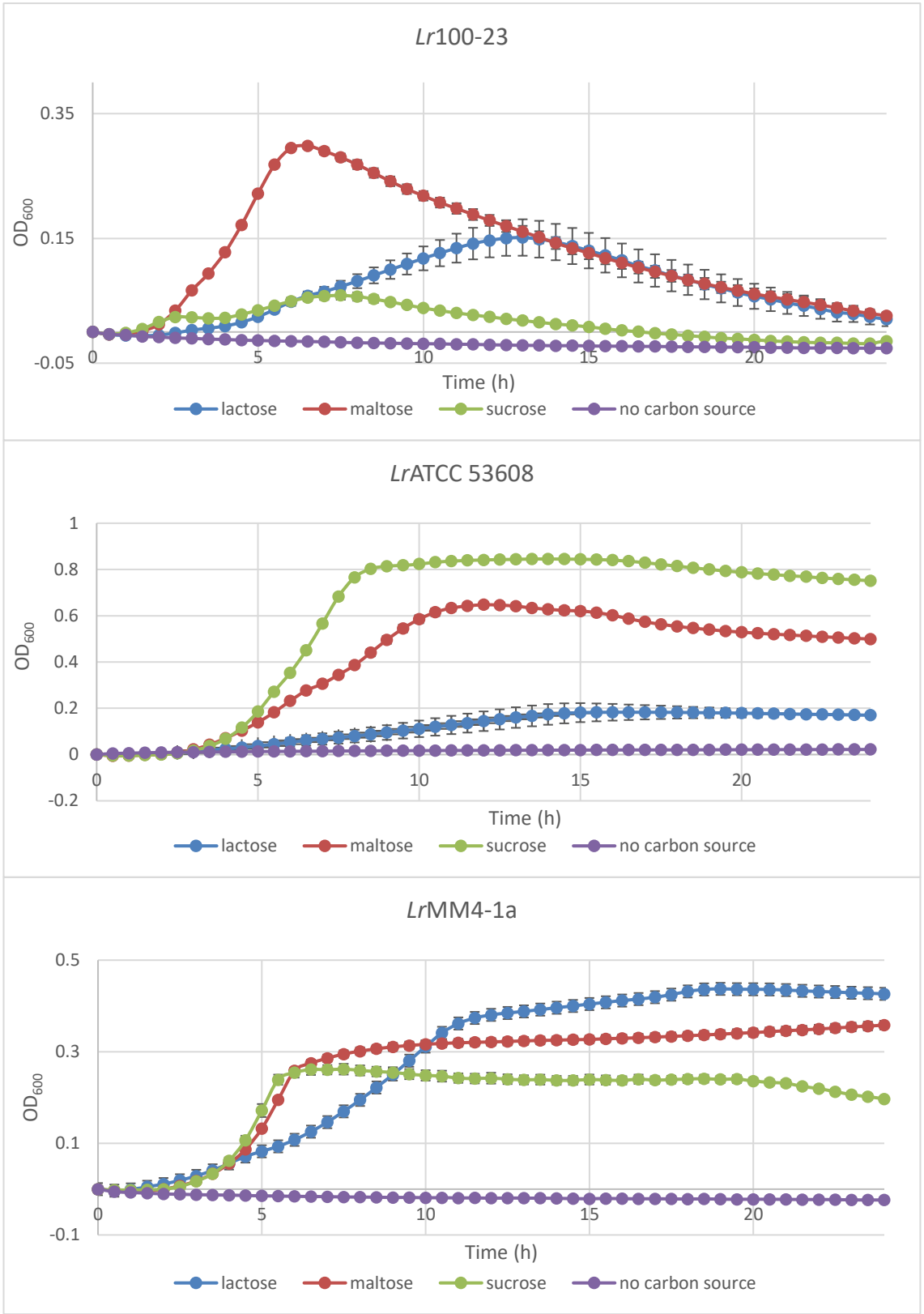

Figure 16 Growth rates of *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a when grown on different disaccharides. Growth rate and maximum OD_{600} is dependent on the disaccharide used for each strain.

Notably, none of the strains grew when Xyl, GlcNAc, GalNAc, 2'FL, 3'FL or 6'SL were used as sole carbon source (Figure 17). These *in vitro* results supported the *in silico* analysis that did not

reveal any HexNAc kinase or specific fucosidase or sialidase encoding genes. The inability of *Lr*100-23 and *Lr*ATCC 53608 to utilise Xyl contrasted with the presence of a predicted Xyl utilisation pathway in these two strains. However, these two strains were shown to grow in the presence of xylo-oligosaccharides or Xyl in MRS media (Dr. Ravindra Pal-Singh, The John Innes Centre, personal communication), although growth in Xyl was impaired, suggesting differences in the transport of Xyl and short xylo-oligosaccharides. In addition, as these strains did not grow in LDM-II when Xyl was used as sole carbon source, this suggests that there are additional components in MRS that promote growth of the bacteria on this substrate.

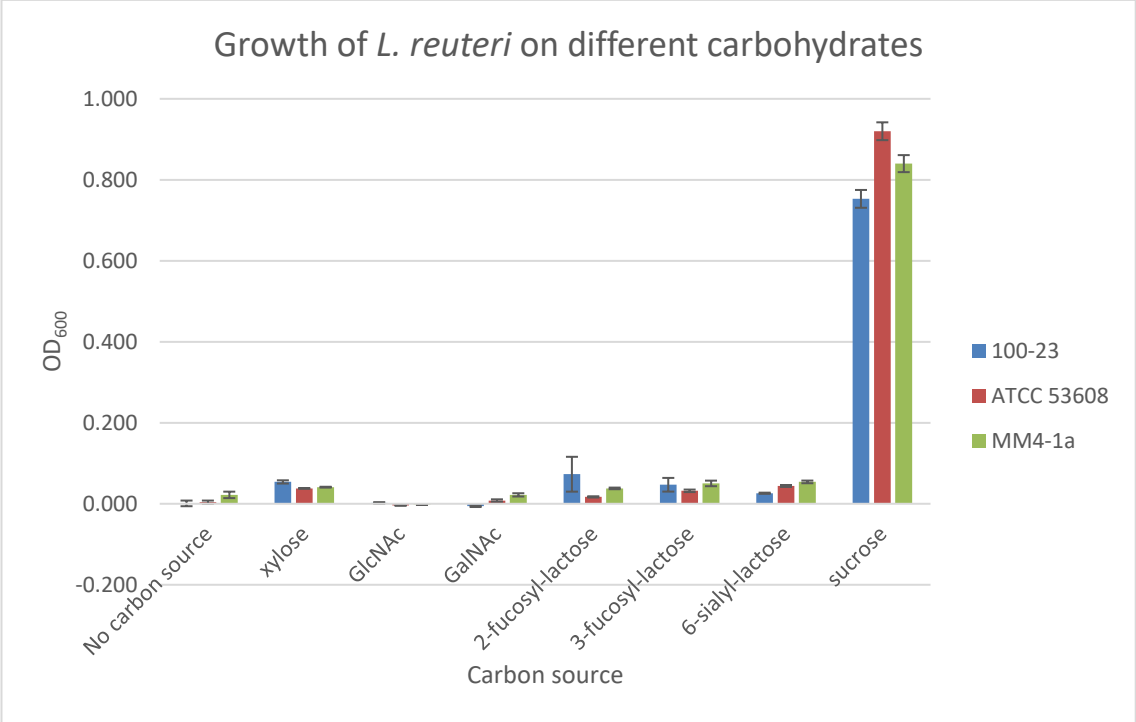

Figure 17 Growth of *L. reuteri* strains in HexNAc as sole carbon source. None of the three strains was able to utilise free HexNAcs, fucosyl-, or sialyl-lactose present in the growth medium, as compared to sucrose (used as control).

3.2.3. Bioinformatics analysis of glycosyltransferase clusters

3.2.3.1. Identification of putative glycosyltransferases

To identify putative protein glycosylation systems, the genomes of *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a were interrogated for the presence of clusters containing GTs: 24 annotated GTs were identified in *Lr*100-23, 21 in *Lr*ATCC 53608 and 12 in *Lr*MM4-1a. Most of the annotated GTs found in all three strains belong to GT families 2 and 4, as per the CAZy database classification (www.cazy.org). The GT2 family contains enzymes that employ an inverting mechanism of glycosylation, suggesting the formation of mainly β-glycosidic bonds. Known substrates for members of this family include UDP-Glc, UDP-GlcNAc, UDP-GalNAc, UDP-Gal_f, dTDP-Rha and GDP-Man. In contrast, members of the GT4 family employ a retaining mechanism that leads to the formation of α-glycosidic linkages. Possible substrates for GT4 members include UDP-Glc, UDP-GlcNAc, UDP-Xyl and GDP-Man. Some predicted multidomain proteins in the three *L. reuteri* strains contain a GT1 domain at the C-terminus, in addition to the GT4 domain at the N-terminus. The GT1 domain uses an inverting mechanism but seems to be active primarily on small molecules. Also, all three strains contain a single gene encoding a predicted GT28, which has an inverting activity and is probably involved in the biosynthesis of peptidoglycan. In addition, *Lr*100-23 and *Lr*ATCC 53608 harbour genes coding for proteins with a GT8 domain. The retaining mechanism of GT8 members mostly leads to the formation of α-glycosidic linkages, using primarily UDP-Glc and UDP-Gal. A summary of the GTs identified in *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a is presented in Table 16.

.

Table 16 Summary of the GTs identified in *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a, as part of clusters. Genes found on the same row were found to be homologues with an E-value < 1.0^{-70}.

	***Lr*100-23**		***Lr*ATCC 53608**		***Lr*MM4-1a**	
Cluster	**Gene ID**	**Domain**	**Gene ID**	**Domain**	**Gene ID**	**Domain**
1	2500069824	GT4/GT1	LRATCC53608_0691	GT4/GT1	2502290820	GT4/GT1
	2500069825	GT4/GT1	LRATCC53608_0690	GT4/GT1	2502290819	GT4/GT1
2 (EPS)	2500069849	GT2				
	2500069850	Not classified				
	2500069851	Not classified				
	2500069853	GT2				
	2500069854	(DUF4422)				
	2500069855	GT2				
	2500069856	(YfhO)				
	2500069857	Not classified (GtrII)				
			LRATCC53608_0666	GT2	2502290049	GT2
			LRATCC53608_0655	Not classified	2502290056	Not classified
			LRATCC53608_0654	GT4	2502290057	GT4
	2500069867	Bacterial GT	LRATCC53608_0651	Bacterial GT	2502290062	Bacterial GT
	2500069868	GT2	LRATCC53608_0650	GT2	2502290063	GT2
	2500069869	(DUF4422)	LRATCC53608_0649	(DUF4422)	2502290064	(DUF4422)
3	2500070567	GT4	LRATCC53608_0089	GT4	2502291957	GT4
	2500070568	Gtf2	LRATCC53608_0090	Gtf2	2502291958	Gtf2
4 ($SecA_2/Y_2$)	2500070887	GT4	LRATCC53608_0907	GT4		
	2500070893	GT4	LRATCC53608_0913	GT4		
	2500070894	Gtf2	LRATCC53608_0914	Gtf2		
	2500070896	GT8/DUF1792				
	2500070897	GT8/DUF1792				
	2500070898	GT8/DUF1792				
	2500070900	DUF1792 (split)				
5	2500070918	Bacterial GT	LRATCC53608_0938	Bacterial GT		
			LRATCC53608_0939	GT4/GT1		
	2500070922	GT8				
	2500070923	GT2				
	2500070924	GT2	LRATCC53608_0940	GT2		
			LRATCC53608_0941	GT4/GT1		
			LRATCC53608_0945	GT2		
N/A	2500071234	GT28/GT28	LRATCC53608_1356	GT28/GT28	2502291442	GT28/GT28
	2500069382	GT2	LRATCC53608_0673	GT2	2500070923	GT2
	2500069655	GT1 or GT4	LRATCC53608_0770	GT1 or GT4	2502290385	GT1 or GT4
			LRATCC53608_1078	GT8		
			LRATCC53608_1691	GT2	2502290343	GT2

3.2.3.2. Glycosylation clusters in L. reuteri

Five glycosylation clusters were identified in *L. reuteri* strains. The first glycosylation cluster is conserved between *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a and consists of two putative GTs (Figure 18). Both GTs contain a GT4 domain at the N-terminus and a GT1 domain at the C-terminus. These genes are part of a predicted four-gene operon, which also encodes a putative lysylphosphatidylglycerol synthase and an uncharacterised YkuJ-like protein. This predicted operon is located directly downstream of the genes encoding HPr and EI, the sugar-independent components of the PTS system.

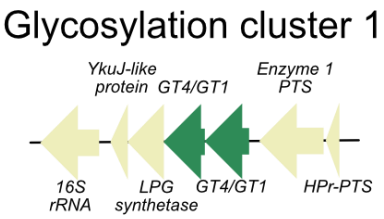

Figure 18 Glycosylation cluster 1 is conserved in *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a and harbours two genes encoding multidomain GTs. Both GTs have a GT4 domain in the N-terminus and a GT1 domain in the C-terminus.

The second cluster is the EPS biosynthetic cluster which is also found in all three strains (Figure 19). This cluster is significantly larger in *Lr*100-23 with 11 predicted GTs, while the clusters in *Lr*ATCC 53608 and *Lr*MM4-1a consist of six putative GTs. The conserved GTs between the EPS clusters of *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a include the priming GT, containing a bacterial GT domain, a GT (GT family 2) and a protein containing a DUF4422, which has been identified in GTs of *C. jejuni subsp*. *jejuni*. *Lr*ATCC 53608 and *Lr*MM4-1a also harbour genes encoding a GT2 and a GT4, which have no homologues in *Lr*100-23. Instead, the *Lr*100-23 EPS cluster contains four additional GT2 and one DUF4422 containing protein, as well as two proteins that are similar to putative GTs from *Oenococcus oeni* (2500069850 and 2500069851) and a YfhO-like protein, which in some cases has been annotated as a GT, but which function remains to be determined experimentally (Figure 19). Additional proteins predicted to be expressed by this cluster include a putative polymerase, a protein involved in the export of polysaccharides, and a chain length

determinant protein, further supporting that this cluster is involved in the biosynthesis of EPS (Figure 19).

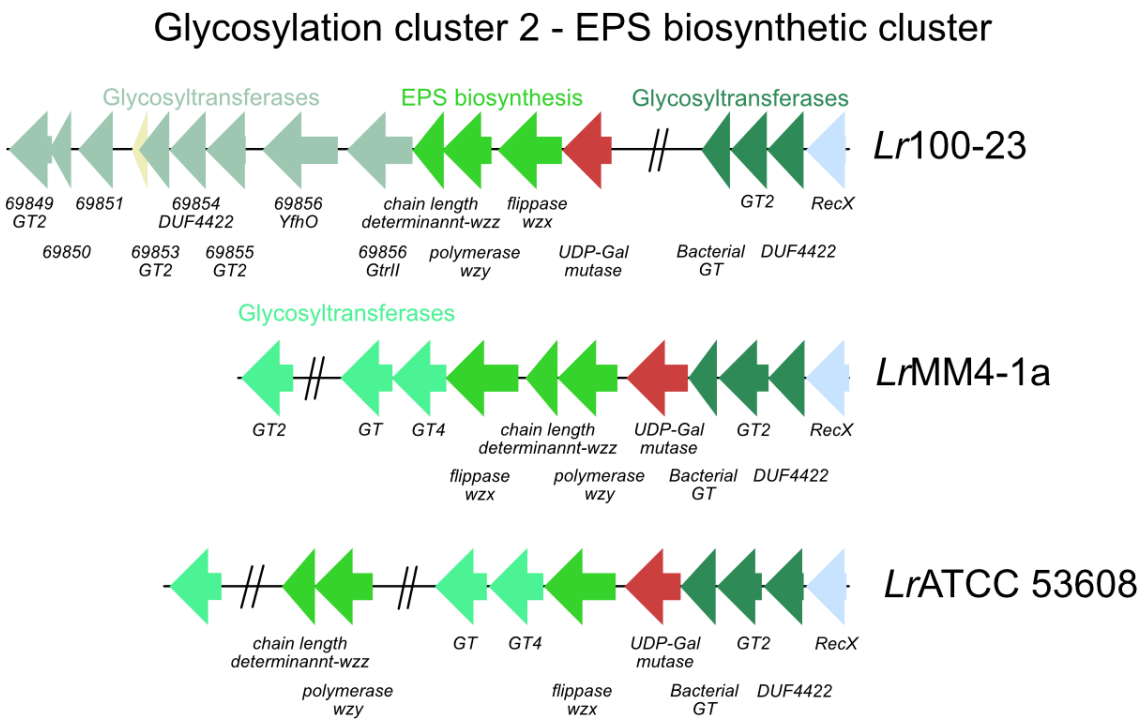

Figure 19 The EPS biosynthetic cluster in *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a. The clusters differ in the organisation of the genes, as well as the number of GT-encoding genes present in each cluster.

A predicted two-GT operon constitutes the third glycosylation cluster, identified in *L. reuteri* strains (Figure 20). Each GT has a conserved DUF1975 at the N-terminus and a GT4 domain at the C-terminus. This operon is homologous to the *gtf1-gtf2* operon in *L. plantarum* WCFS1, which was shown to be involved in a general protein O-glycosylation system, as deletion of either *gtf1* or *gtf2* resulted in the loss of protein glycosylation in more than seven identified glycoproteins in *L. plantarum* (277). The high similarity of the *Lr*100-23 *gtf1* and *gtf2* with their homologues from *L. plantarum* (E-value = 1.0^{-146} and 1.0^{-134}, respectively) suggests that they may play a similar role in *L. reuteri.* In *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a, several predicted surface proteins are located downstream of the GTs, which may represent putative protein glycosylation targets.

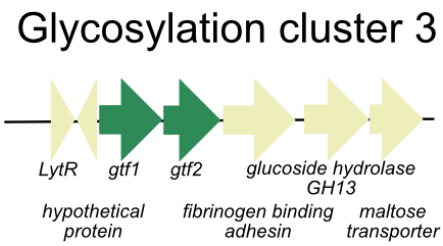

Figure 20 Glycosylation cluster 3 contains two putative GTs, similar to gtfA and gtfB from the *secA$_2$/Y$_2$* cluster, and homologous to gtf1 from *L. plantarum*, that may be involved in protein glycosylation.

The fourth group of GTs is only found in *Lr*100-23 and *Lr*ATCC 53608 and is part of the SecA$_2$/Y$_2$ cluster (Figure 21). Both clusters contain genes encoding the priming pair of GTs (GtfA and GtfB), and GtfC, in addition to the accessory secretion proteins (Asp) 1-3, the translocases SecA$_2$/Y$_2$ and SecY2, as well as the target serine rich repeat protein (SRRP) (see section 1.3.3.3.2). *Lr*100-23 also carries four additional GT-encoding genes. Of these, genes 250070898 and 250070900 appear to be part of a single original gene, which became split by the insertion of a transposase. These putative GTs have a GT8 domain at the N-terminus and a DUF1792 at the C-terminus, showing a reverse organisation compared to the well-studied homologue dGT1 from *Streptococcus parasanguinis* FW213 (262), which contains a DUF1792 at the N-terminus and a GT domain at the C-terminus and was shown to mediate the third and fourth glycosylation steps of the fimbriae associated protein 1 (Fap1) in *S. parasanguinis* FW213 (264). However, while *S. parasanguinis* encodes one protein containing a DUF1792, *Lr*100-23 has four genes encoding similar proteins, suggesting a more variable glycosylation of the target serine rich repeat protein (SRRP). In contrast, the *Lr*ATCC 53608 cluster only contains the putative GtfA, B and C. It is therefore more likely to glycosylate SRR with a disaccharide, as GtfA and B mediate the first glycosylation step and GtfC mediate the second one, as shown for *S. parasanguinis* FW213 and *S. pneumoniae* TIGR4 (191,268).

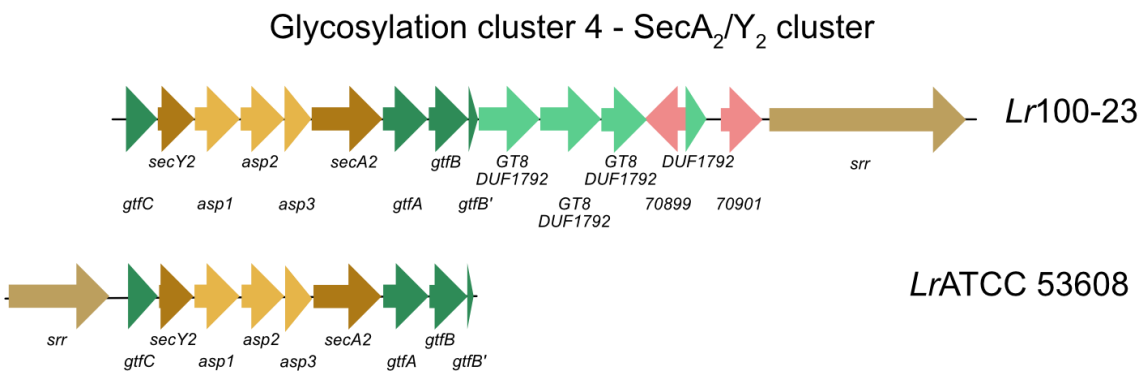

Figure 21 The $SecA_2/Y_2$ cluster in *Lr*100-23 and *Lr*ATCC 53608. Both systems contain genes that form the core of the cluster (srr, secA2, secY2, asp1-3 and gtfA-B) as well as gtfC. The *Lr*100-23 system contains four additional multidomain GTs.

Another glycosylation cluster found only in *Lr*100-23 and *Lr*ATCC 53608 is located downstream of the $SecA_2/Y_2$ cluster and contains four, or five GTs, respectively (Figure 22). It shows some organisational similarity with the EPS glycosylation cluster, as it contains a chain length determinant protein, a polymerase and several putative membrane proteins. However, no annotated proteins involved in export of the EPS subunits have been identified (Figure 22). In addition, a TDP-Rha cluster is found in that locus in *Lr*100-23.

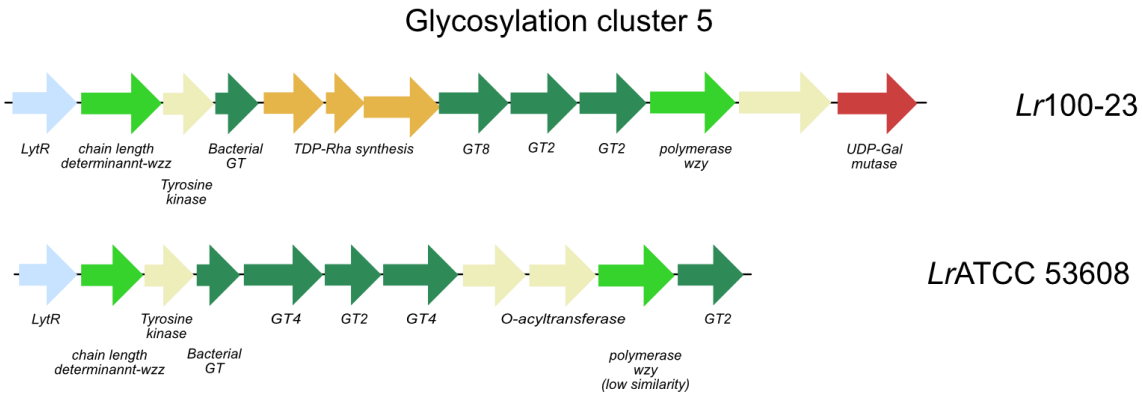

Figure 22 Glycosylation cluster 5 is similar to the EPS cluster, in the nature of GTs that it contains, but is not found in *Lr*MM4-1a and also lacks a putative flippase

In addition to the GTs in the clusters described above, there are other GTs spread across the genome of the three *L. reuteri* strains that do not appear to be part of any predicted glycosylation cluster. Of these, the most notable genes are 250069655 in *Lr*100-23 and its homologues LRATCC53608_0770 and 2502290385 in *Lr*ATCC 53608 and *Lr*MM4-1a, respectively, which all have a DUF1975 at the N-terminus and a GT domain at the C-terminus,

as found in GtfA from the $SecA_2/Y_2$ cluster and Gtf1 from *L. plantarum* WCFS1. However, there is no neighbouring gene similar to GtfB or Gtf2. Deletion of the homologous gene in *L. plantarum* WCFS1 did not alter the protein glycosylation profile (274).

Taken together this *in silico* analysis suggests that there are at least two putative protein glycosylation systems in *Lr*100-23 and *Lr*ATCC 53608 (glycosylation clusters 3 and 4), and at least one in *Lr*MM4-1a (glycosylation cluster 3).

3.3. Summary and Discussion

The range of monosaccharides used in protein glycosylation is often dictated by the nature of the sugar nucleotides found in an organism. To identify the sugar nucleotides that *L. reuteri* strains can synthesise, we searched for genes involved in the synthesis and modification of sugar nucleotides. The results of the *in silico* analysis suggested that the three *L. reuteri* strains *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a are able to synthesise UDP-Glc, UDP-Gal, UDP-Gal_f, UDP-GlcNAc, UDP-ManNAc and GDP-Man using Glc, Fru or Gal, as starting material; *Lr*100-23 may be able to additionally synthesise TDP-Rha. Microbes have been reported to synthesise glycans that mimic the structure of mammalian ones (271), but this does not appear to be the case in *L. reuteri* as the three strains appear to lack the necessary enzymes to utilise GalNAc, Fuc or Neu5Ac, sugars commonly found in mucin O-glycans. However, these *in silico* results would need to be confirmed by profiling the sugar nucleotides synthesised by *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a using HPLC based techniques.

The lack of genes involved in HexNAc metabolism, as well as utilisation of carbohydrates often found in host glycans, was confirmed by growth assays on various mono-, di- or trisaccharides used as sole carbon source. The results showed that the three *L. reuteri* strains could utilise lactose, maltose and sucrose with different efficiency, but none of the HexNAcs or modified lactose tested, thus confirming the results from the bioinformatics analysis. This suggests that the *L. reuteri* strains tested rely mostly on diet-derived oligosaccharides rather than on host-

derived glycans, which is consistent with their localisation in the GI tract (104). Transcriptomics studies would be required to gain further insights into the observed differential growth profile of the *L. reuteri* strains, despite sharing similar genetic potential for the utilisation of these carbohydrates.

To identify putative systems involved in protein glycosylation, we searched for gene clusters encoding putative GTs. It is of note that as the *in silico* analysis focused on gene clusters containing GTs, the levansucrase (fructosyltransferase) of *Lr*100-23 (287) or *Lr*ATCC 53608 was not identified. This enzyme is classified as a glycoside hydrolase (GH family 68) and is involved in the formation of levan, an extracellular fructose polymer, by transglycosylation activity (287). Five clusters containing two or more putative GTs were identified in *Lr*100-23 and *Lr*ATCC 53608 and three in the *Lr*MM4-1a genome. These included a cluster responsible for EPS formation and a cluster of unknown function, as well as a cluster containing two GT-encoding genes highly similar to *gtf1* and *gtf2* from *L. plantarum,* which are involved in protein glycosylation (274). In addition, a $SecA_2/Y_2$cluster containing varying number of GTs was identified in *Lr*100-23 and *Lr*ATCC 53608, as previously described in (103) and (159), respectively.

The identified glycosylation clusters *gtf1/gtf2* and $secA_2/Y_2$ are both predicted to be involved in protein O-glycosylation. While the presence of an *en bloc* N-glycosylation system similar to that found in *Campylobacter* species is unlikely to be identified in Gram-positive bacteria, due to their different cell surface structure, the presence of a N-glycosylation system similar to that of *Haemophilus influenza* cannot be excluded. However, no genes encoding for proteins that could be involved in a N-glycosylation system could be identified, based on homology to the proteins involved in protein N-glycosylation in Gram-negative bacteria, suggesting that *L. reuteri* can probably perform O-glycosylation.

In *L. plantarum,* Gtf1 and Gtf2 act together to add a α-GlcNAc moiety to various proteins including a Acm2 muramidase (274). As these two GTs have highly conserved homologues in *L.*

reuteri (E-value < 1.0^{-130}), they could act in a similar way. Similarly to *L. plantarum,* no neighbouring GTs were identified, suggesting that this modification is not further extended by other monosaccharides.

The $secA_2/Y_2$ clusters in *Lr*100-23 and *Lr*ATCC 53608 shared a similar organisation but differed in the number of GTs constituting the cluster, as well as the localisation of the *srr* gene (at the 5'-end in *Lr*ATCC 53608 and at the 3'-end in *Lr*100-23). Both systems contained GtfA, B, and C, but the *Lr*100-23 cluster contained three additional GTs and there is an insertion of a putative transposase to the last GT, suggesting that it is inactive. The additional GTs have a reversed organisation to dGT1 from *S. parasanguinis* FW213, which mediates the third and fourth glycosylation steps of Fap1 (191). These results suggest that $SRRP_{53608}$ is probably glycosylated with a disaccharide, whereas $SRRP_{100\text{-}23}$ may be glycosylated with glycans ranging from two to ten monosaccharides long, depending on the activity of the additional GTs present in the $SecA_2/Y_2$ cluster.

In conclusion, based on the *in silico* analysis of *L. reuteri* genomes, two putative protein glycosylation systems were identified in *Lr*100-23 and *Lr*ATCC 53608 (*gtf1/gtf2* and the $secA_2/Y_2$ cluster) and one putative protein glycosylation system in *Lr*MM4-1a (*gtf1/gtf2* cluster). These glycosylation systems may utilise the sugar nucleotides identified by the bioinformatics analysis. As this analysis is based on annotation of the genomes, the presence of additional sugar nucleotides or further glycan modifications cannot be excluded in the *L. reuteri* strains analysed in this work.

Chapter 4

Analysis and identification of L. reuteri glycoproteins and glycosylation pathways

4.1. Introduction

Protein O-glycosylation in bacteria is often associated with surface or secreted proteins. For example, *en bloc* O-glycosylation in Gram-negative bacteria occurs in the periplasm and many of the target proteins are found on the membrane (237,272). Similarly, sequential O-glycosylation identified in Gram-positive bacteria such as *L. plantarum, Streptococcus parasanguinis* and *S. pneumoniae* (260,267,277,278) occurs intracellularly and the target proteins are either secreted in the environment, or anchored on the surface, *via* specific cell-wall anchoring motifs (see section 1.3). Our bioinformatics analysis showed that *L. reuteri* strains harbour several gene clusters containing glycosyltransferases (GTs). For example, glycosylation cluster 3 (see section 3.2.3.2) encodes two putative GTs, homologous to the Gtf1 and Gtf2 from *L. plantarum* WCFS1, which have been shown to be involved in protein glycosylation (274). In addition, *L. reuteri* 100-23 (*Lr*100-23) and ATCC 53608 (*Lr*ATCC 53608) have been found to harbour a functional auxiliary $SecA_2/Y_2$ system (109,159), similar to that found in pathogenic Gram-positive bacteria, where it has been shown to be involved in the glycosylation of large, serine rich repeat proteins (SRRPs) (193). These data suggest that *L. reuteri* has the capacity to carry out protein glycosylation. However, no biochemical or structural information is available to date on the nature of glycosylated proteins in *L. reuteri* strains.

The aim of this chapter is to determine the ability of *L. reuteri* to perform protein glycosylation and to identify the major glycoproteins secreted by *L. reuteri* 100-23 (*Lr*100-23), ATCC 53608 (*Lr*ATCC 53608) and MM4-1a (*Lr*MM4-1a) using lectin screening and monosaccharide composition analysis.

4.2. Results

4.2.1. Lectin screening of *L. reuteri* secreted proteins (SM)

To determine whether *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a are capable of performing protein glycosylation, the strains were grown in MRS or LDM-II, and the spent media (SM) proteins were analysed by SDS-PAGE followed by western blot using a range of fluorescein (*f*-)-labelled lectins (summarised in Table 17).

Lectins are proteins that recognise and bind on specific carbohydrate structures and they provide an invaluable tool in glycoprotein analysis. Most plant lectins are specific to mono-, or disaccharides and can be used to identify characteristic epitopes on glycoconjugates. In addition, mammalian lectins form part of the immune system, and some of them recognise microbial carbohydrate structures.

Table 17 Specificity of plant lectins used to identify putative SM glycoproteins from *L. reuteri* strains, after SDS-PAGE and western blot.

Lectin	Abbreviation	Specificity
Concanavalin A	ConA	α-mannose
Lotus tetragonolobus lectin	LTL	α-L-fucose
Peanut agglutinin	PNA	Gal-(β-1,3)-GalNAc (T-antigen)
Ricinus communis agglutinn	RCA	Gal or GalNAc
Sambucus nigra agglutinin	SNA	α-2,6-linked sialic acid (α-2,3-linked sialic acid with lower affinity)
Ulex europaeus agglutinin	UEA	α-L-fucose (differs from LTL)
Wheat germ agglutinin	WGA	GlcNAc or sialic acid

*Lr*100-23, and *Lr*ATCC 53608 SM proteins shared a similar lectin recognition profile with specific binding of f-WGA, f-RCA and f-SNA (Figure 23A, B, C). *Lr*MM4-1a SM proteins showed a different lectin recognition profile with binding restricted to f-RCA and f-SNA (Figure 23B, C). No binding was observed with f-ConA, f-LTL, f-PNA, or f-UEA against SM proteins from the three strains (data not shown). More specifically, f-WGA recognised a large protein with an apparent molecular weight (MW) >300 kDa in *Lr*100-23 and *Lr*ATCC 53608, in addition to a smaller protein of ~130 kDa and 50 kDa, respectively (Figure 23). The high MW protein band in *Lr*100-23 was mainly detected when the strain was grown in the MRS medium, whereas it was predominant when *Lr*ATCC 53608 was grown in LDM-II (Figure 23A), suggesting that the growth medium may influence protein secretion and/or glycosylation. In addition, western blot analysis of the *Lr*ATCC 53608 SM proteins with anti-SRRP53608 antibodies revealed that SRRP migrates at the same size as the protein recognised by f-WGA (Figure 23D), suggesting that this protein may be glycosylated with GlcNAc or sialic acid.

f-RCA recognised several protein bands in *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a, ranging from 40 to 130 kDa (Figure 23). Protein bands at 75 kDa and 55 kDa were shown to be recognised by f-RCA across all three L. reuteri strains, which could suggest that these strains share the same target proteins and glycosylation system. In addition to these proteins, f-RCA also recognised a high MW band at ~500 kDa, which may correspond to a mucus binding protein (MUB_{53608}), the major adhesin in *Lr*ATCC 53608 (151,154). This protein was recognised by an anti-MUB_{53608}-antibody (Figure 23E), suggesting that MUB_{53608} may be glycosylated with glycans containing galactose residues.

f-SNA recognised SM proteins in all three strains grown in LDM-II, ranging from 20 to ~100 kDa (Figure 23C). Samples from MRS cultures gave similar results (data not shown), suggesting that the growth medium does not affect this glycosylation process.

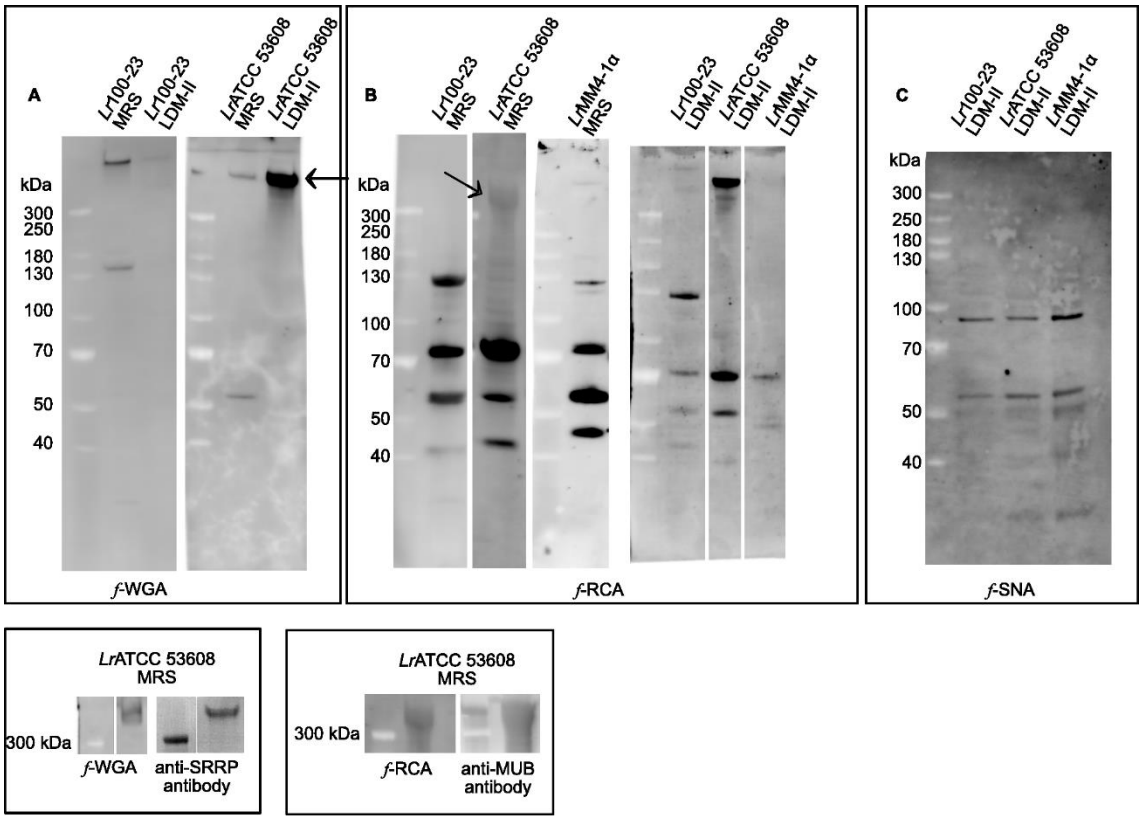

Figure 23 Western blot analysis of SM proteins from *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a cultures in MRS or LDM-II. The blots were probed with various FITC-labelled lectins. A) f-WGA, B) f-RCA, C) f-SNA. D) *Lr*ATCC 53608 MRS SM probed with f-WGA or anti-SRRP antibody. E) *Lr*ATCC 53608 MRS SM probed with f-RCA or anti-MUB_{53608} antibody.

Taken together, these results suggest that the glycans present on *L. reuteri* SM proteins may contain galactose (Gal) or N-acetyl-galactosamine (GalNAc), N-acetyl-glucosamine (GlcNAc) or sialic acid, based on RCA, WGA and SNA binding, respectively. The presence of GlcNAc and Gal was predicted by bioinformatics analysis (see section 3.2.1), however, no enzyme involved in sialic acid metabolism was identified. As there is no sialic acid in the LDM-II growth media, this result suggests that there may be un-annotated genes that encode proteins involved in the biosynthesis of sialic acid and its utilisation in protein glycosylation or that SNA recognises modified sugars other than sialic acid, or that the interaction is not specific. To further investigate if sialic acid is present in *L. reuteri* glycoproteins, these were treated with neuraminidase A from *Arthrobacter ureafaciens*, an enzyme that cleaves sialic acid, irrespective

of linkage, and the reaction products were analysed by SDS-PAGE and western blot, using *f*-SNA or *f*-WGA. The results showed that recognition of these two lectins towards SM proteins remained unaltered (Figure 24).

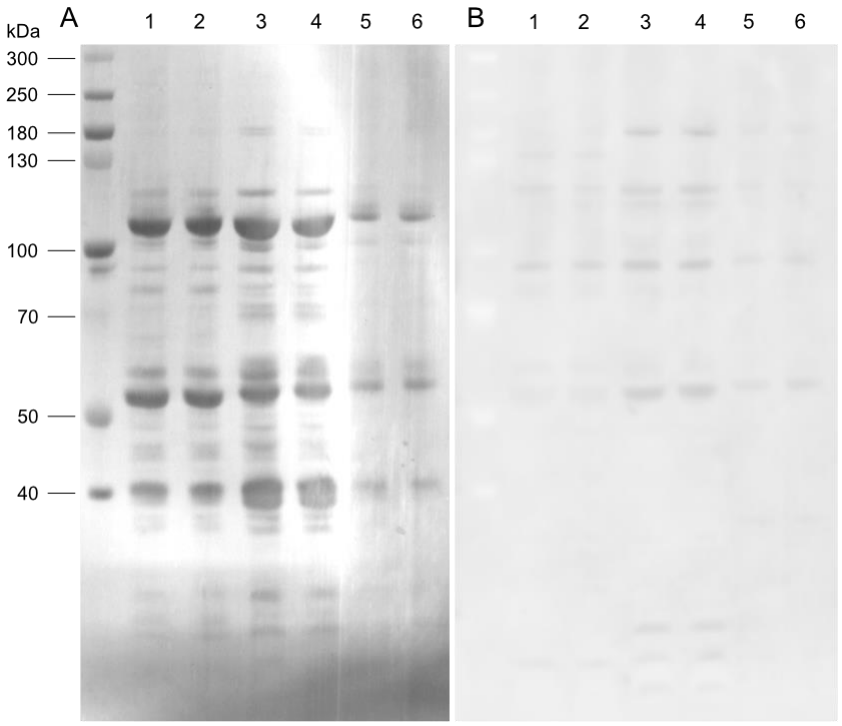

Figure 24 Treatment of *L. reuteri* SM proteins with neuraminidase A. 1) *Lr*100-23 control, 2) *Lr*100-23 treated, 3) *Lr*ATCC 53608 control, 4) *Lr*ATCC 53608 treated, 5) *Lr*MM4-1a control, 6) *Lr*MM4-1a treated. A) InstantBlue stained blot, B) f-SNA probed blot. The results show that neuraminidase A has no effect on SNA recognition of *L. reuteri* SM proteins.

This suggests that either the sialic acid present on *L. reuteri* glycoproteins cannot be cleaved by neuraminidase A due to some modification, or that SNA interaction is not glycan mediated. To address this, chemical release of sialic acid was employed to SM proteins from *Lr*100-23, *Lr*ATCC53608 and *Lr*MM4-1a cultures. Briefly, SM proteins were precipitated, mildly hydrolysed in acetic acid and the reaction products were labelled with 1,2-diamino-4,5-methylenedioxybenzene (DMB). Following analysis by High Performance Liquid Chromatography (HPLC), no sialic acid could be detected in the samples, whereas control samples showed the elution and detection of sialic acids between 15 and 35 min, under the conditions analysed (data not shown).

These data, confirmed our bioinformatics analysis that did not identified any sialic acid biosynthetic pathway, suggesting that SNA recognition of *L. reuteri* SM proteins was not specific and that no sialic acid is present in these proteins.

4.2.2. Monosaccharide composition analysis of *L. reuteri* SM proteins

Another approach to assess whether *L. reuteri* proteins are glycosylated was to carry out monosaccharide composition analysis of SM proteins by gas chromatography – mass spectrometry (GC-MS). This technique allows the identification of monosaccharides present on glycoproteins. To that end, *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a were grown in LDM-II medium for 18 h, SM proteins were precipitated from the SM by methanol/chloroform precipitation to remove free sugars and methanolysed. The released monosaccharides were re-N-acetylated, derivatised with trimethylsilane and analysed by GC-MS.

The results are summarised in Table 18. The major monosaccharides identified in this analysis were rhamnose (Rha), Fuc, xylose (Xyl), Gal, glucose (Glc) and GlcNAc (Figure 25). The strongest peak in all chromatograms corresponds to Glc, but shows large variation in concentration between the different samples. The second most intense peak corresponded to Gal, in SM proteins from all three *L. reuteri* strains. Rha was one of the major peaks in *Lr*100-23 SM proteins but was not detected in the *Lr*ATCC53608 and *Lr*MM4-1a samples. GlcNAc was also found in all three strains with highest concentration in *Lr*ATCC 53608. Fuc and Xyl were identified in *L. reuteri* 100-23 and ATCC 53608, but not in *Lr*MM4-1a.

Table 18 Monosaccharides identified by GC-MS, after methanolysis of L. reuteri glycoproteins and derivatisation of the released monosaccharides with TMS.

Monosaccharide	**Retention time (min)**	**Concentration (µg / mg of protein)**		
		***Lr*100-23 WT**	***Lr*ATCC 53608**	***Lr*MM4-1a**
Rhamnose	5.5	37.3±7.5	-	-
Fucose	6.3	2.6±0.2	Trace	-
Xylose	7.8	0.2±0.1	Trace	-
Galactose	10.5/12.1	42.9±6.5	15.8±3.3	65±28
Glucose	11.2/11.4	324.3±170	63±17.5	108.2±57
GlcNAc	29.3	1.4±0.3	4.9±1.5	0.26±0.1

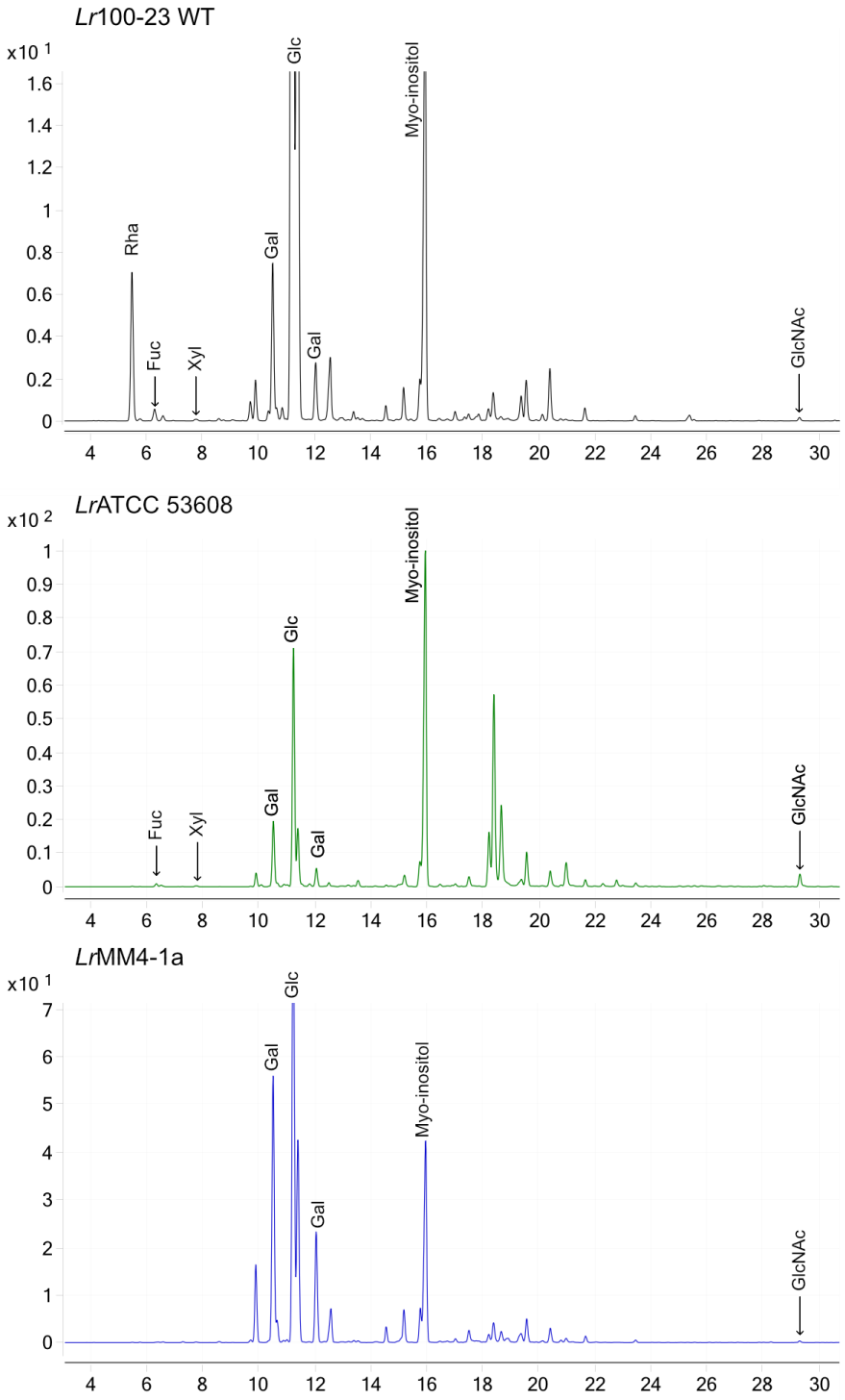

Figure 25 GC-MS chromatogram of monosaccharide composition analysis of glycoproteins from *Lr*100-23 WT, *Lr*100-23 Δgtf1, *Lr*ATCC 53608 and *Lr*MM4-1a. The results showed the presence of Glc, Gal and GlcNAc in each strain, in addition to Fuc and Xyl in *Lr*100-23 and *Lr*ATCC 53608 and Rha in *Lr*100-23.

4.2.3. Identification of *L. reuteri* glycoprotein candidates

To identify the nature of the putative glycoproteins secreted by *Lr*ATCC 53608, protein bands indicated with arrows in Figure 23 were excised from the polyacrylamide gel and subjected to in-gel trypsin digest. The proteins were identified by mass spectrometry (MS). The major protein band from *Lr*ATCC 53608 that interacted with WGA was identified as the serine rich repeat protein ($SRRP_{53608}$). As $SRRP_{53608}$ has no trypsin digestion site within the serine rich repeat regions, the coverage of the tryptic peptides was limited to 21%. MUB_{53608} was also identified in this sample, but the overall score was lower. The high MW protein from *Lr*ATCC 53608 that interacted with RCA was identified by

MS as MUB_{53608}. The MS analysis of the high MW proteins from *Lr*ATCC 53608 that interacted with RCA or WGA confirmed the immunoblotting data obtained from the anti-MUB_{53608} and anti-$SRRP_{53608}$ antibodies, respectively, indicating that MUB_{53608} and $SRRP_{53608}$ are putative glycoproteins. Lastly, the predominant protein in the ~75 kDa protein band from all three strains was found to be a predicted muramidase, homologous to Acm2, a known glycoprotein in *L. plantarum* WCFS1 (275).

To investigate whether $SRRP_{100\text{-}23}$ is glycosylated, the lectin binding profile of three *L. reuteri* 100-23 mutants with an insertion mutation in *asp2, gtfB* or *srr* was determined. *Asp2* encodes an accessory secretion protein which is involved in the secretion of SRRPs and the modification of GlcNAc residues found on the adhesin, whereas *gtfB* encodes a chaperon required for successful glycosylation of SRRP. Briefly, *Lr*100-23 WT strain and the insertion mutants were grown in MRS for 18 h and the secreted proteins were concentrated by spin filtration and analysed by SDS-PAGE and western blot, using fluorescently-labelled lectins. While no major differences in the protein pattern were observed with InstantBlue stain (Figure 26A), differences were noted when the proteins were probed with *f*-WGA after western blot analysis. A protein band >300 kDa was apparent in the WT strain but was missing in the *Lr*100-23 *Δsrr* mutant, as well as the other mutants, suggesting that this protein is $SRRP_{1000\text{-}23}$ (marked with an arrow in Figure 26B). The interaction with WGA, in combination with $SRRP_{100\text{-}23}$'s apparent MW of >300 kDa on SDS-PAGE (predicted MW is 224 kDa) suggest that $SRRP_{100\text{-}23}$ is glycosylated. In addition, other protein bands that were recognised by *f*-WGA were consistently found across all samples and no other differences were observed when the proteins were probed with *f*-RCA (Figure 26C) or *f*-SNA (data not shown). These results suggest that Asp2 and GtfB are dedicated to the processing of $SRRP_{100\text{-}23}$ in *L. reuteri* 100-23 strain.

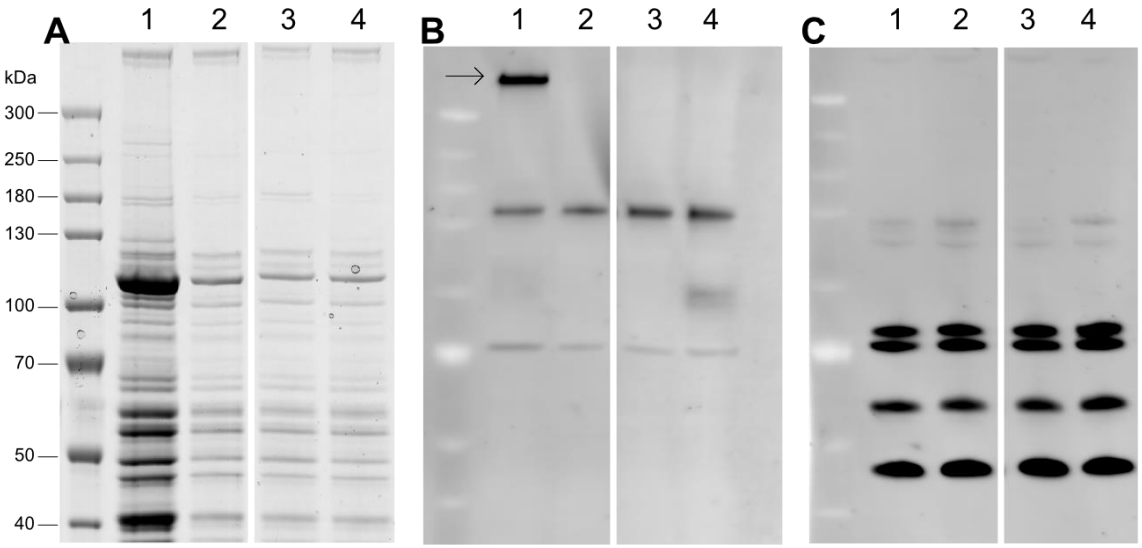

Figure 26 A) SDS-PAGE, followed by western blot analysis with B) f-WGA or C) f-RCA of secreted proteins from 1) *Lr*100-23 WT, 2) Δ*asp2*, 3) Δ*gtfB* 4) Δ*srr*. The arrow indicates the $SRRP_{100\text{-}23}$.

4.2.4. Identification and characterisation of *L. reuteri* glycosylation pathways

4.2.4.1. $Gtf1_{100\text{-}23}$ is involved in protein glycosylation

*Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a harbour a glycosylation cluster composed of two predicted GTs (see section 3.2.3.2), homologous to the Gtf1 / Gtf2 general protein glycosylation system in *L. plantarum* WCFS1 (274).

To investigate the role of Gtf1 in *Lr*100-23, a *Lr*100-23 *Δgtf1* deletion mutant was generated by Dr. Rebbeca Duar (University of Alberta, Edmonton, Canada), following the protocol described in (303). The role of Gtf1 in protein glycosylation was then assessed by western blot analysis of SM proteins from *Lr*100-23 WT and *Lr*100-23 *Δgtf1* grown in MRS (Figure 27). The results showed that while SNA and WGA recognition remained unaffected in the *Lr*100-23 *Δgtf1* mutant compared to the WT strain, RCA binding was abolished when *gtf1* was deleted.

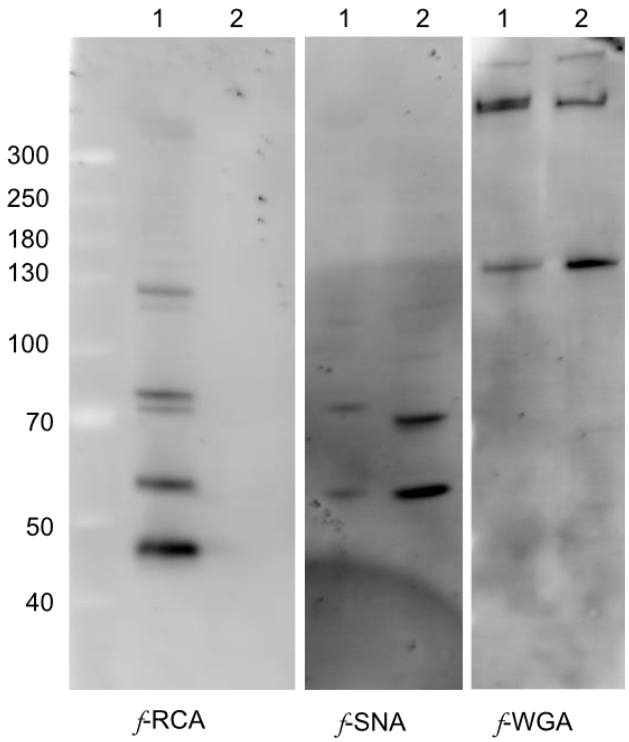

Figure 27 Western blot analysis of SM proteins from *Lr*100-23 WT (lane 1) and *Lr*100-23 Δgtf1 (lane 2) and probing with various lectins. The results show that RCA recognition of *l. reuteri* proteins is lost in the *gtf1* mutant, but the recognition of SNA and WGA was not affected.

Taken together, these results suggest that Gtf1 is involved in a general protein glycosylation system in *L. reuteri* 100-23, similar to the one found in *L. plantarum* WCFS-1 (274), targeting several secreted proteins. However, in contrast to *L. plantarum,* where glycoproteins interact with WGA and are suggested to be glycosylated predominantly with a single GlcNAc molecule on each glycosylation site (277), Gtf1 in *Lr*100-23 appears to be involved in the glycosylation of proteins that are specifically recognised by RCA, *via* a Gal or a GalNAc moiety. As no GalNAc was identified in the monosaccharide composition analysis, the monosaccharide that mediates RCA recognition is more likely to be Gal. Taken together, these data suggest that $Gtf1_{100\text{-}23}$ is involved in the synthesis of glycans containing Gal residues. It is possible that Gtf1 transfers Gal residues directly, or that it adds monosaccharides that precede Gal deposition, leading to a loss of RCA recognition, when *gtf1* is deleted.

4.2.4.2. *Gtf1$_{100\text{-}23}$ is involved in the synthesis of glycans containing both GlcNAc and Gal*

To identify the monosaccharides used by the Gtf1 glycosylation system, monosaccharide composition analysis was performed on SM proteins from *Lr*100-23 *Δgtf1* strain, as described in 4.2.2. Table 19 summarises the identified monosaccharides and their respective concentrations. The mutant strain showed reduced levels of Gal, Glc and GlcNAc (Figure 28); interestingly, both Gal and Glc showed a ~5-fold decrease while GlcNAc was decreased ~10-fold in the mutant strain. A small reduction (>1.5-fold) was observed in the concentration of Fuc and Rha.

Table 19 Monosaccharide composition analysis of *Lr*100-23 *Δgtf1*, compared to *Lr*100-23 WT.

Monosaccharide	**Retention time (min)**	**Concentration (µg / mg of protein)**	
		***Lr*100-23 WT**	***Lr*100-23 *Δgtf1***
Rhamnose	5.5	37.3±7.5	24.52±5.4
Fucose	6.3	2.6±0.2	1.75±0.6
Xylose	7.8	0.2±0.1	0.31±0.1
Galactose	10.5/12.1	42.9±6.5	8.15±1.7
Glucose	11.2/11.4	324.3±170	62.75±26.7
GlcNAc	29.3	1.4±0.3	0.13±0.05

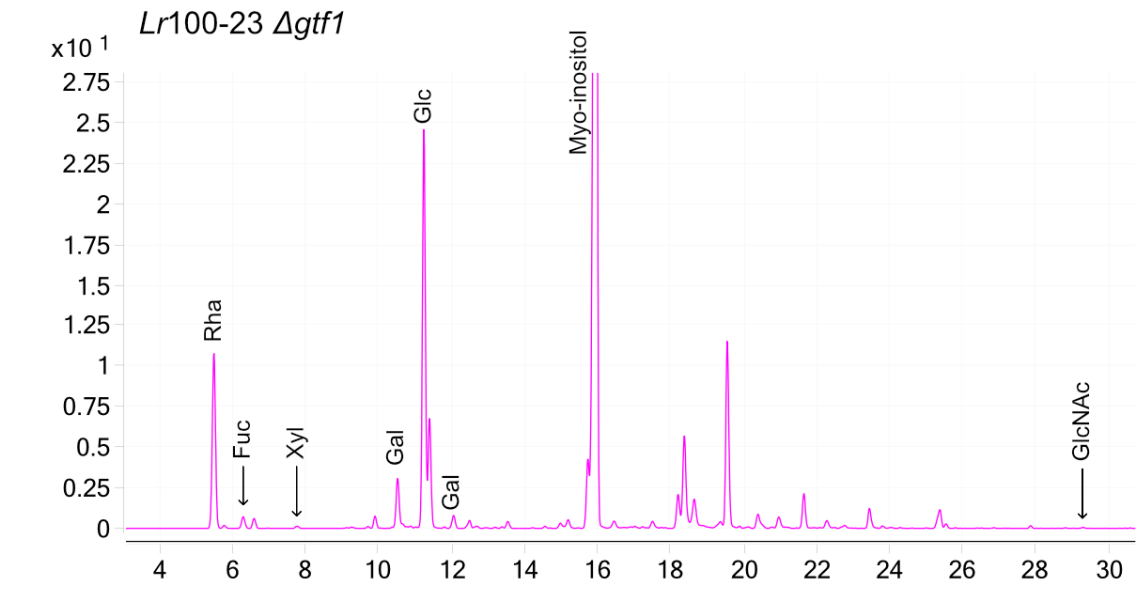

Figure 28 GC-MS chromatogram of monosaccharide composition analysis of glycoproteins from *Lr*100-23 *Δgtf1*. The results showed decreased concentration of Gal, Glc and GlcNAc.

These results suggest that $Gtf1_{100\text{-}23}$ is part of a glycosylation system that generates glycan structures containing Glc, Gal and GlcNAc residues. Based on the similarity of Gtf1 from *Lr*100-23, with its homologous enzyme from *L. plantarum* WCFS1 (43,28%, E-value=10^{146}), it is possible that Gtf1 serves as the priming GT, depositing GlcNAc residues onto the target protein, while additional GTs outside the gene cluster extend the glycan with Glc or Gal moieties. The presence of other modified sugars that were not identified by GC-MC cannot be excluded. The presence of Rha, Fuc and Xyl seems to be independent of Gtf1 and possibly part of glycans synthesised by a different glycosylation system.

4.2.4.3. Gtf1 plays a role in aggregation of Lr100-23

To test if the deletion of *gtf1* affected the growth of *Lr*100-23, the OD_{600} of *Lr*100-23 WT and *Lr*100-23 Δ*gtf1* cultures was monitored for 24 h. Both *Lr*100-23 WT and the isogenic *Lr100-23 Δgtf1* mutant grew to the same level under these conditions. The wild type (WT) strain showed a slight increase in growth rate at 3-4 h, (α=0.05), but no other statistically significant differences were observed over the course of the experiment (Figure 29). This suggests that Gtf1 it is not required for growth in *Lr*100-23, under the conditions tested.

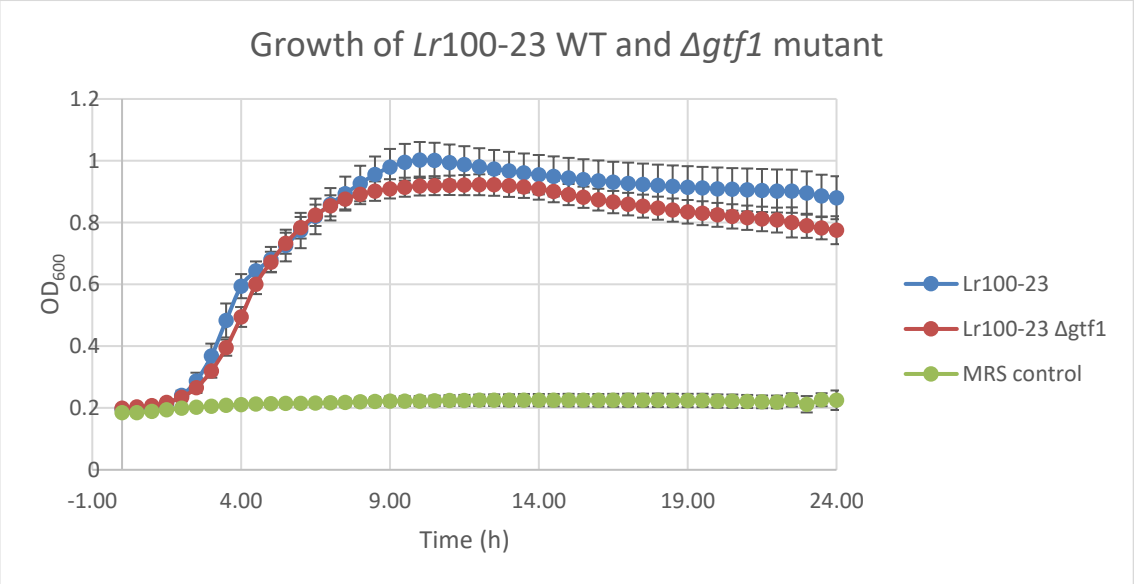

Figure 29 Growth of *Lr*100-23 WT and *Lr*100-23 *Δgtf1* mutant.

However, it was observed that while the parent strain sedimented after prolonged static incubation, the *Δgtf1* mutant remained in suspension, which could suggest impaired aggregation. To further assess the aggregation properties of the *Lr*100-23 *Δgtf1* mutant, the sedimentation of the bacteria was determined by spectrophotometry (304), in comparison to the WT strain. *Lr*ATCC 53608 and its isogenic mutant *L. reuteri* 1063N (*Lr*1063N) were used as controls. *Lr*1063N carries an insertion in the *mub* gene encoding the surface adhesin MUB_{53608}, which resulted in an early stop codon, leading to the synthesis of a truncated MUB_{53608} ($tMUB_{1063N}$), which lacks the C-terminus LPxTG anchoring motif and is thus secreted to the growth medium but not attached onto the cell surface. As *Lr*1063N lacks its major adhesin, it fails to form aggregates or biofilms, as previously determined by flow-cytometry (151). Here, to measure sedimentation, *Lr*100-23 *Δgtf1* and *Lr*1063N were suspended in PBS and the OD_{600} was monitored regularly for 8 h, using *Lr*100-23 WT or *Lr*ATCC56308 as reference, respectively.

The results showed that the WT strains (*Lr*100-23 and *Lr*ATCC 53608) in the reference cell sedimented faster, leading to an increased OD_{600} measurement. This suggests that they form more or larger aggregates. In contrast, the mutant strains Lr*100-23 Δgtf1* and *Lr*1063N remained in suspension, suggesting impaired aggregation. This observation was also confirmed by macroscopic visual observation of the cultures, which showed a more turbid solution for the mutant, compared to the WT, after 8 h (Figure 30E). The reduced aggregation capability of *Lr*1063N compared to *Lr*ATCC 53608 is consistent with previous observations by flow cytometry, showing that MUB_{53608} is involved in cell aggregation (151). When *Lr*100-23 WT or *Lr*ATCC53608 sedimentation was measured, the OD_{600} measurements fluctuated around the baseline and resulted in overall similar sedimentation rate between the measuring and reference sample, as determined by the slope of the trendline ($m = 10^{-4}$; Figure 30A and C). These results also suggest that the proteins targeted by Gtf1 may be located on the cell surface and involved in cell-cell interactions.

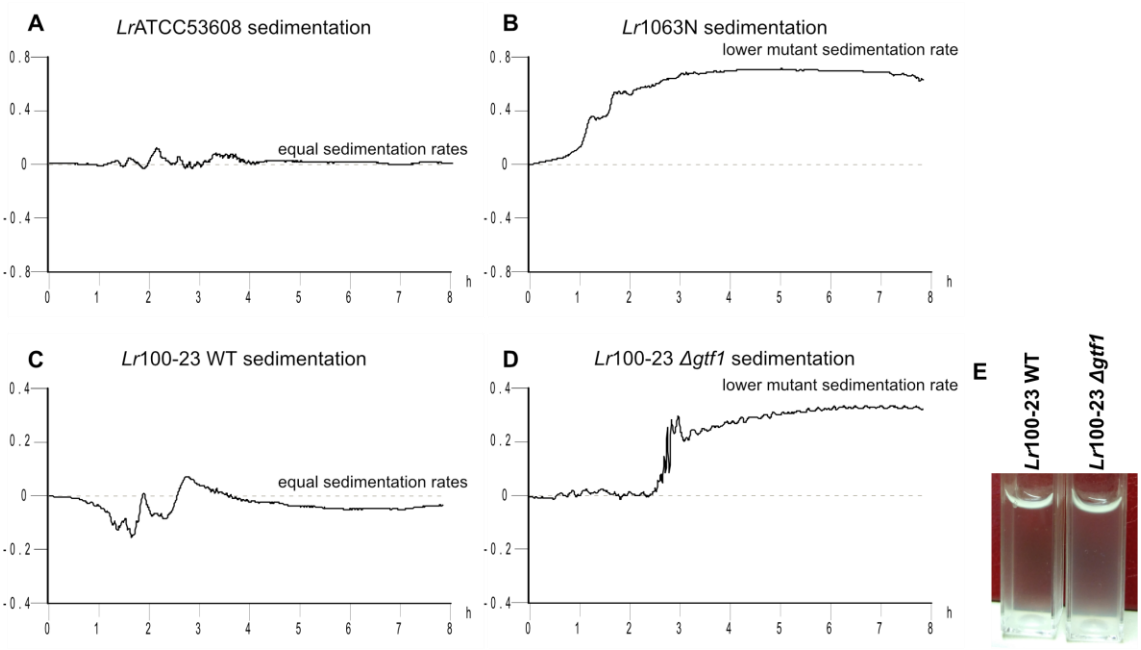

Figure 30 Measurement of the change in ΔOD_{600} of A) *Lr*100-23 WT, B) *Lr*100-23 Δgtf1, C) *Lr*ATCC 53608 and D) *Lr*1063N, using *Lr*100-23 WT as reference for A and B and *Lr*ATCC 53608 for C and D. E) Visual observation of the sedimentation of *Lr*100-2 WT and *Lr*100-23 *Δgtf1*, after 8 h of stationary incubation.

4.2.4.4. Overexpression of putative glycosyltransferases involved in glycosylation of MUB

Deletion of Gtf1 in *Lr*100-23 abolished the recognition of SM proteins by RCA. The high similarity between Gtf1 from 100-23 ($Gtf1_{100\text{-}23}$) and Gtf1 from *Lr*ATCC 53608 ($Gtf1_{53608}$) (98% identities, E-value = 0.0), strongly suggests that this enzyme may have a similar role in *Lr*ATCC 53608. As MUB_{53608} was shown to be recognised by RCA, it may be possible to be a glycosylation target for the $Gtf1_{53608}$/$Gtf2_{53608}$ glycosylation system.

To confirm the role of these enzymes in *Lr*ATCC 53608 (and since it is difficult to transform this strain), $Gtf1_{53608}$ and $Gtf2_{53608}$ were heterologously expressed in *E. coli* BL21 (DE3) using the pOPINF overexpression vector. This vector allows the overexpression of N-terminal His_6-tagged recombinant proteins, and their purification by immobilised metal ion affinity chromatography (IMAC). Upon induction, the proteins were successfully produced, as shown by the overexpressed protein bands at ~60 kDa on SDS-PAGE ($Gtf1_{53608}$ and $Gtf2_{53608}$ have a predicted MW of 59.7 kD and 56.8 kDa, respectively (Figure 31). After cell lysis, $Gtf2_{53608}$ was found in the soluble fraction, whereas $Gtf1_{53608}$ was found in the insoluble fraction. Gtf2 was purified by

IMAC, however, elution with 0.25 M imidazole led to the precipitation of the protein (data not shown). Therefore, the protein was eluted using EDTA, leading to depletion of Ni^{2+} from the IMAC column and elution of $Gtf2_{53608}$ as a soluble protein. Attempts were made to recover Gtf1 from the inclusion bodies, following denaturation with 6 M guanidinium chloride, and re-folding. However, upon removal of the chaotropic agent, the protein precipitated again, preventing any further purification (data not shown).

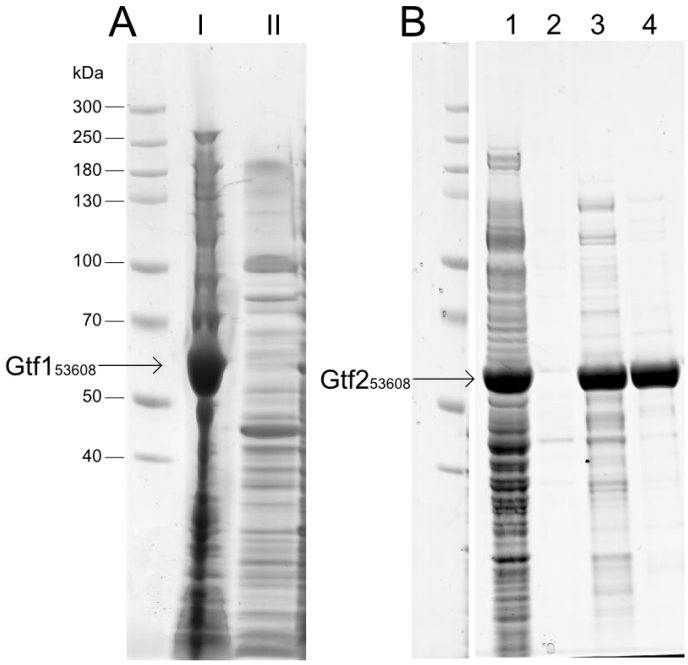

Figure 31 SDS-PAGE analysis of A) *E. coli* pOPINF_gtf1 lysate after induction and B) *E. coli* pOPINF_gtf2 lysates during IMAC purification. I) insoluble fraction, II) soluble fraction, 1) flow-through, 2) wash, 3) first elution fraction, 4) second elution fraction. $Gtf1_{53608}$ was found in the insoluble fraction and was not further purified, whereas $Gtf2_{53608}$ was found in the soluble fraction and purified by IMAC.

Alternatively, attempts were made to co-express $Gtf1_{53608}$ and $Gtf2_{53608}$ in *E. coli* BL21 (DE3) using a pRSFDuet-1 expression system, since co-expression of insoluble proteins with their natural partner can help maintain proteins in solution. Upon induction, $Gtf1_{53608}$ and $Gtf2_{53608}$ were successfully overexpressed (Figure 32), as shown by the intense protein bands at ~58 kDa after SDS-PAGE, however they were both found in the insoluble fraction, after cell lysis, thus preventing further characterisation.

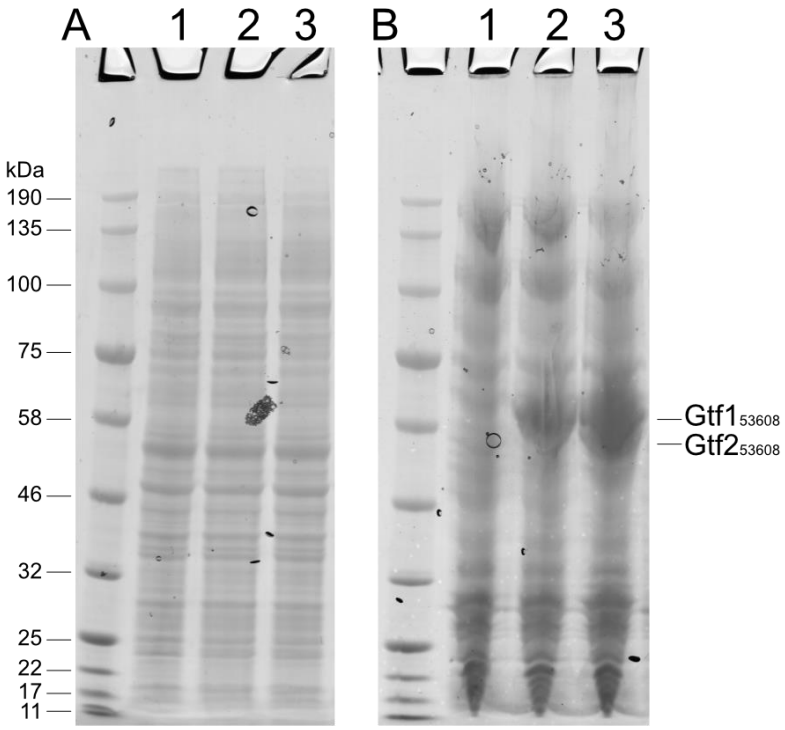

Figure 32 SDS-PAGE analysis of the A) soluble and B) insoluble fractions from 1) *E. coli* BL21 (DE3) pRSFDuet-1, 2) *E. coli* BL21 (DE3) pRSFDuet-1_gtf1 and 3) *E. coli* BL21 (DE3) pRSFDuet-1_gtf1/2 cell lysates. Overexpression of the proteins was successful, however, both proteins were found in the insoluble fraction.

4.2.4.5. The Lr100-23 $SecA_2/Y_2$ cluster is organised into two operons

Protein analysis of *Lr*100-23 insertional mutants with disrupted $SecA_2/Y_2$-associated genes showed that *Lr*100-23 harbours a functional $SecA_2/Y_2$ cluster (109). This includes eight predicted GTs (Figure 33A). GtfA, GtfB and GtfC have been annotated based on homology with other studied proteins (see section 1.4.1.2.4). GtfD and GtfE contain two GT domains, a GT4 at the N-terminus and a DUF1792 at the C-terminus. Genes 2500070898 and 2500070900 have a GT4 and a DUF 1792 domain, respectively, and are probably part of the same gene, which has been split by a transposase, hence they have been annotated as *gtfF1* and *gtfF2*. To determine the transcriptional organisation of the cluster and to identify the expressed GTs, RT-PCR was employed. Briefly, RNA extracted from *Lr*100-23was used to synthesise cDNA. The cDNA was then used as a template in PCR reactions, whereas RNA was used as a negative control. The primers used in the PCR reactions were specifically designed to target the intergenic regions of the genes in the cluster (Figure 33A, Table 14).

Initially, random hexameric nucleotides were used for the synthesis of cDNA. When this cDNA was used a template for the RT-PCR, reactions 1-3, 5, and 9-12 gave bands of the expected size (Figure 33B). This would suggest the presence of three operons (*gtfC-asp2, asp3-secA*$_2$/*Y*$_2$ and *gtfB*$_{C\text{-}term}$-*Tnase*) and five independent genes (*gtfA, gtfB, gtfF2, RTase* and *srr*). In all secA$_2$/Y$_2$ clusters studied so far, the transcriptional organisation varies, but the core genes (*secY*$_2$-*gtfB*) are always linked into a single operon (189,305), in contrast to the findings for the *Lr*100-23 cluster. As it is possible that the random primers used for the cDNA synthesis did not accurately target the entire cluster, gene specific primers were used for the synthesis of cDNA from the genes that did not produce RT-PCR products. When the gene-specific cDNA was used as a template, reactions 4, 6-8 and 12-14 gave bands of the expected size, whereas reaction 15 was the only one that did not yield any product. Taken together, these results suggest that the SecA$_2$/Y$_2$ cluster from *Lr*100-23 is organised into two operons: one spanning genes *gtfC-RTase* and a monocistronic operon containing *srr* alone. This was unexpected as the insertion of two exogenous genes (*Tnase* and *RTase*) could have introduced transcriptional terminator sequences. These results suggest that all GTs are expressed, which could lead to a complex glycosylation pathway.

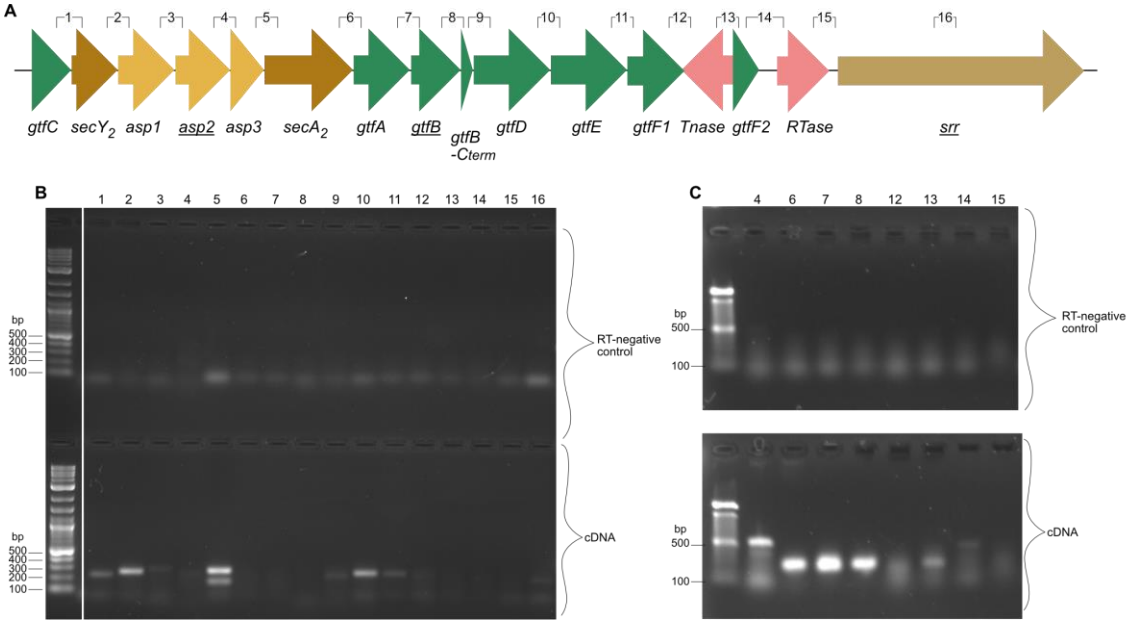

Figure 33 A) schematic representation of the *Lr*100-23 $SecA_2/Y_2$ cluster. B) RT-PCR analysis of the *Lr*100-23 $SecA_2/Y_2$ cluster, using cDNA synthesised by random hexameric primers or C) gene-specific reverse primers. The lane numbers correspond to the intergenic regions noted on A

4.2.4.6. The LrATCC 53608 $SecA_2/Y_2$ cluster is organised into a single operon

RT-PCR analysis of the $SecA_2/Y_2$ cluster from *Lr*100-23 showed that the genes in this cluster are expressed, while the change in the recognition of $SRRP_{100\text{-}23}$ by WGA in *Lr*100-23 insertion mutants of *$secA_2$*, *asp2* and *srr* shows that the glycosylation system is active. Similarly, the interaction of $SRRP_{53608}$ with WGA suggests that the *Lr*ATCC 53608 $SecA_2/Y_2$ glycosylation system is also functional. To characterise the transcriptional organisation of the $SecA_2/Y_2$ cluster, RT-PCR was used to amplify the intergenic regions of the genes composing the cluster, as described for *Lr*100-23 above. The primers used in the PCR reactions were specifically designed to target the intergenic regions of the genes in the cluster (Figure 34A).

The PCR reactions yielded products of the expected size for each targeted sequence, apart from the reaction targeting the intergenic region between *$secY_2$* and *asp1* (Figure 34B, lane 4). This would suggest the presence of two distinct operons, one spanning genes from *srr* to *$SecY_2$* and a second one spanning from *asp1* to *gtfB*. However, given that the intergenic region between *$secY_2$* and *asp1* is only two nucleotides long and does not contain a transcription termination sequence and a promoter for the second operon, it is more likely that a complex secondary structure in this region prevented cDNA synthesis. It is worth noting that the primer pair targeting the intergenic region between *asp1* and *asp2* gave a non-specific band in the RNA control, although the product intensity was significantly higher in the cDNA sample. All other reactions gave a negative result in the RNA control, suggesting that this is unlikely to be due to DNA contamination. Taken together, these results suggest that the $SecA_2/Y_2$ cluster in *Lr*ATCC 53608 is organised into a single operon. In addition, based on the number of GTs expressed in the operon, these data suggest that $SRRP_{53608}$ is probably glycosylated with disaccharide moieties, as, based on previous studies, GtfA and GtfB are predicted to initiate glycosylation, whereas GtfC is the only GT present in the cluster that is predicted to elongate the glycan (see

section 1.4.1.2.4). When the cDNA from the stationary phase was used, no PCR product was obtained (data not shown), suggesting downregulation of the expression of this cluster under these conditions.

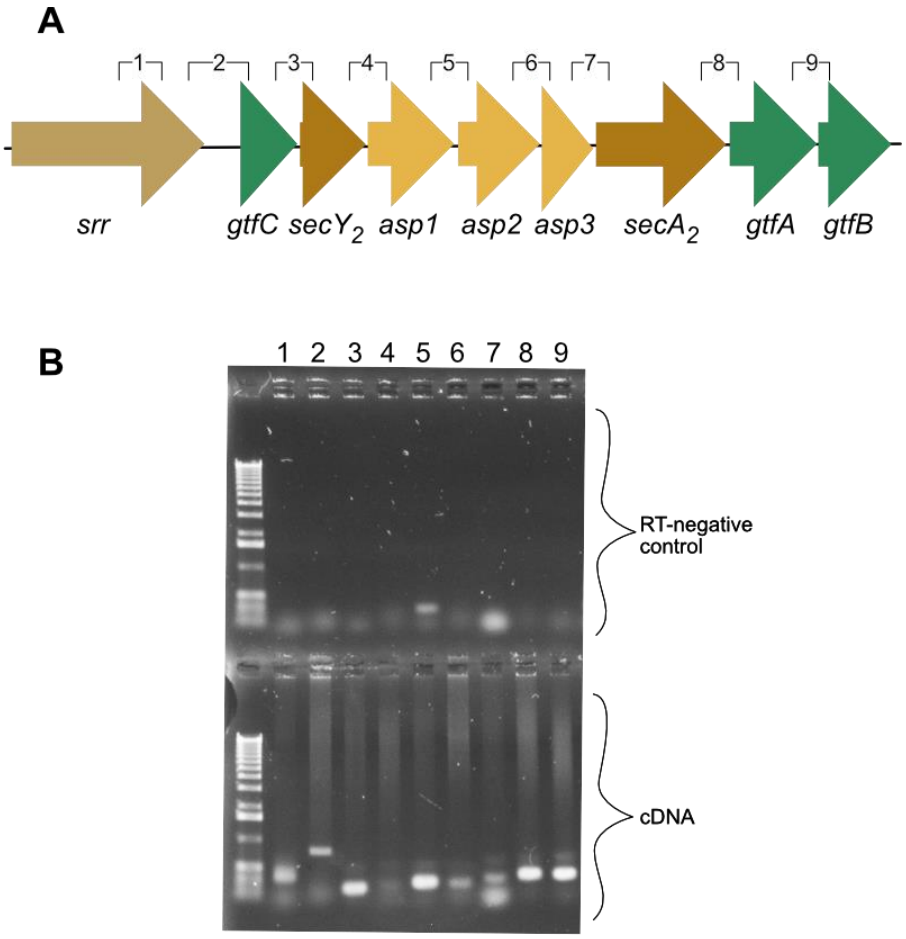

Figure 34 Analysis of the *secA₂/Y₂* gene cluster in *Lr*ATCC 53608 A) Schematic representation of the *secA₂/Y₂* gene cluster. The numbers correspond to the PCR reactions analysed by electrophoresis. B) Agarose electrophoresis of the intergenic PCR products. The results suggest that the cluster is composed by a single operon.

4.2.4.7. *Biochemical characterisation of GTs from the SecA₂/Y₂ cluster by differential scanning fluorimetry (DSF)*

4.2.4.7.1. Overexpression of the $SecA_2/Y_2$ associated glycosyltransferases

Based on studies on homologous glycosylation clusters in streptococcal and staphylococcal systems, GtfA and GtfB are predicted to work together to initiate glycosylation of the SRRP by the addition of GlcNAc residues, whereas GtfC is predicted to mediate the second glycosylation step, by adding a Glc residue onto the extending glycan (191,305,306). To characterise their

activity, GtfA GtfB and GtfC from *Lr*ATCC 53608 were heterologously expressed in *E. coli* BL21 (DE3), using the pOPINF vector. After induction, all three proteins were successfully produced, as determined by SDS-PAGE (Figure 35). $GtfB_{53608}$ was found in the insoluble fraction, preventing further purification and characterisation. $GtfA_{53608}$ and $GtfC_{53608}$ were found in the soluble fraction and further purified by IMAC, using EDTA for the elution, to avoid protein precipitation.

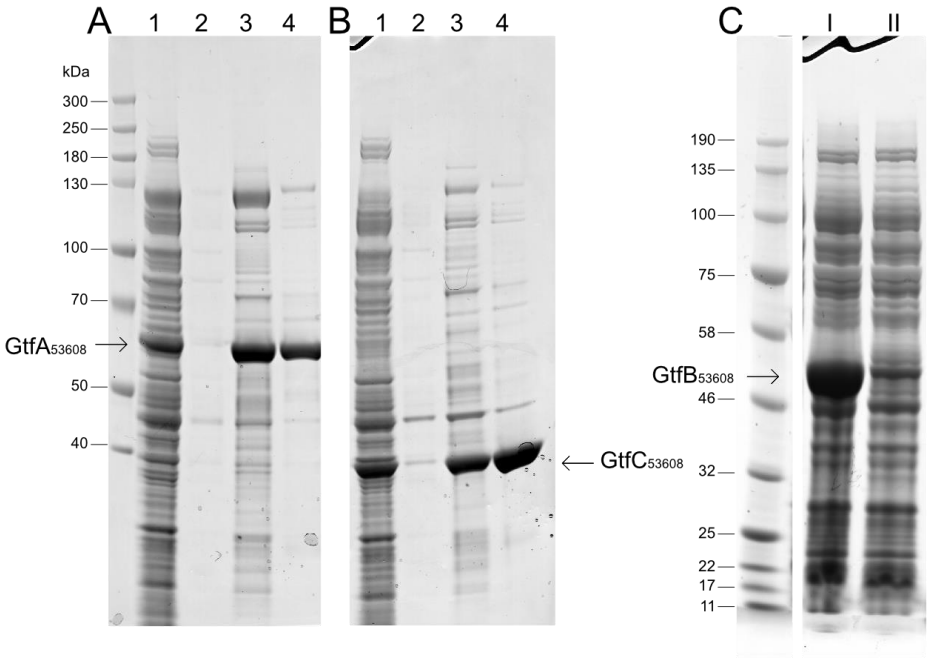

Figure 35 SDS-PAGE analysis of A) *E. coli* BL21 (DE3) pOPINF_gtfA, B) *E. coli* BL21 (DE3) pOPINF_gtfC lysates during IMAC purification and C) *E. coli* BL21 (DE3) pOPINF_gtfB lysate after induction of the overexpression system. 1) flow-through fraction, 2) wash fraction, 3) first elution fraction, 4) second elution fraction. I) insoluble fraction, II) soluble fraction.

4.2.4.7.2. UDP-GlcNAc stabilises $GtfC_{53608}$

Based on homology with functionally characterised homologues, GtfA and GtfC are expected to mediate the first and second step in glycosylation of $SRRP_{53608}$. In the studied staphylococcal and pneumococcal $SecA_2/Y_2$ systems, GtfA adds a GlcNAc residue onto the SRRP and based on homology (~44% identity, E-value > 10^{-150}), GtfA from *Lr*ATCC 53608 is expected to have a similar activity. In addition, since $SRRP_{53608}$ was found to interact with WGA, it is possible that the glycans that modify the adhesin have a terminal GlcNAc moiety. As $GtfC_{53608}$ is the only GT present in the cluster that is predicted to extend the glycan, UDP-GlcNAc is likely to be the donor substrate for both $GtfA_{53608}$ and $GtfC_{53608}$. To test this hypothesis, $GtfA_{53608}$ and $GtfC_{53608}$

interaction with UDP-GlcNAc was assessed by differential scanning fluorimetry (DSF), which can be used to measure the melting temperature (Tm) of a protein. Interactions of proteins with their ligands often lead to increased stabilisation of the protein, and this is reflected by an increased Tm. This approach indicated an interaction between UDP-GlcNAc and $GtfC_{53608}$, but not $GtfA_{53608}$. $GtfC_{53608}$ showed a UDP-GlcNAc-concentration dependent increase in Tm, from 42°C in the absence of the ligand to 47°C in presence of UDP-GlcNAc 4 mM (Figure 36B). No change was recorded for $GtfA_{53608}$, showing a constant Tm of ~38°C across all UDP-GlcNAc concentrations tested (Figure 36A). The absence of recognition of UDP-GlcNAc by $GtfA_{53608}$ is intriguing, as crystal structures of GtfA from *S. pneumoniae* and *S. gordonii* show UDP-GlcNAc binding, independent of GtfB (266,267). As GtfA acts synergistically with GtfB, it is possible that, in solution, binding of UDP-GlcNAc to GtfA requires both proteins, or that the binding of UDP-GlcNAc to GtfA may not be sufficient to stabilise the enzyme under the conditions tested by DSF.

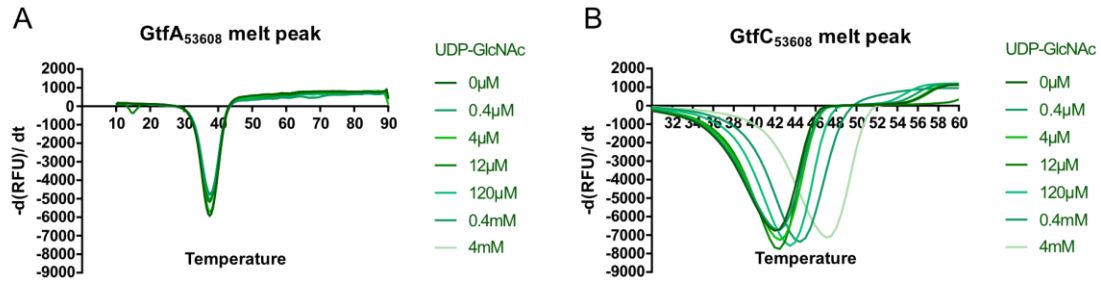

Figure 36 Melt curves of A) $GtfA_{53608}$ and B) $GtfC_{53608}$, in the presence of increasing concentration of UDP-GlcNAc. The ligand appears to stabilise $GtfC_{53608}$, but not $GtfA_{53608}$.

The specificity of GtfC interaction was further tested against UDP, UDP-Gal, and UDP-Glc. The results showed a concentration dependent increase in Tm for all ligands tested (Figure 37). However, the interaction with UDP-GlcNAc at concentrations up to 4 mM caused the greater increase in Tm compared to the other ligands, suggesting that there is a preference of the enzyme towards this sugar nucleotide. At concentrations > 4 mM, UDP caused the greatest increase in Tm. As UDP is smaller than the other ligands, it may be able at high concentrations to bind allosterically to sites that are not accessible to the other ligands, thus stabilising the enzyme further.

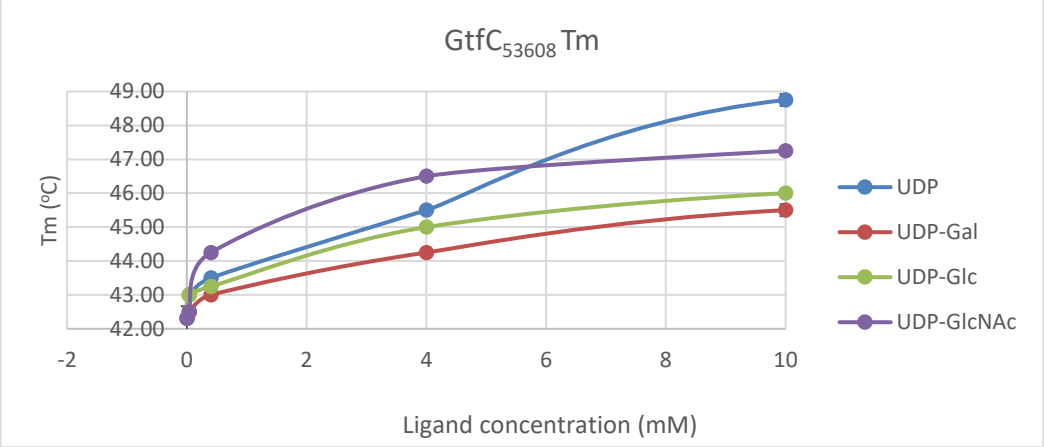

Figure 37 Increase in the Tm of $GtfC_{53608}$ from *Lr*ATCC 53608 in the presence of increasing concentrations of different ligands.

DSF was also used to investigate the binding dependency of $GtfC_{53608}$ activity to metal ions, in the presence or absence of UDP-GlcNAc. In the presence of 4 mM UDP-GlcNAc, the Tm of $GtfC_{53608}$ was increased by 3°C (Figure 38B), whereas in the presence of 5 mM of the divalent ions (Mg^{2+}, Mn^{2+}, Ca^{2+}), the Tm increased by 2.5°C (Figure 38A). However, when both the sugar ligand and the ions were present, the Tm of $GtfC_{53608}$ increased by 7°C, suggesting that divalent ions may be required for optimum binding. The requirement of divalent ions is well established in Leloir GTs, and some examples have been recently reported prokaryotic systems. For example, the dGT1-mediated glycosylation of Fap1 in *S. parasanguinis* is enhanced by divalent ions (262), whereas, using DSF, a glucosyltransferase from *Clostridium difficile* showed a 11°C increase in Tm, when both the sugar substrate and Mn^{2+} were present, also suggesting the dependency of the GT enzymatic activity on divalent ions (307). However, no divalent ions have been identified enzymatically or in the crystal structures of GtfC from other microorganisms so far (263). It is worth noting that $GtfC_{53608}$ does not contain a DxD motif; it does, however contain a DxE motif, which could facilitate binding of the divalent ions, as shown for a GT from the *S. parasangunis* FW213 $SecA_2/Y_2$ cluster (262).

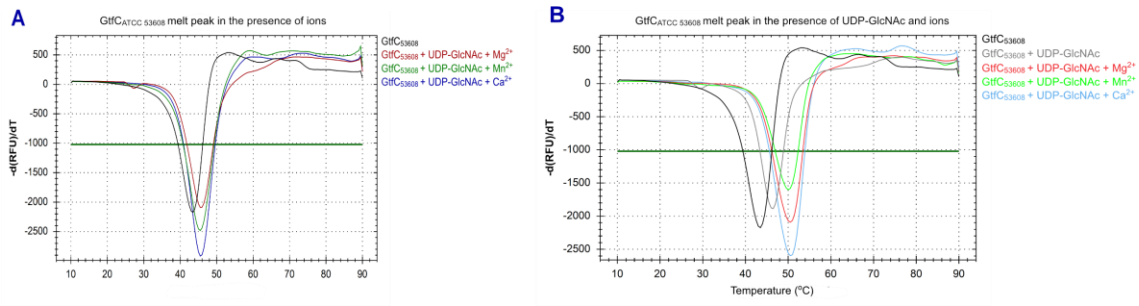

Figure 38 Melt peaks of $GtfC_{53608}$ in the presence of A) divalent ions only or B) UDP-GlcNAc and divalent ions.

4.2.4.7.3. $GtfC_{100\text{-}23}$ is preferentially stabilised by UDP-Glc

$GtfC_{100\text{-}23}$ is predicted to mediate the second step in $SRRP_{100\text{-}23}$ glycosylation, similarly to $GtfC_{53608}$. To determine the acceptor specificity of $GtfC_{100\text{-}23}$ (cloned and expressed by Prozomix, UK), DSF was employed, as described above for $GtfC_{53608}$. Incubation of $GtfC_{100\text{-}23}$ with UDP-Glc led to an increase of the Tm of up to 3°C, whereas as incubation with other ligands had a milder effect, when their concentration was < 4 mM, (Figure 39), indicating a stronger interaction between the enzyme and UDP-Glc. In contrast, at concentrations > 4 mM, UDP led to the highest increase in Tm, as previously observed for $GtfC_{53608}$.

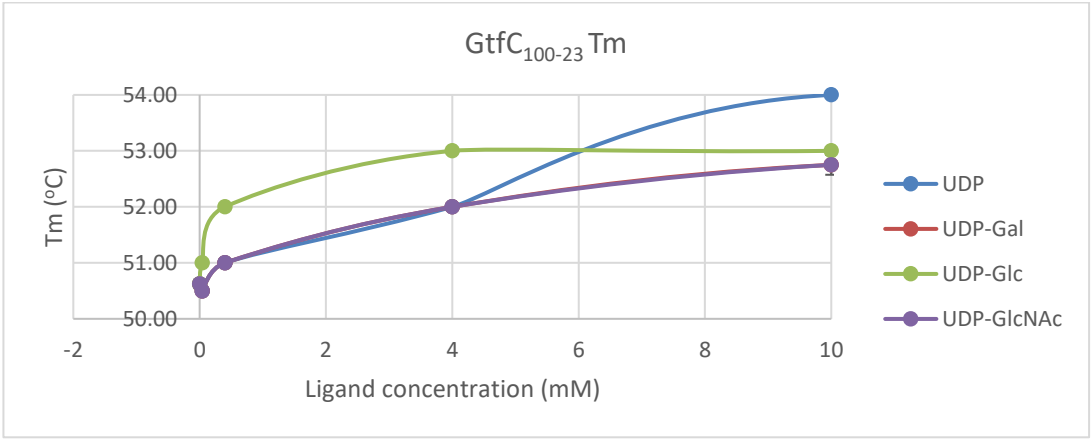

Figure 39 Increase in the Tm of $GtfC_{100\text{-}23}$ in the presence of increasing concentrations of different UDP ligands. The results show a preference for UDP-Glc.

DSF was also employed to determine the effect of divalent ions in the stability of $GtfC_{100\text{-}23}$. The enzyme was incubated with Mg^{2+}, Mn^{2+}, or Ca^{2+} alone, or in the presence of UDP-Glc and the melt peaks showed that the ions had very small effect on the stability of the protein, as determined by the small shift - < 1°C - in Tm (Figure 40), suggesting that divalent ions may not be crucial for

enzymatic activity of $GtfC_{100\text{-}23}$. Interestingly, the DxE motif identified in $GtfC_{53608}$, is also conserved here, but the effect of ions in the stability of $GtfC_{100\text{-}23}$ is not as strong as for $GtfC_{53608}$.

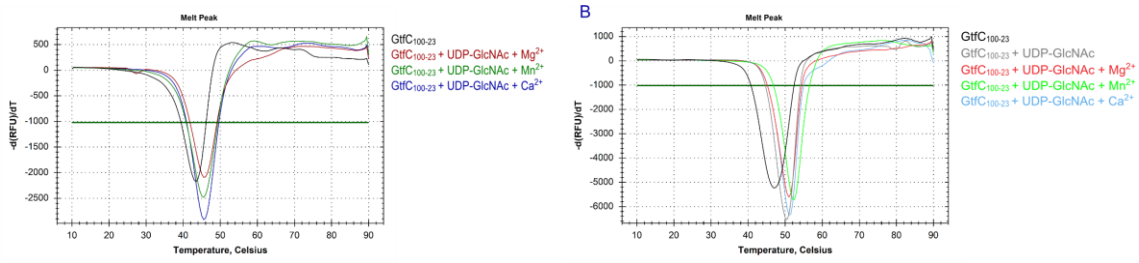

Figure 40 Melt peaks of $GtfC_{100\text{-}23}$ in the presence of A) divalent ions or B) UDP-Glc and divalent ions.

4.3. Summary and discussion

The presence of glycosylation clusters homologous to those involved in protein glycosylation in pathogenic Gram-positive bacteria and *L. plantarum* WCFS1 in the genome of *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a suggests that these strains are capable of performing protein sequential O-glycosylation. These include the $SecA_2/Y_2$ glycosylation system which is dedicated to the glycosylation of SRRPs and a general glycosylation system composed of two putative GTs, Gtf1 and Gtf2. To assess the ability of *L. reuteri* strains to carry out protein glycosylation, lectin screening of SM proteins from *L. reuteri* cultures was first used. The interaction of SM proteins from *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a with lectins such as RCA and WGA suggests that these strains can modify their proteins with glycans containing GlcNAc and/or Gal. SNA, a sialic acid-specific lectin, was also found to bind SM proteins from *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a, however, further investigation with neuraminidase A and HPLC analysis showed that this interaction was not carbohydrate mediated. In addition, monosaccharide composition analysis of secreted SM proteins revealed the presence of Gal, Glc and GlcNAc in the three *L. reuteri* strains analysed, as well as Xyl and Fuc in SM proteins from *Lr*100-23 and *Lr*ATCC 53608, and Rha in *Lr*100-23. Glc was identified as the strongest peak in all three samples. The high abundance of Gal in *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a is in agreement with the lectin screening showing strong interaction of RCA with SM proteins from all three *L. reuteri* strains. In addition, the absence of GalNAc suggests that RCA-binding is Gal-mediated. The presence of Fuc

and Xyl was not expected, as no genes involved in the synthesis of GDP-Fuc or UDP-Xyl were identified in the bioinformatics analysis. This suggests that there may be un-annotated genes involved in the biosynthesis of these sugar nucleotides or that their presence in SM proteins is due to a different mechanism, such as a transglycosylation reaction. In addition, no binding was observed when UEA was tested against SM proteins from *Lr*100-23 or *Lr*ATCC 53608. This could be due to the low concentration of Fuc, which may be below the detection limit of *f*-UEA. The presence of GlcNAc is in agreement with the bioinformatics analysis that revealed enzymes involved in the biosynthesis of UDP-GlcNAc in these three *L. reuteri* strains. The low concentration of GlcNAc in *Lr*100-23 and *Lr*MM4-1a compared to *Lr*ATCC 53608 is also in agreement with the lectin screening that showed weak or no WGA binding against SM proteins from *Lr*100-23 and *Lr*MM4-1a, respectively. The presence of Rha in *Lr*100-23 SM proteins is consistent with the bioinformatics analysis, that showed the presence of a TDP-Rha biosynthesis cluster in this strain.

To further identify the nature of the putative glycoproteins, immunoblot and MS analysis of proteins recognised by lectins was carried out. MUB_{53608}, a high MW and the major adhesin from *Lr*ATCC 53608 (151,154), and $SRRP_{53608}$ were identified as putative glycoproteins recognised by RCA and WGA, respectively. As there was no antibody against $SRRP_{100\text{-}23,}$ lectin blotting analysis of *Lr*100-23 WT and *Lr*100-23 *Δasp2, ΔgtfB or Δsrr* mutants was used, demonstrating that $SRRP_{100\text{-}23}$ is also glycosylated. Studies on homologous clusters from Streptococci or Staphylococci have shown that SRRPs are the target proteins of this cluster (see section 1.3.3.3.2). Similarly, our analysis of the *Lr*100-23 *ΔgtfB* mutant showed that the Gtf is dedicated to the glycosylation of SRRP. The $secA_2/Y_2$ cluster of *Lr*100-23 and *Lr*ATCC 53608, differs in the length of the SRRP encoded from each strain, as well as the number of GTs present (Figure 33A and Figure 34A). To determine the transcriptional organisation of the two clusters, RT-PCR was performed to amplify the intergenic regions from cDNA. The results showed that the $SecA_2/Y_2$ cluster in *Lr*100-23 is organised into two operons, one containing the genes required for

glycosylation and secretion of SRRP, and a second one containing the target protein SRRP. In contrast, the $SecA_2/Y_2$ cluster from *Lr*ATCC 53608 is organised into a single operon. Studies on $SecA_2/Y_2$ clusters from *Streptococcus salivarius* JIM8777 and *S. pneumoniae* TIGR4 show that the transcriptional organisation of this cluster is highly variable between strains (189,305). However, it is clear that the core genes encoding the proteins for the secretion system, as well as the priming GTs, GtfA and GtfB, are usually organised in the same operon.

While the *Lr*100-23 cluster expresses seven GTs, four of which contain two distinct GT folds each, the *Lr*ATCC53608 one contains only three GTs. Considering that GtfA and GtfB, present in both clusters are required for the first glycosylation step with GlcNAc, $SRRP_{53608}$ is expected to be glycosylated with a disaccharide, while glycosylation of $SRRP_{100\text{-}23}$ appears to be more complex. In order to get some information on the specificity of GtfC, the enzyme that is predicted to mediate the second glycosylation step of SRRP, DSF was employed. The results suggested that despite their high sequence similarity, $GtfC_{53608}$ and $GtfC_{100\text{-}23}$ have different activities: $GtfC_{53608}$ showed stronger interaction with UDP-GlcNAc, in the presence of divalent ions, while $GtfC_{100\text{-}23}$ interacted stronger with UDP-Glc, and the divalent ions tested did not have a significant effect on the binding of the enzyme. Taken together, these results suggested that $SRRP_{53608}$ is glycosylated with a di-GlcNAc, while the core disaccharide of $SRRP_{100\text{-}23}$ is likely to be GlcNAc-Glc.

In addition to MUB_{53608} and SRRPs, a muramidase, homologous to the glycosylated Acm2 from *L. plantarum* WCFS1 was identified as a putative glycoprotein recognised by RCA in all three *L. reuteri* strains, suggesting that this protein may be glycosylated by a conserved glycosylation system.

To assess the role of Gtf1 in *Lr*100-23, SM proteins from a *Lr*100-23 *Δgtf1* deletion mutant were tested for interaction with RCA or WGA. The results showed a complete loss of RCA recognition in the mutant, while the recognition of WGA remained unaffected. This suggests that Gtf1 is also involved in protein glycosylation in *Lr*100-23. Monosaccharide composition analysis revealed a

decrease in Gal, Glc and GlcNAc in the *Lr*100-23 *Δgtf1* mutant. This suggests that Gtf1 is involved in the synthesis of glycans containing these three monosaccharides. The presence of modified monosaccharides that were not identified by GC-MS cannot be excluded. The reduction in Gal concentration is in agreement with the western blot analysis, showing that the affinity of RCA (a Gal-specific lectin) towards *Lr*100-23 glycoproteins is lost when *gtf1* is deleted. Gtf1 has high homology with the *L. plantarum* WCFS1 homologue that is involved in protein glycosylation. In *L. plantarum* WCFS1*,* the synergistic action of two GTs, Gtf1 and Gtf2, is required for protein glycosylation with GlcNAc residues (274), therefore it is likely that the function is conserved in *L. reuteri* 100-23, with Gtf1 transferring GlcNAc residues onto protein targets. Other GTs may be responsible for the subsequent attachment of Gal residues. In addition, while Gtf1 was not essential for bacterial growth under controlled, laboratory conditions in *Lr*100-23, it was shown to be involved in bacterial aggregation. This suggests that either the glycans found on surface glycoproteins act as ligands for surface, lectin-type bacterial adhesins, or that loss of glycosylation may lead to impaired secretion of proteins involved in aggregation and/or biofilm formation. Attempts to characterise heterologously expressed Gtf1 and Gtf2 from *Lr*ATCC 53608 were not successful, as Gtf1 was found to be insoluble when expressed in *E. coli* independently, while both enzymes were found to be insoluble when co-expressed. Further experiments are required to determine the enzymatic activity of these proteins, and confirm their glycosylation targets.

MUB, SRRP and muramidase have been previously shown to mediate the adhesion of Lactobacilli to mucin and/or the epithelium of their host (109,144,151,156). Glycosylation of these adhesins may suggest a role for the glycans in the adhesion mechanism of *L. reuteri* strains, as reported for pathogenic bacteria (see section 1.4.1.2.4), underscoring the importance of protein glycosylation for bacterial colonisation.

Chapter 5

Characterisation of the *L. reuteri* adhesin glycosylation

5.1. Introduction

In the previous chapter, we identified $SRRP_{100\text{-}23}$, $SRRP_{53608}$ and MUB_{53608}, the major adhesins of *L. reuteri* 100-23 (*Lr*100-23) and ATCC 53608 (*Lr*1ATCC 53608), respectively, as putative glycoproteins.

MUB_{53608} is a long, multidomain protein of a predicted molecular weight (MW) of 353 kDa. However, it migrates at ~500 kDa on SDS-PAGE (154). This aberrant migration is indicative of post-translational modifications. In addition, it has been shown to interact with C-Type lectins, such as Dectin-2 and DC-SIGN, that are part of the mammalian immune system (175). Both Dectin-2 and DC-SIGN recognise mannose-containing glycans (308), while DC-SIGN can also recognise fucose (309,310), suggesting that MUB_{53608} is glycosylated.

SRRPs are the target proteins of the $SecA_2/Y_2$ glycosylation and secretion system see section 1.3.3.3.2). Often, more than one SRRPs are found in each cluster, however, in most cases, only one *srr* gene is considered to be functional, while the other ones are annotated as pseudo-genes. While this cluster is well studied in pathogenic strains of streptococci and staphylococci, little is known about its role in the physiology of commensal bacteria. SRRPs are long surface proteins composed of an unusually long N-terminal signal peptide (~90 aa), followed by a short alanine-serine-threonine (AST) domain, a first serine rich repeat region (SRR-1), the binding region (or basic region; BR), a second, longer serine rich repeat region (SRR-2) and lastly an LPXTG anchoring motif. Many differences can be observed between SRRPs of different origins, as well as within the same organism. These differences are mostly located in the number and composition of the serine rich repeats in SRRP-1 and SRRP-2, as well as the composition of the BR region. For instance, $SRRP_{100\text{-}23}$ has 101 10-amino acid long serine-rich repeats, whereas $SRRP_{53608}$ only contains 22 (Figure 41) (159). In addition, the BR regions of the two SRRPs only have 48% similarity, with an E-value = $3x10^{126}$.

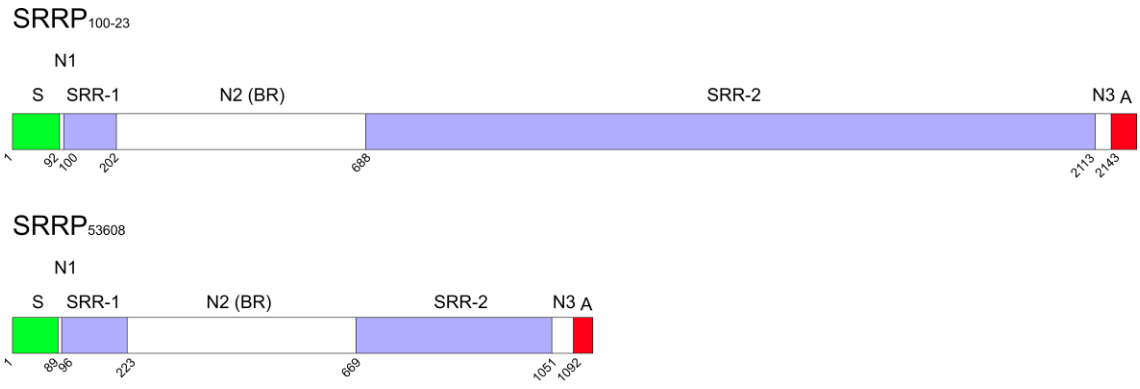

Figure 41 Schematic representation of SRRP from *Lr*100-23 and *Lr*ATCC 53608.

These adhesins are both predicted to be surface anchored by an LPXTG anchoring motif. Sortase A, the surface-bound enzyme responsible for the anchoring of the proteins onto the surface, cleaves the peptide bond between Thr and Gly in the LPXTG motif and forms a peptide bond between the Thr and the N-terminal Gly in the Gly_5 motif found in the cell wall (Figure 42).

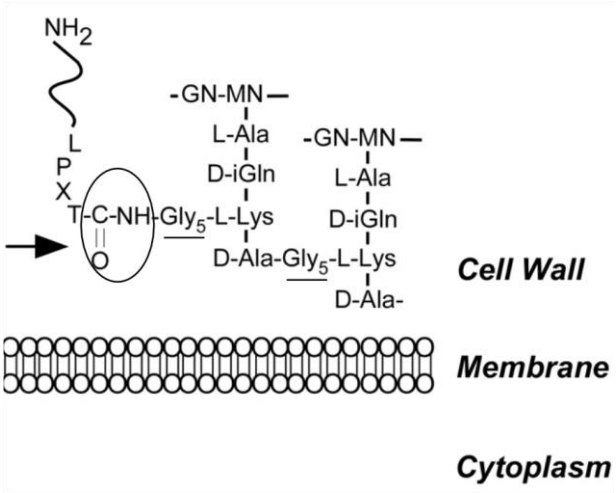

Figure 42 Anchoring of surface proteins via the LPXTG anchoring motif. The arrow points at the peptide bond formed between Thr and Gly. The Gly_5 motif is underlined. Adapted from *(174)*.

The aim of this work is to characterise the glycans found on *L. reuteri* adhesins identified as putative glycoproteins in the previous chapter, namely $SRRP_{100\text{-}23}$, $SRRP_{53608}$ and MUB_{53608}.

5.2. Results and discussion

5.2.1. Characterisation of $SRRP_{100\text{-}23}$ glycosylation pattern

5.2.1.1. Purification of $SRRP_{100\text{-}23}$

To determine the nature of the glycans present on $SRRP_{100\text{-}23}$, the adhesin was purified from *L. reuteri* culture. Briefly, *L. reuteri* 100-23 was grown in LDM-II, supplemented with maltose, for 18 h. After concentration and dialysis of the SM, $SRRP_{100\text{-}23}$ was purified by affinity chromatography, using an agarose-bound WGA column, taking advantage of the lectin recognition results described in section 4.2.3. After washing of the column, bound proteins were eluted with 0.5 M GlcNAc. The collected fractions were analysed by SDS-PAGE (Figure 43). The major purified protein migrated at a MW >300 kDa. MS analysis of the gel revealed that $SRRP_{100\text{-}23}$ was the fifth most strongly identified protein, with a score of 472 and coverage of 8%. Other surface proteins, including Lsp (Large Surface Protein, Accession no. 2500070581, score: 1225, coverage: 22%) were also identified. However, the score for $SRRP_{100\text{-}23}$ is underestimated, as the serine-rich repeat regions (accounting for ~65% of the total protein) contain no trypsin digestion sites, preventing their identification by MS due to the size of the resulting peptides.

These results suggest that either the identified surface proteins are also glycosylated with GlcNAc residues, but the abundance of the glycans is below the detection limit of WGA on a western blot, or that these surface proteins interact with SRRP, leading to co-purification by WGA-affinity chromatography. A smaller protein at 130 kDa was also co-purified in much lower yield. This was expected from earlier results showing that WGA recognised a protein band of this size after western blot analysis of the secreted proteins from *Lr*100-23 (see section 4.2.1). MS analysis identified this protein as an LPXTG- anchored peptidase (accession no. 2500070874, score: 1088, coverage: 40%) of a predicted MW of 100 kDa, suggesting that this may also be a putative glycosylation target. The elution fractions containing $SRRP_{100\text{-}23}$ were pooled and dialysed against ammonium bicarbonate. Purification of $SRRP_{100\text{-}23}$ by WGA affinity

chromatography supported the results showing that WGA interacts with $SRRP_{100\text{-}23}$ in a glycan-mediated interaction.

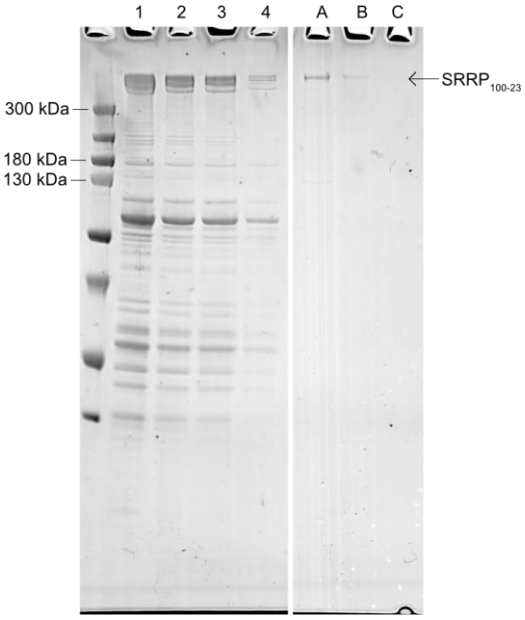

Figure 43 Purification of $SRRP_{100\text{-}23}$ by WGA affinity chromatography. Lane 1) starting material, Lanes 2-4) wash fractions, Lanes A-C) elution fractions.

5.2.1.2. $SRRP_{100\text{-}23}$ is glycosylated with $HexNAc_1Hex_2$ moieties

Previous studies on SRRPs from pathogenic organisms have shown that these proteins are O-glycosylated on serine residues (191,269). To identify the glycans present on $SRRP_{100\text{-}23}$, the purified protein was subjected to β-elimination, and the released glycans were permethylated and analysed by MALDI-ToF. The spectra showed a characteristic peak at 738 Da, that corresponds to $HexNAc_1Hex_2$, and fragmentation of this ion species suggests it is a linear glycan structure (Figure 44). As the MS analysis of the purified $SRRP_{100\text{-}23}$ identified multiple large surface proteins in the sample, it is possible that this glycan structure was also present in these proteins. To confirm that this is an $SRRP_{100\text{-}23}$-specific glycan, the same glycomics analysis was performed on SM proteins from *Lr*100-23 WT and *Lr*100-23 *ΔgtfB* mutant, shown to have lost the WGA affinity to $SRRP_{100\text{-}23}$. The results showed that the peak at 738 Da was missing in the mutant sample, while the rest of the spectra was identical (data not shown). Taken together, these data suggest that $SRRP_{100\text{-}23}$ is glycosylated with $HexNAc_1Hex_2$ moieties.

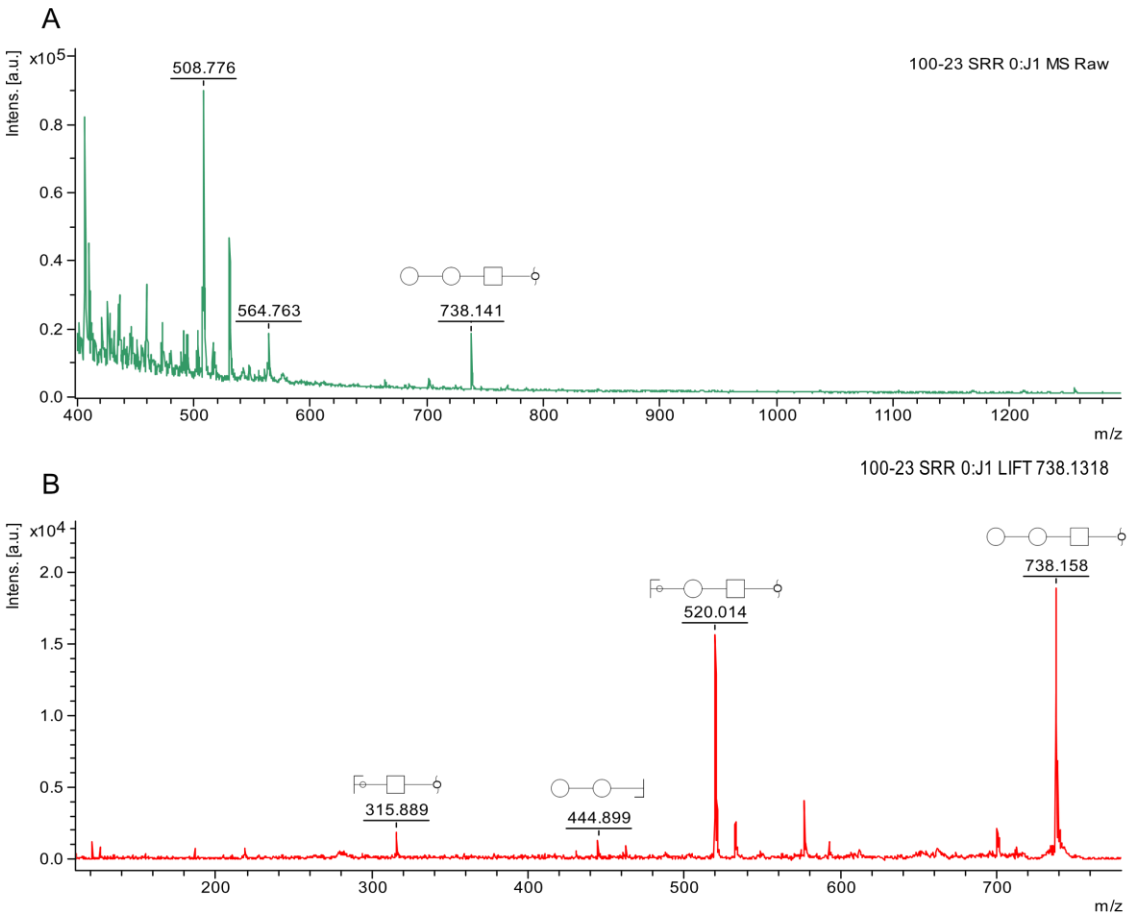

Figure 44 A: MS spectrum of the released and permethylated O-glycans from $SRRP_{100-23}$. B) Fragmentation spectrum of the peak at 738 Da. The data suggest the presence of $HexNAc_1Hex_2$ moieties on $SRRP_{100-23}$.

5.2.1.3. Monosaccharide composition analysis of $SRRP_{100-23}$

To determine the nature of the monosaccharides constituting $SRRP_{100-23}$ glycosylation, sugar analysis was performed on purified proteins by GC-MC, after methanolysis, N-acetylation and TMS-derivatisation of the released methyl-glycosides. The results showed mainly the presence of Glc, as well as smaller amounts of Gal, GlcNAc and Rha (Figure 45). Based on the MALDI-ToF analysis, it was expected that Hex and HexNAc would be found in a 2:1 ratio. However, the total amount of Hex molecules is more than 100 times higher than GlcNAc. This could suggest the glycosylation of $SRRP_{100-23}$ with single Glc or Gal residues, which are not retained by the graphitised carbon cartridges used for desalting the samples during preparation and therefore cannot be detected by MALDI-ToF analysis.

In addition, no deoxy-hexose was identified in the MALDI-ToF analysis, so the presence of Rha was not expected. It may be that only a small fraction of the glycans carries a Rha residue, which will be below the detection limits of MALDI-ToF, under the conditions tested.

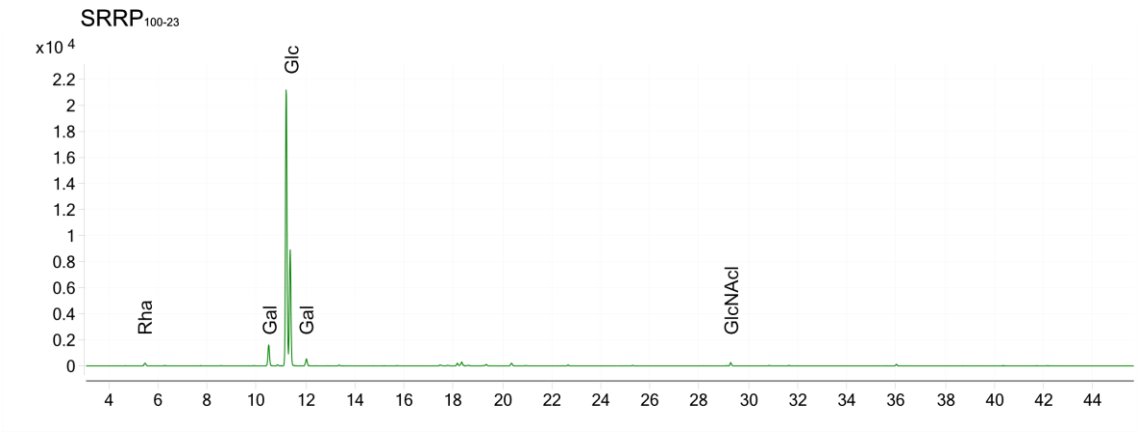

Figure 45 Extracted chromatogram of ions 173.1 and 204.1 (main ions produced upon fragmentation of HexNAc and Hex, respectively) after GC-MS analysis of the monosaccharides found on $SRRP_{100-23}$. The results show the excess of Glc, and the presence of Gal, GlcNAc and Rha.

5.2.2. Characterisation of $SRRP_{53608}$ glycosylation pattern

5.2.2.1. Purification of $SRRP_{53608}$ and MUB_{53608}

In the previous chapter, we identified MUB_{53608} and $SRRP_{53608}$ as putative glycoproteins in *L. reuteri* ATCC 53608. These are high molecular weight proteins, both migrating at >300 kDa on SDS-PAGE. MUB_{53608} generally shows higher expression levels in MRS, whereas $SRRP_{53608}$ production is enhanced in LDM-II growth medium. However, as MRS is a semi-defined medium containing yeast derived glycans, LDM-II was chosen for bacterial growth and purification of both proteins. MUB_{53608}, in addition to other glycoproteins from *L. reuteri* ATCC 53608 was found to interact with RCA, whereas WGA was found to specifically bind $SRRP_{53608}$. To purify the two adhesins, *L. reuteri* was grown for 24 h in LDM-II media and proteins were precipitated in ammonium sulphate. After resuspension, proteins were extracted in a triple phase partitioning system using tert-butanol and an increasing concentration of ammonium sulphate. After three cycles of extraction, the proteins were analysed by SDS-PAGE. MUB_{53608} and $SRRP_{53608}$ were both found in the second fraction (Figure 46A), which was then further separated by gel filtration (see 2.8.2)78. Despite having a predicted molecular weight (MW) of 353 kDa and 160 kDa,

respectively, MUB_{53608} and $SRRP_{53608}$ both co-eluted in the void volume of the column (approximately 2 MDa) (Figure 46B). This may be due to their extended protein fold, which will increase their hydrodynamic volume in solution (311). The two proteins were separated by affinity chromatography, using an agarose-bound WGA column. MUB_{53608} was recovered from the flow-through and wash fractions, whereas $SRRP_{53608}$ was retained on the column and eluted with high concentration of GlcNAc, the cognitive sugar for WGA, therefore confirming the ability of $SRRP_{53608}$ to interact with WGA in a glycan-mediated manner. Slot blot analysis of the wash and elution fractions showed that the two glycoproteins could be successfully separated (Figure 46C), as differentially recognised by *f*-WGA or *f*-RCA. The identity of the two proteins was further confirmed by anti-$SRRP_{56308}$ and anti-MUB_{53608} antibodies, as well as mass spectrometry (MS) (data not shown).

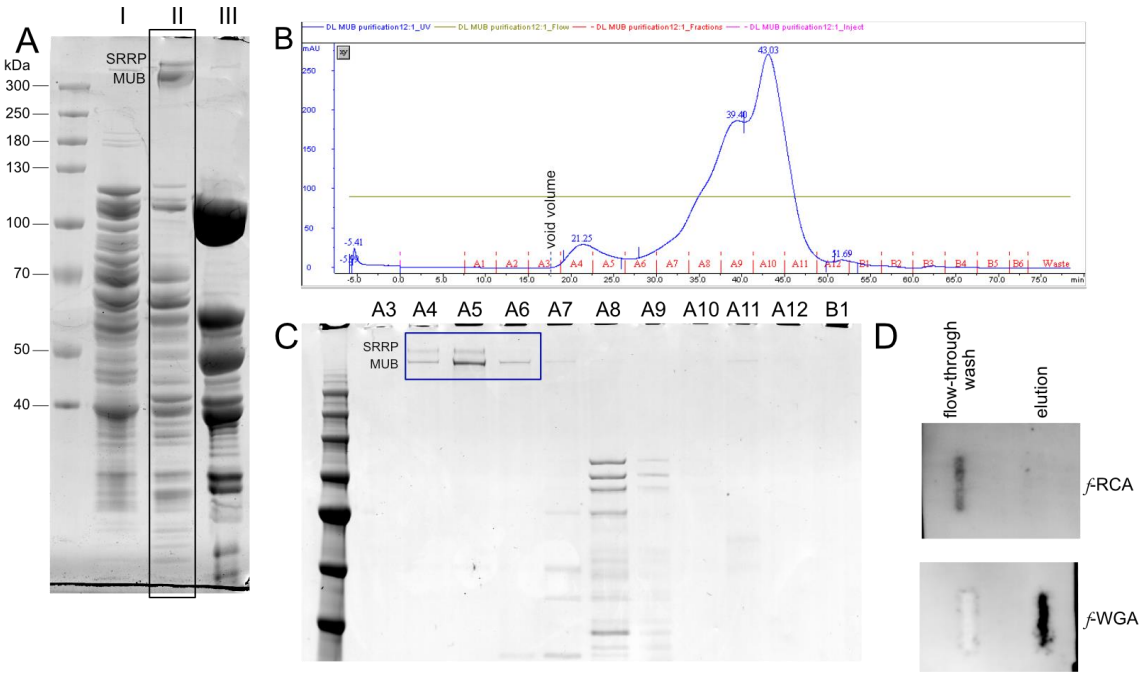

Figure 46 Purification of MUB_{53608} and $SRRP_{53608}$ from *Lr*ATCC 53608 bacterial culture SM. A) SDS-PAGE analysis of protein extracts collected after three rounds of triple phase partitioning (TPP) (I-III). B) Gel filtration chromatogram of the protein separation from fraction II. C) SDS-PAGE of fractions collected after separation of fraction II from the TPP extraction by gel filtration. D) Slot blot analysis of wash and elution fractions after WGA affinity chromatography.

5.2.2.2. *SRRP$_{53608}$ is glycosylated with di-HexNAc moieties*

To identify the glycans present in SRRP$_{53608}$, purified SRRP$_{53608}$ was subjected to reductive β-elimination. The chemically released glycans were then permethylated and analysed by Matrix Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry (MALDI-ToF). The analysis revealed a single peak at 575 Da, which corresponds to the mass of a reduced, permethylated sodiated di-N-acetylhexosamine (HexNAc) (Figure 47A). Further fragmentation of this species confirmed the nature of the glycan, as it produced two main peaks at 282 and 316 Da, corresponding to a non-reducing and a reducing terminal HexNAc (Figure 47B).

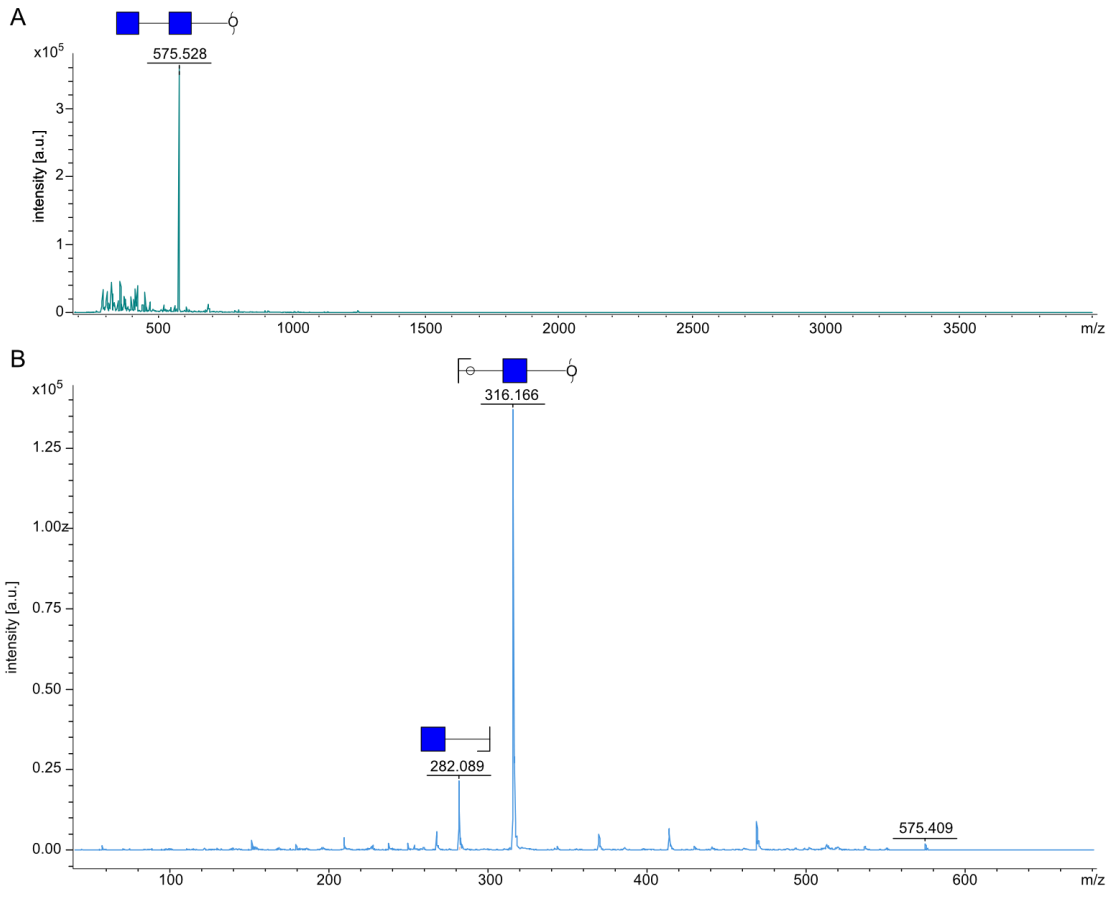

Figure 47 A) MS spectrum of the released and permethylated O-glycans from SRRP$_{53608}$. B) Fragmentation spectrum of the main peak at 575 Da. The data suggest the presence of diHexNAc moieties on SRRP$_{53608}$.

5.2.2.3. *The SRRP$_{53608}$ glycans are comprised of GlcNAc molecules*

The bioinformatics analysis, as well as the monosaccharide composition analysis of the *Lr*ATCC53608 SM proteins, and the interaction of WGA with SRRP$_{53608}$, all suggest that the di-HexNAc moiety identified by MS may be comprised of GlcNAc residues. To determine the exact

nature of the glycan residues, the carbohydrate content of purified $SRRPP_{53608}$ was further analysed by GC-MS. After methanolysis of $SRRP_{53608}$, the released sugars were re-N-acetylated and derivatised with trimethylsilane (TMS). The GC-MS chromatogram showed a single HexNAc peak with a retention time (~29 min) corresponding to that of GlcNAc (Figure 48). Glucose was also detected (data not shown), but as no hexose was identified in the MALDI-ToF analysis, this was attributed to contamination. The GC-MS results supported the previous bioinformatics and monosaccharide composition analyses that identified no HexNAc other than GlcNAc (see section 4.2.2). In addition, this result suggests that GtfC from *L. reuteri* ATCC 53608 is distinct from GtfCs characterised from other organisms, as in all cases studied so far, GtfC was shown to use UDP-Glc as the sugar donor for the second glycosylation step of SRRPs (see section 1.4.1.2.4).

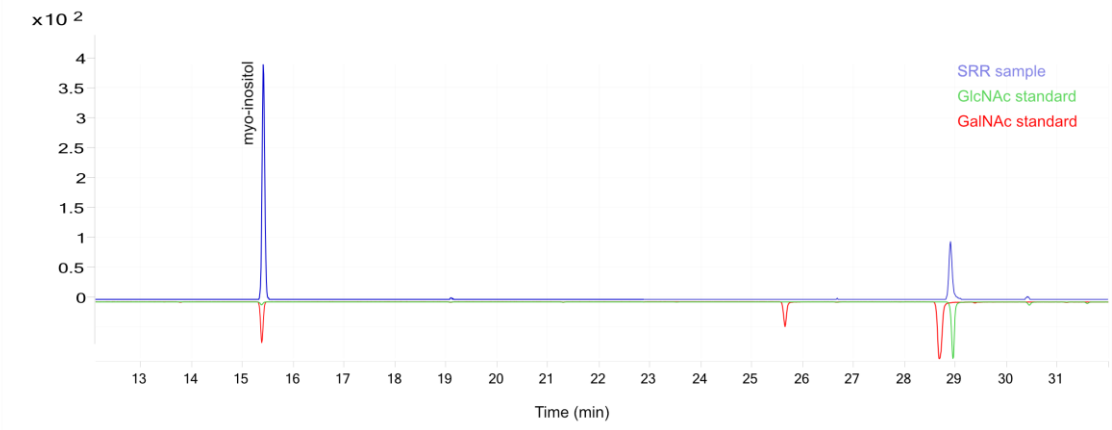

Figure 48 GC-MS analysis of $SRRP_{53608}$ monosaccharides. MS trace of monosaccharide-characteristic ions at 173 and 204 Da.

5.2.2.4. The GlcNAc moieties of $SRRP_{53608}$ are α-linked

In homologous glycosylation clusters, GtfC was shown to belong to the GT family 4, which contains GTs with a retaining mechanism (www.cazy.org) (306), therefore generating α-glycosidic bonds (Figure 14).

To determine the conformation of the GlcNAc residue in $SRRP_{53608}$, purified $SRRP_{53608}$ was treated with enzymes specific for α- or β-linkages, *i.e.* a commercially available, β-N-acetyl-hexosaminidase$_f$, or a GH89 α-N-acetyl-glucosaminidase (α-GlcNAcase) from *Bacteroides*

thetaiotaomicron, produced in-house. Briefly, an overexpression vector containing the GH89 gene fused to a His_6 tag sequence, was transformed into *E. coli* BL21 (DE3). After induction of the gene expression in Auto Induction Media, soluble recombinant α-GlcNAcase was obtained and purified by IMAC (Figure 49).

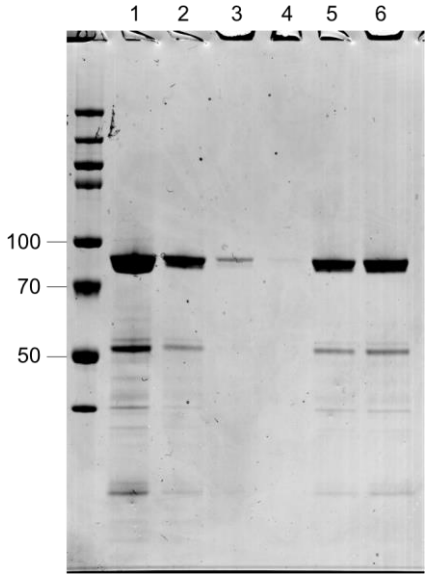

Figure 49 Purification of GH89 α-GlcNAcase by IMAC. 1) starting material, 2) flow-through, 3) wash fraction 1, 4) wash fraction 2, 5) elution fraction 1, 6) elution fraction 2.

The specific activity of the IMAC-purified enzyme was assessed using oNP-α-GlcNAC and found to be ~350 U/ml of purified protein (one unit of enzyme hydrolyses 1.0 nmole of o-nitrophenyl-α-D-GlcNAc to o-nitrophenol and D-GlcNAc in 1 h, at pH 6.5 and 37°C).

The reaction products of the enzymatically treated $SRRP_{53608}$ were analysed by western blot, probed with *f*-WGA or anti-$SRRP_{53608}$ antibodies. When $SRRP_{53608}$ was treated with α-GlcNAcase at 25°C, 37°C or 42°C, deglycosylation products were observed (Figure 50A), suggesting that the GlcNAc moieties found on $SRRP_{53608}$ are α-linked. The deglycosylation products were recognised by the anti-$SRRP_{53608}$ antibodies, but not by *f-WGA*, suggesting that they are fully deglycosylated. Treatment with an excess of β-N-acetyl-hexosaminidase$_f$ under the same conditions did not produce any deglycosylated products (Figure 50B) whereas treatment with both enzymes yielded products similar to the treatment with α-GlcNAcase only (Figure 50C).

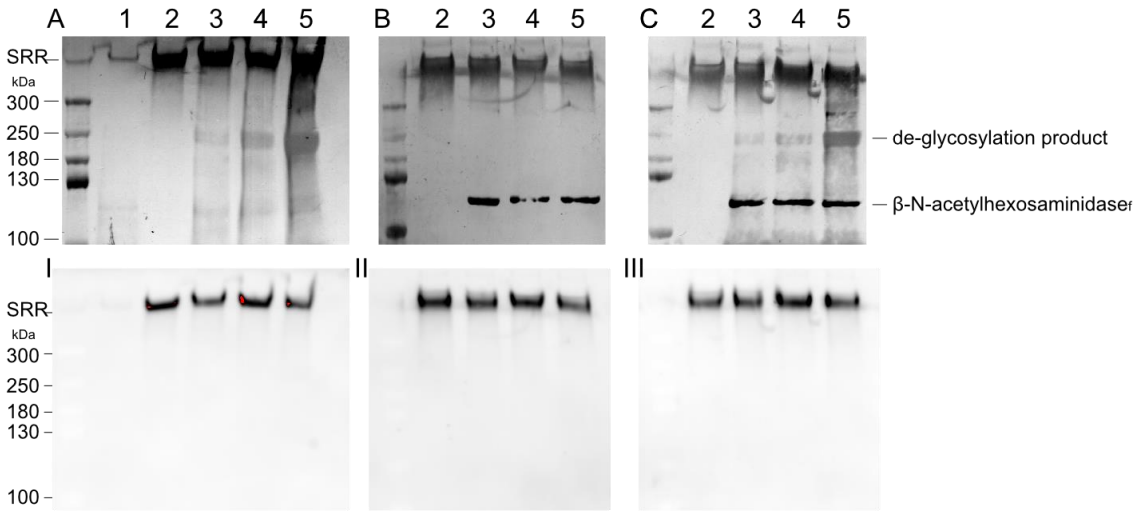

Figure 50 Enzymatic deglycosylation of $SRRP_{53608}$ using α-GlcNAcase, β-N-acetyl-hexosaminidase$_f$ or both. Top set: Western blot analysis of enzymatic deglycosylation products using anti-$SRRP_{53608}$ antibody. Bottom set: Western blot analysis of enzymatic deglycosylation products using f-WGA. A,I) Treatment with α-GlcNAcase, B,II) Treatment with β-N-acetyl-hexosaminidase$_f$, C,III) Treatment with both enzymes. 1) α-GlcNAcase control, 2) $SRRP_{53608}$ control, 3) reaction performed at 25°C,4) reaction performed at 37°C, 5) reaction performed at 42°C. The results show deglycosylation products when $SRRP_{53608}$ is treated with α-GlcNAcase. Fully glycosylated $SRRP_{53608}$ is still present after enzymatic deglycosylation.

However, the α-GlcNAcase treated protein migrated at ~220 kDa, therefore higher than the fully deglycosylated $SRRP_{53608}$ which has a predicted MW 116 kDa, based on amino acid composition. This could be due to the presence of sugars that are not recognised by WGA, or cleaved by the N-acetyl-glucosaminidases used in this assay. For instance, it was recently shown in *S. gordonii* M99 that Asp2 could modify GlcNAc residues on SRRP with an additional O-acetyl group (O-AcGlcNAC) and that the modified sugar could not be identified by WGA (204). The use of similar monosaccharides has been reported in *S. agalactiae* H36b and *S. salivarius* JIM8777 (189,269). As the catalytic residues are conserved in Asp2 from *L. reuteri* ATCC 53608 (Figure 51), it is possible that $SRRP_{53608}$ also carries O-AcGlcNAc residues, which could account for the aberrant mobility of the deglycosylated protein on SDS-PAGE. However, this modification is lost under the conditions used for glycomics or sugar composition analysis, so O-AcGlcNAc it is not detectable by MALDI-ToF or GC-MS.

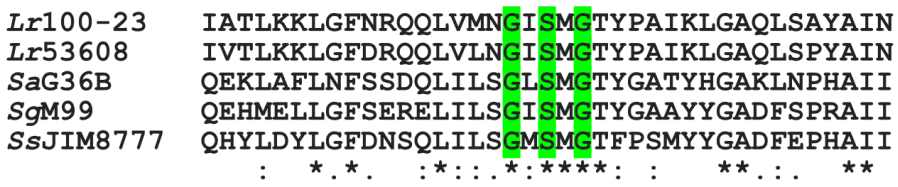

Figure 51 Alignment of the Asp2 sequence from *L. reuteri* and *Streptococcus sp*. flanking the catalytic residues (highlighted in green).

In addition, despite the detection of deglycosylation products, *f*-WGA was still found to bind to protein bands corresponding to native $SRRP_{53608}$ (Figure 50 I, III). The lack of complete deglycosylation with α-GlcNAcase may be due to the substrate specificity of the enzyme, which may not be optimum for this epitope, or to high levels of glycosylation, providing little accessibility to the enzyme.

GlcNAc is one of the constituents of the bacteria cell wall, and WGA is often used to label Gram-positive bacteria. In order to exclude the possibility that WGA recognises cell wall debris that could be attached to the C-terminus of $SRRP_{53608,}$ the adhesin was treated with lysostaphin, a peptidase that targets the Gly_5 anchor motif (see Figure 42) between the cell wall and the surface proteins (174). Treatment with lysostaphin alone did not alter the mobility of $SRRP_{53608}$ in SDS-PAGE (Figure 52A). In addition, lysostaphin treatment did not change the recognition of $SRRP_{53608}$ by *f*-WGA (Figure 52B), suggesting that the interaction is mediated by the $SRRP_{53608}$ glycans and not by cell wall debris. This is also supported by the fact that WGA binding is specific to $SRRP_{53608}$ and does not recognise other predicted cell surface anchored proteins from *Lr*ATCC 53608 SM. Taken together, these results suggest that $SRRP_{53608}$ is glycosylated with GlcNAcα-GlcNAc moieties.

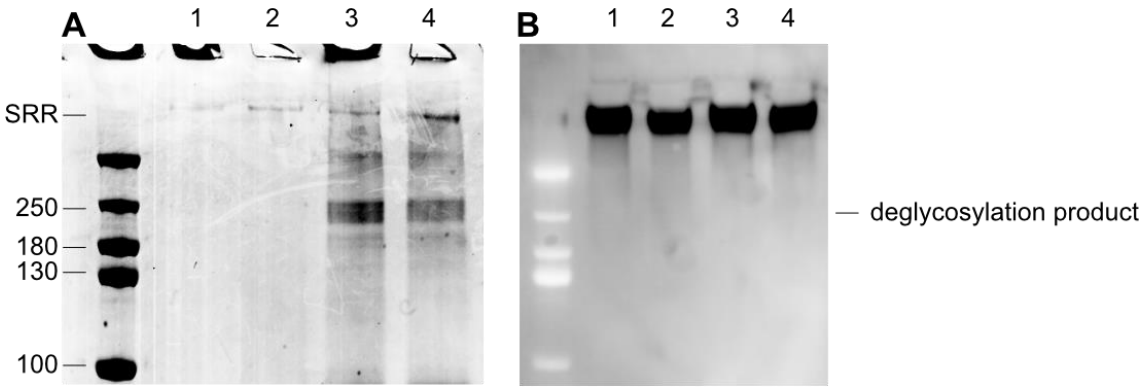

Figure 52 A) SDS-PAGE analysis of $SRRP_{53608}$ (1) following treatment with lysostaphin (2), α-GlcNAcase (3) or both (4). B) Western blot analysis of the same reactions, using f-WGA. The results show deglycosylation products when $SRRP_{53608}$ is treated with α-GlcNAcase, but no effect with lysostaphin.

5.2.2.5. $SRRP_{53608}$ is resistant to proteolysis by GluC

Having identified that $SRRP_{53608}$ is glycosylated, we were interested in identifying the glycosylation sites, using liquid chromatography-mass spectrometry (LC-MS) after proteolytic digestion. In most proteomics studies, trypsin is used as the protease of choice, as it generally generates peptides long enough to provide information for sequencing and characterisation of the proteins and their modifications. However, *in silico* digest of $SRRP_{53608}$ with trypsin showed that the reaction generates large peptides, up to 71 kDa, as there are no lysine residues within the SRRP regions. In contrast, each repeat terminates with a glutamic acid, which would make it a preferential substrate for the endoproteinase GluC (312). However, upon treatment of a $SRRP_{53608}$ with GluC in solution, the analysis of the reaction products by western blotting, using *f*-WGA and anti-$SRRP_{53608}$ antibodies showed that the adhesin remained intact, showing increased resistance to proteolysis. In contrast, treatment of a truncated MUB_{53608} from *L. reuteri* 1063N ($tMUB_{1063N}$) under the same conditions led to completely digestion (Figure 53). This prevented further characterisation of $SRRP_{53608}$ glycosylation sites. Glycans have been widely implicated in protection of proteins from proteolysis, and this seems to be the case for $SRRP_{53608}$.

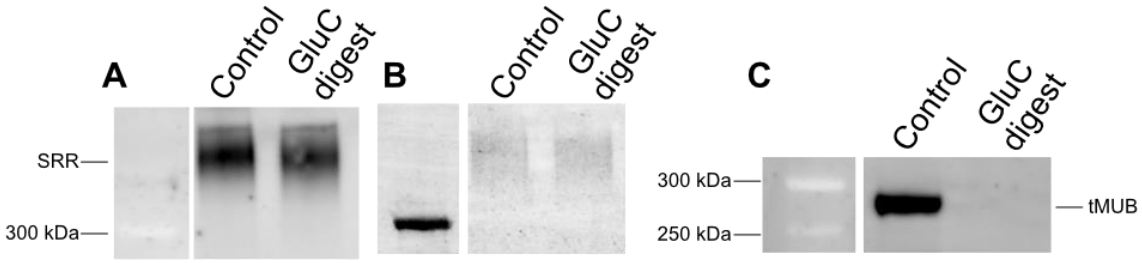

Figure 53 Western blot analysis of purified SRRP digested with GluC, using A) f-WGA, or B) anti-SRRP$_{53608}$ antibodies. C) Western blot analysis of purified tMUB1063N digested with GluC, using f-GSL-1 B4.

5.2.3. Characterisation of MUB$_{53608}$ glycosylation pattern

5.2.3.1. The interaction of MUB$_{53608}$ with RCA is glycan mediated

MUB$_{53608}$ was previously shown to interact with RCA, a galactose specific lectin. To confirm that the interaction between RCA and MUB$_{53608}$ is glycan-mediated, SM proteins from *Lr*ATCC 53608 culture were loaded onto an RCA affinity column. After washing of the unbound proteins, lactose was used to elute proteins that interacted with RCA. All unbound proteins eluted in the first two wash fractions. MUB$_{53608}$ was identified by immunoblot and MS analysis in all fractions (wash and elution) collected (data not shown), but was found to be the predominant protein in the elution fractions, as determined by SDS-PAGE (Figure 54). The fact that MUB$_{53608}$ was retained by RCA and eluted by lactose suggests that the interaction between the two proteins is glycan-mediated. MUB$_{53608}$ found in the flow-through and wash fraction was also recognised by RCA (data not shown). The presence of MUB$_{53608}$ in the wash fractions therefore may be due to a saturation of the RCA column, or to insufficient time allowed for the interaction between the two proteins. These parameters would need to be optimised to increase the separation efficiency of the affinity chromatography.

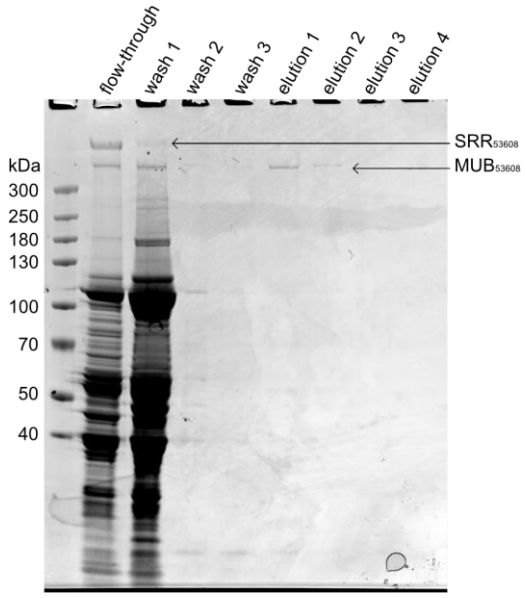

Figure 54 SDS-PAGE analysis of fractions collected after RCA affinity chromatography of *Lr*ATCC 53608 SM proteins. Most proteins are eluted in the flow-through and wash fractions, while MUB_{53608} is retained and eluted after addition of lactose.

Force spectroscopy was also used to directly measure the interaction of RCA with MUB_{53608}. The analysis showed interactions between MUB_{53608} and RCA between 150 and 250 pN. When galactose was added as a competitive ligand, the frequency of adhesion events was decreased. This is in agreement with the affinity chromatography results with RCA that showed glycan mediated interactions between MUB_{53608} and RCA.

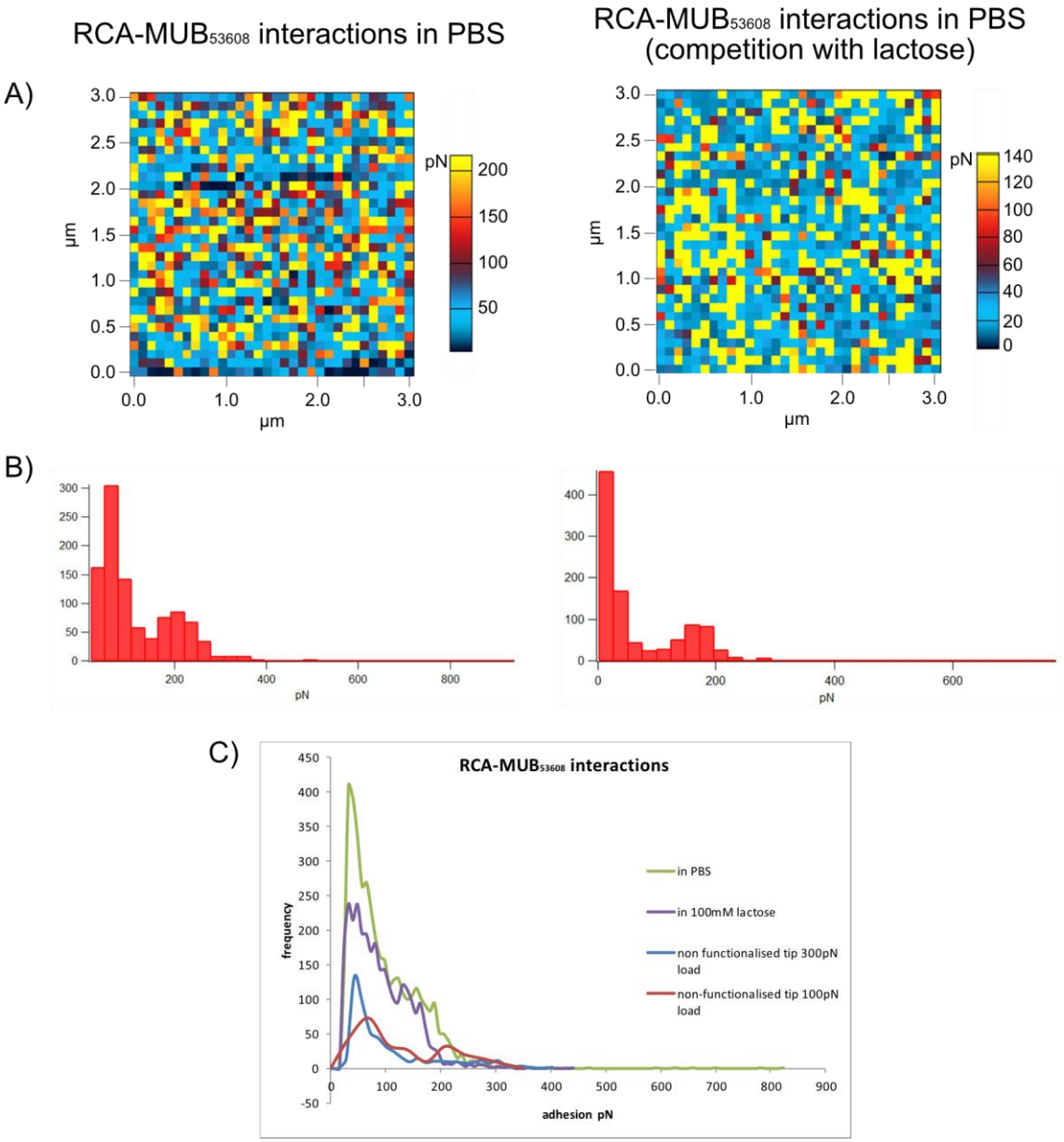

Figure 55 A) Force maps, B) Frequency histograms and C) graphic representation of the interaction between MUB_{53608} and RCA, as measured by force spectroscopy. The results show reduced interaction upon the presence of lactose as a competing sugar.

5.2.3.2. MUB_{53608} carries a terminal α-galactose moiety

The interaction of MUB_{53608} with RCA suggests that the adhesin is glycosylated with Gal moieties. To identify the conformation of this sugar, purified MUB_{53608} was treated with α-, or β-galactosidase, exoglycosidases specific for Gal residues with the specified conformation. The reaction products were then analysed by western blot, using *f*-RCA. RCA binding was abolished when MUB_{53608} was treated with α-, but not β-galactosidase (Figure 56), suggesting the presence of α-galactose (α-Gal).

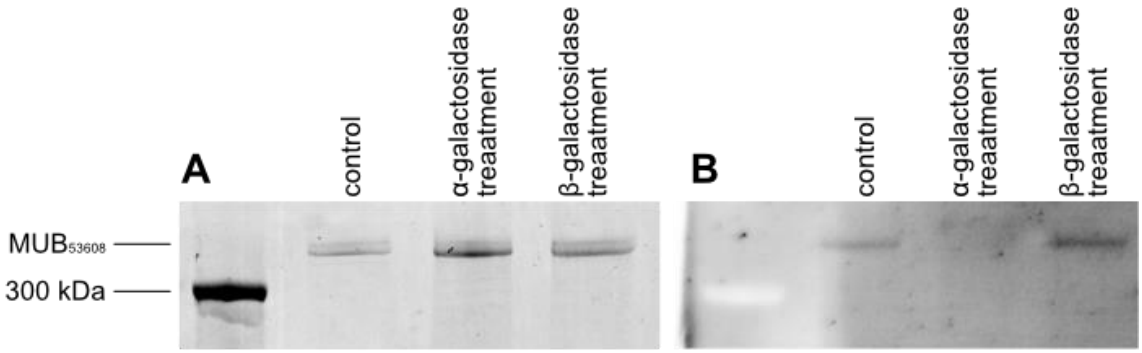

Figure 56 A) SDS-PAGE analysis or B) western blot analysis of purified MUB_{53608} using f-RCA, after treatment with α- or β- galactosidase,

However, as RCA has only been reported to bind β-Gal moieties, equal amounts of MUB_{53608} was also probed with other galactose-specific lectins, such as *Griffonia simplicifolia* Lectin I isolectin B4 (GSL-I B4), *Maackia amurensis* lectin I (MAL I) and *Maclura pomifera* lectin (MPL) (Table 20).

Table 20 Gal-specific, FITC-labelled plant lectins used to probe MUB_{53608}, after western blot analysis.

Lectin	**Abbreviation**	**Specificity**	**Fluorescein/ protein ratio**
***Ricinus communis* agglutinn**	RCA	Gal or GalNAc	6:1
***Griffonia simplicifolia* Lectin I isolectin B4**	GSL-I B4	α-Gal	2.5:1
***Maackia amurensis* lectin I**	MAL I	Galβ-1,4-GlcNAc	8.9:1
***Maclura pomifera* lectin**	MPL	Galβ-1,3-GalNAc	3.9:1

In addition to RCA, only the α-Gal specific lectin GSL-I B4 was found to bind MUB_{53608}, whereas no binding was observed with the β-Gal specific MAL I and MPL (Figure 57). Probing with GSL-I B4 gave a stronger intensity, despite a lower fluorescein/protein ratio compared to RCA, suggesting a stronger interaction.

Together, these data suggest that MUB_{53608} carries terminal α-Gal moieties. α-Galactose is the terminal sugar found on the "α-Gal epitope" (Galα-1,3-Galβ-1,4-GlcNAc), a widely expressed epitope in mammals that has also been found in bacteria and parasites and is highly immunogenic in humans (313). However, MUB_{53608} was not recognised by an anti-α-Gal antibody, raised against the mammalian "α-Gal epitope" (data not shown), suggesting that the epitope present in *L. reuteri* is different. This is further supported by the fact that α-galactosidase

treatment abolishes the binding of RCA to MUB_{53608}. If the Galα-1,3-Galβ-1,4-GlcNAc epitope was present, treatment of α-galactosidase would have exposed a β-linked Gal, leading to greater interaction between MUB_{53608} and RCA.

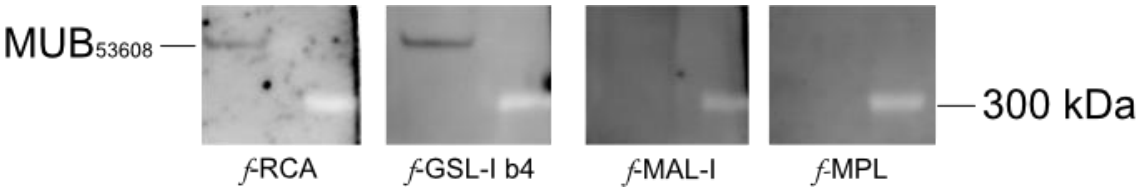

Figure 57 Western blot analysis of purified MUB_{53608} using fluorescein labelled lectins specific for galactose.

5.2.3.3. MUB_{53608} interacts with human intelectin-1 (h-Int1)

Human intelectin 1 (h-Int1) is a calcium dependent, X-type lectin that recognises carbohydrates carrying terminal 1,2-diol groups (314). As *Lr*ATCC 53608 harbours genes that are necessary for the synthesis of UDP- Gal_f, a known ligand of h-Int1, and MUB_{53608} was shown to contain Gal residues, hInt-1 was tested for its ability to bind MUB_{53608}, by force spectroscopy or slot blot. In the presence of Ca^{2+} ions, hInt-1 produced multiple adhesion events which were reduced upon addition of EGTA, as shown in the force maps (Figure 58A). In particular, interaction of hInt-1 with MUB_{53608} resulted in adhesion events ranging from 100 to 500 pN (Figure 58B and C, green line). When EGTA was added to the mixture, the frequency of the adhesion events decreased (Figure 58B and C, purple line), but did not go down to baseline levels (Figure 58A, red-blue). These results suggest that binding of h-Int1 to MUB_{53608} is mainly mediated by glycans found on MUB_{53608}, but that there is also protein-protein interaction, although to a lesser degree. Slot blot analysis also showed binding of MUB_{53608} to h-Int1 in the presence of Ca^{2+} or EGTA, with stronger binding in the presence of Ca^{2+} (Figure 58D), supporting the data obtained by AFM.

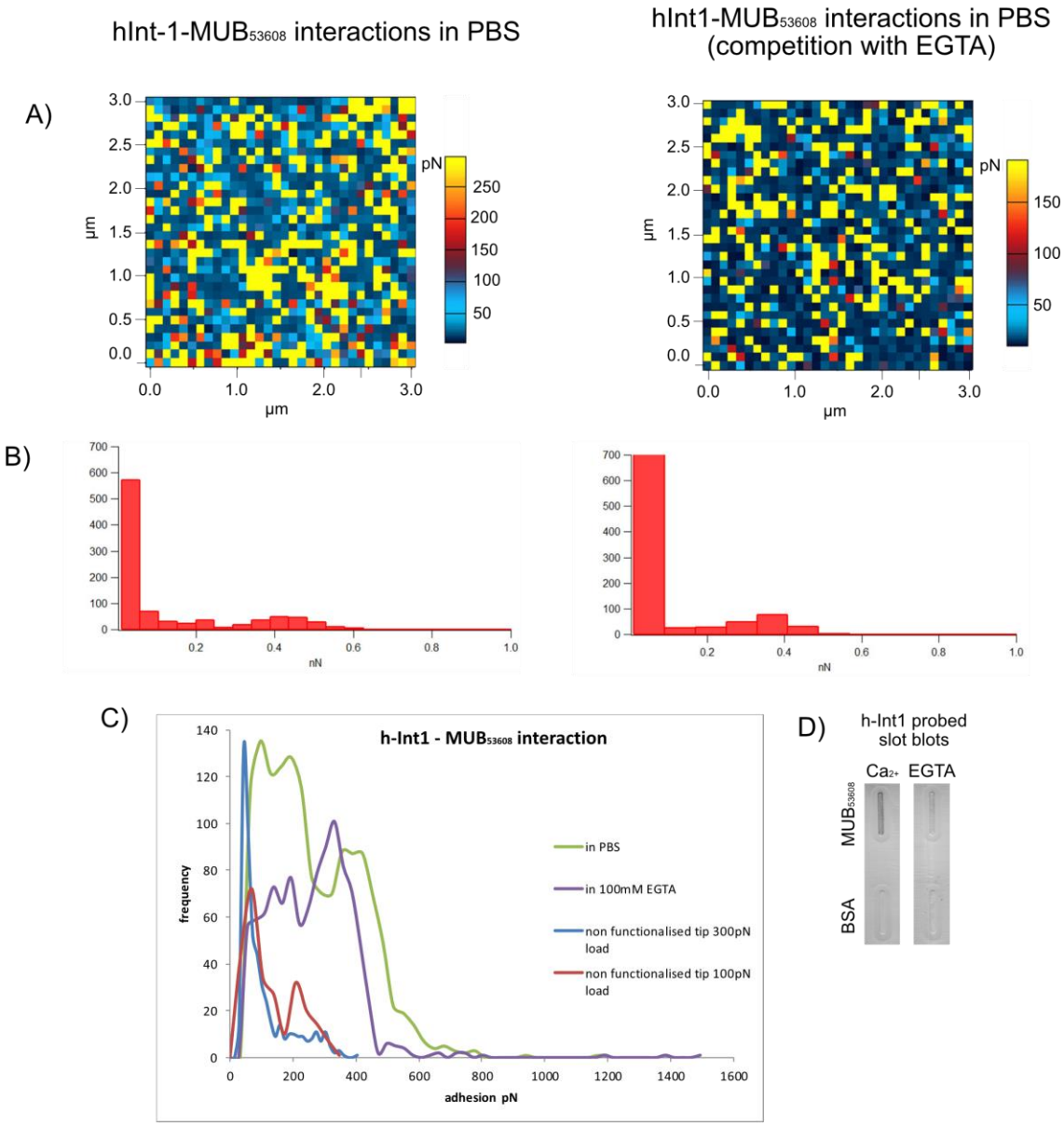

Figure 58 A) Force maps of the interaction of hInt-1 with MUB_{53608}, in PBS with $CaCl_2$ or EGTA. B) Force histograms. C) Force spectroscopy curves, showing the interaction of MUB_{53608} with h-Int1 in the presence of Ca^{2+}. When EGTA is added, the frequency and force of the adhesion events dropped. D) Slot blot analysis of the interaction of h-Int1 with MUB_{53608}. In presence of Ca^{2+} the interaction is strong, while addition of EGTA reduced the amount of lectin that binds onto the protein.

5.2.3.4. Glycomics analysis of MUB_{53608}

To structurally characterise the glycans found on MUB_{53608}, these were chemically released by β-elimination and analysed by MALDI-ToF, after permethylation. The generated MS spectra were not reproducible between experiments, and fragmentation of the major peaks did not yield ions or fragmentation patterns characteristic to glycans making it difficult to identify potential glycan structures. A peak at 534 Da, corresponding to a $HexNAc_1Hex_1$ structure, was occasionally observed (Figure 59), but a fragmentation spectrum could not be obtained to

confirm the annotation. In addition, the peak at 534 Da was often masked by a dominant ion at 531 Da, making it difficult to conclusively identify the ion at 534 Da as a glycan. It is possible that the linkages formed between the glycan and MUB_{53608} are different to those formed during O-glycosylation, making the chemical release and subsequent analyses not a suitable approach to determine the glycosylation of MUB_{53608}. However, if MUB_{53608} is a glycosylation target of the Gtf1/Gtf2 glycosylation system, as suggested in section 4.2.4.4, it is more likely that glycans are attached to the adhesin with an O-linked GlcNAc.

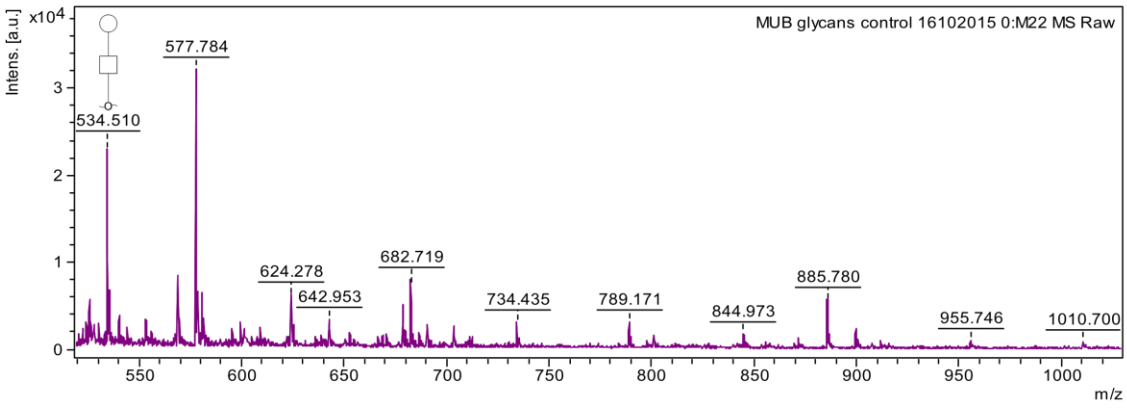

Figure 59 Example of MS spectrum obtained from the analysis of MUB_{53608} glycans after chemical release and permethylation.

To identify the monosaccharides present in MUB_{53608} glycans, these were analysed by GC-MS, as described for the SRRPs. The analysis showed that the MUB_{53608} fraction contains Gal, Glc, GlcNAc and traces of Xyl and Fuc (Figure 60). The presence of Gal is in agreement with the lectin assays that showed recognition of MUB_{53608} by RCA, whereas Xyl, Gal, Glc and GlcNAc had been earlier identified in the analysis of proteins from the spent media of *Lr*ATCC 53608 cultures. Therefore, MUB_{53608} is likely to be glycosylated with glycans containing GlcNAc, Glc and Gal, while Xyl and Fuc may also be present in smaller amounts, or due to insufficient purification of the adhesin.

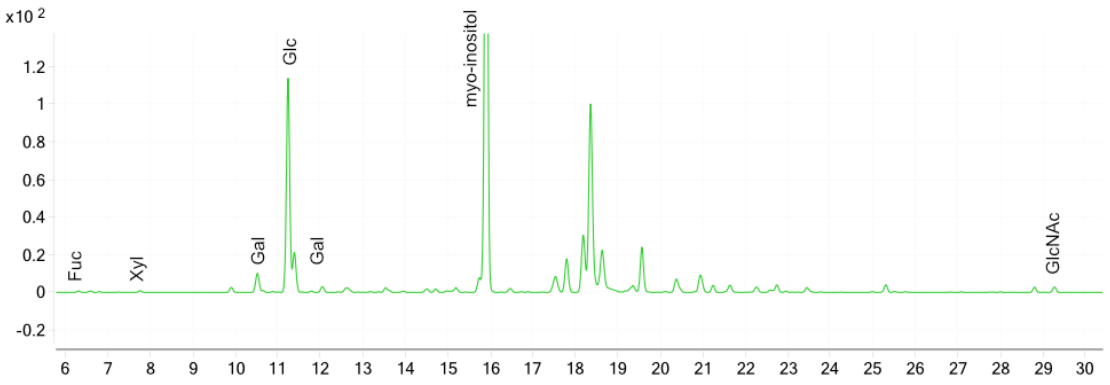

Figure 60 Extracted chromatogram of ions 173.1 and 204.1 (main ions produced upon fragmentation of HexNAc and Hex, respectively) after GC-MS analysis of the monosaccharides found on MUB_{53608}. The results show the presence of Fuc, Xyl, Glc, Gal, and GlcNAc.

5.3. Summary and discussion

Microbial appendages like flagella and pili are used by bacteria to adhere onto host tissue. These structures have been shown to be glycosylated in most cases, however, these are absent in *L. reuteri,* an important member of the gut microbiota. Instead, *L. reuteri* species produces long, fibre-like adhesins, to mediate binding onto the host's surface. *L. reuteri* ATCC 53608 has been shown to produce a long multidomain mucus binding protein, MUB_{53608} (154,156), in addition to a serine rich repeat protein, $SRRP_{53608}$ (159), that belongs to a class of adhesins often found in Gram-positive bacteria (176). $SRRP_{100\text{-}23}$ was also identified in *L. reuteri* 100-23, where it was shown to be involved in biofilm formation and required for successful colonisation in mice (109). In this work, we identified MUB_{53608}, $SRRP_{53608}$ and $SRRP_{100\text{-}23}$ as the main glycoproteins in *Lr*ATCC 53608 and *Lr*100-23 strains. Here, we aimed to characterise the glycans found on these adhesins.

Our glycomics analysis of the purified SRRPs suggested that $SRRP_{53608}$ is glycosylated with di-HexNAc moieties, while $SRRP_{100\text{-}23}$ is glycosylated with linear $Hex_2HexNAc_1$ glycans. Monosaccharide composition analysis of $SRRP_{53608}$ identified GlcNAc as the sole HexNAc. This confirmed the hypothesis that $SRRP_{53608}$ is glycosylated with di-GlcNAc residues and is in agreement with the DSF analysis of $GtfC_{53608}$, the enzyme that mediates the second step in SRRP glycosylation. This is unlike any other studied SRRPs, as in most cases the second monosaccharide was found to be Glc (see section 1.4.1.2.4). In addition, treatment of $SRRP_{53608}$

with an α-N-acetyl-glucosaminidase resulted in the formation of deglycosylated products, suggesting that the GlcNAc residues are α-linked. Although the exact linkage is unknown, the GlcNacα-GlcNac epitope has been previously identified in the O-antigen of *E. coli* O101 (315) and *Shigella dysenteriae* type 4 (316) and a surface polysaccharide of *Acinetobacter baumannii* O16 (317). This epitope has also been reported in structures deposited in the GlyTouCan glycan repository (318), although no information on their origin is available.

Taken together, the results from the monosaccharide composition and western blot analysis of $SRRP_{100\text{-}23}$, and the DSF analysis of $GtfC_{100\text{-}23}$ suggest that the glycan present on $SRRP_{100\text{-}23}$ is Glc-Glc-GlcNAc. Based on the DSF assay with $GtfC_{100\text{-}23}$, that showed greater interaction of the enzyme with UDP-Glc, than with UDP-Gal, as well as monosaccharide composition analysis of SM proteins from *Lr*100-23 showing Gal and Glc as the only hexoses present, and lectin affinity assays that show no interaction between $SRRP_{100\text{-}23}$ and RCA. Such trisaccharide structure is also found in the reducing end of the SRRP glycans from *S. parasanguinis* FW213 (191)). It is worth noting that the nature of this trisaccharide contradicts the recognition of $SRRP_{2100\text{-}23}$ by WGA. It is possible that glycosylation is not always complete, and that WGA binds to GlcNAc residues that have not been further extended by GtfC. Further studies with a combination of glucosidases and N-acetyl-glucosaminidases are therefore needed to confirm the exact nature and structure of these glycans. In addition, the identification of a trisaccharide for $SRRP_{100\text{-}23}$ was surprising, as the GT content of the $SecA_2/Y_2$ cluster in *Lr*100-23 strain suggested a more complex glycan. This may be due to inactive enzymes, but further studies are required to analyse the role of the additional GTs present in the $secA_2/Y_2$ cluster.

Purified MUB_{53608} was found to interact with RCA, a Gal specific lectin, similarly to other secreted proteins from *Lr*ATCC 53608, suggesting it may be a target protein of a general glycosylation system. This interaction was shown to be glycan-mediated by force spectroscopy and RCA affinity chromatography, using competition assays with lactose. To gain more insights into the structure

of the glycan, MUB_{53608} was treated with galactosidases, and western blot analysis of the products using lectins revealed that the Gal residues were α-linked. Terminal α-Gal is part of a highly immunogenic epitope in humans, that is abundant in other mammalian tissues (319). In addition, h-Int1 was found to recognise MUB_{53608}. h-Int1 is a Ca^{2+}-dependent secreted lectin, expressed in intestinal goblet and Paneth cells (320), able to distinguish microbial from host cells, as it recognises carbohydrates not naturally found in mammals (314). Apart from Gal, sugar analysis of MUB_{53608} revealed the presence of Glc, GlcNAc, Xyl and Fuc. However, glycomics analysis of MUB_{53608} failed to characterise the glycans present. This could be due to the presence of a linkage that is not cleaved chemically by β-elimination, the method of choice for chemical deglycosylation of proteins, or the presence of further modified glycans that may interfere with the analysis of the data.

Table 21 summarises the glycans identified in $SRRP_{100\text{-}23}$, $SRRP_{53608}$ and MUB_{53608}.

Table 21 Summary of the identified glycosylated adhesins and their glycans

Strain	**Adhesin**	**Glycan identified**	**Putative pathway**
*Lr*100-23	$SRRP_{100\text{-}23}$	Glc-Glc-GlcNAc	$SecA_2/Y_2$
*Lr*ATCC 53608	$SRRP_{53608}$	GlcNAcα-GlcNAc	$SecA_2/Y_2$
*Lr*ATCC 53608	MUB_{53608}	N.D. Contains, GlcNAc, Gal, Glc and Gal_f	Gtf1/Gtf2

Chapter 6

General discussion, conclusions and perspectives

To date, the role of protein glycosylation in bacteria has been extensively studied in pathogens, underscoring the importance of glycans in colonisation, virulence and survival of bacterial pathogens (250,321,322). However, knowledge on protein glycosylation in commensal bacteria is still at its infancy. Here we used *L. reuteri* as a model organism to gain insights into the protein glycosylation pathways of gut commensal probiotic strains. *L. reuteri* is a widespread gut symbiont found in many vertebrate hosts and one of the first to colonise the human GI tract (323). As such, it plays a pivotal role in the host's health and physiology. Many studies have used *L. reuteri* strains as probiotics in humans, mice and pigs, demonstrating their beneficial effects in the health and well-being of individuals (see section 1.2). Adhesion of the bacteria to the host's tissue is considered a key step in colonisation. This adhesion is often mediated by adhesins, which specifically bind to components of the host's surface. The aim of this project was to identify putative glycoproteins expressed by *L. reuteri* 100-23 (rat isolate), ATCC 53608 (pig isolate) and MM4-1a (human isolate) and to characterise the nature of the glycans present on these proteins.

To that end, a variety of approaches were employed. First, a bioinformatics analysis was carried out to identify putative enzymes involved in the synthesis of sugar nucleotides, the building blocks of glycans, as well as putative glycosylation pathways in *L. reuteri* strains. Following this *in silico* analysis, lectin affinity screening combined with mass spectrometry was used to identify putative glycoproteins secreted by *L. reuteri* strains. This included the analysis of deletion or insertion *L. reuteri* mutants of putative glycosyltransferases (GTs) or genes encoding proteins with a putative role in protein glycosylation. Following the identification of putative glycoproteins, the adhesins were purified and the glycans were determined by analytical techniques such as gas chromatography - mass spectrometry (GC-MS) or matrix assisted laser desorption/ionisation - time of flight mass spectrometry (MALDI-ToF), force spectroscopy and other biochemical assays.

The bioinformatics analysis suggested the ability of *Lr*100-23, *Lr*ATCC 53608 *Lr*MM4-1a to utilise a very limited number of carbohydrates, mainly simple sugars, as carbon source. Differences were observed in between these *L. reuteri* strains in their ability and efficiency to utilise certain monosaccharides, which may be the result of co-evolution between the host and the bacteria, reflecting a difference in the diet of the respective host. It is also worth noting that no enzymes involved in the utilisation of host glycans were identified, suggesting that *L. reuteri* is dependent on dietary carbohydrates, in agreement with studies showing that this species has evolved and diversified from free-living ancestors (104).

The bioinformatics analyses also suggested that using these simple sugars, *L. reuteri* had the potential to produce a number of activated sugars (*i.e.* sugar nucleotides) that could be involved in cell wall and exopolysaccharide (EPS) synthesis, or protein glycosylation. While the biosynthetic pathways of sugar nucleotides were conserved among *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a strains, *Lr*100-23 was also found to harbour genes responsible for the expression of enzymes involved in TDP-rhamnose (Rha), which the other strains lacked. However, further analysis is required to provide biochemical evidence on the range of sugar nucleotides used by these *L. reuteri* strains.

In addition, the analysis of the *Lr*100-23, *Lr*ATCC 53608 and *Lr*MM4-1a genomes revealed the presence of five putative glycosylation pathways. Four of them were conserved across the three *L. reuteri* strains, including an EPS glycosylation cluster and a two-glycosyltransferase (GT) cluster (*gtf1/gtf2*) that could be involved in protein glycosylation, based on its homology with a general glycosylation system found in *L. plantarum* WCFS1 (274). The fifth glycosylation cluster was identified as the $SecA_2/Y_2$ gene cluster which is conserved in *Lr*100-23 and *Lr*ATCC53608, but absent from *Lr*MM4-1a. Such clusters have been shown to be dedicated to the expression of proteins involved in the glycosylation and secretion of serine-rich repeat proteins (SRRPs) in pathogens (180,190). SRRPs are large surface adhesins characterised primarily in Gram-positive

pathogenic bacteria, but which have recently been identified in commensal species (109,176). The $SecA_2/Y_2$ clusters in *Lr*100-23 and *Lr*ATCC 53608 differ by the number of GT-encoding genes, as well as the localisation of the *srr* gene (encodes the SRRP) within the cluster, which may suggest that the two clusters were acquired independently, or that the cluster was acquired by a common ancestor and diversified during evolution events (159). The different number of GTs also suggests a diverse SRRP glycosylation pattern between *Lr*100-23 and *Lr*ATCC 53608 SRRPs.

To test the hypothesis that the products of the *gtf1/gtf2* and of the *$secA_2/Y_2$* clusters were involved in protein glycosylation, a deletion mutant of g*tf1* and insertion mutants of *gtfB, asp2* and *srr* (genes found in the *$secA_2/Y_2$ cluster* encoding a GT, an accessory secretion protein and SRRP, respectively) were generated in *Lr*100-23 and analysed for their ability to perform protein glycosylation. Western blot analysis of the proteins found in the spent media (SM) after growth of *Lr*100-23 WT and the mutants using lectins suggested that these genes are involved in protein glycosylation. In particular, deletion of *gtf1* resulted in loss of RCA recognition (a galactose-mediated interaction) to several SM proteins, supporting the hypothesis that Gtf1 is involved in protein glycosylation. In addition, it suggested that glycans synthesised by this glycosylation system contain a terminal Gal residues. This is in contrast to the reported role of Gtf1 from *L. plantarum* WCFS1, where its target proteins are recognised by WGA (a GlcNAc specific lectin) (274). In addition, protein glycosylation mediated by Gtf1 in *Lr*100-23 appears to affect aggregation of the cells with the WT strain showing higher aggregation capability than the *Lr*100-23 *Δgtf1* deletion mutant. While further experiments are required to elucidate the cause of this phenotype, it may be that the glycans normally found on the secreted glycoproteins are involved in cell-cell interactions and bacterial aggregation, or that the lack of glycosylation leads to impaired secretion or attachment to the cell-surface of components required for aggregation. It has been previously shown that protein glycosylation is important in bacterial aggregation and biofilm formation, but as glycosylation normally precedes secretion of the target protein, it is

difficult to uncouple the events leading to this phenotype. As close homologues of this gene are found in *Lr*ATCC 53608 and *Lr*MM4-1a, it is probable that Gtf1 has a similar role in these strains.

In contrast, western blot analysis of the *Lr*100-23 insertional mutants of *gtfB, asp2* and *srr* showed that WGA recognition of a high molecular weight protein, corresponding to $SRRP_{100\text{-}23}$, was abolished, while RCA recognition remained unaffected. This suggests that the proteins encoded from *gtfB* and *asp2* are dedicated to the glycosylation and secretion of the $SRRP_{100\text{-}23}$, respectively. This is in agreement to $secA_2/Y_2$ clusters from other organisms, where the enzymes encoded by this system are exclusively involved in the processing of their respective SRRPs (190,199,324). Interestingly, an LPxTG-surface anchored protein, found to be recognised by WGA, was not affected by either deletion of *gtf1* or mutation of *gtfB,* also suggesting the presence of a different glycosylation pathway in *Lr*100-23.

Western blot analysis of SM proteins from cultures of *Lr*ATCC 53608 and *Lr*MM4-1a using RCA showed a similar lectin recognition pattern as for *Lr*100-23. In contrast, only proteins from the *Lr*100-23 and *Lr*ATCC 53608 SM were recognised by WGA. These high molecular weight proteins from *Lr*100-23 and *Lr*ATCC 53608 were further identified by MS as the $SRRP_{100\text{-}23}$ and $SRRP_{53608}$ (recognised by WGA) and MUB_{53608} (recognised by RCA). MUB_{53608} and SRRPs are large surface adhesins and protein structures like pili and flagella that have a similar role in adhesion and colonisation are often found to be glycosylated in pathogens (166,240,325,326). Additionally, a protein identified as a muramidase was recognised by RCA across the three *L. reuteri* strains. This protein had the highest similarity to Acm2, a known glycoprotein from *L. plantarum* WCFS1 and a known target of the Gtf1/Gtf2 glycosylation system (275).

$SecA_2/Y_2$ clusters have been primarily identified and studied in oral pathogenic bacteria (see section 1.3.3.3.2), and this is the first time that such a glycosylation system is studied in gut microbes. RT-PCR analysis of the $secA_2/Y_2$ clusters from *Lr*ATCC 53608 and *Lr*100-23 revealed a different transcriptional organisation of the two clusters. The *Lr*ATCC 53608 cluster is organised

into a single operon that spans from *srr* to g*tfB*. Notably, this cluster contains only three putative GTs. In contrast, the *Lr*100-23 cluster is organised into two operons, one spanning from *gtfC*, the first gene of the cluster, to *Lr_70901*, a gene encoding a reverse transcriptase, and containing seven GTs, and a second one that consists of the *srr* gene alone. The transcriptional organisation of the two clusters, although different, suggests a tight and coordinated control of the gene expression in both organisms.

Following identification of MUB_{53608}, $SRRP_{53608}$ and $SRRP_{100\text{-}23}$ as putative glycoproteins, the adhesins were purified by affinity chromatography, taking advantage of their specific binding to lectins, and their glycans were analysed by MALDI-ToF and GC-MS. The glycan present on $SRRP_{53608}$ was identified by MALDI-ToF and GC-MS analysis as a GlcNAc-GlcNAc disaccharide (Figure 61). GlcNAc was found to be in α-configuration, after enzymatic treatment of $SRRP_{100\text{-}23}$ with a α- or β-specific N-acetyl-glucosaminidases. $SRRP_{100\text{-}23}$ was found to be glycosylated by linear $GlcNAc_1Hex_2$ moieties. GC-MS analysis of $SRRP_{100\text{-}23}$ suggested that the hexoses are likely to be Glc or Gal (Figure 61).

The presence of a disaccharide in $SRRP_{53608}$ is in agreement with the number of GTs encoded by the *Lr*ATCC 53608 $SecA_2/Y_2$ cluster. In all $secA_2/Y_2$ clusters from pathogenic bacteria studied to date, GlcNAc has been found at the reducing end, as the product of the combined activity of GtfA and GtfB, the most conserved and essential GTs of the cluster. However, in the pathogenic systems, Glc has been identified as the second residue added to the extending glycan, which contrasts with the observation of a second GlcNAc being present in the $SRRP_{53608}$ glycans. This finding suggests that $GtfC_{53608}$, the enzyme mediating the second glycosylation step uses UDP-GlcNAc as the donor substrate, in contrast to other studied GtfCs that use UDP-Glc. To investigate this hypothesis, and as no suitable acceptor was available, a differential scanning fluorimetry (DSF) assay was employed. These data showed greater stabilisation of $GtfC_{53608}$ in

the presence of UDP-GlcNAc compared to UDP-Glc, supporting the nature of the GlcNAc-GlcNAc glycan structure identified by MALDI-ToF.

Although MS analysis suggested that $SRRP_{100\text{-}23}$ was glycosylated with trisaccharide moieties, a larger glycan was expected based on the number of GTs encoded by the *Lr*100-23 $secA_2/Y_2$ cluster. Using DSF, we showed that $GtfC_{100\text{-}23}$ had stronger interactions with UDP-Glc, compared to UDP-Gal, suggesting that GtfC adds a Glc residue onto the extended glycan. The third glycosylation step of SRRP in *S. parasanguinis* FW213 and *S. pneumoniae* TIGR4 is mediated by a GT containing a DUF1972 (268). The *Lr*100-23 $secA_2/Y_2$ contains three genes encoding such enzymes. In addition, GlyD, the respective GT in *S. pneumoniae* TIGR4 has been shown to use both UDP-Glc and UDP-Gal as donor substrates for the glycosylation of the SRRP (268). This may also be the case for the *Lr*100-23 DUF1972-containing genes. However, further experiments are required to identify the exact substrate specificity of each enzyme, to allow a thorough characterisation of the *Lr*100-23 machinery leading to $SRRP_{100\text{-}23}$ glycosylation.

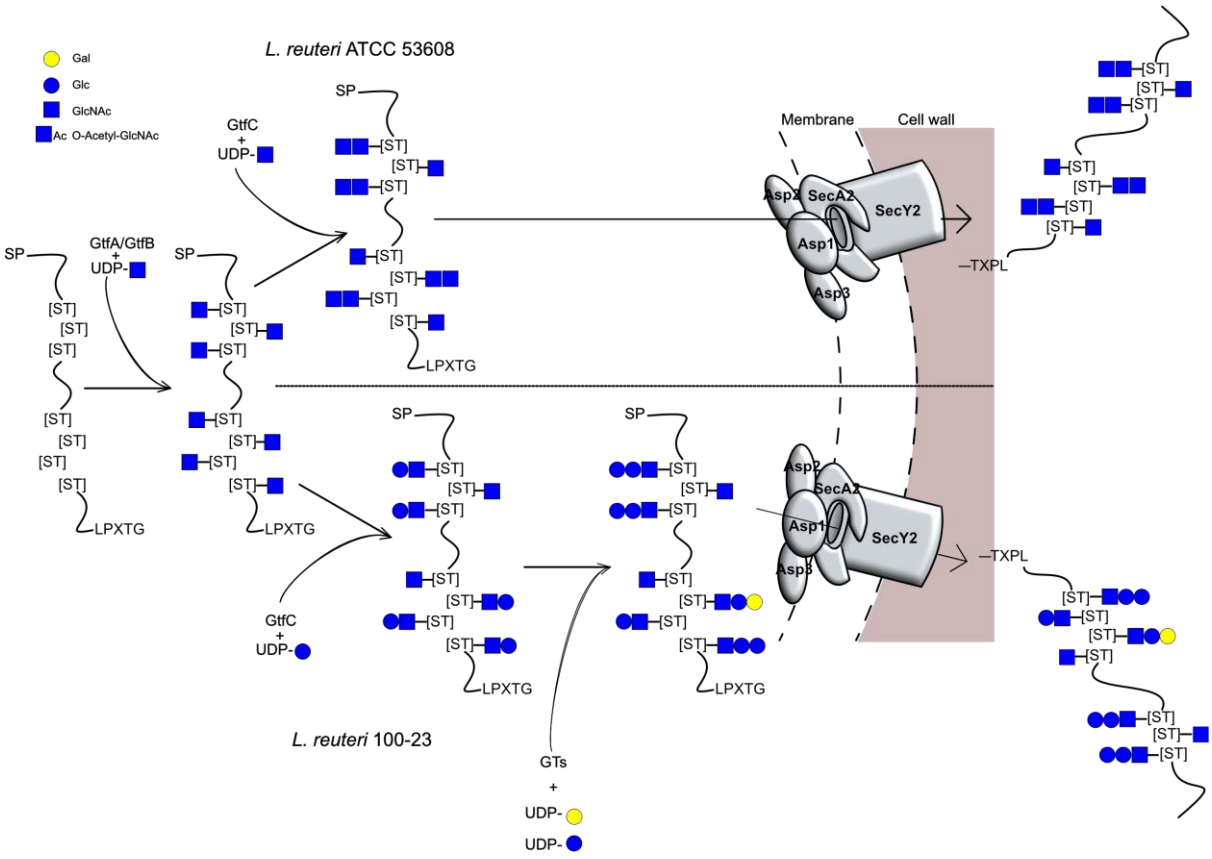

Figure 61 Model of pathways leading to $SRRP_{53608}$ and $SRRP_{100\text{-}23}$ glycosylation.

Here, it is worth noting that DSF was shown to be able to distinguish the different specificities of GtfC from *Lr*100-23 and *Lr*ATCC 53608, without the requirement of an acceptor substrate. This could have a useful application when analysing unknown GTs or other carbohydrate binding enzymes for binding specificity.

While the mechanism of SRRP glycosylation appears to be conserved among bacterial species, their glycan structures are very diverse. For example, SraP, the staphylococcal SRRP, is predicted to be glycosylated with single GlcNAc residues (190), while Fap1 in *S. parasanguinis* FW213 is glycosylated with Glc-GlcNAc-(Rha-Glc)-Glc-GlcNAc moieties (191) (see section 1.4.1.2.4). In contrast, $SRRP_{100\text{-}23}$ and $SRRP_{53608}$ are glycosylated with GlcNAc-Glc-Glc (or -Gal) and GlcNAc-GlcNAc moieties, respectively. However, it is not clear what drives this diversification of the glycans. It could be the localisation of the bacteria, as streptococcal and staphylococcal species reside predominantly in the oral cavity, in contrast to the gut-adapted Lactobacilli, or a

difference in pathogenic or immune-tolerant responses that leads to modifications of the glycans. In addition, future work could focus on elucidating the role of glycosylation in the binding specificity and interaction of $SRRP_{100\text{-}23}$ and $SRRP_{53608}$ with the host.

MUB_{53608} was identified as a glycoprotein that was recognised by RCA. Thus, it was hypothesised that the Gtf1/Gtf2 general glycosylation system was also responsible for the glycosylation of this adhesin in *Lr*ATCC 53608. To further investigate this hypothesis, $Gtf1_{53608}$ and $Gtf2_{53608}$ were cloned and heterologously expressed in *E. coli.* However, as $Gtf1_{53608}$ was found to be insoluble under the conditions tested, the glycosylation system and its putative target proteins could not be further characterised. The interaction between MUB_{53608} and RCA suggests the presence of a Gal residue, which present was also confirmed by GC-MS. However, the glycan structure of the glycans present on MUB_{53608} could not be elucidated by MALDI-ToF mass spectrometry. Treatment with conformation-specific galactosidases showed that this Gal residue is α-linked. α-Linked Gal is part of an immunogenic epitope that is recognised by IgE antibodies and induces allergic reactions in humans (327). In addition, using force spectroscopy and slot blot assays, MUB_{53608} was shown to interact with human intelectin 1 (h-Int1, a human lectin that recognises galactofuranose (Gal_f) residues. *Lr*ATCC 53608 harbours the necessary genes responsible for the synthesis of UDP- Gal_f, which could be used for protein glycosylation. Gal_f, often found in microbial EPS, was found to be highly immunogenic when interacted with monocyte-derived dendritic cells, in a DC-SIGN dependent manner (328). Therefore, the presence of Gal_f residues in MUB_{53608} is in line with recent work showing that this adhesin interacts with DC-SIGN (175). Surface adhesins are important contributors of host-microbe interactions (175,329–331). Recent studies showed that gut commensal and probiotic bacteria also utilise these adhesins to interact with components of the host immune system and regulate the immune responses in a strain-dependent manner (91,116,175,329,332). While the exact nature of these interactions remains unknown, the involvement of host lectins suggests that bacterial carbohydrates play an important role in this process. In line with this, a recent study showed that the glycosylated pili

from *L. rhamnosus* interacted with DC-SIGN *via* the glycan moieties (280). The carbohydrate epitopes identified as part of this work could, in part, account for the immunomodulatory properties of *Lr*ATCC 53608.

Many efforts are geared towards engineering bacteria to synthesise glycoproteins. For this reason, glycosylation systems are extensively studied and characterised for their sugar donor specificity, the number and nature of acceptor proteins, as well as the glycosylation sites of the target proteins. Most knowledge so far comes from studies carried out with Gram-negative, pathogenic bacteria, such as *C. jejuni* (see section 1.4.1.2). Here we showed the presence of Gtf1/Gtf2 and $SecA_2/Y_2$ glycosylation systems in *Lr*100-23 and *lr*ATCC 53608. The work with the *Lr*100-23 mutants showed that the target proteins for each system are distinct, with no redundancy or overlapping observed. As the priming GT pairs (Gtf1/Gtf2 and GtfA/GtfB) interact with their target proteins *via* a DUF1975 in order to initiate the glycosylation (274,333,334), it would be interesting to identify the amino acids that dictate the substrate specificity. Understanding how this priming GTs work, will boost the exploitation of Gram-positive bacteria in the field of glycoengineering. The O-glycosylation mechanism mediated by the Gtf1/Gtf2 and the $SecA_2/Y_2$ cluster offers certain advantages compared to *en bloc* N-, or O- glycosylation, which may increase its applicability in the future. Firstly, it can be functional in both Gram-negative and Gram-positive bacteria. The different types of *en bloc* glycosylation require the inner membrane of the Gram-negative bacteria, where the glycan is assembled on a lipid carrier, and the periplasmic space, where protein glycosylation takes place (208,209). In contrast, this type of O-glycosylation occurs in the cytoplasm, directly onto the target protein (275,335). While identified in Gram-positive bacteria, this glycosylation machinery has been shown to be functional in the Gram-negative *E. coli* (191). In addition, the genetic manipulation of engineered microorganisms may be easier, due to the decreased size of exogenous DNA transferred, as less genes are required for this type of glycosylation (no flippase, or oligosaccharyl-transferase are involved). Finally, by using the Gtf1/Gtf2 glycosylation system, more proteins can be targeted

for glycosylation, compared to the *en bloc* glycosylation systems which only target proteins in the periplasmic space. The study of such glycosylation systems in gut commensal bacteria can provide invaluable tools in the engineering of novel glycoproteins for industrial and pharmaceutical applications. In addition, future work is required to identify the sites of glycosylation in the glycoproteins identified in this work, by glycoproteomics analysis, in order to understand the acceptor specificity of the enzymes involved in the glycosylation process and the requirements of the target glycoproteins.

The growing interest around gut microbiota and probiotics has led to important insights into the mechanisms underpinning the interaction between bacteria and their host. Carbohydrate structures surrounding the bacterial cells play a crucial role in this interaction and are often used by the host to distinguish between symbiont and pathogenic species. However, it is important to also acknowledge the importance of glycans found on adhesins, as these proteins are one of the first bacterial components to come in contact with the host. Studies on the role played by the glycosylation of the adhesins in successful colonisation and in mediating pro- or anti-inflammatory properties are therefore required to further our mechanistic understanding of the health benefits of these bacterial species. In addition, unravelling the nature of the glycans used in protein glycosylation could provide novel ways of designing or selecting probiotic strains. The knowledge generated from these studies can also be used to improve ongoing efforts to generate bacterial systems capable of producing proteins with well-defined glycan structures.

6. References

1. Kleerebezem M, Vaughan EE. Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity. Annu Rev Microbiol. 2009;63:269–290.

2. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012 Sep 13;489(7415):220–230.

3. Jiménez E, Marín ML, Martín R, Odriozola JM, Olivares M, Xaus J, et al. Is meconium from healthy newborns actually sterile? Res Microbiol. 2008 Apr;159(3):187–193.

4. Matamoros S, Gras-Leguen C, Le Vacon F, Potel G, de La Cochetiere M-F. Development of intestinal microbiota in infants and its impact on health. Trends Microbiol. 2013 Apr;21(4):167–173.

5. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015 Jun 10;17(6):852.

6. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature. 2010 Jul 15;466(7304):334–338.

7. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009 Jan 22;457(7228):480–484.

8. Tschöp MH, Hugenholtz P, Karp CL. Getting to the core of the gut microbiome. Nat Biotechnol. 2009 Apr;27(4):344–346.

9. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. Gut. 2014 Apr;63(4):559–566.

10. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics. 2006 Aug;118(2):511–521.

11. Zivkovic AM, German JB, Lebrilla CB, Mills DA. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1:4653–4658.

12. Fallani M, Young D, Scott J, Norin E, Amarri S, Adam R, et al. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. J Pediatr Gastroenterol Nutr. 2010 Jul;51(1):77–84.

13. Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). Anaerobe. 2011 Dec;17(6):478–482.

14. Yu Z-T, Chen C, Newburg DS. Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes. Glycobiology. 2013 Nov;23(11):1281–1292.

15. Favier CF, Vaughan EE, De Vos WM, Akkermans ADL. Molecular monitoring of succession of bacterial communities in human neonates. Appl Environ Microbiol. 2002 Jan;68(1):219–226.

16. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr. 2012 Sep;96(3):544–551.

17. Hunt KM, Foster JA, Forney LJ, Schütte UME, Beck DL, Abdo Z, et al. Characterization of the diversity and temporal stability of bacterial communities in human milk. PLoS ONE. 2011 Jun 17;6(6):e21313.

18. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1:4578–4585.

19. Thursby E, Juge N. Introduction to the human gut microbiota. Biochemical Journal [Internet]. 2017 Jun 1; Available from: http://www.biochemj.org/content/474/11/1823

20. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. Nat Rev Microbiol. 2016 Jan;14(1):20–32.

21. Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CCGM, Troost FJ, et al. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J. 2012 Jul;6(7):1415–1426.

22. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A. 2010 Aug 17;107(33):14691–14696.

23. Hildebrandt MA, Hoffmann C, Sherrill-Mix SA, Keilbaugh SA, Hamady M, Chen Y-Y, et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. Gastroenterology. 2009 Nov;137(5):1716–24.e1.

24. Binder HJ, Filburn B, Floch M. Bile acid inhibition of intestinal anaerobic organisms. Am J Clin Nutr. 1975 Feb;28(2):119–125.

25. Islam KBMS, Fukiya S, Hagio M, Fujii N, Ishizuka S, Ooka T, et al. Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. Gastroenterology. 2011 Nov;141(5):1773–1781.

26. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011 Oct 7;334(6052):105–108.

27. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature. 2012 Jul 5;487(7405):104–108.

28. Ketabi A, Dieleman LA, Gänzle MG. Influence of isomalto-oligosaccharides on intestinal microbiota in rats. J Appl Microbiol. 2011 May;110(5):1297–1306.

29. Rodríguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, et al. The composition of the gut microbiota throughout life, with an emphasis on early life. Microb Ecol Health Dis. 2015 Feb 2;26(0):26050.

30. Hu Y, Ketabi A, Buchko A, Gänzle MG. Metabolism of isomalto-oligosaccharides by Lactobacillus reuteri and bifidobacteria. Lett Appl Microbiol. 2013 Aug;57(2):108–114.

31. Ventura M, O'Flaherty S, Claesson MJ, Turroni F, Klaenhammer TR, van Sinderen D, et al. Genome-scale analyses of health-promoting bacteria: probiogenomics. Nat Rev Microbiol. 2009 Jan;7(1):61–71.

32. Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niazi F, et al. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci U S A. 2010 Apr 20;107(16):7503–7508.

33. Erwin G. Zoetendal, Antoon D. L. Ak. The host genotype affects the bacterial community in the human gastronintestinal tract. Microb Ecol Health Dis. 2001 Jan;13(3):129–134.

34. Martin F-PJ, Dumas M-E, Wang Y, Legido-Quigley C, Yap IKS, Tang H, et al. A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. Mol Syst Biol. 2007 May 22;3:112.

35. Salzman NH, Underwood MA, Bevins CL. Paneth cells, defensins, and the commensal microbiota: a hypothesis on intimate interplay at the intestinal mucosa. Semin Immunol. 2007 Apr;19(2):70–83.

36. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A. 2005 Aug 2;102(31):11070–11075.

37. Zhang C, Zhang M, Wang S, Han R, Cao Y, Hua W, et al. Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. ISME J. 2010 Feb;4(2):232–241.

38. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011 Apr;9(4):279–290.

39. Ouwerkerk JP, de Vos WM, Belzer C. Glycobiome: bacteria and mucus at the epithelial interface. Best Pract Res Clin Gastroenterol. 2013 Feb 1;27(1):25–38.

40. Juge N. Microbial adhesins to gastrointestinal mucus. Trends Microbiol. 2012 Jan;20(1):30–39.

41. Tailford LE, Owen CD, Walshaw J, Crost EH, Hardy-Goddard J, Le Gall G, et al. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. Nat Commun. 2015 Jul 8;6:7624.

42. Sommer F, Adam N, Johansson MEV, Xia L, Hansson GC, Bäckhed F. Altered mucus glycosylation in core 1 O-glycan-deficient mice affects microbiota composition and intestinal architecture. PLoS ONE. 2014 Jan 9;9(1):e85254.

43. Kashyap PC, Marcobal A, Ursell LK, Smits SA, Sonnenburg ED, Costello EK, et al. Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. Proc Natl Acad Sci U S A. 2013 Oct 15;110(42):17059–17064.

44. Mäkivuokko H, Lahtinen SJ, Wacklin P, Tuovinen E, Tenkanen H, Nikkilä J, et al. Association between the ABO blood group and the human intestinal microbiota composition. BMC Microbiol. 2012 Jun 6;12:94.

45. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, et al. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. Nature. 2013 Oct 3;502(7469):96–99.

46. Crost EH, Tailford LE, Le Gall G, Fons M, Henrissat B, Juge N. Utilisation of mucin glycans by the human gut symbiont Ruminococcus gnavus is strain-dependent. PLoS ONE. 2013 Oct 25;8(10):e76341.

47. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol. 2017 Jan;19(1):29–41.

48. McGuckin MA, Lindén SK, Sutton P, Florin TH. Mucin dynamics and enteric pathogens. Nat Rev Micro [Internet]. 2011;9:265–278. Available from: http://dx.doi.org/10.1038/nrmicro2538

49. Hooper LV, Macpherson AJ. Immune adaptations that maintain homeostasis with the intestinal microbiota. Nat Rev Immunol. 2010 Mar;10(3):159–169.

50. Cash HL, Whitham CV, Behrendt CL, Hooper LV. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. Science. 2006 Aug 25;313(5790):1126–1130.

51. Macpherson AJ, McCoy KD, Johansen FE, Brandtzaeg P. The immune geography of IgA induction and function. Mucosal Immunol. 2008 Jan;1(1):11–22.

52. Begley M, Gahan CGM, Hill C. The interaction between bacteria and bile. FEMS Microbiol Rev. 2005 Sep;29(4):625–651.

53. De Boever P, Wouters R, Verschaeve L, Berckmans P, Schoeters G, Verstraete W. Protective effect of the bile salt hydrolase-active Lactobacillus reuteri against bile salt cytotoxicity. Appl Microbiol Biotechnol. 2000 Jun 13;53(6):709–714.

54. Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell. 2013 Jan 17;152(1-2):39–50.

55. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, et al. Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest. 2010 Dec;120(12):4332–4341.

56. Lewis BB, Buffie CG, Carter RA, Leiner I, Toussaint NC, Miller LC, et al. Loss of Microbiota-Mediated Colonization Resistance to Clostridium difficile Infection With Oral Vancomycin Compared With Metronidazole. J Infect Dis. 2015 Nov 15;212(10):1656–1665.

57. Langdon A, Crook N, Dantas G. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. Genome Med. 2016 Apr 13;8(1):39.

58. Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut microbiota in health and disease. Physiol Rev. 2010 Jul;90(3):859–904.

59. Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze X, et al. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J. 2011 Feb 1;5(2):220–230.

60. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Micro [Internet]. 2008;6:121–131. Available from: http://dx.doi.org/10.1038/nrmicro1817

61. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014 Jan 23;505(7484):559–563.

62. Bergman EN. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Physiol Rev. 1990 Apr;70(2):567–590.

63. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell. 2016 Nov 17;167(5):1339–1353.e21.

64. Earle KA, Billings G, Sigal M, Lichtman JS, Hansson GC, Elias JE, et al. Quantitative imaging of gut microbiota spatial organization. Cell Host Microbe. 2015 Oct 14;18(4):478–488.

65. Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, et al. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A. 2004 Nov 2;101(44):15718–15723.

66. Ellekilde M, Selfjord E, Larsen CS, Jakesevic M, Rune I, Tranberg B, et al. Transfer of gut microbiota from lean and obese mice to antibiotic-treated mice. Sci Rep. 2014 Aug 1;4(1):5922.

67. Martens JH, Barg H, Warren MJ, Jahn D. Microbial production of vitamin B12. Appl Microbiol Biotechnol. 2002 Mar;58(3):275–285.

68. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr Opin Biotechnol. 2013 Apr;24(2):160–168.

69. Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. Front Genet. 2015 Apr 20;6:148.

70. Karl JP, Fu X, Wang X, Zhao Y, Shen J, Zhang C, et al. Changes in Fecal Vitamin K Content are Associated with the Gut Microbiota. FASEB J [Internet]. 2015 Jan 4; Available from: http://www.fasebj.org/content/29/1_Supplement/262.1.short

71. O'Keefe SJD, Ou J, Aufreiter S, O'Connor D, Sharma S, Sepulveda J, et al. Products of the colonic microbiota mediate the effects of diet on colon cancer risk. J Nutr. 2009 Nov;139(11):2044–2048.

72. Den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud D-J, Bakker BM. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J Lipid Res. 2013 Sep;54(9):2325–2340.

73. Macfarlane S, Macfarlane GT. Regulation of short-chain fatty acid production. Proc Nutr Soc. 2003 Feb;62(1):67–72.

74. Chambers ES, Viardot A, Psichas A, Morrison DJ, Murphy KG, Zac-Varghese SEK, et al. Effects of targeted delivery of propionate to the human colon on appetite

regulation, body weight maintenance and adiposity in overweight adults. Gut. 2015 Nov;64(11):1744–1754.

75. Lin Y, Ma C, Liu C, Wang Z, Yang J, Liu X, et al. NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. Oncotarget. 2016 May 17;7(20):29454–29464.

76. Donohoe DR, Collins LB, Wali A, Bigler R, Sun W, Bultman SJ. The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. Mol Cell. 2012 Nov 30;48(4):612–626.

77. Robinson AM, Gondalia SV, Karpe AV, Eri R, Beale DJ, Morrison PD, et al. Fecal Microbiota and Metabolome in a Mouse Model of Spontaneous Chronic Colitis: Relevance to Human Inflammatory Bowel Disease. Inflamm Bowel Dis [Internet]. 2016 Dec 1; Available from: http://insights.ovid.com/pubmed?pmid=27824648

78. Corrêa-Oliveira R, Fachi JL, Vieira A, Sato FT, Vinolo MAR. Regulation of immune cell function by short-chain fatty acids. Clin Transl Immunology. 2016 Apr 22;5(4):e73.

79. Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, de Los Reyes-Gavilán CG, Salazar N. Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. Front Microbiol. 2016 Feb 17;7:185.

80. Bouskra D, Brézillon C, Bérard M, Werts C, Varona R, Boneca IG, et al. Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. Nature. 2008 Nov 27;456(7221):507–510.

81. Purchiaroni F, Tortora A, Gabrielli M, Bertucci F, Gigante G, Ianiro G, et al. The role of intestinal microbiota and the immune system. European Review for Medical and Pharmacological Sciences. 2013;17:323–333.

82. Rautava S, Kalliomäki M, Isolauri E. New therapeutic strategy for combating the increasing burden of allergic disease: Probiotics-A Nutrition, Allergy, Mucosal Immunology and Intestinal Microbiota (NAMI) Research Group report. J Allergy Clin Immunol. 2005 Jul;116(1):31–37.

83. Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G, Takahashi D, et al. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature. 2013 Dec 19;504(7480):446–450.

84. Di Giacinto C, Marinaro M, Sanchez M, Strober W, Boirivant M. Probiotics ameliorate recurrent Th1-mediated murine colitis by inducing IL-10 and IL-10-dependent TGF-beta-bearing regulatory cells. J Immunol. 2005 Mar 15;174(6):3237–3246.

85. Mazmanian SK, Round JL, Kasper DL. A microbial symbiosis factor prevents intestinal inflammatory disease. Nature. 2008 May 29;453(7195):620–625.

86. Cervantes-Barragan L, Chai JN, Tianero MD, Di Luccia B, Ahern PP, Merriman J, et al. Lactobacillus reuteri induces gut intraepithelial CD4(+)CD8αα(+) T cells. Science. 2017 Aug 25;357(6353):806–810.

87. Das G, Augustine MM, Das J, Bottomly K, Ray P, Ray A. An important regulatory role for CD4+CD8 alpha alpha T cells in the intestinal epithelial layer in the prevention of inflammatory bowel disease. Proc Natl Acad Sci U S A. 2003 Apr 29;100(9):5324–5329.

88. Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, et al. Gut immune maturation depends on colonization with a host-specific microbiota. Cell. 2012 Jun 22;149(7):1578–1593.

89. Swanson PA, Kumar A, Samarin S, Vijay-Kumar M, Kundu K, Murthy N, et al. Enteric commensal bacteria potentiate epithelial restitution via reactive oxygen species-mediated inactivation of focal adhesion kinase phosphatases. Proc Natl Acad Sci U S A. 2011 May 24;108(21):8803–8808.

90. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, et al. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci U S A. 2013 May 28;110(22):9066–9071.

91. Freitas M, Axelsson L-G, Cayuela C, Midtvedt T, Trugnan G. Indigenous microbes and their soluble factors differentially modulate intestinal glycosylation steps in

vivo. Use of a “lectin assay” to survey in vivo glycosylation changes. Histochem Cell Biol. 2005 Nov 3;124(5):423–433.

92. Varyukhina S, Freitas M, Bardin S, Robillard E, Tavan E, Sapin C, et al. Glycan-modifying bacteria-derived soluble factors from Bacteroides thetaiotaomicron and Lactobacillus casei inhibit rotavirus infection in human intestinal cells. Microbes Infect. 2012 Mar;14(3):273–278.

93. Kamada N, Kim Y-G, Sham HP, Vallance BA, Puente JL, Martens EC, et al. Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. Science. 2012 Jun 8;336(6086):1325–1329.

94. Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, et al. Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol. 2007 Oct;5(10):2177–2189.

95. Walsham ADS, MacKenzie DA, Cook V, Wemyss-Holden S, Hews CL, Juge N, et al. Lactobacillus reuteri Inhibition of Enteropathogenic Escherichia coli Adherence to Human Intestinal Epithelium. Front Microbiol. 2016 Mar 1;7:244.

96. Itoh T, Fujimoto Y, Kawai Y, Toba T, Saito T. Inhibition of food-borne pathogenic bacteria by bacteriocins from Lactobacillus gasseri. Lett Appl Microbiol. 1995 Sep;21(3):137–141.

97. Hevia A, Martínez N, Ladero V, Alvarez MA, Margolles A, Sánchez B. An extracellular Serine/Threonine-rich protein from Lactobacillus plantarum NCIMB 8826 is a novel aggregation-promoting factor with affinity to mucin. Appl Environ Microbiol. 2013 Oct;79(19):6059–6066.

98. Wilks J, Beilinson H, Golovkina TV. Dual role of commensal bacteria in viral infections. Immunol Rev. 2013 Sep;255(1):222–229.

99. Sorg JA, Sonenshein AL. Inhibiting the initiation of Clostridium difficile spore germination using analogs of chenodeoxycholic acid, a bile acid. J Bacteriol. 2010 Oct;192(19):4983–4990.

100. Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, et al. Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat Rev Gastroenterol Hepatol. 2014 Aug;11(8):506–514.

101. Sinkiewicz G, Ljunggren L. Occurrence ofLactobacillus reuteri in human breast milk. Microb Ecol Health Dis. 2008 Jan;20(3):122–126.

102. Walter J. Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. Appl Environ Microbiol. 2008 Aug;74(16):4985–4996.

103. Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, et al. The evolution of host specialization in the vertebrate gut symbiont Lactobacillus reuteri. PLoS Genet. 2011 Feb 17;7(2):e1001314.

104. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Pérez-Muñoz ME, et al. Lifestyles in transition: evolution and natural history of the genus Lactobacillus. FEMS Microbiol Rev. 2017 Jun 30;

105. Lin JH-C, Savage DC. Host specificity of the colonization of murine gastric epithelium by lactobacilli. FEMS Microbiol Lett [Internet]. 1984;24:67–71. Available from: http://www.sciencedirect.com/science/article/pii/0378109784903422

106. Wesney E, Tannock GW. Association of rat, pig, and fowl biotypes of lactobacilli with the stomach of gnotobiotic mice. Microb Ecol. 1979 Mar;5(1):35–42.

107. Duar RM, Frese SA, Lin XB, Fernando SC, Burkey TE, Tasseva G, et al. Experimental evaluation of host adaptation of Lactobacillus reuteri to different vertebrate species. Appl Environ Microbiol. 2017 Apr 7;83(12).

108. Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, et al. Diversification of the gut symbiont Lactobacillus reuteri as a result of host-driven evolution. ISME J. 2010 Mar;4(3):377–387.

109. Frese SA, Mackenzie DA, Peterson DA, Schmaltz R, Fangman T, Zhou Y, et al. Molecular characterization of host-specific biofilm formation in a vertebrate gut symbiont. PLoS Genet. 2013 Dec 26;9(12):e1004057.

110. Reid G. The Scientific Basis for Probiotic Strains ofLactobacillus. Appl Environ Microbiol [Internet]. 1999;65:3763–3766. Available from: http://aem.asm.org/content/65/9/3763.short

111. Szajewska H, Gyrczuk E, Horvath A. Lactobacillus reuteri DSM 17938 for the management of infantile colic in breastfed infants: a randomized, double-blind, placebo-controlled trial. J Pediatr. 2013 Feb;162(2):257–262.

112. Savino F, Cordisco L, Tarasco V, Palumeri E, Calabrese R, Oggero R, et al. Lactobacillus reuteri DSM 17938 in infantile colic: a randomized, double-blind, placebo-controlled trial. Pediatrics. 2010 Sep;126(3):e526–33.

113. Di Nardo G, Oliva S, Menichella A, Pistelli R, De Biase RV, Patriarchi F, et al. Lactobacillus reuteri ATCC55730 in cystic fibrosis. J Pediatr Gastroenterol Nutr. 2014 Jan;58(1):81–86.

114. Anukam KC, Osazuwa EO, Osadolor HB, Bruce AW, Reid G. Yogurt containing probiotic Lactobacillus rhamnosus GR-1 and L. reuteri RC-14 helps resolve moderate diarrhea and increases CD4 count in HIV/AIDS patients. J Clin Gastroenterol. 2008 Mar;42(3):239–243.

115. Liu Y, Fatheree NY, Dingle BM, Tran DQ, Rhoads JM. Lactobacillus reuteri DSM 17938 changes the frequency of Foxp3+ regulatory T cells in the intestine and mesenteric lymph node in experimental necrotizing enterocolitis. PLoS ONE. 2013 Feb 20;8(2):e56547.

116. Liu Y, Fatheree NY, Mangalat N, Rhoads JM. Human-derived probiotic Lactobacillus reuteri strains differentially reduce intestinal inflammation. Am J Physiol Gastrointest Liver Physiol. 2010 Nov;299(5):G1087–96.

117. Mackos AR, Eubank TD, Parry NMA, Bailey MT. Probiotic Lactobacillus reuteri attenuates the stressor-enhanced severity of Citrobacter rodentium infection. Infect Immun. 2013 Sep;81(9):3253–3263.

118. Dicksved J, Schreiber O, Willing B, Petersson J, Rang S, Phillipson M, et al. Lactobacillus reuteri maintains a functional mucosal barrier during DSS treatment despite mucus layer dysfunction. PLoS ONE. 2012 Sep 27;7(9):e46399.

119. Schreiber O, Petersson J, Phillipson M, Perry M, Roos S, Holm L. Lactobacillus reuteri prevents colitis by reducing P-selectin-associated leukocyte- and platelet-endothelial cell interactions. Am J Physiol Gastrointest Liver Physiol. 2009 Mar;296(3):G534–42.

120. Voravuthikunchai SP, Bilasoi S, Supamala O. Antagonistic activity against pathogenic bacteria by human vaginal lactobacilli. Anaerobe. 2006 Dec;12(5-6):221–226.

121. Zhang D, Li R, Li J. Lactobacillus reuteri ATCC 55730 and L22 display probiotic potential in vitro and protect against Salmonella-induced pullorum disease in a chick model of infection. Res Vet Sci. 2012 Aug;93(1):366–373.

122. Mobini R, Tremaroli V, Ståhlman M, Karlsson F, Levin M, Ljungberg M, et al. Metabolic effects of Lactobacillus reuteri DSM 17938 in people with type 2 diabetes: A randomized controlled trial. Diabetes Obes Metab. 2017 Apr;19(4):579–589.

123. Francavilla R, Polimeno L, Demichina A, Maurogiovanni G, Principi B, Scaccianoce G, et al. Lactobacillus reuteri strain combination in Helicobacter pylori infection: a randomized, double-blind, placebo-controlled study. J Clin Gastroenterol. 2014 Jun;48(5):407–413.

124. Huang C-H, Lin Y-C, Jan T-R. Lactobacillus reuteri' ' induces intestinal immune tolerance against food allergy in mice. J Funct Foods. 2017 Apr;31:44–51.

125. Sengupta R, Altermann E, Anderson RC, McNabb WC, Moughan PJ, Roy NC. The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. Mediators Inflamm. 2013 Mar 13;2013:237921.

126. Nwodo UU, Green E, Okoh AI. Bacterial exopolysaccharides: functionality and prospects. Int J Mol Sci. 2012 Oct 30;13(11):14002–14015.

127. Ruas-Madiedo P, Gueimonde M, Margolles A, de los Reyes-Gavilán CG, Salminen S. Exopolysaccharides produced by probiotic strains modify the adhesion of probiotics and enteropathogens to human intestinal mucus. J Food Prot. 2006 Aug;69(8):2011–2015.

128. Kos B, Susković J, Vuković S, Simpraga M, Frece J, Matosić S. Adhesion and aggregation ability of probiotic strain Lactobacillus acidophilus M92. J Appl Microbiol. 2003;94(6):981–987.

129. Kolberg J, Hammerschmidt S, Frank R, Jonák J, Sanderová H, Aase A. The surface-associated elongation factor Tu is concealed for antibody binding on viable pneumococci and meningococci. FEMS Immunol Med Microbiol. 2008 Jul;53(2):222–230.

130. Dertli E, Mayer MJ, Narbad A. Impact of the exopolysaccharide layer on biofilms, adhesion and resistance to stress in Lactobacillus johnsonii FI9785. BMC Microbiol. 2015 Feb 4;15(1):8.

131. Thamadilok S, Roche-Håkansson H, Håkansson AP, Ruhl S. Absence of capsule reveals glycan-mediated binding and recognition of salivary mucin MUC7 by Streptococcus pneumoniae. Mol Oral Microbiol. 2016 Apr;31(2):175–188.

132. Schade J, Weidenmaier C. Cell wall glycopolymers of Firmicutes and their role as nonprotein adhesins. FEBS Lett. 2016 Nov;590(21):3758–3771.

133. Schneewind O, Missiakas D. Lipoteichoic acids, phosphate-containing polymers in the envelope of gram-positive bacteria. J Bacteriol. 2014 Mar;196(6):1133–1142.

134. Granato D, Perotti F, Masserey I, Rouvet M, Golliard M, Servin A, et al. Cell surface-associated lipoteichoic acid acts as an adhesion factor for attachment of Lactobacillus johnsonii La1 to human enterocyte-like Caco-2 cells. Appl Environ Microbiol. 1999 Mar;65(3):1071–1077.

135. Abachin E, Poyart C, Pellegrini E, Milohanic E, Fiedler F, Berche P, et al. Formation of D-alanyl-lipoteichoic acid is required for adhesion and virulence of Listeria monocytogenes. Mol Microbiol. 2002 Jan;43(1):1–14.

136. Macías-Rodríguez ME, Zagorec M, Ascencio F, Vázquez-Juárez R, Rojas M. Lactobacillus fermentum BCS87 expresses mucus- and mucin-binding proteins on the cell surface. J Appl Microbiol. 2009 Dec 1;107(6):1866–1874.

137. Jensen H, Roos S, Jonsson H, Rud I, Grimmer S, van Pijkeren J-P, et al. Role of Lactobacillus reuteri cell and mucus-binding protein A (CmbA) in adhesion to intestinal epithelial cells and mucus in vitro. Microbiology (Reading, Engl). 2014 Apr;160(Pt 4):671–681.

138. Kinoshita H, Ohuchi S, Arakawa K, Watanabe M, Kitazawa H, Saito T. Isolation of lactic acid bacteria bound to the porcine intestinal mucosa and an analysis of their moonlighting adhesins. Biosci Microbiota Food Health. 2016 Aug 2;35(4):185–196.

139. Dhanani AS, Bagchi T. The expression of adhesin EF-Tu in response to mucin and its role in Lactobacillus adhesion and competitive inhibition of enteropathogens to mucin. J Appl Microbiol. 2013 Aug;115(2):546–554.

140. Granato D, Bergonzelli GE, Pridmore RD, Marvin L, Rouvet M, Corthésy-Theulaz IE. Cell surface-associated elongation factor Tu mediates the attachment of Lactobacillus johnsonii NCC533 (La1) to human intestinal cells and mucins. Infect Immun. 2004 Apr;72(4):2160–2169.

141. Nishiyama K, Ochiai A, Tsubokawa D, Ishihara K, Yamamoto Y, Mukai T. Identification and characterization of sulfated carbohydrate-binding protein from Lactobacillus reuteri. PLoS ONE. 2013 Dec 31;8(12):e83703.

142. Kinoshita H, Uchida H, Kawai Y, Kawasaki T, Wakahara N, Matsuo H, et al. Cell surface Lactobacillus plantarum LA 318 glyceraldehyde-3-phosphate dehydrogenase (GAPDH) adheres to human colonic mucin. J Appl Microbiol. 2008 Jun;104(6):1667–1674.

143. Patel DK, Shah KR, Pappachan A, Gupta S, Singh DD. Cloning, expression and characterization of a mucin-binding GAPDH from Lactobacillus acidophilus. Int J Biol Macromol. 2016 May 11;91:338–346.

144. Martín R, Sánchez B, Suárez JE, Urdaci MC. Characterization of the adherence properties of human Lactobacilli strains to be used as vaginal probiotics. FEMS Microbiol Lett. 2012 Mar;328(2):166–173.

145. Bergonzelli GE, Granato D, Pridmore RD, Marvin-Guy LF, Donnicola D, Corthésy-Theulaz IE. GroEL of Lactobacillus johnsonii La1 (NCC 533) is cell surface associated: potential role in interactions with the host and the gastric pathogen Helicobacter pylori. Infect Immun. 2006 Jan;74(1):425–434.

146. Watanabe M, Kinoshita H, Huang I-N, Eguchi K, Tsurumi T, Kawai Y, et al. An adhesin-like protein, Lam29, from Lactobacillus mucosae ME-340 binds to histone H3 and blood group antigens in human colonic mucus. Biosci Biotechnol Biochem. 2012 Sep 7;76(9):1655–1660.

147. Etzold S, MacKenzie DA, Jeffers F, Walshaw J, Roos S, Hemmings AM, et al. Structural and molecular insights into novel surface-exposed mucus adhesins from Lactobacillus reuteri human strains. Mol Microbiol. 2014 May;92(3):543–556.

148. Miyoshi Y, Okada S, Uchimura T, Satoh E. A mucus adhesion promoting protein, MapA, mediates the adhesion of Lactobacillus reuteri to Caco-2 human intestinal epithelial cells. Biosci Biotechnol Biochem. 2006 Jul;70(7):1622–1628.

149. Von Ossowski I, Satokari R, Reunanen J, Lebeer S, De Keersmaecker SCJ, Vanderleyden J, et al. Functional characterization of a mucus-specific LPXTG

surface adhesin from probiotic Lactobacillus rhamnosus GG. Appl Environ Microbiol. 2011 Jul;77(13):4465–4472.

150. Pretzer G, Snel J, Molenaar D, Wiersma A, Bron PA, Lambert J, et al. Biodiversity-based identification and functional characterization of the mannose-specific adhesin of Lactobacillus plantarum. J Bacteriol. 2005 Sep;187(17):6128–6136.

151. Mackenzie DA, Jeffers F, Parker ML, Vibert-Vallet A, Bongaerts RJ, Roos S, et al. Strain-specific diversity of mucus-binding proteins in the adhesion and aggregation properties of Lactobacillus reuteri. Microbiology (Reading, Engl). 2010 Nov;156(Pt 11):3368–3378.

152. Gunning AP, Kavanaugh D, Thursby E, Etzold S, MacKenzie DA, Juge N. Use of Atomic Force Microscopy to Study the Multi-Modular Interaction of Bacterial Adhesins to Mucins. Int J Mol Sci. 2016 Nov 8;17(11).

153. Boekhorst J, Helmer Q, Kleerebezem M, Siezen RJ. Comparative analysis of proteins with a mucus-binding domain found exclusively in lactic acid bacteria. Microbiology (Reading, Engl). 2006 Jan;152(Pt 1):273–280.

154. Roos S, Jonsson H. A high-molecular-mass cell-surface protein from Lactobacillus reuteri 1063 adheres to mucus components. Microbiology (Reading, Engl). 2002 Feb;148(Pt 2):433–442.

155. MacKenzie DA, Tailford LE, Hemmings AM, Juge N. Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. J Biol Chem. 2009 Nov 20;284(47):32444–32453.

156. Etzold S, Kober OI, Mackenzie DA, Tailford LE, Gunning AP, Walshaw J, et al. Structural basis for adaptation of lactobacilli to gastrointestinal mucus. Environ Microbiol. 2014 Mar;16(3):888–903.

157. Devi SM, Halami PM. Diversity and evolutionary aspects of mucin binding (MucBP) domain repeats among Lactobacillus plantarum group strains through comparative genetic analysis. Syst Appl Microbiol. 2017 Apr 12;

158. Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, Partanen P, et al. Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human- mucus binding protein. Proc Natl Acad Sci U S A. 2009 Oct 6;106(40):17193–17198.

159. Wegmann U, MacKenzie DA, Zheng J, Goesmann A, Roos S, Swarbreck D, et al. The pan-genome of Lactobacillus reuteri strains originating from the pig gastrointestinal tract. BMC Genomics. 2015 Dec 1;16(1):1023.

160. Weijland A, Harmark K, Cool RH, Anborgh PH, Parmeggiani A. Elongation factor Tu: a molecular switch in protein biosynthesis. Mol Microbiol. 1992 Mar;6(6):683–688.

161. Mayer F. Cytoskeletons in prokaryotes. Cell Biol Int. 2003 May;27(5):429–438.

162. Dallo SF, Zhang B, Denno J, Hong S, Tsai A, Haskins W, et al. Association of Acinetobacter baumannii EF-Tu with cell surface, outer membrane vesicles, and fibronectin. ScientificWorldJournal. 2012 May 15;2012:128705.

163. Wolff DG, Castiblanco-Valencia MM, Abe CM, Monaris D, Morais ZM, Souza GO, et al. Interaction of Leptospira elongation factor Tu with plasminogen and complement factor H: a metabolic leptospiral protein with moonlighting activities. PLoS ONE. 2013 Nov 27;8(11):e81818.

164. Nikolic M, López P, Strahinic I, Suárez A, Kojic M, Fernández-García M, et al. Characterisation of the exopolysaccharide (EPS)-producing Lactobacillus paraplantarum BGCG11 and its non-EPS producing derivative strains as potential probiotics. Int J Food Microbiol. 2012 Aug 17;158(2):155–162.

165. Erdem AL, Avelino F, Xicohtencatl-Cortes J, Girón JA. Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing Escherichia coli. J Bacteriol. 2007 Oct;189(20):7426–7435.

166. Guerry P. Campylobacter flagella: not just for motility. Trends Microbiol. 2007 Oct 24;15(10):456–461.

167. Neville BA, Forde BM, Claesson MJ, Darby T, Coghlan A, Nally K, et al. Characterization of pro-inflammatory flagellin proteins produced by Lactobacillus ruminis and related motile Lactobacilli. PLoS ONE. 2012 Jul 10;7(7):e40592.

168. Roos S, Aleljung Pã¤, Robert N, Lee B, WadstrÃ¶m T, Lindberg M, et al. A collagen binding protein fromLactobacillus reuteri is part of an ABC transporter system? FEMS Microbiol Lett. 1996 Oct;144(1):33–38.

169. Sillanpää J, Martínez B, Antikainen J, Toba T, Kalkkinen N, Tankka S, et al. Characterization of the collagen-binding S-layer protein CbsA of Lactobacillus crispatus. J Bacteriol. 2000 Nov;182(22):6440–6450.

170. Toba T, Virkola R, Westerlund B, Bjorkman Y, Sillanpaa J, Vartio T, et al. A collagen-binding S-layer protein in Lactobacillus crispatus. Appl Environ Microbiol. 1995;61:2467–2471.

171. Buck BL, Altermann E, Svingerud T, Klaenhammer TR. Functional analysis of putative adhesion factors in Lactobacillus acidophilus NCFM. Appl Environ Microbiol. 2005 Dec;71(12):8344–8351.

172. Hynönen U, Westerlund-Wikström B, Palva A, Korhonen TK. Identification by flagellum display of an epithelial cell- and fibronectin-binding function in the SlpA surface protein of Lactobacillus brevis. J Bacteriol. 2002 Jun;184(12):3360–3367.

173. Mobili P, Serradell M de LA, Trejo SA, Avilés Puigvert FX, Abraham AG, De Antoni GL. Heterogeneity of S-layer proteins from aggregating and non-aggregating Lactobacillus kefir strains. Antonie Van Leeuwenhoek. 2009 May;95(4):363–372.

174. Marraffini LA, DeDent AC, Schneewind O. Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. Microbiol Mol Biol Rev. 2006 Mar 1;70(1):192–221.

175. Bene KP, Kavanaugh DW, Leclaire C, Gunning AP, MacKenzie DA, Wittmann A, et al. Lactobacillus reuteri Surface Mucus Adhesins Upregulate Inflammatory Responses Through Interactions With Innate C-Type Lectin Receptors. Front Microbiol. 2017 Mar 7;8:321.

176. Lizcano A, Sanchez CJ, Orihuela CJ. A role for glycosylated serine-rich repeat proteins in gram-positive bacterial pathogenesis. Mol Oral Microbiol. 2012 Aug;27(4):257–269.

177. Seo HS, Xiong YQ, Sullam PM. Role of the serine-rich surface glycoprotein Srr1 of Streptococcus agalactiae in the pathogenesis of infective endocarditis. PLoS ONE. 2013 May 23;8(5):e64204.

178. Zhou M, Wu H. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. Microbiology (Reading, Engl). 2009 Feb;155(Pt 2):317–327.

179. Sullam PM, Bayer AS, Foss WM, Cheung AL. Diminished platelet binding in vitro by Staphylococcus aureus is associated with reduced virulence in a rabbit model of infective endocarditis. Infect Immun. 1996 Dec;64(12):4915–4921.

180. Bensing BA, Sullam PM. An accessory sec locus of Streptococcus gordonii is required for export of the surface protein GspB and for normal levels of binding to human platelets. Mol Microbiol. 2002 May;44(4):1081–1094.

181. Siboo IR, Chambers HF, Sullam PM. Role of SraP, a Serine-Rich Surface Protein of Staphylococcus aureus, in binding to human platelets. Infect Immun. 2005 Apr;73(4):2273–2280.

182. Takamatsu D, Bensing BA, Prakobphol A, Fisher SJ, Sullam PM. Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. Infect Immun. 2006 Mar;74(3):1933–1940.

183. Xiong YQ, Bensing BA, Bayer AS, Chambers HF, Sullam PM. Role of the serine-rich surface glycoprotein GspB of Streptococcus gordonii in the pathogenesis of infective endocarditis. Microb Pathog. 2008 Oct;45(4):297–301.

184. Seo HS, Minasov G, Seepersaud R, Doran KS, Dubrovska I, Shuvalova L, et al. Characterization of fibrinogen binding by glycoproteins Srr1 and Srr2 of Streptococcus agalactiae. J Biol Chem. 2013 Dec 13;288(50):35982–35996.

185. Wu H, Mintz KP, Ladha M, Fives-Taylor PM. Isolation and characterization of Fap1, a fimbriae-associated adhesin of Streptococcus parasanguis FW213. Mol Microbiol. 1998 May;28(3):487–500.

186. Shivshankar P, Sanchez C, Rose LF, Orihuela CJ. The Streptococcus pneumoniae adhesin PsrP binds to Keratin 10 on lung cells. Mol Microbiol. 2009 Aug;73(4):663–679.

187. Schulte T, Mikaelsson C, Beaussart A, Kikhney A, Deshmukh M, Wolniak S, et al. The BR domain of PsrP interacts with extracellular DNA to promote bacterial aggregation; structural insights into pneumococcal biofilm formation. Sci Rep. 2016 Sep 1;6(1):32371.

188. Sanchez CJ, Shivshankar P, Stol K, Trakhtenbroit S, Sullam PM, Sauer K, et al. The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. PLoS Pathog. 2010 Aug 12;6(8):e1001044.

189. Couvigny B, Lapaque N, Rigottier-Gois L, Guillot A, Chat S, Meylheuc T, et al. Three glycosylated serine-rich repeat proteins play a pivotal role in adhesion and colonization of the pioneer commensal bacterium, Streptococcus salivarius. Environ Microbiol. 2017 Jul 11;

190. Siboo IR, Chaffin DO, Rubens CE, Sullam PM. Characterization of the accessory Sec system of Staphylococcus aureus. J Bacteriol. 2008 Sep;190(18):6188–6196.

191. Zhu F, Zhang H, Yang T, Haslam SM, Dell A, Wu H. Engineering and Dissecting the Glycosylation Pathway of a Streptococcal Serine-rich Repeat Adhesin. J Biol Chem. 2016 Dec 30;291(53):27354–27363.

192. Rigel NW, Braunstein M. A new twist on an old pathway--accessory Sec [corrected] systems. Mol Microbiol. 2008 Jul;69(2):291–302.

193. Feltcher ME, Braunstein M. Emerging themes in SecA2-mediated protein export. Nat Rev Microbiol. 2012 Nov;10(11):779–789.

194. Ramboarina S, Garnett JA, Zhou M, Li Y, Peng Z, Taylor JD, et al. Structural insights into serine-rich fimbriae from Gram-positive bacteria. J Biol Chem. 2010 Oct 15;285(42):32446–32457.

195. Takahashi Y, Yajima A, Cisar JO, Konishi K. Functional analysis of the Streptococcus gordonii DL1 sialic acid-binding adhesin and its essential role in bacterial binding to platelets. Infect Immun. 2004 Jul;72(7):3876–3882.

196. Pyburn TM, Bensing BA, Xiong YQ, Melancon BJ, Tomasiak TM, Ward NJ, et al. A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. PLoS Pathog. 2011 Jul 7;7(7):e1002112.

197. Wu H, Bu S, Newell P, Chen Q, Fives-Taylor P. Two gene determinants are differentially involved in the biogenesis of Fap1 precursors in Streptococcus parasanguis. J Bacteriol. 2007 Feb;189(4):1390–1398.

198. Bensing BA, Siboo IR, Sullam PM. Glycine residues in the hydrophobic core of the GspB signal sequence route export toward the accessory Sec pathway. J Bacteriol. 2007 May;189(10):3846–3854.

199. Bensing BA, Yen YT, Seepersaud R, Sullam PM. A Specific interaction between SecA2 and a region of the preprotein adjacent to the signal peptide occurs during transport via the accessory Sec system. J Biol Chem. 2012 Jul 13;287(29):24438–24447.

200. Takamatsu D, Bensing BA, Sullam PM. Two additional components of the accessory sec system mediating export of the Streptococcus gordonii platelet-binding protein GspB. J Bacteriol. 2005 Jun;187(11):3878–3883.

201. Seepersaud R, Bensing BA, Yen YT, Sullam PM. Asp3 mediates multiple protein-protein interactions within the accessory Sec system of Streptococcus gordonii. Mol Microbiol. 2010 Oct;78(2):490–505.

202. Yen YT, Seepersaud R, Bensing BA, Sullam PM. Asp2 and Asp3 interact directly with GspB, the export substrate of the Streptococcus gordonii accessory Sec System. J Bacteriol. 2011 Jul;193(13):3165–3174.

203. Seepersaud R, Bensing BA, Yen YT, Sullam PM. The accessory Sec protein Asp2 modulates GlcNAc deposition onto the serine-rich repeat glycoprotein GspB. J Bacteriol. 2012 Oct;194(20):5564–5575.

204. Seepersaud R, Sychantha D, Bensing BA, Clarke AJ, Sullam PM. O-acetylation of the serine-rich repeat glycoprotein GspB is coordinated with accessory Sec transport. PLoS Pathog. 2017 Aug 21;13(8):e1006558.

205. Chen Q, Sun B, Wu H, Peng Z, Fives-Taylor PM. Differential roles of individual domains in selection of secretion route of a Streptococcus parasanguinis serine-rich adhesin, Fap1. J Bacteriol. 2007 Nov;189(21):7610–7617.

206. Prabudiansyah I, Driessen AJM. The Canonical and Accessory Sec System of Gram-positive Bacteria. Curr Top Microbiol Immunol. 2017;404:45–67.

207. Bastos PAD, da Costa JP, Vitorino R. A glimpse into the modulation of post-translational modifications of human-colonizing bacteria. J Proteomics. 2017 Jan 30;152:254–275.

208. Dell A, Galadari A, Sastre F, Hitchen P. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. Int J Microbiol. 2010;2010:148178.

209. Schäffer C, Messner P. Emerging facets of prokaryotic glycosylation. FEMS Microbiol Rev. 2017 Jan;41(1):49–91.

210. Aebi M. N-linked protein glycosylation in the ER. Biochim Biophys Acta. 2013 Nov;1833(11):2430–2437.

211. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

212. Munro S. What can yeast tell us about*N* -linked glycosylation in the Golgi apparatus? FEBS Lett. 2001 Jun 8;498(2-3):223–227.

213. Wang Y-C, Peterson SE, Loring JF. Protein post-translational modifications and regulation of pluripotency in human stem cells. Cell Res. 2014 Feb;24(2):143–160.

214. Hang HC, Bertozzi CR. The chemistry and biology of mucin-type O-linked glycosylation. Bioorg Med Chem. 2005 Sep 1;13(17):5021–5034.

215. Tran DT, Ten Hagen KG. Mucin-type O-glycosylation during development. J Biol Chem. 2013 Mar 8;288(10):6921–6929.

216. Fu C, Zhao H, Wang Y, Cai H, Xiao Y, Zeng Y, et al. Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen. HLA. 2016 Dec;88(6):275–286.

217. Rose MC, Voynow JA. Respiratory tract mucin genes and mucin glycoproteins in health and disease. Physiol Rev. 2006 Jan;86(1):245–278.

218. Fukuda M. Roles of mucin-type O-glycans in cell adhesion. Biochimica et Biophysica Acta (BBA) - General Subjects. 2002 Dec;1573(3):394–405.

219. Kelly J, Jarrell H, Millar L, Tessier L, Fiori LM, Lau PC, et al. Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. J Bacteriol. 2006 Apr;188(7):2427–2434.

220. Young NM, Brisson J-R, Kelly J, Watson DC, Tessier L, Lanthier PH, et al. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. J Biol Chem. 2002 Nov 8;277(45):42530–42539.

221. Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, et al. Definition of the bacterial N-glycosylation site consensus sequence. EMBO J. 2006 May 3;25(9):1957–1966.

222. Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, et al. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science. 2006 Nov 17;314(5802):1148–1150.

223. Schwarz F, Lizak C, Fan Y-Y, Fleurkens S, Kowarik M, Aebi M. Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. Glycobiology. 2011 Jan;21(1):45–54.

224. Feldman MF, Wacker M, Hernandez M, Hitchen PG, Marolda CL, Kowarik M, et al. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc Natl Acad Sci U S A. 2005 Feb 22;102(8):3016–3021.

225. Schwarz F, Huang W, Li C, Schulz BL, Lizak C, Palumbo A, et al. A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat Chem Biol. 2010 Apr 1;6(4):264–266.

226. Nothaft H, Scott NE, Vinogradov E, Liu X, Hu R, Beadle B, et al. Diversity in the protein N-glycosylation pathways within the Campylobacter genus. Mol Cell Proteomics. 2012 Nov;11(11):1203–1219.

227. Hopf PS, Ford RS, Zebian N, Merkx-Jacques A, Vijayakumar S, Ratnayake D, et al. Protein glycosylation in Helicobacter pylori: beyond the flagellins? PLoS ONE. 2011 Sep 30;6(9):e25722.

228. Grass S, Lichti CF, Townsend RR, Gross J, St Geme JW. The Haemophilus influenzae HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. PLoS Pathog. 2010 May 27;6(5):e1000919.

229. Gross J, Grass S, Davis AE, Gilmore-Erdmann P, Townsend RR, St Geme JW. The Haemophilus influenzae HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. J Biol Chem. 2008 Sep 19;283(38):26010–26015.

230. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009 Jan;37(Database issue):D233–8.

231. McCann JR, St Geme JW. The HMW1C-like glycosyltransferases--an enzyme family with a sweet tooth for simple sugars. PLoS Pathog. 2014 Apr 10;10(4):e1003977.

232. Naegeli A, Neupert C, Fan Y-Y, Lin C-W, Poljak K, Papini AM, et al. Molecular analysis of an alternative N-glycosylation machinery by functional transfer from Actinobacillus pleuropneumoniae to Escherichia coli. J Biol Chem. 2014 Jan 24;289(4):2170–2179.

233. Nothaft H, Szymanski CM. Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol. 2010 Nov;8(11):765–778.

234. Daubenspeck JM, Jordan DS, Simmons W, Renfrow MB, Dybvig K. General N-and O-Linked Glycosylation of Lipoproteins in Mycoplasmas and Role of Exogenous Oligosaccharide. PLoS ONE. 2015 Nov 23;10(11):e0143362.

235. Tan FYY, Tang CM, Exley RM. Sugar coating: bacterial protein glycosylation and host-microbe interactions. Trends Biochem Sci. 2015 Jul;40(7):342–350.

236. Aas FE, Vik A, Vedde J, Koomey M, Egge-Jacobsen W. Neisseria gonorrhoeae O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. Mol Microbiol. 2007 Aug;65(3):607–624.

237. Vik A, Aas FE, Anonsen JH, Bilsborough S, Schneider A, Egge-Jacobsen W, et al. Broad spectrum O-linked protein glycosylation in the human pathogen Neisseria gonorrhoeae. Proc Natl Acad Sci U S A. 2009 Mar 17;106(11):4447–4452.

238. Hartley MD, Morrison MJ, Aas FE, Børud B, Koomey M, Imperiali B. Biochemical characterization of the O-linked glycosylation pathway in Neisseria gonorrhoeae responsible for biosynthesis of protein glycans containing N,N'-diacetylbacillosamine. Biochemistry. 2011 Jun 7;50(22):4936–4948.

239. Cuccui J, Wren BW. Bacteria like sharing their sweets. Mol Microbiol. 2013 Sep;89(5):811–815.

240. DiGiandomenico A, Matewish MJ, Bisaillon A, Stehle JR, Lam JS, Castric P. Glycosylation of Pseudomonas aeruginosa 1244 pilin: glycan substrate specificity. Mol Microbiol. 2002 Oct;46(2):519–530.

241. Horzempa J, Dean CR, Goldberg JB, Castric P. Pseudomonas aeruginosa 1244 pilin glycosylation: glycan substrate recognition. J Bacteriol. 2006 Jun;188(12):4244–4252.

242. Balonova L, Mann BF, Cerveny L, Alley WR, Chovancova E, Forslund A-L, et al. Characterization of protein glycosylation in Francisella tularensis subsp. holarctica: identification of a novel glycosylated lipoprotein required for virulence. Mol Cell Proteomics. 2012 Jul;11(7):M111.015016.

243. Samuel G, Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res. 2003 Nov 14;338(23):2503–2519.

244. Faridmoayer A, Fentabil MA, Mills DC, Klassen JS, Feldman MF. Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. J Bacteriol. 2007 Nov;189(22):8088–8098.

245. Lithgow KV, Scott NE, Iwashkiw JA, Thomson ELS, Foster LJ, Feldman MF, et al. A general protein O-glycosylation system within the Burkholderia cepacia complex is involved in motility and virulence. Mol Microbiol. 2014 Apr;92(1):116–137.

246. Qutyan M, Henkel M, Horzempa J, Quinn M, Castric P. Glycosylation of pilin and nonpilin protein constructs by Pseudomonas aeruginosa 1244. J Bacteriol. 2010 Nov;192(22):5972–5981.

247. Harding CM, Nasr MA, Kinsella RL, Scott NE, Foster LJ, Weber BS, et al. Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. Mol Microbiol. 2015 Jun;96(5):1023–1041.

248. Musumeci MA, Hug I, Scott NE, Ielmini MV, Foster LJ, Wang PG, et al. In vitro activity of Neisseria meningitidis PglL O-oligosaccharyltransferase with diverse

synthetic lipid donors and a UDP-activated sugar. J Biol Chem. 2013 Apr 12;288(15):10578–10587.

249. Gebhart C, Ielmini MV, Reiz B, Price NL, Aas FE, Koomey M, et al. Characterization of exogenous bacterial oligosaccharyltransferases in Escherichia coli reveals the potential for O-linked protein glycosylation in Vibrio cholerae and Burkholderia thailandensis. Glycobiology. 2012 Jul;22(7):962–974.

250. Zebian N, Merkx-Jacques A, Pittock PP, Houle S, Dozois CM, Lajoie GA, et al. Comprehensive analysis of flagellin glycosylation in Campylobacter jejuni NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization. Glycobiology. 2016 Apr;26(4):386–397.

251. Zampronio CG, Blackwell G, Penn CW, Cooper HJ. Novel glycosylation sites localized in Campylobacter jejuni flagellin FlaA by liquid chromatography electron capture dissociation tandem mass spectrometry. J Proteome Res. 2011 Mar 4;10(3):1238–1245.

252. Valiente E, Bouché L, Hitchen P, Faulds-Pain A, Songane M, Dawson LF, et al. Role of Glycosyltransferases Modifying Type B Flagellin of Emerging Hypervirulent Clostridium difficile Lineages and Their Impact on Motility and Biofilm Formation. J Biol Chem. 2016 Dec 2;291(49):25450–25461.

253. Bouché L, Panico M, Hitchen P, Binet D, Sastre F, Faulds-Pain A, et al. The Type B Flagellin of Hypervirulent Clostridium difficile Is Modified with Novel Sulfonated Peptidylamido-glycans. J Biol Chem. 2016 Dec 2;291(49):25439–25449.

254. Twine SM, Reid CW, Aubry A, McMullin DR, Fulton KM, Austin J, et al. Motility and flagellar glycosylation in Clostridium difficile. J Bacteriol. 2009 Nov;191(22):7050–7062.

255. Faulds-Pain A, Twine SM, Vinogradov E, Strong PCR, Dell A, Buckley AM, et al. The post-translational modification of the Clostridium difficile flagellin affects motility, cell surface properties and virulence. Mol Microbiol. 2014 Oct;94(2):272–289.

256. Twine SM, Paul CJ, Vinogradov E, McNally DJ, Brisson J-R, Mullen JA, et al. Flagellar glycosylation in Clostridium botulinum. FEBS J. 2008 Sep;275(17):4428–4444.

257. Maes E, Krzewinski F, Garenaux E, Lequette Y, Coddeville B, Trivelli X, et al. Glycosylation of BclA Glycoprotein from Bacillus cereus and Bacillus anthracis Exosporium Is Domain-specific. J Biol Chem. 2016 Apr 29;291(18):9666–9677.

258. Schirm M, Kalmokoff M, Aubry A, Thibault P, Sandoz M, Logan SM. Flagellin from Listeria monocytogenes is glycosylated with beta-O-linked N-acetylglucosamine. J Bacteriol. 2004 Oct;186(20):6721–6727.

259. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. Annu Rev Biochem. 2011;80:825–858.

260. Bu S, Li Y, Zhou M, Azadin P, Zeng M, Fives-Taylor P, et al. Interaction between two putative glycosyltransferases is required for glycosylation of a serine-rich streptococcal adhesin. J Bacteriol. 2008 Feb;190(4):1256–1266.

261. Zhou M, Zhu F, Dong S, Pritchard DG, Wu H. A novel glucosyltransferase is required for glycosylation of a serine-rich adhesin and biofilm formation by Streptococcus parasanguinis. J Biol Chem. 2010 Apr 16;285(16):12140–12148.

262. Zhang H, Zhu F, Yang T, Ding L, Zhou M, Li J, et al. The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. Nat Commun. 2014 Jul 15;5:4339.

263. Zhu F, Erlandsen H, Ding L, Li J, Huang Y, Zhou M, et al. Structural and functional analysis of a new subfamily of glycosyltransferases required for glycosylation of serine-rich streptococcal adhesins. J Biol Chem. 2011 Jul 29;286(30):27048–27057.

264. Zhang H, Zhou M, Yang T, Haslam SM, Dell A, Wu H. A New Helical Binding Domain Mediates a Unique Glycosyltransferase Activity of a Bifunctional Protein. J Biol Chem. 2016 Aug 17;

265. Peng Z, Wu H, Ruiz T, Chen Q, Zhou M, Sun B, et al. Role of gap3 in Fap1 glycosylation, stability, in vitro adhesion, and fimbrial and biofilm formation of Streptococcus parasanguinis. Oral Microbiol Immunol. 2008 Feb;23(1):70–78.

266. Chen Y, Seepersaud R, Bensing BA, Sullam PM, Rapoport TA. Mechanism of a cytosolic O-glycosyltransferase essential for the synthesis of a bacterial adhesion protein. Proc Natl Acad Sci U S A. 2016 Mar 1;113(9):E1190–9.

267. Shi W-W, Jiang Y-L, Zhu F, Yang Y-H, Shao Q-Y, Yang H-B, et al. Structure of a novel O-linked N-acetyl-D-glucosamine (O-GlcNAc) transferase, GtfA, reveals insights into the glycosylation of pneumococcal serine-rich repeat adhesins. J Biol Chem. 2014 Jul 25;289(30):20898–20907.

268. Jiang Y-L, Jin H, Yang H-B, Zhao R-L, Wang S, Chen Y, et al. Defining the enzymatic pathway for polymorphic O-glycosylation of the pneumococcal serine-rich repeat protein PsrP. J Biol Chem. 2017 Feb 28;

269. Chaze T, Guillot A, Valot B, Langella O, Chamot-Rooke J, Di Guilmi A-M, et al. O-Glycosylation of the N-terminal region of the serine-rich adhesin Srr1 of Streptococcus agalactiae explored by mass spectrometry. Mol Cell Proteomics. 2014 Sep;13(9):2168–2182.

270. Bernard E, Rolain T, Courtin P, Guillot A, Langella P, Hols P, et al. Characterization of O-acetylation of N-acetylglucosamine: a novel structural variation of bacterial peptidoglycan. J Biol Chem. 2011 Jul 8;286(27):23950–23958.

271. Coyne MJ, Reinap B, Lee MM, Comstock LE. Human symbionts use a host-like pathway for surface fucosylation. Science. 2005 Mar 18;307(5716):1778–1781.

272. Fletcher CM, Coyne MJ, Villa OF, Chatzidaki-Livanis M, Comstock LE. A general O-glycosylation system important to the physiology of a major human intestinal symbiont. Cell. 2009 Apr 17;137(2):321–331.

273. Fletcher CM, Coyne MJ, Comstock LE. Theoretical and experimental characterization of the scope of protein O-glycosylation in Bacteroides fragilis. J Biol Chem. 2011 Feb 4;286(5):3219–3226.

274. Lee I-C, van Swam II, Tomita S, Morsomme P, Rolain T, Hols P, et al. GtfA and GtfB are both required for protein O-glycosylation in Lactobacillus plantarum. J Bacteriol. 2014 May;196(9):1671–1682.

275. Fredriksen L, Mathiesen G, Moen A, Bron PA, Kleerebezem M, Eijsink VGH, et al. The major autolysin Acm2 from Lactobacillus plantarum undergoes cytoplasmic O-glycosylation. J Bacteriol. 2012 Jan;194(2):325–333.

276. Rolain T, Bernard E, Beaussart A, Degand H, Courtin P, Egge-Jacobsen W, et al. O-glycosylation as a novel control mechanism of peptidoglycan hydrolase activity. J Biol Chem. 2013 Aug 2;288(31):22233–22247.

277. Fredriksen L, Moen A, Adzhubei AA, Mathiesen G, Eijsink VGH, Egge-Jacobsen W. Lactobacillus plantarum WCFS1 O-linked protein glycosylation: an extended spectrum of target proteins and modification sites detected by mass spectrometry. Glycobiology. 2013 Dec;23(12):1439–1451.

278. Lebeer S, Claes IJJ, Balog CIA, Schoofs G, Verhoeven TLA, Nys K, et al. The major secreted protein Msp1/p75 is O-glycosylated in Lactobacillus rhamnosus GG. Microb Cell Fact. 2012 Feb 1;11:15.

279. Lebeer S, Claes I, Tytgat HLP, Verhoeven TLA, Marien E, von Ossowski I, et al. Functional analysis of Lactobacillus rhamnosus GG pili in relation to adhesion and immunomodulatory interactions with intestinal epithelial cells. Appl Environ Microbiol. 2012 Jan;78(1):185–193.

280. Tytgat HLP, van Teijlingen NH, Sullan RMA, Douillard FP, Rasinkangas P, Messing M, et al. Probiotic Gut Microbiota Isolate Interacts with Dendritic Cells via Glycosylated Heterotrimeric Pili. PLoS ONE. 2016 Mar 17;11(3):e0151824.

281. Kajikawa A, Midorikawa E, Masuda K, Kondo K, Irisawa T, Igimi S, et al. Characterization of flagellins isolated from a highly motile strain of Lactobacillus agilis. BMC Microbiol. 2016 Mar 22;16(1):49.

282. Sleytr UB, Schuster B, Egelseer E-M, Pum D. S-layers: principles and applications. FEMS Microbiol Rev. 2014 Sep;38(5):823–864.

283. Avall-Jääskeläinen S, Palva A. Lactobacillus surface layers and their applications. FEMS Microbiol Rev. 2005 Aug;29(3):511–529.

284. Anzengruber J, Pabst M, Neumann L, Sekot G, Heinl S, Grabherr R, et al. Protein O-glucosylation in Lactobacillus buchneri. Glycoconj J. 2014 Feb;31(2):117–131.

285. Konstantinov SR, Smidt H, de Vos WM, Bruijns SCM, Singh SK, Valence F, et al. S layer protein A of Lactobacillus acidophilus NCFM regulates immature dendritic cell and T cell functions. Proc Natl Acad Sci U S A. 2008 Dec 9;105(49):19474–19479.

286. Sequeira S, Kavanaugh D, MacKenzie DA, Šuligoj T, Walpole S, Leclaire C, et al. Structural basis for the role of Serine-Rich Repeat Proteins from Lactobacillus reuteri in niche adaptation and biofilm formation. 2017;

287. Sims IM, Frese SA, Walter J, Loach D, Wilson M, Appleyard K, et al. Structure and functions of exopolysaccharide produced by gut commensal Lactobacillus reuteri 100-23. ISME J. 2011 Jul;5(7):1115–1124.

288. Heavens D, Tailford LE, Crossman L, Jeffers F, Mackenzie DA, Caccamo M, et al. Genome sequence of the vertebrate gut symbiont Lactobacillus reuteri ATCC 53608. J Bacteriol. 2011 Aug;193(15):4015–4016.

289. Kotarski SF, Savage DC. Models for study of the specificity by which indigenous lactobacilli adhere to murine gastric epithelia. Infect Immun. 1979 Dec;26(3):966–975.

290. Hinterdorfer P, Gruber HJ, Kienberger F, Kada G, Riener C, Borken C, et al. Surface attachment of ligands and receptors for molecular recognition force microscopy. Colloids and Surfaces B: Biointerfaces. 2002 Feb;23(2-3):115–123.

291. Gunning AP, Chambers S, Pin C, Man AL, Morris VJ, Nicoletti C. Mapping specific adhesive interactions on living human intestinal epithelial cells with atomic force microscopy. FASEB J. 2008 Jul;22(7):2331–2339.

292. Wadström T, Andersson K, Sydow M, Axelsson L, Lindgren S, Gullmar B. Surface properties of lactobacilli isolated from the small intestine of pigs. J Appl Bacteriol. 1987 Jun;62(6):513–520.

293. Saier MH. Bacterial phosphoenolpyruvate: sugar phosphotransferase systems: structural, functional, and evolutionary interrelationships. Bacteriol Rev [Internet]. 1977;41:856–871. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC414030/

294. Francl AL, Thongaram T, Miller MJ. The PTS transporters of Lactobacillus gasseri ATCC 33323. BMC Microbiol. 2010 Mar 12;10:77.

295. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol. 2008 Aug;6(8):613–624.

296. Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol. 2003 Apr 25;328(2):307–317.

297. Chaillou S, Pouwels PH, Postma PW. Transport of D-xylose in Lactobacillus pentosus, Lactobacillus casei, and Lactobacillus plantarum: evidence for a mechanism of facilitated diffusion via the phosphoenolpyruvate:mannose phosphotransferase system. J Bacteriol. 1999 Aug;181(16):4768–4773.

298. Shibata N, Okawa Y. Chemical structure of beta-galactofuranose-containing polysaccharide and O-linked oligosaccharides obtained from the cell wall of pathogenic dematiaceous fungus Fonsecaea pedrosoi. Glycobiology. 2011 Jan;21(1):69–81.

299. Dykhuizen EC, May JF, Tongpenyai A, Kiessling LL. Inhibitors of UDP-galactopyranose mutase thwart mycobacterial growth. J Am Chem Soc. 2008 May 28;130(21):6706–6707.

300. Soltero-Higgin M, Carlson EE, Gruber TD, Kiessling LL. A unique catalytic mechanism for UDP-galactopyranose mutase. Nat Struct Mol Biol. 2004 Jun;11(6):539–543.

301. Oppenheimer M, Valenciano AL, Sobrado P. Biosynthesis of galactofuranose in kinetoplastids: novel therapeutic targets for treating leishmaniasis and chagas' disease. Enzyme Res. 2011 May 25;2011:415976.

302. Uehara T, Park JT. The N-acetyl-D-glucosamine kinase of Escherichia coli and its role in murein recycling. J Bacteriol. 2004 Nov;186(21):7273–7279.

303. Su MS, Schlicht S, Gänzle MG. Contribution of glutamate decarboxylase in Lactobacillus reuteri to acid resistance and persistence in sourdough fermentation. Microb Cell Fact. 2011 Aug 30;10 Suppl 1:S8.

304. Koop HM, Valentijn-Benz M, Nieuw Amerongen AV, Roukema PA, De Graaff J. Comparison of different assays for the aggregation of oral bacteria by human whole saliva. Antonie Van Leeuwenhoek. 1989 Jun;55(2):109–122.

305. Lizcano A, Akula Suresh Babu R, Shenoy AT, Saville AM, Kumar N, D'Mello A, et al. Transcriptional organization of pneumococcal psrP-secY2A2 and impact of GtfA and GtfB deletion on PsrP-associated virulence properties. Microbes Infect. 2017 Apr 10;

306. Zhu F, Zhang H, Wu H. A conserved domain is crucial for acceptor substrate binding in a family of glucosyltransferases. J Bacteriol. 2015 Feb;197(3):510–517.

307. D'Urzo N, Malito E, Biancucci M, Bottomley MJ, Maione D, Scarselli M, et al. The structure of Clostridium difficile toxin A glucosyltransferase domain bound to Mn2+ and UDP provides insights into glucosyltransferase activity and product release. FEBS J. 2012 Sep;279(17):3085–3097.

308. Wittmann A, Lamprinaki D, Bowles KM, Katzenellenbogen E, Knirel YA, Whitfield C, et al. Dectin-2 Recognizes Mannosylated O-antigens of Human Opportunistic Pathogens and Augments Lipopolysaccharide Activation of Myeloid Cells. J Biol Chem. 2016 Aug 19;291(34):17629–17638.

309. Su SV, Hong P, Baik S, Negrete OA, Gurney KB, Lee B. DC-SIGN binds to HIV-1 glycoprotein 120 in a distinct but overlapping fashion compared with ICAM-2 and ICAM-3. J Biol Chem. 2004 Apr 30;279(18):19122–19132.

310. Garcia-Vallejo JJ, van Kooyk Y. The physiological role of DC-SIGN: a tale of mice and men. Trends Immunol. 2013 Oct;34(10):482–486.

311. Taylor AM, Boulter J, Harding SE, Cölfen H, Watts A. Hydrodynamic properties of human erythrocyte band 3 solubilized in reduced Triton X-100. Biophys J. 1999 Apr;76(4):2043–2055.

312. Drapeau GR, Boily Y, Houmard J. Purification and properties of an extracellular protease of Staphylococcus aureus. J Biol Chem. 1972 Oct 25;247(20):6720–6726.

313. Pothoulakis C, Galili U, Castagliuolo I, Kelly CP, Nikulasson S, Dudeja PK, et al. A human antibody binds to alpha-galactose receptors and mimics the effects of Clostridium difficile toxin A in rat colon. Gastroenterology. 1996 Jun;110(6):1704–1712.

314. Wesener DA, Wangkanont K, McBride R, Song X, Kraft MB, Hodges HL, et al. Recognition of microbial glycans by human intelectin-1. Nat Struct Mol Biol. 2015 Aug;22(8):603–610.

315. Staaf M, Urbina F, Weintraub A, Widmalm G. Structure determination of the O-antigenic polysaccharide from the enterotoxigenic Escherichia coli (ETEC) O101. Carbohydr Res. 1997 Jan;297(3):297–299.

316. Knirel IA, Kochetkov NK. [Structure of lipopolysaccharides from gram-negative bacteria. III. Structure of O-specific polysaccharides]. Biokhimiia. 1994 Dec;59(12):1784–1851.

317. Haseley SR, Diggle HJ, Wilkinson SG. Structure of a surface polysaccharide from Acinetobacter baumannii O16. Carbohydr Res. 1996 Oct;293(2):259–265.

318. Aoki-Kinoshita K, Agravat S, Aoki NP, Arpinar S, Cummings RD, Fujita A, et al. GlyTouCan 1.0--The international glycan structure repository. Nucleic Acids Res. 2016 Jan 4;44(D1):D1237–42.

319. Huai G, Qi P, Yang H, Wang Y. Characteristics of α-Gal epitope, anti-Gal antibody, α1,3 galactosyltransferase and its clinical exploitation (Review). Int J Mol Med. 2016 Jan;37(1):11–20.

320. Voehringer D, Stanley SA, Cox JS, Completo GC, Lowary TL, Locksley RM. Nippostrongylus brasiliensis: identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. Exp Parasitol. 2007 Aug;116(4):458–466.

321. Hanuszkiewicz A, Pittock P, Humphries F, Moll H, Rosales AR, Molinaro A, et al. Identification of the flagellin glycosylation system in Burkholderia cenocepacia and the contribution of glycosylated flagellin to evasion of human innate immune responses. J Biol Chem. 2014 Jul 4;289(27):19231–19244.

322. Mistou M-Y, Dramsi S, Brega S, Poyart C, Trieu-Cuot P. Molecular dissection of the secA2 locus of group B Streptococcus reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence. J Bacteriol. 2009 Jul;191(13):4195–4206.

323. Walter J, Britton RA, Roos S. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the Lactobacillus reuteri paradigm. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1:4645–4652.

324. Bensing BA, Sullam PM. Characterization of Streptococcus gordonii SecA2 as a paralogue of SecA. J Bacteriol. 2009 Jun;191(11):3482–3491.

325. Kolappan S, Coureuil M, Yu X, Nassif X, Egelman EH, Craig L. Structure of the Neisseria meningitidis Type IV pilus. Nat Commun. 2016 Oct 4;7:13015.

326. Iwashkiw JA, Vozza NF, Kinsella RL, Feldman MF. Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. Mol Microbiol. 2013 Jul;89(1):14–28.

327. Galili U. The alpha-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. Immunol Cell Biol. 2005 Dec;83(6):674–686.

328. Chiodo F, Marradi M, Park J, Ram AFJ, Penadés S, van Die I, et al. Galactofuranose-coated gold nanoparticles elicit a pro-inflammatory response in human monocyte-derived dendritic cells and are recognized by DC-SIGN. ACS Chem Biol. 2014 Feb 21;9(2):383–389.

329. Wells JM. Immunomodulatory mechanisms of lactobacilli. Microb Cell Fact. 2011 Aug 30;10 Suppl 1:S17.

330. Call EK, Goh YJ, Selle K, Klaenhammer TR, O'Flaherty S. Sortase-deficient lactobacilli: effect on immunomodulation and gut retention. Microbiology (Reading, Engl). 2015 Feb;161(Pt 2):311–321.

331. Remus DM, Bongers RS, Meijerink M, Fusetti F, Poolman B, de Vos P, et al. Impact of Lactobacillus plantarum sortase on target protein sorting, gastrointestinal persistence, and host immune response modulation. J Bacteriol. 2013 Feb;195(3):502–509.

332. Fanning S, Hall LJ, Cronin M, Zomer A, MacSharry J, Goulding D, et al. Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. Proc Natl Acad Sci U S A. 2012 Feb 7;109(6):2108–2113.

333. Wu R, Wu H. A molecular chaperone mediates a two-protein enzyme complex and glycosylation of serine-rich streptococcal adhesins. J Biol Chem. 2011 Oct 7;286(40):34923–34931.

334. Li Y, Chen Y, Huang X, Zhou M, Wu R, Dong S, et al. A conserved domain of previously unknown function in Gap1 mediates protein-protein interaction and is required for biogenesis of a serine-rich streptococcal adhesin. Mol Microbiol. 2008 Dec;70(5):1094–1104.

335. Takamatsu D, Bensing BA, Sullam PM. Four proteins encoded in the gspB-secY2A2 operon of Streptococcus gordonii mediate the intracellular glycosylation of the platelet-binding protein GspB. J Bacteriol. 2004 Nov;186(21):7100–7111.