GLYCOBIOLOGY
OF
COMMENSAL BACTERIA
with emphasis on cell surface adhesins
and exopolysaccharides

A thesis submitted to the National University of Ireland, Galway for the degree of Doctor of Philosophy (PhD)

By

Benoît Houeix, BSc, MSc,

Glycoscience group,

National Centre for Biomedical Engineering Science,

National University of Ireland, Galway

September 2014

Supervisor: Prof. Lokesh Joshi

Co-supervisors: Prof. Fergus Shanahan and Prof. Douwe Van Sinderen

“The role of the infinitely small in nature is infinitely great.” Louis Pasteur (1822-1895).

Contents

Table of Contents

Declaration

I certify that this thesis has not previously been submitted as an exercise for a degree at the National University of Ireland, or at any other university, and I further declare that the work embodied in it is my own.

Benoît Houeix

Abstract

The human gastro-intestinal tract (GIT) is an elaborate ecosystem specific to an individual and his physiological stage that involves inter- and intra-relationships between the bacteria composing the microbiota and between the host and the microbes. The idea of a communication system or a 'crosstalk' between the host and the microbiota has emerged after bacterial-derived molecules, such as polysaccharide A in *Bacteroides fragilis*, were shown to be essential for the maturation of the host immune system (Mazmanian et al., 2007). Understanding the molecular nature of the interaction between host gut epithelial cells and commensal microbes is essential for promotion of the mutualistic relationship with all its associated benefits for the host and prevention of the deleterious effects of infection by pathogens. Considerable efforts have been directed at the exploration of the lectin - adhesins of pathogenic bacteria and their relevant glycan receptors, with a view to identification of potential candidates for vaccines. To date, very little is known about adhesion mechanisms in gut commensal species. In particular, the lectin-like adhesins and their associated carbohydrate ligands have not been described. Recent technological advances in glycomics (omic databases, synthesis of glyco-conjugates and glycan microarrays) have provided new tools to study lectin/adhesin interactions with glycans.

In this thesis, an *in silico* approach was used to identify surface-exposed proteins and putative adhesins of human commensal species, from the genera, *Faecalibacterium*, *Bifidobacterium* and *Lactobacillus*, that could represent potential probiotic strains. *In vitro* expression studies supported the *in silico* work, especially for the *Lactobacillus* species. Selected strains and their extracted pili were also incubated on mucin and neo-glycoconjugate microarrays and glycan ligands identified in *F. prausnitzii*, *B. bifidum* PRL2010 and *L. paracasei* subsp. *paracasei*. The potential of these novel approaches for the study of adhesins was thus demonstrated.

The beneficial effects attributed to commensals in the GIT are in many cases exerted through exopolysaccharides (EPSs). Bacteria assemble a variety of glycoconjugates and polysaccharides as part of their cell walls. These include anchored molecules such as bacterial peptidoglycan, capsules and EPSs. Several functions, including health benefits for the host, have been attributed to EPSs. However, their isolation and analysis is a lengthy process that requires several analytical steps. The study of mammalian glycosylation has

benefited significantly from recent developments in microarray technology, which enables the simultaneous analysis of the interactions between glycans and immobilised carbohydrate-binding proteins, such as plant lectins. Here, lectin microarray technology was successfully used for the first time to profile and differentiate fluorescently-labelled EPSs from commensal strains, as confirmed by monosaccharide analyses. Thus will help to move our understanding of EPS molecules ahead more quickly and promote their exploitation.

Acknowledgements

This work was funded by Science Foundation Ireland (SFI) under the Alimentary Glycoscience Research Cluster (AGRC).

I would, first, like to thank my supervisor Prof. Lokesh Joshi for offering me the opportunity to take up this challenge. I am also deeply grateful to Dr Marian Kane for providing direction and support. Sincere thanks to Dr Michael Cairns for all the comments, advices and for reviewing this thesis. I would also like to thank Dr Conor O'Byrne, Microbiology Dept, National University of Ireland, Galway, member of the Graduate Research Committee, for his expertise in Microbiology. I would also take this opportunity to acknowledge all the people without whom this project would not have been completed:

Prof. Fergus Shanahan, Prof. Douwe Van Sinderen, Dr Saranna Fanning and Dr Mary O'Connell-Motherway, Alimentary Pharmabiotic Centre, University College Cork, Ireland;

Dr Michelle Kilcoyne, Dr Jared Gerlach, Dr Marie Le Berre, Dr Iain Shaw, Dr Tinaig Le Costaouec and Ms Catherine Loughrey, Glycoscience Group, National Centre for Biomedical Engineering Science, National University of Ireland, Galway;

Prof. Vincent O'Flaherty, Dr Aoife Boyd, Dr Eoin Gugglin, Dr Nicky O'Boyle, Dr Marcus Twomey, Microbiology Dept, National University of Ireland, Galway;

Dr Aideen Timmins, Centre for Microscopy and Imaging, National Centre for Biomedical Engineering Science, National University of Ireland, Galway;

Prof. Maarten Van de Guchte and Mr Valentin Loux, Unité Mathématique, Informatique et Génome, INRA, Jouy en Josas, France ;

All the staff from the Glycoscience Group and the Martin Ryan Institute, National University of Ireland, Galway;

Thank you also to Declan and Aoife, biomedical science students, who completed their fourth year project with me.

List of Figures

Chapter 1: General introduction

Chapter 2: Materials and methods

Chapter 3 : *In silico* identification of putative adhesins and experimental validation

Chapter 4 : Glycan ligands of selected commensal adhesins

Chapter 5 : Commensal exopolysaccharides

Appendix

List of tables

Chapter 1: General introduction

Chapter 2: Materials and methods

Chapter 3 : *In silico* analysis of surface-exposed proteins

Chapter 4 : Glycan ligands of selected commensal adhesins

Chapter 5 : Commensal exopolysaccharides

List of abbreviations

AAR	:	adhesion adaptive response
bEPS	:	cell-bound EPS
CBP	:	carbohydrate-binding protein
CD	:	Crohn's disease
CD4+	:	naïve lymphocyte T cell
cDNA	:	complementary to DNA
CNS	:	central nervous system
CP	:	cytoplasm
CRD	:	carbohydrate recognition domain
CS	:	chondroitin sulfate
CW	:	cell wall
DAEC	:	entero-adherent *E. coli*
DC	:	dendritic cell
EAEC	:	entero-aggregativec *E. coli*
ECM	:	extracellular matrix
ECP	:	*E. coli* common pili
EHEC	:	entero-hemorragic *E. coli*
EIEC	:	entero-invasive *E. coli*
EPEC	:	entero-pathogenic *E. coli*
EPS	:	exopolysaccharide
ER	:	endoplasmic reticulum
ETEC	:	entero-toxigenic *E. coli*
FAE	:	follicle-associated epithelium
Fuc	:	fucose
FUT2	:	fucosyltransferase 2
GAG	:	glycosaminoglycan
Gal	:	galactose
GalNAc:		*N*-acetylgalactosamine
GALT	:	gut-associated lymphoid tissue
Gc	:	goblet cell
GF	:	germ free mice
GH	:	glycosylhydrolase
GIT	:	gastro-intestinal tract
Glc	:	glucose
GlcNAc:		*N*-acetylglucosamine
GPI	:	glycosylphosphatidylinositol
GRAS	:	generally considered as safe
GT	:	glycosyltransferase
HA	:	hyaluronan
HBG	:	histo-blood group
HMO	:	human milk oligosaccharide
HPLC	:	high-performance liquid chromatography
HS	:	heparan sulfate

IBD : inflammatory bowel diseases
IEC : intestinal epithelial cell
IM : inner membrane
INV : inverted EPS mutant
KS : keratan sulfate
LAB : lactic acid bacteria
Lac : lactose
LGG : *Lactobacillus rhamnosus* GG
LOS : lipooligosaccharide
LPS : lipopolysaccharide
LTA : lipoteichoic acid
Man : mannose
Mc : microfold cell
mRNA : messenger RNA
MSCRAMM: microbial surface component recognising adhesive matrix molecule
NCBI : National Center for Biotechnology Information
NGC : neo-glycoconjugate
NMR : nuclear magnetic resonance
OM : outer membrane
PCR : polymerase chain reaction
PG : peptidoglycan
pGT : priming glycosyltransferase
PGM : porcine gastric mucin
PL : phospholipid
PRR : pattern-recognising receptor
PSA : polysaccharide A
PSA : periplasmic space
PTM : post-translational modification
PTS : proline-threonine-serine tandem repeat
qPCR : quantitative polymerase chain reaction
rEPS : media-released EPS
Rha : rhamnose
RT : reverse transcription
SCFA : short chain fatty acid
SEP : surface-exposed protein
SIgA : secretory immunoglobulin A
SLP : surface layer protein
Spa : sortase-mediated pilus assembly
Srt : sortase
TLR : Toll-like receptor
TMH : transmembrane helix
vWF : von Willebrand factor
WT : wild type

Chapter 1: Introduction to gut and commensal glycobiology

1.1 The gut and its inhabitants

1.1.1 The microbiota

'Thanks to sour milk, we shall live one hundred years'. Elie Metchnikoff (1845-1916), a Russian microbiologist, was one of the first scientists to link microbial life to health and disease (Metchnikoff 1903)**.** The human gastro-intestinal tract (GIT) houses a considerable microbial biomass, termed the microbiota. It is estimated to contain 100 trillion foreign cells from over 1000 species belonging to four major bacterial phyla (Bacteroidetes, Firmicutes, Actinobacteria and Proteobacteria) with over 500 Giga bases of prokaryotic DNA composing the microbiome (Xu, Bjursell et al. 2003; Rajilic-Stojanovic, Smidt et al. 2007; Qin, Li et al. ; Qin, Li et al.).

Through generations of co-evolution, microbes and humans have developed different degrees of inter-specific relationships that range from parasitism to mutualism (Ventura 2007). The microbiota expands in the GIT (Figure 1.1) from the oesophagus, through the acidic stomach, to the small intestine then the colon, where microbes are the most numerous, reaching 10^{12} cells/g of luminal content (Dethlefsen, Eckburg et al. 2006; Sartor 2008). The lumen of the human digestive system is surrounded by the mucosa, composed of a mucus layer secreted by specialised cells (goblet and Paneth cells) covering the epithelium, the connective tissue (*lamina propria*) and the smooth muscle coating. The mucus layer is a complex structure that increases in thickness along the intestine. Two parts can be distinguished: an inner layer which is virtually microbe-free and a thicker layer, loosely attached to the inner layer but densely colonised by bacteria. The main structural components of mucus are mucin glycoproteins. The inner mucus layer is composed of mucins tightly adhering to the epithelial cells (such as MUC1) and part of the glycocalyx, whereas the outer layer is formed by gel-like mucins (such as MUC2) (Atuma, Strugala et al. 2001; Johansson, Phillipson et al. 2008).

Here, microbes from various genera have been able to settle and adapt to their host's environment, creating ecological niches or biotopes between the microvilli and the crypts of the mucosal barrier. Thus, a well-established complex, diverse and dynamic ecosystem from tolerated to symbiotic microbial species has emerged and is believed to be host and

tissue specific (Ley, Peterson et al. 2006). These species are referred to as commensals, which is the Latin term for ‘sharing a table’. Although most commensal strains have developed a mutualistic relationship with their host (symbionts), some resident species can turn pathogenic (pathobionts) in particular conditions such as when the environment is altered (Mazmanian, Round et al. 2008; Chow and Mazmanian 2010). Bacterial communities in the GIT are part of a trophic chain and interact with each other through quorum sensing (Zhu, Miller et al. 2002) and it is recognised they exert significant impact in human health.

The gut contains the highest population of immune cells in the human body and the second greatest of neurons outside of the central nervous system (CNS) (Ley, Turnbaugh et al. 2006; Turnbaugh, Ley et al. 2007; Kau, Ahern et al. 2011). Understanding how the gut microbiota affects health and disease has been of increasing interest in the recent years. As in all ecosystems, species live together in a homeostatic state, where the interactions between the host and the microbes are optimal. The microbiota is then stable and the mucosal barrier intact. Under certain circumstances, including ageing, diet change, illnesses, medication such as antibiotics or trauma, the balance gets disrupted and, as a result, the barrier can no longer fully play its role. The intestinal microbiota gets disturbed with overgrowth of some species, an increase in the number of pathogens, an elevation of pH and a decrease of butyrate in the intestinal lumen. This results in an inflammation state, opportunistic infections, troubled intestinal motility and an aberrant immune response in the host (Vrieze, Holleman et al. ; Molloy, Grainger et al. 2013). Disruptions to the normal balance between the gut microbiota and the host (dysbiosis) have been linked with various diseases such as inflammatory bowel diseases (IBD), obesity, metabolic syndromes, cancer and also neurological or behavioural disorders (Hawrelak and Myers 2004; Mayer, Savidge et al. 2014).

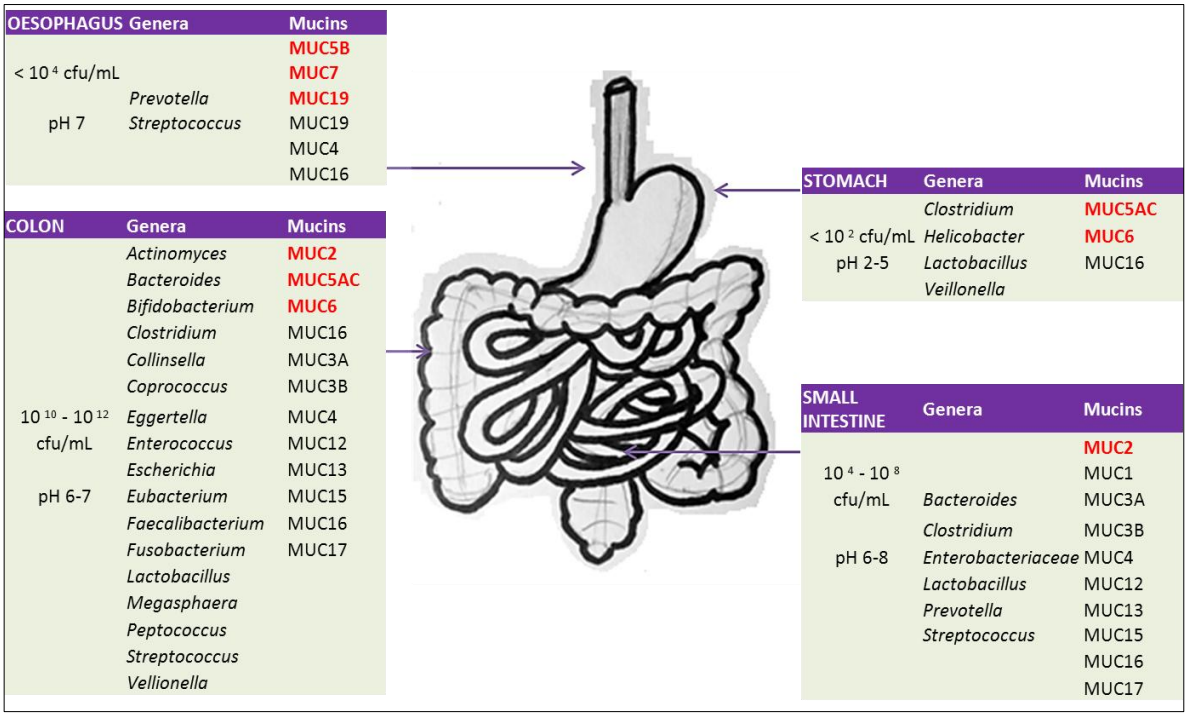

Figure 1.1: Bacterial communities in the human adult gastro-intestinal tract

The major habitats for commensal bacteria featuring pH condition, density of bacteria, main population and mucins secreted by the intestinal cells with the gel forming mucins in **red**; adapted from (Dethlefsen, Eckburg et al. 2006; Sartor 2008)

1.1.2 Development of the gut microbiota

The development of metagenomics has enabled the study of unculturable micro-organisms and represents a powerful 'microscope' to profile microbial gene abundance in the human GIT. Consortia focussing on microbiomes linked to health and disease, such as METAHIT (http://www.metahit.eu), the Human Microbiome Project (http://www.hmpdacc.org/) or ELDERMET (http://eldermet.ucc.ie/), have established an extensive reference catalogue of microbial genes present in the human intestine using fecal samples.

The gut of the new-born infant is considered sterile at birth but is gradually colonised by the microbes it encounters firstly in the vaginal tract of the mother during labour, then on the skin and the milk microbiota. Staphylococci, streptococci, enterococci, lactobacilli and especially bifidobacteria are the dominant genera amongst these first colonisers, and their presence coincides with the development of the infant's immune system (Martín, Langa et al. 2004; Wells, Rossi et al. 2011; Cahenzli, Balmer et al. 2013). External factors that influence the development of the infant's microbiota have been reviewed recently

(Matamoros, Gras-Leguen et al. 2013). Translocation from the mother's microbiota, delivery modes (caesarean or vaginal), types of feeding (breast milk or formula), familial environment (such as geographical location) and antibiotic treatments influence the development of the microbiota. A rapid increase in abundance and diversity of bacterial species takes place in the following 100 days of life, by contact with micro-organisms from the surrounding environment, shaping the microbiota population (Palmer, Bik et al. 2007; Round and Mazmanian 2009).

Little is known about the variation in terms of composition of the microbiota throughout human life. The initial colonisation during early life by microbes creates an electrochemically reducing environment in the GIT, favourable for the establishment of strict anaerobes such as *Faecalibacterium prausnitzii*, a predominant human commensal not found in neonates (O'Toole and Claesson 2010; Jost, Lacroix et al. 2012). A high stability and resilience characterise then the adult microbiota (Franks, Harmsen et al. 1998). Recently, researchers have combined 22 metagenomic analyses of human fecal samples from six developed countries, resulting in the identification of three robust clusters (Arumugam, Raes et al. 2011). The authors suggested that three types of microbiomes or ecosystems (enterotypes) are encountered in human adults, comprised of a similar core of bacterial populations and their associated metabolic functions. Enterotype 1 is rich in the genus *Bacteroides*, whereas *Prevotella* dominates the enterotype 2, and enterotype 3 is highly abundant in *Ruminococcus* species. In another study, Wu et al. were able to correlate two of these ecosystems with long term diet (Wu, Chen et al. 2011). Enterotype 1 seems to correlate with individuals having a high animal fat diet, whereas enterotype 2 associated with a carbohydrate-rich diet.

Despite the relative stability of the microbiota during adult life, recent studies highlighted an alteration of the intestinal flora with age. Differences in the composition of the microbial population have been reported in elderly people with a marked decrease in bifidobacteria, lactobacilli and *F. prausnitzii* as well as extreme variability between individuals (Mariat, Firmesse et al. 2009; Biagi, Nylund et al. 2010; Claesson, Cusack et al. 2011).

1.1.3 Roles of commensal bacteria

The roles attributed to commensal bacteria in the gut are expanding with advances in the understanding of the interactions between the intestinal microbiota and the host. These have been extensively reviewed in recent years (Lebeer, Vanderleyden et al. 2010; Littman and Pamer 2011; Tlaskalova-Hogenova, Stepankova et al. 2011; Cryan and Dinan 2012; Martin, Miquel et al. 2013) and are summarised in Table 1.1.

Table 1.1: Main roles attributed to commensal bacteria

Function	Action
Digestion/Nutrition of host	Fermentation of dietary components
	Production of essential nutrients (short chain fatty acids (SCFA), vitamins)
Pathogen inhibition	Occupation of ecological niche
	Luminal pH decrease
	Bacteriocins secretion
	Stimulation of defensins production
	Acetate production
	pH, proteolysis, binding
Improve intestinal barrier	Growth of intestinal cells via SCFA production
	Mucosal blood flow enhancement
	Up-regulation of mucin genes
	Stimulation of mucus secretion
Development/maintenance of immune response	Stimulation of secretory IgA (SIgA)
	Macrophage function enhancement
	T cell modulation
Brain development and behaviour	Microbial (neuro)metabolites

Adapted from (Lebeer, Vanderleyden et al. 2010; Bravo, Forsythe et al. 2011; Cryan and Dinan 2012; Martin, Miquel et al. 2013)

The first function of commensal bacteria corresponds to a trophic role for both the host and the bacterial communities in the GIT (Lupp and Finlay 2005; Mueller and Macpherson 2006; Koropatkin, Cameron et al. 2012). The vast repertoire of bacterial enzymes enables commensals to metabolise macromolecules from the host diet that are otherwise indigestible by the host. Oligosaccharides from milk and particularly those found in human breast milk, human milk oligosaccharides (HMOs), represent complex structures that are resistant to human digestive enzymes. It is estimated that HMOs consist of approximately

200 oligosaccharides, including sialylated but especially rich in fucosylated structures (LoCascio, Ninonuevo et al. 2007). Recent reports showed that early colonisers of the infant's gut such as bifidobacterial species can grow on HMOs as the sole carbon source (Ward, Ninonuevo et al. 2006; Kitaoka 2012).

Several studies have also reported the characterisation of glycosylhydrolases (GHs) including sialidases (Sela, Li et al. 2011), galactosidases (Yoshida, Sakurama et al. 2012) and fucosidases (Sela, Garrido et al. 2012) active on milk oligosaccharides in *Bifidobacterium longum* subsp. *infantis.* Prominent symbionts such as *Bacteroidetes thetaiotaomicron* display a remarkable adaptation to the human GIT through their capacity to degrade a variety of macromolecules including HMOs, plant polysaccharides, as well as host mucus glycans (Backhed, Ley et al. 2005; Sonnenburg, Xu et al. 2005). More recently, *Akkermansia muciniphila* was isolated from the human intestine. The strain is highly abundant in healthy mucosa and is able to grow on mucin as the sole source of carbon and nitrogen (van Passel, Kant et al. 2011), thus holds a selective advantage over other microbes (Belzer and de Vos 2012). These nutrients originating from the host and their diet are metabolised into short chain fatty acids (SCFAs), mainly butyrate, propionate and acetate that generate hydrogen and carbon dioxyde (Wang, Zhu et al. 2004). SCFAs are absorbed by the host intestinal epithelial cells (IECs). It has been reported also that butyrate, produced by species such as *F. prausnitzii*, constitutes the main energy source of colonocytes (Duncan, Hold et al. 2002; Wong, de Souza et al. 2006). It has also been shown that gut commensals including bifidobacteria are able to synthesize vitamin K as well as most of the B-group vitamins (Mueller and Macpherson 2006; LeBlanc, Milani et al. 2013)

As a consequence of the establishment of commensals in the GIT, the microbiota protects against the colonisation of opportunistic pathogens by provoking micro-niche exclusion and also through the secretion of anti-microbial substances such as bacteriocins (Van der Waaij, Berghuis-de Vries et al. 1971; Vollaard and Clasener 1994; Reiff and Kelly 2010). Bifidobacteria including *Bifidobacterium longum* and *B. infantis* were shown recently to prevent infection by entero-haemorrhagic *Escherichia coli* through the production of acetate (Fukuda, Toh et al. 2011). Intestinal cells such as Paneth cells secrete antimicrobial

peptides (enteric defensins), which in turn limit bacterial overgrowth (Mukherjee, Vaishnava et al. 2008; Jarczak, Kościuczuk et al. 2013).

By interacting with the mucosa, commensal bacteria improve the intestinal barrier function by enhancing tight junction formation (Sultana, McBain et al. 2013). Species such as *Bacteroides thetaiotaomicron* trigger signalling pathways in the host that promote angiogenesis and the development of epithelial cells (Rakoff-Nahoum, Paglino et al. 2004; Rhee, Sethupathi et al. 2004), thus contributing to the architecture of the gut (Hooper, Midtvedt et al. 2002). Various studies have also shown the upregulation of mucin genes in the host by the presence of commensal bacteria (Mattar, Teitelbaum et al. 2002; Otte and Podolsky 2004). Caballero et al. also reported that the MUC2 gene was highly upregulated in the presence of lactobacilli (Caballero-Franco, Keller et al. 2007). MUC2 codes for the main gel-forming mucin present in the human intestine and where commensal species reside.

Several studies mentioned earlier have demonstrated the involvement of commensal bacteria in the development of the immune system in infants. There is also clear evidence that disruption of the balance between these well adapted microbes and the host results in an aberrant immune response in the host and several related inflammatory disorders. The role of commensals in the maintenance of the immune system is developed in section 1.3 in details.

Furthermore, the concept of microbiota-gut-brain-axis has recently emerged with a bi-directional cross-talk between the microbiota and the nervous system (Collins and Bercik 2009; Cryan and Dinan 2012). Studies on murine models showed that the microbiota influences the brain development (Diaz Heijtz, Wang et al. 2011; Hsiao, McBride et al. 2013) and that gut symbionts such as some lactobacilli strains exert beneficial effects on reducing stress, pain and regulating emotional behaviour through their action on the enteric neural cells and, consequently, to the brain (Rousseaux, Thuru et al. 2007; Bravo, Forsythe et al. 2011). Reports also suggest the presence of an altered microbiota with a decrease of bifidobacteria in children suffering from autistic related diseases (Parracho, Bingham et al. 2005; De Angelis, Piccolo et al. 2013).

1.2 Commensals and the immune system

The idea of a communication system or cross-talk between the host and the microbiota has gained support after a bacteria-derived molecule, polysaccharide A (PSA), isolated from the prominent mammal commensal pathobiont *Bacteroides fragilis*, was shown to be essential to the maturation of the murine host immune system (Mazmanian, Round et al. 2008). Recently, Bjerke et al. showed a direct transmission from mother to infant of *B. fragilis* (Bjerke, Wilson et al. 2011).Whether discoveries of such effector molecules in mice are translated in the case of humans has yet to be shown. The highly mutualistic interplay that takes place between the host immune system and the microbiota has been reviewed recently and is summarised in Figure 1.2 (Ventura, Turroni et al. 2012; Cahenzli, Balmer et al. 2013).

Commensals play a major role in the differentiation and maturation of the immune cells (Wells, Rossi et al. 2011). These interactions during early life have been studied *in vitro* and *in vivo* by comparing germ free (GF) and colonised mice to mimic a human baby acquiring a bacterial population (Corthésy, Gaskins et al. 2007). GF mice possess a poorly developed immune system structurally and functionally. The Peyer's patches and gut-associated lymphoid tissues are underdeveloped with few germinal centers and the level of circulating lymphocytes is reduced, with very few IgA-secreting plasma cells (Szeri, Anderlik et al. 1976; Umesaki, Setoyama et al. 1993). Histological and molecular biological observations of GF rodents, as they become colonised, show that the presence of the microbiota causes extensive adaptation of the host (Macpherson, Geuking et al. 2012). The introduction of commensal bacteria induces the development of lymphoid tissues and leads to an increase in the number of CD4+ naive T lymphocytes.

Innate and acquired immune systems have to tolerate the presence of the microbes that colonise the gut. The mutualistic relationship between the host and some commensal microbes dictates that the bacteria in the luminal space or within the mucus layer should be tolerated. However, some microbes are able to penetrate the epithelial barrier and translocate inside the host (Berg 1999). These need to be rapidly eliminated by the mucosal immune system. Dendritic cells (DCs), through their protrusions present at the top of the microvilli of the IECs or through specialised 'micro fold cells' above lymphoid follicles (M cells), get conditioned by commensals to activate the immune system. Toll-like receptors

(TLRs) present on these specialised cells play a key role in bacterial recognition and the innate immune response. A selective local induction of the mucosal immune response occurs through (i) the stimulation of secretory immunoglobulin A production against the microbe (SIgA) (Brandtzaeg 2013) and (ii) SIgA antibody-coated commensals get excluded from penetrating the intestinal epithelium, thus modulating the bacterial densities (He, Xu et al. 2007; Macpherson and Slack 2007).

However, small numbers of commensals are allowed to penetrate through the epithelial barrier into specialised inductive sites known as Peyer's patches where they are picked up by DCs or phagocytosed by macrophages. Microbes, through antigen-presenting cells (APCs) such as DCs, induce the differentiation of naïve CD4+ T cells into various specialised lymphocytes in the intestinal lamina propria. These include interferon-γ-producing T-helper cells (Th1), interleukin-4-producing T-helper cells (Th2), interleukin-17-producing T-helper cells (Th17) or regulatory T-cells (T*reg*) hence inducing the adaptive response (Feng and Elson 2011). A flexible continuum between innate and acquired immune function thus results in the containment of commensal microbes (Nagler-Anderson 2001).

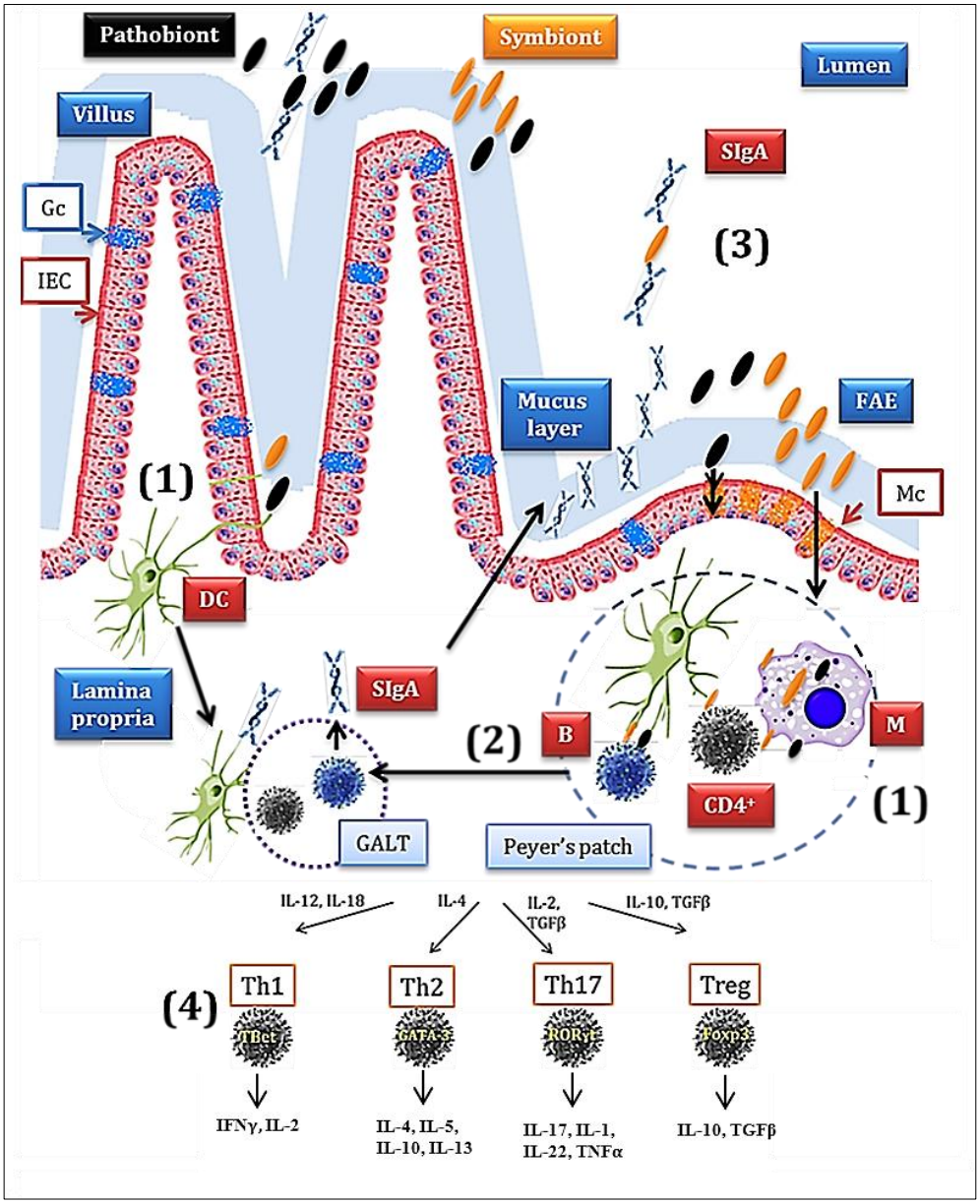

Figure 1.2: Cross-talk between the microbiota and the host immune system

Abbreviations are as follows: GALT: Gut-associated lymphoid tissue; FAE: Follicle-associated epithelium; Gc: Mucus secretion cell (Goblet cell); IEC: Enterocyte; Mc: Microfold cell; DC: Dendritic cell; B: Lymphocyte B; M: Macrophage; $CD4^+$: Naive T lymphocyte; SIgA: Secretory immunoglobulin A

(1) Detection of microbe by antigen presentation cells and activation of the innate immune system (e.g phagocytosis);

(2) Recruitment of lymphocytes and activation of the acquired immunity;

(3) Production of SIgA by B lymphocytes and secretion to the intestinal lumen;

(4) Differentiation of naïve CD4+ T cells into various specialised T helper lymphocytes producing chemokines and cytokines.

Adapted from (Wells, Rossi et al. 2011; Ventura, Turroni et al. 2012; Cahenzli, Balmer et al. 2013).

1.3 The glycobiome

Both bacteria and the gut are coated by glycans that play a critical role in the interaction between the host and the commensal microbes. The genes involved in the assembly as well as in the metabolism of dietary and host-derived glycans are referred to as the glycobiome (Zivkovic, German et al. 2010; Ouwerkerk, de Vos et al. 2013)

1.3.1 Mammalian glycosylation

Glycosylation is the process that involves the covalent attachment of glycans to another molecule such as a lipid but particularly to an amino-acid on a newly synthesised peptide. It takes place in the endoplasmic reticulum (*N*-glycosylation) and/or the Golgi apparatus (*O*-glycosylation) in mammalian cells. The biosynthesis of glycoconjugates is based and involves various conserved glyco-related enzymes, glycosylhydrolases (GHs) and glycosyltransferases (GTs), which use sugar nucleotides as donors (Ohtsubo and Marth 2006; Paulson, Blixt et al. 2006).

The linkages between monosaccharides on the oligosaccharide chain and the branching enable an enormous variety of three-dimensional structures. Four types of glycans are commonly encountered in mammals (Figure 1.3). *N*-linked glycans are covalently bound to the carboxamido Nitrogen on asparagine (Asn) residues. *N*-glycosylation is the prevalent form of post-translational modifications (PTMs) and often involves the attachment of a precursor glycan to the peptidic chain as it is being translated and transported into the endoplasmic reticulum (ER). *N*-glycosylation is initially similar for all proteins but through the action of GHs and GTs, glycans undergo various steps of trimming and maturation. An initial trimming operates in the ER, then the glycoprotein is trafficked to the Golgi apparatus, where a mannosidase (Man-I) removes multiple mannose residues. Glycans that do not go through further steps are called high-mannose oligosaccharides. The further addition of monosaccharides by GTs leads to complex structures with fucose and sialic acid branching. *N*-glycans can thus be classified as high-mannose, complex or hybrid oligosaccharides (Paulson, Blixt et al. 2006).

Most cell surface and extracellular proteins are *O*-linked glycosylated. This second type of glycosylation is essential in the biosynthesis of mucins and extracellular matrix components. *O*-linked glycosylation occurs post-translationally on serine (Ser) and

threonine (Thr) in the Golgi apparatus (Peter-Katalinic 2005). It mostly involves *N*-acetylgalactosaminyltransferase, which adds a single GalNAc residue by α-glycosidic linkage to the β-OH group of Ser or Thr. The resulting structure is called the Tn antigen. Other monosaccharides can also be involved (e.g. β-linked *O*-GlcNAc to Ser/Thr). Eight types of *O*-linked glycan core structures have been described (cores 1 to 8), with cores 1 to 4 especially found in the GIT (Juge 2012), the most common being Gal-β-(1→3)-GalNAc-Ser/Thr (core 1). More structures can be found based on the elongation of the Tn antigen (detailed in section 1.3.2, Figure 1.4).

Glycolipids correspond to membrane molecules where a glycan (chain) is covalently attached to a lipid backbone. One of the predominant groups, the glycosphingolipid family, is essential in cell-cell communication, cell adhesion and has been associated with several diseases including microbial infections (Hakomori and Igarashi 1995). Glycosphingolipids are formed by a glycan linked to a ceramide (Cer), composed of a long-chain amino alcohol (sphingosine) and an amide-linked fatty acid. The elongation and branching of the glycan chain is carried out in a step-wise manner in the ER and Golgi apparatus. Glycosphingolipids display variations in the acyl as well as in the sphingosine moieties but their sub-classification is based on the structure of the carbohydrate residue. Three major groups are distinguished: the ganglio- (GalNAc-β-(1→4)-Gal-β-(1→4)-Glc-β-Cer), the lacto- (GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-β-Cer) and neolacto-(Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-β-Cer), and the globo-series (Gal-α-(1→4)-Gal-β-(1→4)-Glc-β-Cer).

Proteoglycans are heavily glycosylated proteins consisting of a core protein and at least one covalently attached glycosaminoglycan chain (GAG), which are unbranched highly anionic polysaccharides. GAGs correspond to repeating units of a disaccharide composed of an *N*-acetylated or *N*-sulfated hexosamine and either uronic acid or galactose. They include hyaluran (HA), keratan sulfate (KS), chondroitin sulfate (CS) and heparan sulfate (HS). Some studies have linked GAGs to inflammatory and auto-immune diseases (Wang and Roehrl 2002).

Finally, glycosylphosphatidylinositol (GPI) anchors form a distinct family of glycolipids produced in the ER that covalently links proteins or polysaccharides to cell membranes including blood cell surfaces (Ikezawa 2002).

Once synthesised, the various glycoconjugates are transported to different organelles in the cell, to the cell envelope, or secreted out. The process results in the coating of all cells by an array of glycans, which are essential for cell viability. The dysregulation of the glycosylation process affects greatly the health of the cells and the organism, leading to severe dysfunctions. In particular, protein glycosylation exerts crucial roles in most physiological processes including cell signalling, cellular differentiation and adhesion Glycans modulate protein activity at the cell-extracellular interface and could be seen as a 'biochemical braille' allowing cells to recognise and respond to one another.(Holst, Moran et al. 2010).

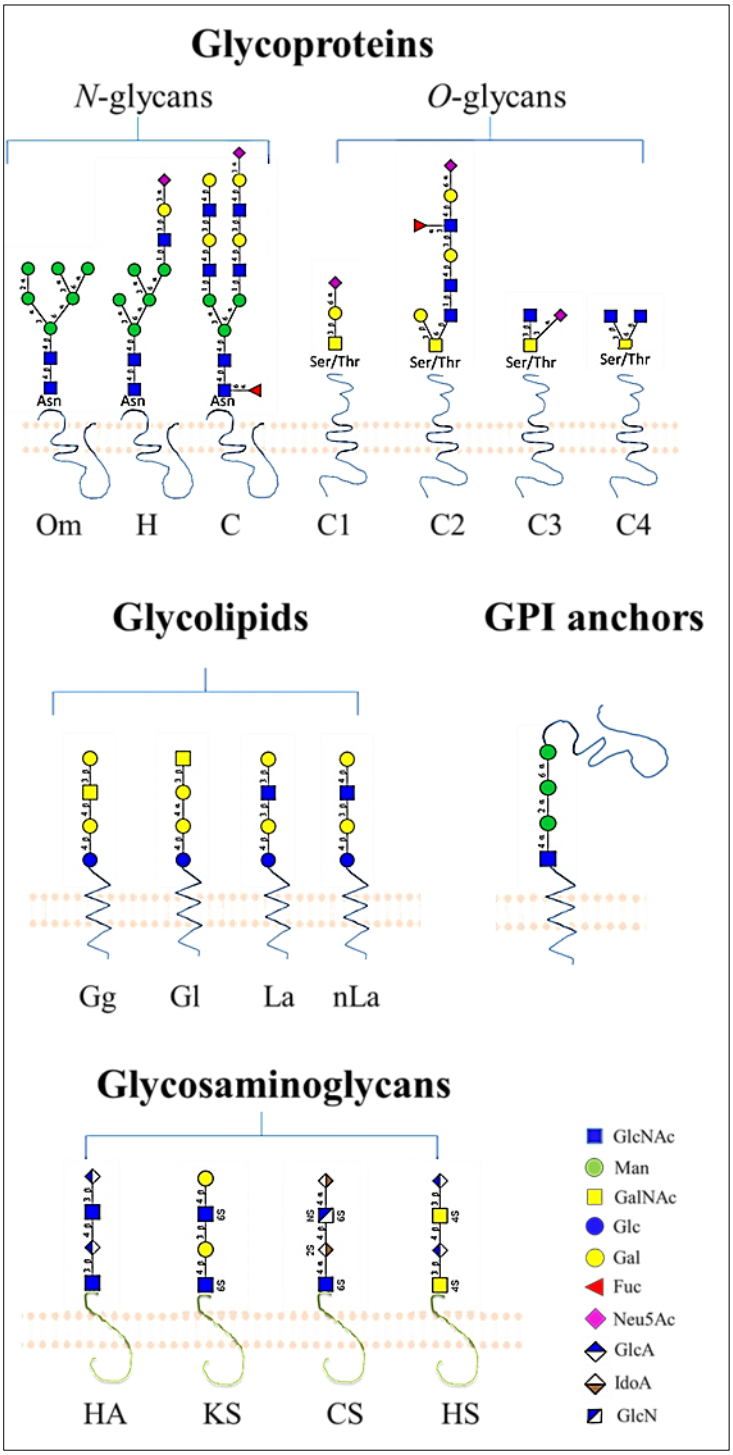

Figure 1.3: Mammalian glycosylation

Classes of glycans encountered in mammalian cells: glycoproteins, glycolipids, glycosaminoglycans and GPI anchors, featuring *N*-glycans (with Om: oligomannose; H: hybrid and C: complex-type) and *O*-glycans (with C1: core 1, C2: core 2, C3: core 3 and C4: core 4); glyco(sphingo)lipids with Gg: ganglio-type, Gl:globo-type , La: lacto-type and nLa: neo-lacto-type; glycosaminoglycans with HA: hyaluronan, KS, keratin, CS: chondroitin sulphate and HS: heparin sulphate) and GPI anchors. Monosaccharides are abbreviated as follows: Gal: galactose; Glc: glucose; GlcA: glucuronic acids;IdoA: iduronic acid, GlcN: glucosamine; GlcNAc: *N*-acetyl glucosamine; GalNAc: *N*-acetyl galactosamine; Fuc: fucose; Neu5Ac: sialic acid. Adapted from Dalziel et al. (Dalziel, Crispin et al. 2014)

1.3.2 Glycosylation and gut mucosal barrier

The mucosal barrier in the gut is composed of a layer of enterocytes, which like all cells is covered by a glycocalyx, a complex sugar coat, comprising the oligosaccharides of membrane glycolipids, glycoproteins and extracellular proteoglycans. As mentioned previously, the gut epithelium is covered by a layer of mucus, which varies in thickness along the length of the gut. In the colon, it comprises an inner almost sterile layer on the enterocytes membrane, beneath a thicker gel which accommodates the microbiota. No inner layer is found in the mucus layer of the small intestine but bacteria are contained to the lumen thanks to the secretion of antimicrobial peptides and SIgA by the host (Macpherson and Slack 2007; Vaishnava, Yamamoto et al. 2011; Macpherson, Geuking et al. 2012).

Mucus, rich in water and mucins, protect the epithelial cells against enteric pathogens by inhibiting direct contact. It is constitutively being produced and replaced since it is lost due to the intestinal clearance under the peristaltic movements in the GIT. Several types of mucin genes (20) are expressed throughout the GIT (see Figure 1.1). The resulting proteins that take the shape of long hairy filaments, share common tandem repeated sequence domains (PTS), rich in proline, serine and threonine, which get extensively glycosylated (*O*-linked glycosylation) (Brockhausen I 1999). These *O*-linked glycans can account to 80% of their molecular mass and can be neutral, sialylated or sulfated. Up to eight core glycan structures have been described, of which cores 1 to 4 are the most common throughout the GIT (Juge 2012). These cores can be further elongated by *N*-acetyllactosamine (LacNAc) units and terminated by histo-blood group (HBG) antigens such as A, B, H, Lewis a (Le^a), Lewis b (Le^b), Lewis x (Le^x), Lewis y (Le^y) and sialylated Lewis structures (Figure 1.4) (Carrington, Clyne et al. 2010).

The glycosylation of mucins not only varies throughout the GIT but it is also influenced by the commensal microbes. The mucus layer is reduced in the gut of GF mice due to a lower number of goblet cells in comparison to conventionally raised mice. Several studies have reported on the influence of the gut microbiota in enhancing the production of mucins such as MUC2. (Uribe, Alam et al. 1994; Belley, Keller et al. 1999; McGuckin, Linden et al. 2011). In addition, commensals such as *Akkermansia sp.* modify directly the mucins' glycans by the action of GHs (van Passel, Kant et al. 2011).

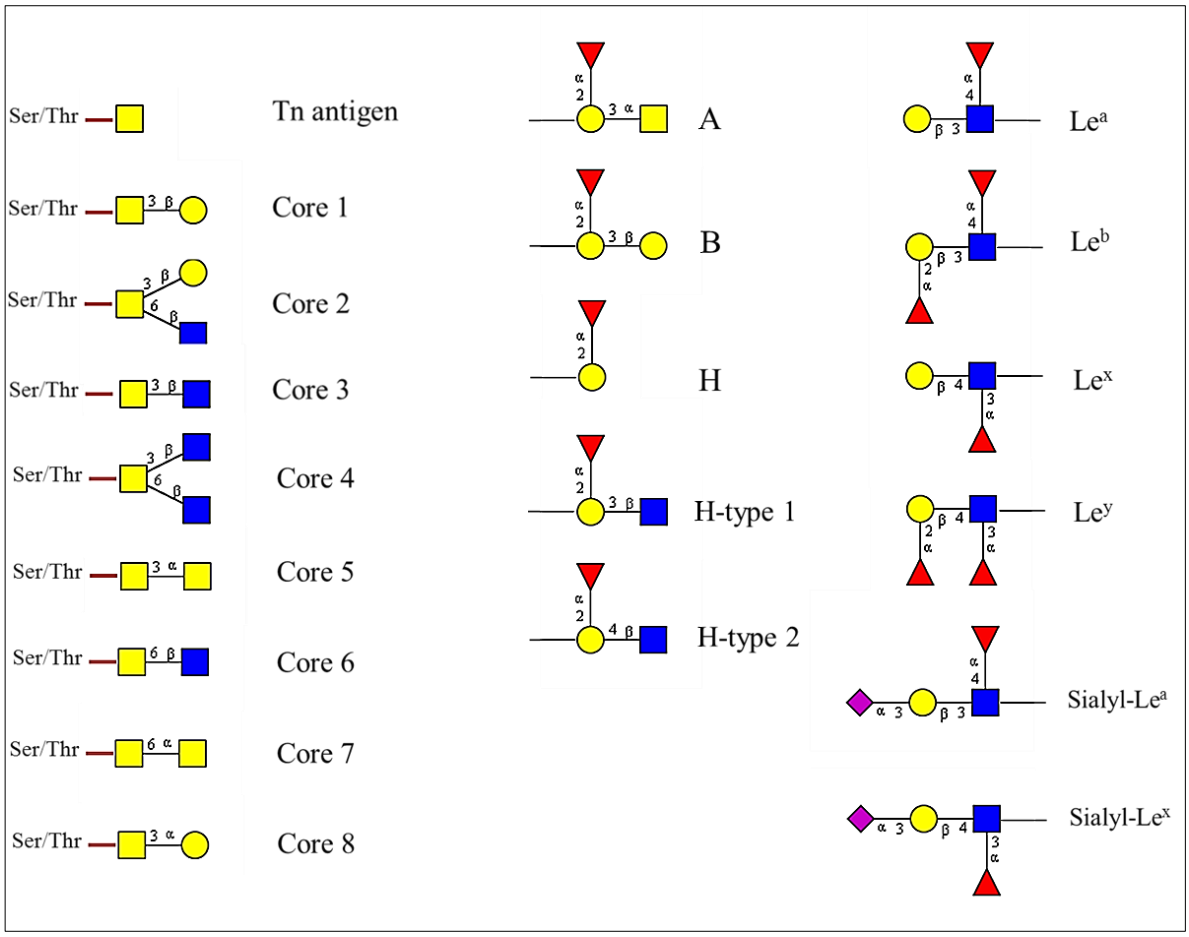

Figure 1.4: Mucin-type *O*-linked glycans

Glycan cores and histo-blood group (HBG) antigens found in mucins: blood groups A, B, H, Lewis a (Le^a), Lewis b (Le^b), Lewis x (Le^x), Lewis y (Le^y), sialylated Lewis a (Sialyl-Le^a) and sialylated Lewis x (Sialyl-Le^x). Yellow square: GalNAc, yellow circle: Gal, blue square: GlcNAc, red triangle: Fuc and purple losange: Neu5Ac.

It is accepted that that the glycoproteins or glycolipids present on the mucosa represent key receptors for bacterial lectins (Kline, Falker et al. 2009). Studies on the composition of the microbiota between twins revealed a higher similarity between monozygotic rather than dizygotic twins, suggesting the involvement of host genetic factors in deciding gut microbial interaction (Tschop, Hugenholtz et al. 2009; Turnbaugh, Hamady et al. 2009). Moreover, the secretor status, defined by the expression of the A, B, H and Lewis HBG antigens on erythrocytes as well as on mucosal epithelium and salivary secretion, affects the mucosal barrier (Linden, Mahdavi et al. 2008; Wacklin, Makivuokko et al. 2011). The secretor phenotype expresses the fucosyltransferase (FUT2) producing the H antigens (Fuc-α-(1→2)-Gal-β-(1→3/4)-GlcNAc) and results in the assembly of type A, B or Lewis-b/y glycans on intestinal mucins. A mutation in the encoding FUT2 gene leads to the non-secretor phenotype (Henry, Oriol et al. 1995). In a recent study, the presence or absence of

a functional FUT2 allele in a Finnish cohort correlated strongly with the composition of the microbiota, with non-secretor individuals rarely colonised by bifidobacteria and more prone to infections or diseases, including Crohn's disease (CD). These studies suggest HBG antigens to represent host genetic determinants that could play a role in shaping the microbial composition (Wacklin, Makivuokko et al. 2011).

1.3.3 Glycosylation of commensal cell surface

Bacterial polysaccharides are carbohydrate polymers joined together by glycosidic bonds, which can be neutral or charged. They differ in the nature of sugar residues, the linkages, the branching and the presence of substituents such as glycerol, acyl, phosphate or sulfate groups. They also differ according to the types of bacterial cell wall. Two types are defined according to the Gram staining, which corresponds to differences in structures (Figure 1.5). A thick peptidoglycan layer is present around the membrane of Gram positive bacteria forming the cell wall. Gram-negative bacteria possess two lipopolysaccharidic membranes, the outer membrane (OM) being surrounded by a thinner coat of peptidoglycan (PG). The cell envelope components form the interface between the microbe and its environment and are likely to be involved in 'cross-talk' with the host. Lipopolysaccharides (LPSs) are the major components of the OM of Gram-negative microbes. Cell surface polysaccharides include polymers covalently attached to the *N*-acetylmuramic acid (MurNAc) of the peptidoglycan (PG) strands. These are referred to as capsular polysaccharides (CPSs), which form a thick shell called capsule or pellicle covering the bacterial cell. Bacteria, and essentially Gram positive species, may release polysaccharides into the extracellular environment. These exopolysaccharides (EPSs) may be loosely attached by electrostatic bonds to the cell envelope (bEPS) or secreted (rEPS). EPS have mainly been studied in the context of pathogenic biofilms responsible for diseases. Many lactic acid bacteria (LAB) also produce EPSs, but the majority of the research relates to dairy strains, which are of industrial importance for the food industry. Physical properties of EPSs depend on the chemical nature but also on the three-dimensional structure of the molecule. The presence of substitutive groups and the branching of the linear chain affect the conformation, the stability and the visco-elasticity or the capacity of the EPS to form gels. A study from 2007 also reported that 17% of human commensal strains from the genera *Lactobacillus* and *Bifidobacterium* are EPS producers (Ruas-Madiedo, Moreno et al. 2007).

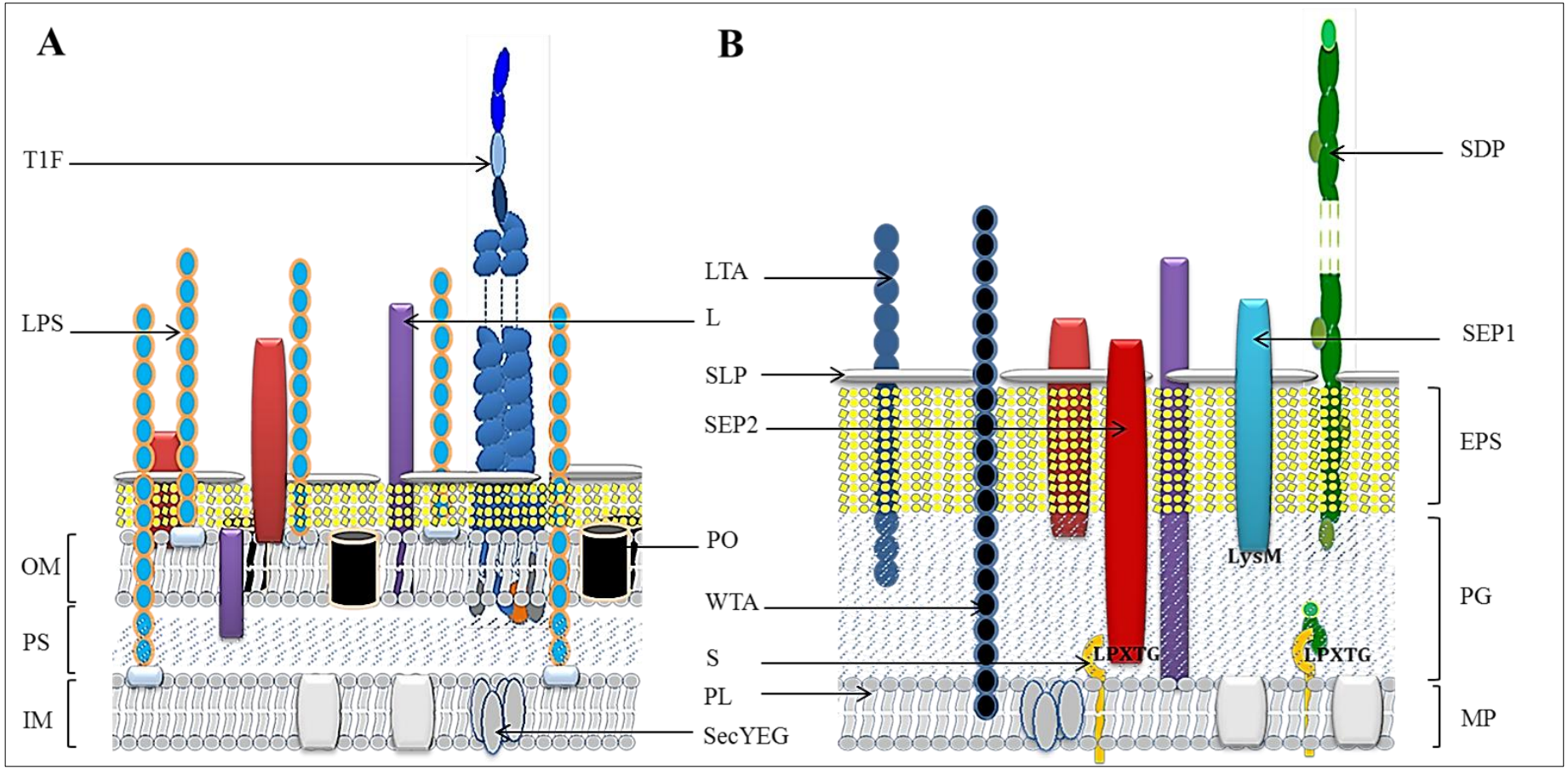

Figure 1.5: Cell wall in Gram-negative (A) and Gram-positive (B) bacteria

Abbreviations are as follows: type-1 fimbria (T1F); lipopolysaccharide (LPS); inner membrane (IM); periplasmic space (PS); outer membrane (OM); sortase-dependent pili (SDP); lipoteichoic acid (LTA); lipoprotein (L); surface layer protein (SLP); non- covalently attached surface-exposed protein (SEP1); covalently attached surface-exposed protein (SEP2); porin (PO); wall teichoic acid (WTA); sortase (S); phospholipid (PL); translocon (SecYEG); peptidoglycaan (PG); membrane (MP); exopolysaccharide (EPS). Adapted from various reviews (Desvaux, Dumas et al. 2006; Scott and Zahner 2006; Hendrickx, Budzik et al. 2011)

Several functions have been attributed to commensal EPSs in the self-interest of the bacterial cell. These include a trophic role as fermentable substrates, cell aggregation, biofilm formation, niche colonisation, adherence to mucus, immune evasion and protection against stress, desiccation or phagocytosis (Jost Wingender 1999; Lebeer, Vanderleyden et al. 2010). Other studies have linked them to being receptors for bacteriophages and conversely as a protection against phages or from antimicrobial agents (Lindberg 1977). Some EPSs are known to provide health benefits to their mammalian host (Figure 1.6), hence the increasing interest in isolating EPS and investigating these effects further (Ruas-Madiedo, Salazar et al. 2010). Some EPSs were reported to affect the immune system of the host by promoting the production of cytokines, the proliferation of lymphocytes and activating macrophages. EPSs have also been shown to serve as prebiotics to promote the growth of other beneficial microbes from the microbiota such as bifidobacteria (Korakli, Gänzle et al. 2002). Some studies have also attributed a direct involvement of EPSs in lowering the level of blood cholesterol, to exert anti-ulcer action and an anti-tumorigenic effect on malignant cells. (Lebeer, Ceuppens et al. 2007; Lebeer, Vanderleyden et al. 2010; Ruas-Madiedo, Salazar et al. 2010; Şengül, Işık et al. 2011).

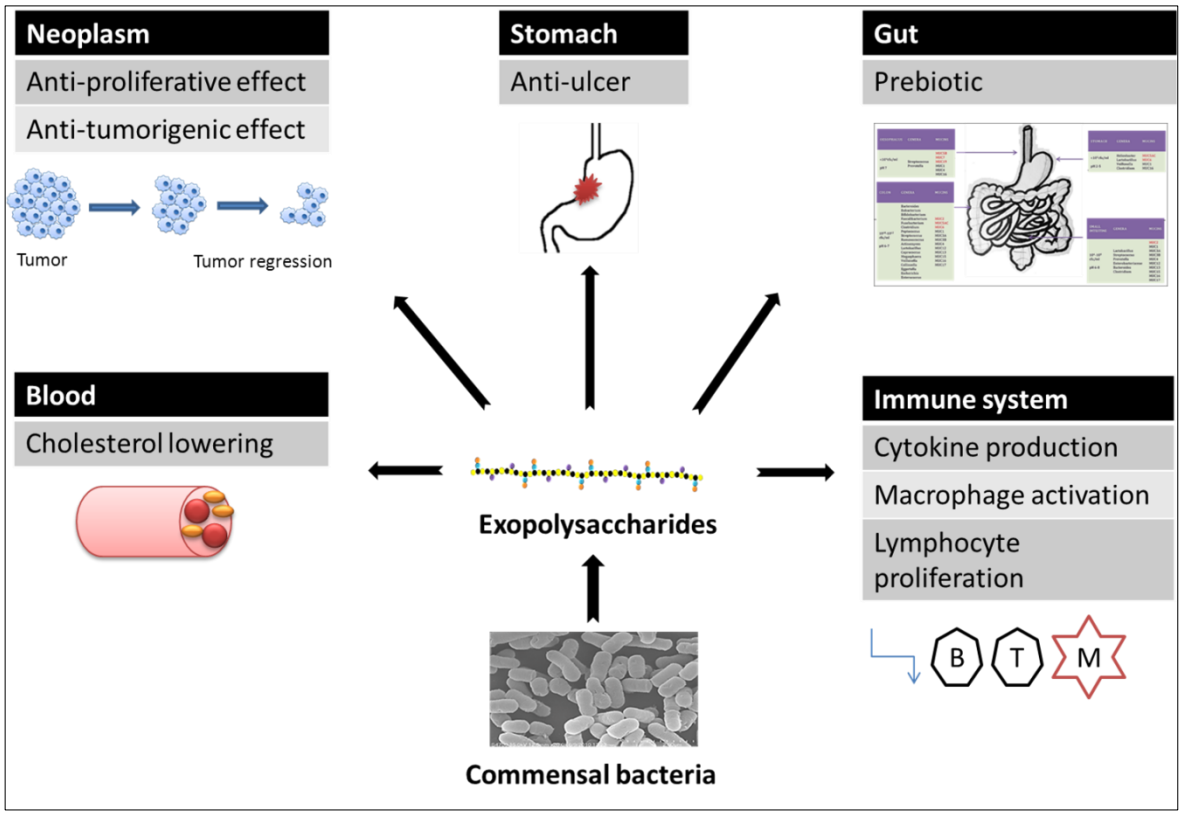

Figure 1.6: Health effects of EPSs from gut commensal bacteria

Adapted from Ruas-Madiedo et al. (Ruas-Madiedo, Salazar et al. 2010)

Synthesis of bacterial EPSs share common steps with PG and lipoteichoic acid (LTA) pathways and their complexity is increased by the presence of multiple strain-specific gene clusters (De Vuyst and Degeest 1999). EPSs are synthesised during the stationary phase of bacterial growth and vary greatly in their chemical structure, residue composition and molecular mass. They can be divided into two classes. Homopolysaccharides have the same sugar residues, such as glucose in the case of glucans, and can be of high molecular mass (10^7 Da). They are synthesised by extracellular transglycosylases (glycansucrases) from sucrose (Monsan, Bozonnet et al. 2001). Heteropolysaccharides are usually smaller (up to 10^6 Da) and generally consist of repeating units of two to nine different bacterial monosaccharides, usually galactose (Gal), glucose (Glc), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), rhamnose (Rha), mannose (Man) and glucuronic acid (GlcA). Biosynthesis of heteropolysaccharides is more complex and not fully understood. It takes place, initially, in the cytoplasm where simple sugars are converted to nucleotide sugars providing the energy for the polymerisation. Subsequently, the monosaccharides repeating units are assembled by GTs on the internal side of the cell membrane and attached to a lipid carrier. The process then involves a flippase to transfer the polymerised repeating units to the outer surface of the membrane (De Vuyst and Degeest 1999; Laws, Gu et al. 2001). As an example, the biosynthesis of a long Gal-rich EPS from *Lactobacillus rhamnosus* GG is illustrated in Figure 1.7. Initially, the priming glycosyltransferase (WelE) catalyses the transfer of a phosphogalactosyl residue from an activated nucleotide sugar to the undecaprenyl phosphate (UndP)-lipid carrier on the cytoplasmic side of the membrane. Subsequently, five glycosyltransferases (WelF to –J) add the remaining monosaccharides composing the EPS subunit, which is translocated across the cytoplasmic membrane by a flippase (Wzx). A polymerase (Wzy) then links the repeating units of the EPS together. The biosynthesis also involves a phosphorylation complex (Wze, Wzb) with a possible regulation function (Lebeer, Verhoeven et al. 2009).

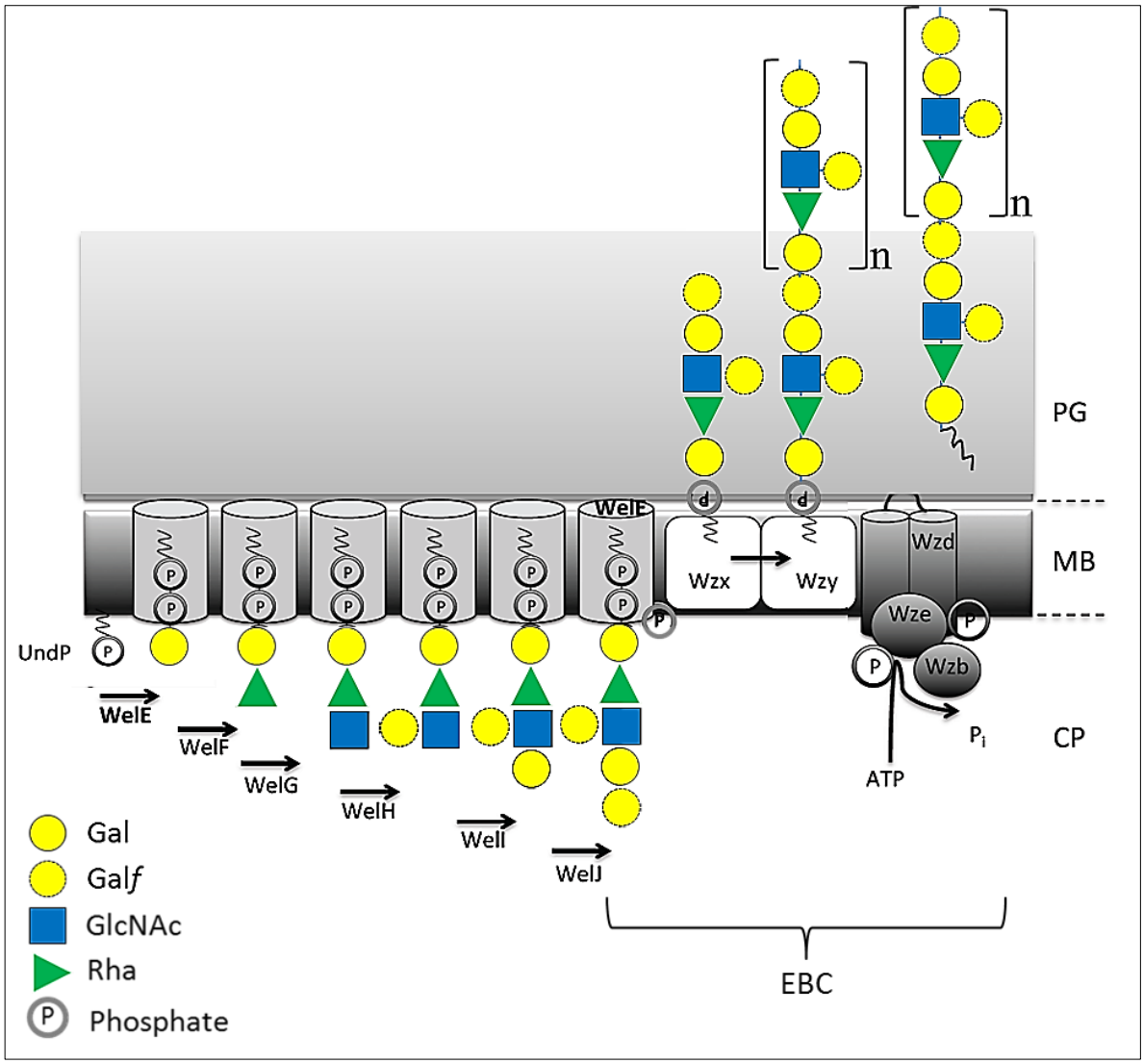

Figure 1.7: Biosynthesis of the long-galactose rich EPS from *L. rhamnosus* GG

Abbreviatons are as follows, PG: peptidoglycan layer; MB: membrane; CP: cytoplasm; EBC: EPS biosynthesis complex; UndP: undecaprenyl phosphate lipid carrier, WelE: priming glycosyltransferase; WelF, WelG, WelH, WelI, WelJ: glycosyltransferases; Gal*f*: galactofuranose; (adapted from Lebeer et al. and Landersjo et al. (Landersjo, Yang et al. 2002; Lebeer, Verhoeven et al. 2009)).

Very few commensal EPSs have been characterised. As of June 2014, 26 EPSs structures isolated from respectively 7 and 11 strains of bifidobacteria and lactobacilli, could be compiled from the Bacterial Carbohydrate Structure Data Base (Table 1.2). The genes involved in the polysaccharides syntheses are mostly found clustered together in genomes and often enclosed in plasmids. Therefore, strategies have been recently developed to identify the EPS clusters *in silico*, then engineer glycosyltransferases knockouts and compare the resulting structures. Another approach is to use glyco-enzymes to alter the EPS polymer and thus create mutants (Ruas-Madiedo and de los Reyes-Gavilán 2005).

Table 1.2: Characterised commensal EPSs

Commensal strain	Glc	Gal	Rha	Others / Comments	Ref
Bifidobacterium adolescentis YIT 4011	3			6 Tal	(Nagaoka, Muto et al. 1988)
Bifidobacterium animalis ssp. lactis IPLA-R1	1	2	3		(Leivers, Hidalgo-Cantabrana et al. 2011)
Bifidobacterium bifidum BIM B-465	4	3			(Zdorovenko, Kachala et al. 2009)
Bifidobacterium bifidum BIM B-465	4				(Zdorovenko, Kachala et al. 2009)
Bifidobacterium bifidum BIM B-465	3				(Zdorovenko, Kachala et al. 2009)
Bifidobacterium breve YIT 4007	1		2		(Habu, Nagaoka et al. 1987)
Bifidobacterium catenulatum YIT 4016		3			(Nagaoka, Hashimoto et al. 1996)
Bifidobacterium longum subsp. *infantis* ATCC 15697	1	2			(Tone-Shimokawa, Toida et al. 1996)
Bifidobacterium longum YIT 4028		3	2		(Nagaoka, Hashimoto et al. 1996)
Lactobacillus casei YIT 9018	3		6	1 GalNAc, 1 GlcNAc	(Nagaoka, Muto et al. 1990)
Lactobacillus johnsonii 142	1	4			(Gorska, Jachymek et al. 2010)
Lactobacillus johnsonii 151	2	2			(Górska-Frączek, Sandström et al. 2011)
Lactobacillus johnsonii FI9785	4	2			(Dertli, Colquhoun et al. 2013)
Lactobacillus paracasei 34-1		2	2		(Robijn, Wienk et al. 1996)

Commensal strain	Glc	Gal	Rha	Others / Comments	Ref
Lactobacillus paracasei 34-1		3		1 Gro, 1 GlcNAc	(Robijn, Wienk et al. 1996)
Lactobacillus paracasei 34-1		3		1 GlcNAc	(Robijn, Wienk et al. 1996)
Lactobacillus paracasei 34-1		2		1 Gro, 1 GalNAc	(Robijn, Wienk et al. 1996)
Lactobacillus reuteri 180	40			No repeating unit	(van Leeuwen, Kralj et al. 2008)
Lactobacillus reuteri 35-5	35			No repeating unit	(van Leeuwen, Kralj et al. 2008)
Lactobacillus rhamnosus ATCC 53103		4	1	1 GlcNAc	(Landersjo, Yang et al. 2002)
Lactobacillus rhamnosus KL37	2	3			(Lipinski, Jones et al. 2003)
Lactobacillus rhamnosus KL37	1	2			(Gorska, Grycko et al. 2007)
Lactobacillus rhamnosus KL37	2	3			(Lipinski, Jones et al. 2003)
Lactobacillus rhamnosus KL37B	3	6			(Lipinski, Jones et al. 2003)
Lactobacillus rhamnosus RW9595M	2	1	4	1 Pyr	(Van Calsteren, Pau-Roblot et al. 2002)
Lactobacillus rhamnosus RW9595M	1		4	1 Gro	(Van Calsteren, Pau-Roblot et al. 2002)

Composition of the repeating unit of EPSs as compiled in the Bacterial Carbohydrate Structure DataBase (http://csdb.glycoscience.ru/bacterial/). Abbreviations are as follows: talose (Tal), glycerol (Gro), pyruvate (Pyr).

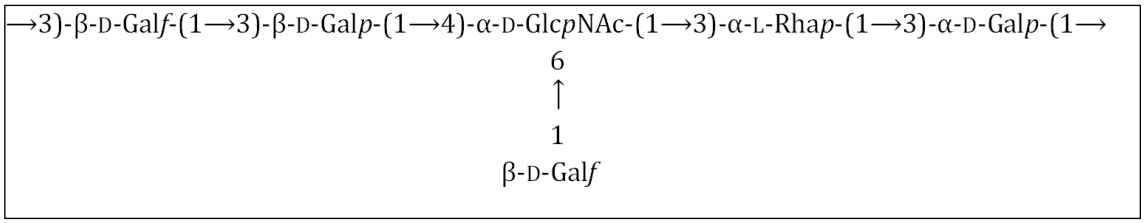

Figure 1.8: Structure of a long Gal-rich EPS from *Lactobacillus rhamnosus* GG (LGG)

Structure of the repeating unit of a long Gal rich neutral EPS from LGG (Landersjo, Yang et al. 2002)

A long Gal-rich EPS from *Lactobacillus rhamnosus* 'GG' strain ATCC 53103 (LGG) was isolated and structurally characterised recently (Figure 1.8) (Landersjo, Yang et al. 2002). This EPS was not produced by the non-adherent dairy strain *Lactobacillus rhamnosus* LC-705, suggesting the long Gal-rich EPS is involved in the adhesion to the GIT. Following sequencing of the LGG genome, Lebeer et al. identified the unique eps cluster corresponding to the long Gal-rich polymer (Lebeer, Verhoeven et al. 2009). However, they found that a knockout mutant for this operon resulted in an increased attachment to mucus and epithelial cells *in vitro.* They suggested that the EPS might play a role in masking the cell surface adhesins that enable bacteria to dock to the gut mucosa. It was proposed that EPS corresponds to an adaptation factor for the LGG strain that promotes its persistence in the GIT and serves as a shield against host innate immunity (Lebeer, Vanderleyden et al. 2010). Similarly, a cell surface EPS from *Bifidobacterium breve* UCC 2003 was recently identified (Fanning, Hall et al. 2012). By comparison to a knockout mutant for the 'eps' operon, it was demonstrated that the wild type strain (WT) was able to evade the host adaptive immune response and increase the *in vivo* persistence in the GIT of mice. The WT was also more tolerable to bile challenge and responsible for reduced infection from a gut pathogen (*Citrobacter rodentium*).

Little is known about structure- function relationships due to the physico-chemical complexity of the polymers. Currently, analysing EPS structure is a lengthy process. It requires repetitive extractions, for which various procedures are used. Isolation can be accomplished chemically or physically from the cells, the media or the whole culture. Several types of media and carbon sources used to grow the micro-organisms influence EPS composition (Ruas-Madiedo and de los Reyes-Gavilán 2005). Various purifications steps and chemical assays are then necessary to determine the EPS composition and structure, which can be very complex in some cases. Indeed, several analytical techniques are necessary to elucidate all the parameters involved in the EPS characterisation. These

include the identity of monosaccharides, the types of carbon rings (pyranose or furanose), their absolute configuration, the stereo-specificity (D or L), the linkages types (α or β), the sequence in the repeating unit, the presence and the position of the substituants, the molecular weight of the polymer. Newer high-throughput microarray methods are now being used for preliminary glycosylation and glycan interaction analysis. Lectin microarrays, for instance, represent interesting tools to rapidly screen bacterial strains or EPSs before a refined analysis. These methods will be described in this thesis.

1.4 The bacterial lectome

Adhesion of bacteria is an important step in their survival and colonisation, whether it is to their host organism, in the case of pathogenic or commensal species that live on/in a living host, or to an inanimate surface in the case of environmental strains (Palmer, Flint et al. 2007). Bacteria express specialised proteins, called adhesins, on their surfaces to facilitate this adhesion. In the gut, adhesins promote interaction of the bacterium with host cell surface or soluble receptors, facilitating firstly the penetration of the gut mucosal barrier and subsequently the invasion of the intestinal epithelial cells (IECs) in the case of pathogenic species (Knutton, Shaw et al. 1999; Klemm and Schembri 2000; Pizarro-Cerda and Cossart 2006). Many adhesins have lectin properties but not all adhesins bind to carbohydrates.

1.4.1 Cell wall structures involved in adherence

Lectins can be defined as non-immunoglobulin-like proteins able to recognise and bind to a carbohydrate specifically and reversibly, without catalysing a reaction. They are therefore neither classified as enzymes nor antibodies and can be grouped in few but large families of proteins. Lectins compensate the weak affinity for their ligand (µM range) through multivalent interactions with glycans (avidity or 'velcro' effect). They can have several carbohydrate recognition domains (CRDs) or form ensembles to increase their attachment capacity (Goldstein, Hughes et al. 1980; Dam and Brewer 2010). The 3D presentation of the carbohydrate ligand is important for lectin binding (Kopitz, Bergmann et al. 2010; Gabius, Andre et al. 2011). In bacteria, lectins can be found soluble in the cytoplasm, can be secreted (e.g. some toxins) and can also be exposed on the cell surface (Imberty, Wimmerova et al. 2004).

Bacterial adhesion to the GIT has to date been studied essentially in pathogens with the view to identify these lectins for the development of vaccines. Two types of adhesive mechanisms can generally be distinguished. Firstly, the docking takes place, which corresponds to a reversible attachment of the bacteria to the receptor. Following this step, anchoring or a more permanent attachment can occur (Dunne 2002). The bacterial adherence follows a tissue tropism and is species-specific. The variety of glycans at the extremities of mucins represents likely receptors for bacterial adhesins (Belley, Keller et al. 1999). Glyco-conjugates on the external surface of the bacterial cell wall may as well have

a reciprocal role in adhesion, acting as targets for epithelial cell surface receptors. Both glyco-strategies of adhesion by pathogens have thus been named the 'zipper attachment-effect' (Dinglasan and Jacobs-Lorena 2005).

In addition to aforementioned structures (section 1.3.3, Figure 1.5), several types of cell surface exposed proteins (SEPs) are decorating or anchored to the bacterial wall (detailed in Chapter 3.1). The most expressed SEPs in Gram-positive bacteria are S-layer proteins, which form a regular monolayer on the cell surface (Sleytr and Beveridge 1999). Some cell surface proteins can also take the shape of long appendages with multiple subunits that extend out of the cell, called fimbriae or pili (although the latter term was initially used for bacterial conjugation). Such structures may be of hetero or homopolymeric nature and have been extensively reviewed in both Gram-negative and Gram-positive bacteria (Telford, Barocchi et al. 2006; Mandlik, Swierczynski et al. 2008; Proft and Baker 2009; Kline, Dodson et al., 2010 ; Korea, Ghigo et al., 2011 ; Burrows, 2012) and summarised in Tables 1.3 and 1.4.

Table 1.3: Main fimbrial structures in Gram-negative bacteria known to date

Fimbrial group	Assembly	Main features	Adhesin example
CU pili	Chaperone-usher (CU)	Non covalent polymers, helical structures	FimH
Type III secretion needle	Secretion system (T3SS)	Flagellum-like nano machine, attaching/effacing mechanism	MxiH
Type IV pili	Secretion system (T2SS)	Filament architecture, twitching motility	PilA
Type IV secretion pili	Secretion system (T4SS)	Polymeric units with an adhesin on the shaft and integrin legs	CagA

Table 1.4: Main fimbrial structures in Gram-positive bacteria known to date

Fimbrial group	Assembly	Main features	Adhesin example
Sortase-mediated pili	Sec and sortase-mediated	Covalent polymerisation, major pilin on the shaft and minor pilins	SpaC
Type IVb pili	Tad secretion system (Type II)	Non-covalent polymerisation of fimbrial low molecular weight pilins and pseudopilins	Flp

Fimbriae display several differences in their structures and biogenesis. They are assembled using bacterial secretory pathways and appear like thin filaments (such as the long polar fimbria) or threads (such as curli or bundle-forming pili). In Gram negative bacteria, they are typically formed by the non-covalent homopolymerisation of the major subunit protein, called pilin, which builds the fimbrial appendage, generally through the chaperone-usher or secretory pathways such as the type IV secretion system. In Gram-positive bacteria, fimbriae are usually formed by the covalent heteropolymerisation of pilin subunits that usually requires a specific sortase enzyme for their assembly as described in the pathogen *Corynebacterium diphteriae*. Genes involved in the expression and mounting of these appendages are organised in genomic islands (pathogenicity islands or operons) containing several genes.

Lectins can be found along the shaft of these polymers but are usually found at the tip of the appendage. Some atypical adhesins are also found in bacteria. Some enteric pathogens such as *Shigella flexneri* have developed needle-like structures (e.g. MxiH) to lock onto epithelial cells and deliver toxins directly in the enterocytes to invade and internalise (Bahrani, Sansonetti et al. 1997). Other enteropathogens use similar molecular mechanisms to intimately adhere to the enterocytes by provoking a cytoskeleton reorganisation or deformation, causing attaching/effacing lesions on the microvilli (e.g *E. coli* ETEC). (Pizarro-Cerda and Cossart 2006). Figure 1.9 represents the major types of afimbrial and fimbrial adhesins present on bacteria are illustrated on the prominent gut bacterium *Escherichia coli*, which is normally commensal but can be an opportunistic pathogen.

1.4.2 Target ligands of bacterial lectins

The glycans present throughout the GIT provide a large range of targets for bacterial lectins to adhere to. Due to the linkages, the branching and the stereoisomery of assembled monosaccharides, the mammalian glycome embraces a variety of three-dimensional structures, resulting in a rich repertoire of epitopes for the microbiota (Kornfeld and Kornfeld 1985; Rademacher, Parekh et al. 1988; Turnbaugh, Hamady et al. 2009). Under suitable conditions, bacteria will express adhesins to conquer a particular niche by interacting with target ligands that surround the IECs. These include naturally the glycan structures available on glycoconjugates from the mucosal barrier, whether glycolipids, *O*-linked or *N*-linked glycans. In the GIT, the majority displays an *O*-linkage due to the glycosylation of mucins.

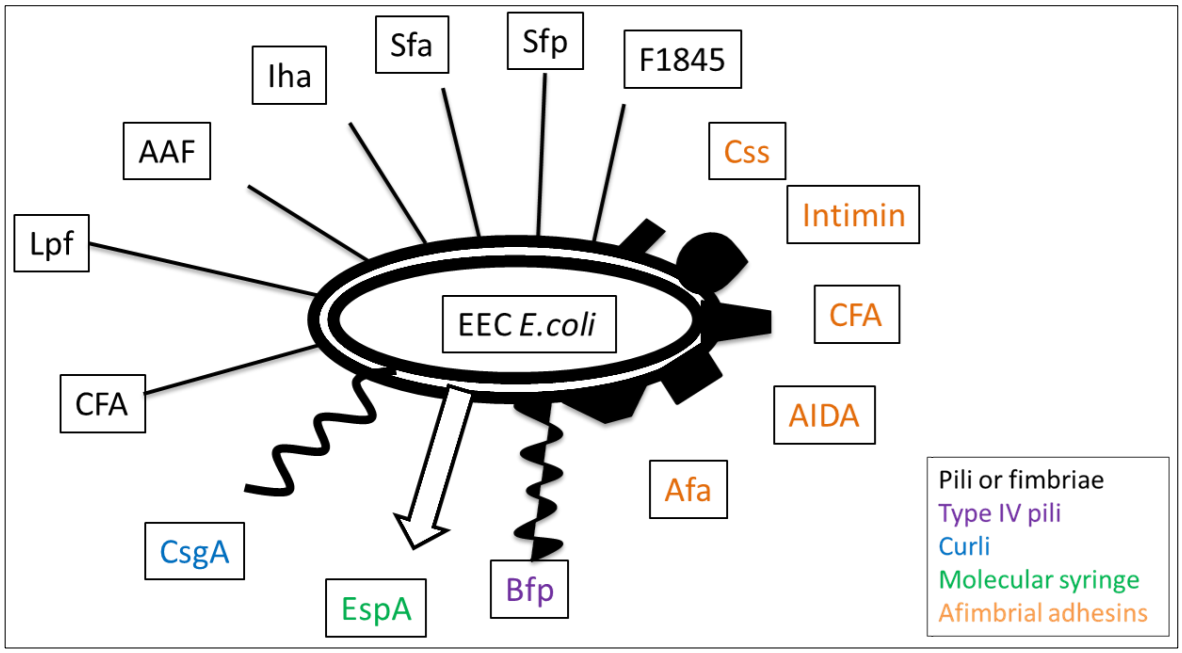

Figure 1.9: Major types of adhesins present on bacterial cell walls

Illustration of fimbrial and afimbrial adhesins found on *E. coli*.
Abbreviations are detailed in Table 1.9. Adapted from various reviews (Gaastra and de Graaf 1982; Archambaud, Courcoux et al. 1988; Bilge, Clausen et al. 1989; Garcia, Labigne et al. 1994; Greene and Klaenhammer 1994; Tarr, Bilge et al. 2000; Gophna, Barlev et al. 2001; Paton, Srimanote et al. 2001; Clarke, Haigh et al. 2003; Neves, Shaw et al. 2003; Jordan, Cornick et al. 2004; Boisen, Struve et al. 2008; Bielaszewska, Prager et al. 2009; Sabui, Ghosal et al. 2010)

Non-mucin target ligands include cell surface integrins and extracellular matrix (ECM) glycoproteins. The latter comprise laminin, which is present in the basal membrane of the IECs, fibronectin (Fn), a substrate for adherence of eukaryotic cells to the matrix and for bacterial attachment, fibrinogen, an acute phase reactant, and, collagen, mainly of type IV, an abundant eukaryotic structural protein. The ECM is a stable macromolecular structure that shapes the epithelium but its components can fall off in the mucus layer (Patti, Allen et al. 1994; Chagnot, Listrat et al. 2012). Additionally, a damaged mucosa would render these proteins exposed to the intestinal lumen and being target ligands for the microbes. Many pathogenic species actually show an ability to adhere to ECM glycoproteins and the surface adhesins binding to ECM molecules are collectively called microbial surface components recognizing adhesive matrix molecules (MSCRAMM) (Patti, Allen et al. 1994; Zhao, Sillanpaa et al. 2009). For instance, the invasion of some species from staphylococci and streptococci involves a fibronectin ‘bridge’ between the bacterial MSCRAMMs and host integrins (Joh, Wann et al. 1999; Hauck and Ohlsen 2006).

But glycosylation is a dynamic process that evolves during human development. Changes occur on the lining of the gut following the maturation of the enterocytes during infancy, from mostly sialylated glycoproteins covering the mucosal epithelium to a shift towards fucosyl moieties (Biol-N'Garagba, Greco et al. 2002). These changes affect the composition of the microbiota and result in the colonisation by new strains as well as the loss of others. It is unclear though if this is the cause or a consequence of the microbes actions (Mariat, Firmesse et al. 2009). Composition of cell surface glycans can also change following illnesses (Roussel 2005). The microbiota plays a role in the maintenance of the mucus layers by upregulating mucin genes such as intestinal MUC2 (Kandori, Hirayama et al. 1996). SCFAs, in particular butyrate, mediate the composition of the mucus coating. Butyrate, the most abundant SCFAs is also a metabolite of the enterocytes and is able to increase the production of mucins (Hamer, Jonkers et al. 2008). *F. prausnitzii*, a predominant strain in the adult microbiota, produces butyrate. Significant counts of *F. prausnitzii* cells have been reported in the microbiota of patients suffering from inflammatory bowel diseases (IBD) such as Crohn's disease (CD) (Sokol, Pigneur et al. 2008; Galecka, Szachta et al. 2013). Adherence of microbes was also reported to trigger glycosylation changes in the intestinal mucosa by interfering with the expression of host GTs and GHs, resulting in modifications of the carbohydrate repertoire of mucins (Hooper and Gordon 2001). Thus, the intestinal microbiota influences the composition of the mucus layer covering the epithelium, and as a consequence increases the mucosal barrier.

1.4.3 Approaches to discovery of bacterial lectin-like adhesins

Research on bacterial adhesins has focussed mostly on pathogens. The adherence of commensals in the GIT and the particular role of adhesins represent an emerging field of interest but has been limited for various reasons. The culturing difficulties of strict anaerobes populating the microbiota constitute naturally a handicap as the vast majority of the species are presumably unculturable or recalcitrant to conventional culture methods (Kau, Ahern et al. 2011). The expression of adhesins might also differ between *in vitro* and *in vivo* conditions and all the more, subject to antigenic variation (O'Connell Motherway, Zomer et al. 2011). An adhesion adaptive response (AAR) has been identified in *L. acidophilus* NCFM. The AAR is linked to the quorum sensing gene LuxS, expressing the auto-inducer 2 and results in an over-expression of adhesive proteins, such as Mub (Buck, Altermann et al. 2005).

In vivo, models to study commensals adhesion include (i) GF mice, which are mice kept in a sterile environment isolated from the outside (with offsprings delivered through caesarian), (ii) gnotobiotic mice, which correspond to GF rodents colonised by known bacterial species, (iii) knockout (KO) mutants, which are transgenic animals (gene knockout), (iv) chemically induced colitis mice, in which chemicals are given to disrupt the tight junctions in the epithelium, and (v) also drosophila. The current approaches rely on the *in silico* identification of putative adhesion molecules.

1.4.4 Search for lectin conserved domains

Distinct amino acid sequences conserved among bacteria share similar roles. Investigating proteins for conserved domains such as peptidoglycan-binding motifs, or more specifically in the case of adherence, LPxTG anchors (TIGR01167), fibronectin-binding domain (FbpA domain, pfam05833), collagen-binding domain (collagen_bind, cl05349) or mucin-binding domains (MucBP, pfam06458), represents a sensible approach to select putative new adhesins. Recent studies have examined sequenced species, including lactobacilli, for potential mucus-binding proteins using the 'MucBP' domain in *in silico* searches (Boekhorst, de Been et al. 2005; Arumugam, Raes et al. 2011). Following a survey of the NCBI database, we found the 'MucBP' motif in a wide variety of bacterial proteins, especially in several families of Firmicutes (1334 protein sequences), including 893 lactobacilli proteins. However, i only found six sequences in Actinomycetes and none related to the bifidobacterial species sequenced to date.

A similar approach could be used to identify the presence of lectin domains and could be relevant to define new candidates involved in carbohydrate binding. A basic *in silico* search was thus performed in the NCBI database to, first, get an overall idea of the number of fully sequenced genomes in bacteria corresponding to the four major commensal phyla (Proteobacteria, Bacteroidetes/Chlorobi group, Firmicutes and Actinobacteria) as well as in the *Lactobacillus* and *Bifidobacterium* genera (Table 1.5).

Large differences were noted between phyla in terms of the number of genomes sequenced, with the Bacteroidetes/Chlorobi division comprising only 111 species (including the commensal prominent bacterium *Bacteroides thetaiotaomicron*) and the Gram negative Proteobacteria phylum containing 1172 species (including *E. coli*, *Helicobacter* or *Salmonella sp.*). Furthermore, to date, more lactobacilli have been sequenced (71 species) than bifidobacteria (24 species), with sequenced genomes

increasing rapidly (almost doubling in the last two years). Therefore, there is definitely enough data available to expand this *in silico* approach to identify commensal adhesins.

Table 1.5: Number of complete bacterial sequenced genomes in GenBank (NCBI)

Groups	Dec-11	Dec-13
Bacteria	1729	2649
Proteobacteria	792	1172
Bacteroidetes/Chlorobi	79	111
Firmicutes	434	619
Actinobacteria	182	280
Lactobacillus	37	71
Bifidobacterium	21	24

All bacterial groups are phyla except *Lactobacillus*, a genus of Firmicutes and *Bifidobacteria*, a genus of Actinomycetes.

The conserved domains database (NCBI CDD) was explored to find lectin motifs. In total, 32 lectin-like motif superfamilies were identified in the database (Table 1.6), of which 17 domains were described in bacterial species. The frequency of occurence of these domains in the major commensal phyla is shown in Table 1.7. Nine domains are found exclusively in the Proteobacteria phylum, which has the highest number of species sequenced. These include for instance the 'Fim-adh_lectin'- like domain (pfam09222) identified solely in *E. coli.* Proteins of this family contain carbohydrate-specific lectin domains found in fimbrial type 1 adhesins including FimH which binds Man and F17-G specific for GlcNAc (Buts, Bouckaert et al. 2003). Other domains unique to Proteobacteria include the 'PA-IL-like' (pfam07828) and PA-IIL–like (pfam07472) domains found in molecules similar to the galactophilic lectin-I (PA-IL) and the Fuc-binding lectin II expressed by *Pseudomonas aeruginosa*. These proteins, involved in the virulence of this pathogen, recognise specific carbohydrates found on the surface of host cells and are known to be involved in the initiation of infections by this organism (Imberty, Wimmerova et al. 2004). Another motif exclusive to Proteobacteria is the 'Intimin_C' (pfam07979). This domain is found at the C-terminus of intimin, an afimbrial adhesion molecule involved in the attachment and invasion of pathogenic strains to mammalian intestinal cells including diarrheagenic *E. coli* (Batchelor, Prasannan et al. 2000).

The von Willebrand factor type A domain (vWFA) was found in all bacterial genomes examined. This was described originally in the large blood coagulation glycoprotein von

Willebrand factor (vWF), but is conserved in a wide variety of proteins involved in important cellular functions, such as signalling or adhesion. The ligand binding is generally mediated by the presence of a metal ion dependent adhesion site (MIDAS motif) that is also a characteristic of several lectins (Booth, Furby et al. 1984). The vWFA motif has been shown to have affinity for polyanionic oligosaccharides (Andrews, Bendall et al. 1995). Interestingly, this domain was found more frequently in the genus *Bifidobacterium* (311 proteins), more so than lactobacilli (101 proteins).

Table 1.6: Lectin-like domains superfamilies listed in NCBI database

Id	Name	Representative proteins	Accession
1	B_lectin	Bulb-type mannose-specific lectin	cd00028
2	Bc2l-C_N	N-Terminal Domain Of Bc2l-C Lectin	cd12211
3	vWFA	Von Willebrand factor type A	cl00057
4	GLECT	Galectin/galactose-binding lectin	cl00071
5	PTX	Pentraxin/Concanavalin A-like lectin/Glucanase	cl00102
6	Thioredoxin_like	Thioredoxin fold	cl00388
7	CLECT	C-type lectin (CTL)/C-type lectin-like (CTLD)	cl02432
8	NTR_like	NTR_like	cl02512
9	Jacalin_like	Jacalin-like lectin	cl03205
10	ISOPREN_C2_like	Class II terpene cyclases	cl08267
11	GH18_chitinase-like	Glycosylhydrolase family 18	cl10447
12	Glyco_tranf_GTA_type	Glycosyltransferase family A	cl11394
13	Ig_domain	Immunoglobulin	cl11960
14	lectin_L-type	Legume lectin	cl14058
15	ChtBD1	Hevein or type 1 chitin binding	cl16916
16	Gal_Lectin	Galactose binding lectin	pfam02140
17	APT	Aerolysin/Pertussis toxin (APT)	pfam03440
18	Lectin_N	Hepatic lectin, N-terminal	pfam03954
19	FB_lectin	Fungal fruit body lectin	pfam07367
20	PA-IIL	Fucose-binding lectin II (PA-IIL)	pfam07472
21	PA-IL	PA-IL-like protein	pfam07828
22	Fungal_lectin	Fungal fucose-specific	pfam07938
23	Intimin_C	Intimin C-type lectin	pfam07979
24	Ly49	Ly49-like protein, N-terminal region	pfam08391
25	Fim-adh_lectin	Fimbrial adhesin F17-AG	pfam09222
26	Fve	Fungal immunomodulatory protein Fve	pfam09259
27	Sial-lect-inser	Vibrio cholerae sialidase, lectin insertion	pfam09264
28	H_lectin	H-type lectin	pfam09458
29	Inhibitor_I48	Peptidase inhibitor clitocypin	pfam10467
30	PP2	Phloem protein 2	pfam14299
31	TECPR	Beta propeller repeats	smart00706
32	RrgB_K2N_iso_D2	Fimbrial isopeptide formation D2	TIGR04226

Table 1.7: Number of bacterial proteins with lectin-like domains

Id	Name	Accession	Bacteria	[1]Proteo	CFB	Firm	Actino	Lacto	Bif
1	B_lectin	cd00028	221	221	0	0	0	0	0
2	Bc2l-C_N	cd12211	29	29	0	0	0	0	0
3	vWFA	cl00057	19310	5756	1165	6146	3302	101	311
4	GLECT	cl00071	97	25	9	25	20	0	0
5	PTX/Laminin_G_3	cl00102	13162	3139	3290	1674	2248	69	230
9	Jacalin_like	cl03205	965	710	18	84	91	10	7
14	lectin_L-type	cl14058	4042	602	153	3068	99	533	8
15	ChtBD1	cl16916	9	0	7	0	0	0	0
16	Gal_Lectin	pfam02140	30	21	0	0	0	0	0
17	APT	pfam03440	143	143	0	0	0	0	0
20	PA-IIL	pfam07472	334	334	0	0	0	0	0
21	PA-IL	pfam07828	234	234	0	0	0	0	0
23	Intimin_C	pfam07979	966	966	0	0	0	0	0
25	Fim-adh_lectin	pfam09222	40	40	0	0	0	0	0
28	H_lectin	pfam10467	217	151	34	4	0	0	0
31	TECPR	smart00706	26	3	0	0	0	0	0
32	RrgB_K2N_iso_D2	TIGR04226	7879	81	9	6699	1004	191	306

[1]Abbreviations are as follows: Proteobacteria (Proteo), Bacteroidetes/Chlorobi (CFB), Firmicutes (Firm), Actinobacteria (Actino), *Lactobacillus* genus (Lacto) and *Bifidobacterium* genus (Bif).

The lectin_L-type domain is found in the legume-like lectins. These lectins comprise numerous families of carbohydrate-binding proteins (CBPs) that generally display no enzymatic activity. Their sugar binding specificities differ widely despite their structural similarities (Varki A 1999). A high number of GenBank entries is associated with this domain in lactobacilli (533 sequences) including the characterised mannose-binding adhesin (msa) from *L. plantarum* (Figure 1.10). Only eight protein sequences had this domain in bifidobacteria, all coming from four strains of *B. dentium*. One of these corresponds to a fimbrial protein, which interestingly also contains a vWFA domain (Figure 1.11).

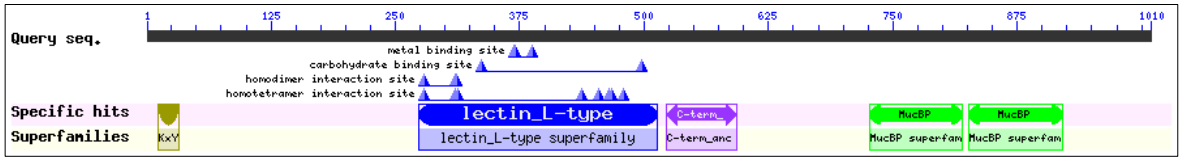

Figure 1.10: Mannose-binding adhesin from *Lactobacillus plantarum*

In silico conserved domain search (NCBI CDD) of sequence [CCC78612] exhibiting membrane signal peptides (in kaki and purple), two mucus-binding domains MucBP (in green) and lectin_L domain (in blue).

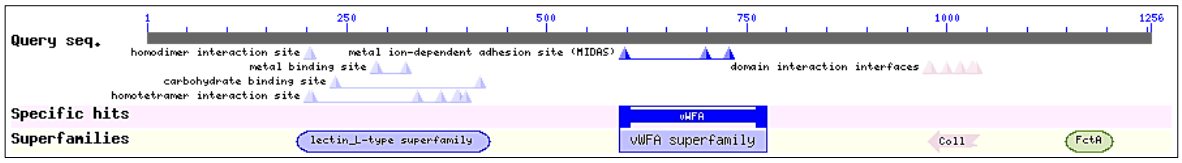

Figure 1.11: Fimbriae protein from *Bifidobacterium dentium* Bd1

Conserved domain analysis of sequence [YP_003359651] featuring a lectin L-type domain (in pale blue), a von Willebrand Factor vWFA domain [cd00198] (in dark blue), a collagenBindB domain [cl15753] (in pink), and a motif found in T surface-antigens [cl16948] (in green).

Lectins bearing the 'Jacalin_like' domain (cl03205) are mostly found in plants. These are divided into two groups depending on their carbohydrate ligands (either Gal or Glc/Man groups) (Barre, Bourne et al. 2001). Seven proteins containing this domain were found in the genus *Bifidobacterium* (from four species) and ten in *Lactobacillus* (originating from six strains).

Finally and most interestingly, a large number of sequences in lactobacilli (191) and especially bifidobacteria (306) contained the fimbrial isopeptide formation domain (TIGR04226). The representative protein corresponds to RrgB, the backbone subunit of *Streptococcus pneumoniae* pilus (Spraggon, Koesema et al. 2010). The majority of the proteins possessing this domain are reported to be involved in surface adhesion to host structures and include lectin-like adhesins (Marchler-Bauer, Lu et al. 2011).

Thus, analysis of existing *in silico* databases shows many potential adhesins in commensal species of interest. This approach represents a good starting point for the further exploration of commensal adhesins carried out in this study.

1.4.5 Other approaches

Software such as SurfG+ or SPAAN (Sachdeva, Kumar et al. 2005; Barinov 2009), have been developed to predict the cellular location of a given protein using signal motifs. Predicted surface-exposed proteins can then undergo further analyses including conserved domains searches or alignments to determine homologies with characterised adhesins, before experimental validation.

Other approaches rely in the transformation and over-expression of pili proteins in strains not constitutively expressing pili *in vitro*, in engineering knockout (KO) mutants or in inducing adhesin-encoding genes in compatible host bacteria such as *Lactococcus lactis* (Buck, Altermann et al. 2005; O'Connell Motherway, Zomer et al. 2011).

Binding studies on recombinant adhesins have also been carried out using the radiolabeled proteins on multi-titer plates coated with mucins (from pig stomach or human intestine) as well as on human cell lines (such as colonic HT29 or Caco-2 cells).

Recent advances in technology such as glycomic databases, synthesis of glycoconjugates and microarrays are providing new tools to study lectin-glycan interactions. In terms of characterising lectin/ligand interactions, the development of high throughput technologies such as mucin or glyco-conjugate microarrays, is a promising path.

1.5 Lectin-like adhesins of gut microbes

1.5.1 Adhesins described to date in *Escherichia coli*.

Escherichia coli are facultative anaerobic Gram negative bacteria considered conspicuous commensal pathobiont in the gut. Thanks to their genomic plasticity, *E. coli* are able to adapt to various environments. Some strains can be opportunistic pathogens and the cause of inconvenience or more serious diseases, especially in the third world or in children. Both *E. coli* commensal and pathogenic strains share a common fimbria, termed *E.coli* pili (ECP). Nevertheless, *E. coli* diarrhoeagenic serotypes synthesise specific surface adhesins to enhance their ability to colonise niches in the GIT (Rendon, Saldana et al. 2007).

Six entero-pathogenic types or pathotypes (Table 1.8) have been described in relation to the adherence mechanism and the enterotoxin products that cause the various symptoms described in humans (Kaper, Nataro et al. 2004). One of the best known pathotypes corresponds to the entero-toxigenic strain (ETEC), which are also referred to as the traveller's diarrhoea. These are bacteria which are not invasive but produce two enterotoxins: a cholera-like toxin termed heat-labile toxin (LT), and a heat-stable toxin (ST) that induces the diarrhoea. A second pathotype include entero-pathogenic strains (EPEC), which are the cause of watery diarrhoea in infants. They produce bundle-forming pili and adhere closely to the enterocytes using afimbrial adhesins (intimin), thus damaging the mucosal barrier.

Another moderately invasive pathotype is the entero-haemorrhagic *E. coli* (EHEC), for which a number of relevant putative adhesins have been identified. As EPEC, they bind to microvilli through actin reorganisation and produce a Shiga-like toxin, leading to colitis associated with an intense inflammatory response in infants. More severe symptoms can be induced by a fourth pathotype: the entero-aggregative *E. coli* strains (EAEC/EAggEC). These induce persistent diarrhoea in children likely caused by an ST-like enterotoxin

(EAST) and a haemolysin. A fifth pathotype corresponds to invasive bacteria. This entero-invasive *E. coli* (EIEC) penetrates enterocytes using afimbrial adhesins, lyses vesicles to multiply within the cells and then moves through the cytoplasm to the adjacent epithelial cells. EIEC strains provoke dysentery-like diarrhoea with high inflammation and fever due to their invading of colonic cells. More recently, the entero-adherent *E. coli* pathotype has been identified (DAEC). It binds to the small bowel in a diffuse adhesion pattern but the mechanism is still largely unknown (Puente, Bieber et al. 1996; Kaper, Nataro et al. 2004; Weintraub 2007). All the adhesin-encoding operons or genomic islands, surveyed in the literature, expressing molecules involved in the adherence of *E. coli* strains to the intestinal epithelial cells have been summarised in Table 1.9.

Table 1.8*:* List of main human enteropathogenic *E. coli*

Serotype	Description	Adherence	Enterotoxin production
ETEC	Entero-toxigenic	Adhere to IECs	LT and/or ST
EPEC	Entero-pathogenic	Moderately invasive in IECs	Shiga-like toxin
EHEC	Entero-hemorrhagic	Moderately invasive in IECs	Shiga-like toxin
EAggEC	Entero-aggregative	Clump to IECs	EAST and a hemolysin
EIEC	Entero-invasive	Invade and multiply within IECs	none
EAEC-DAEC	Entero-adherent	Adhere to IECs	none

Adapted from Kaper et al. and Mainil et al. (Kaper, Nataro et al. 2004; Mainil 2013).

Table 1.9*:* Summary of adhesins involved in the virulence of enteropathogenic *E.coli*

Structure	Adhesin-encoding operon	Name	locus/accession	Strain
Afimbrial	AIDA	Adhesin involved in diffuse adherence	AIDA_ECOLX	ETEC/ EAEC
	Afa	Afimbrial adhesins	INTL_ECOLX	EAEC
	F1845/Daa	Diffuse adherence adhesin	ECODAAF	EAEC
	CS6	Coli surface antigen 6	CAD90926.1	ETEC
	Eae/intimin	*E. coli* attaching and effacing gene	AAD16298	EPEC
Type IV pili	Bfp	Bundle-forming pili	BFPB_ECO11	EPEC
	Saa	STEC autoagglutinating adhesin	AF325220_1	EHEC
Fimbriae	Iha	IrgA homologue adhesin	AF126104	EPEC
	AAF	Aggregative adhesion fimbriae	CAA79242	EAEC
	Sfa	Sialic acid binding fimbrial adhesin	ECO103_1035	EHEC
	Sfp	Sorbitol fermenting plasmid fimbriae	EU980314	EHEC
Curli	CsgA	Amyloid fibrils	AAC74126	EHEC
Chaperone Usher	lpf	Long polar fimbria	AAY18076	EHEC
	CFA/I	Colonisation factor antigen	ECOCFAIA	ETEC
Molecular syringe	LEE/EspA	Filamentous type III secretion protein	CAA91163	EPEC/ EHEC

Summarised from various reviews (Gaastra and de Graaf 1982; Archambaud, Courcoux et al. 1988; Bilge, Clausen et al. 1989; Garcia, Labigne et al. 1994; Greene and Klaenhammer 1994; Tarr, Bilge et al. 2000; Gophna, Barlev et al. 2001; Paton, Srimanote et al. 2001; Clarke, Haigh et al. 2003; Neves, Shaw et al. 2003; Jordan, Cornick et al. 2004; Boisen, Struve et al. 2008; Bielaszewska, Prager et al. 2009; Sabui, Ghosal et al. 2010).

1.5.2 Adhesins described to date in gut commensal species

Very few commensal adhesins have been described so far. Similar structures of attachment exist in pathogenic and commensal strains from the same species, suggesting that commensal adhere in the same way as pathogens. As the ECP in *E. coli*, Opa adhesins in the genus *Neisseria*, are structurally and functionally closely related between commensal and pathogenic strains (Toleman, Aho et al. 2001). The described crystal structures of fimbrial tip adhesins such as FimH, PapGII, and F17G in pathogenic *E. coli*, indicate that each structure is unique and correlate with the host niche environment (Buts, Bouckaert et al. 2003). A highly specific system is thus in place between the eukaryotic glycoconjugates acting as pattern recognition receptors (PRR) and the bacterial lectins.

In silico genomic analysis of LGG revealed the presence of two pili gene clusters, *SpaCBA* and *SpaFED*, expressing sortase-mediated pili (von Ossowski, Satokari et al. 2011). Cloning of the relevant genes expressing the pilins, their recombinant expression into *E. coli* and subsequent purification were performed. The proteins were radiolabeled along with the whole bacteria and the binding to intestinal human mucus assessed in multiwell plates. The experiments resulted in identifying two minor pilins SpaC and SpaF as the tip lectins/adhesins for the two pili in that strain. A very high affinity for mucus was noted by another pilin, SpaB. Competition experiments in this case suggested the binding was most likely due to electrostatic forces between mucins and the pilin (von Ossowski, Satokari et al. 2011).

A list of functional adhesins, characterised by binding experiments, have been described in the literature. In total 16 proteins from ten species of lactobacilli and a single lipoprotein, BopA, from *Bifidobacterium* were identified. Recent findings suggest the latter displays actually a minor role in the adhesion of *B. bifidum* to IECs (Kainulainen, Reunanen et al. 2013). Conserved motifs present in all sequences were searched in the NCBI database (Table 1.10). Msa from *L. plantarum*, is the only adhesin listed here with an L-type lectin domain. The vWFA motif, which is also a lectin-like domain, is present on the mucus binding SpaC from the tip pili *SpaCBA* of LGG. Furthermore, a MucBP domain is present in the sequence of aforementioned Msa. The Mub proteins from *L. acidophilus* NCFM and *L. reuteri* 1063 likewise possess MucBP domains, the latter with 14 repeats throughout its sequence. Other domains featured are signal motifs such as S-layer or YSIRK anchors and periplasmic binding transporter protein modules.

Ertzold et al. recently characterised a novel cell-surface protein, Lar_0958 from *L. reuteri* strain JCM 1112, containing six repeat domains (Etzold, MacKenzie et al. 2014). X-ray crystallography of a purified Lar0958 repeat protein defined the structure as an immunoglobulin-like β-sandwich and a β-grasp fold. Structural similarity was also noted to the internalin protein family of *Listeria*.

In addition, a ribosome-associated protein, the elongaton factor EF-Tu, has been described as an adhesin in *L. johnsonii* NC533, due to its detection on the cell surface, following translocation from the cytoplasm. These 'moonlighting' proteins with functions in several locations of the cell, have been reported for several Gram positive pathogens such as *Streptococcus pyogenes* and are involved in bacterial virulence (Henderson, Nair et al. 2011). Furthermore, cytoplasmic enzymes with moonlighting functions from *L. crispatus* including enolase, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and glutamine synthetase (GS) were found to bind plasminogen. All of these are non-covalently anchored to the bacterial cell surface and possibly released in stress situations (Kainulanen V. et al., 2012 unpubl.).

Novel adhesion molecules have been recently identified and published following *in silico* searches of LPXTG anchors, S-layer motifs, ECM and mucin-binding domains, but as yet have not been experimentally characterised (Boekhorst, de Been et al. 2005; Buck, Altermann et al. 2005; Mobili, Serradell Mde et al. 2009; van Passel, Kant et al.; Van Tassell and Miller 2011; von Ossowski, Satokari et al. 2011). F1C fimbriae, described as adhesin on uroepithelial cells for UPEC, was also found on the probiotic *E. coli* Nissle 1917 strain and described as important adherence molecules for the colonisation of the murine gut (Lasaro, Salinger et al. 2009). A recent transcriptomic study on *Bifidobacterium breve* UCC2003 demonstrated the presence of a type IVb tight adherence pili (Tad). Experiments on re-colonised GF mice demonstrated the Tad pili to be essential for gut colonisation. Furthermore, the Tad locus is conserved in all bifidobacterial genomes sequenced to date, suggesting a common mechanism for colonisation of the GIT (O'Connell Motherway, Zomer et al. 2011). This discovery is a significant achievement towards understanding the interaction between the bifidobacteria and their host and how the microbiota might influence gastro-intestinal health.

Table 1.10: List of characterised adhesins from probiotic strains

Adhesin	Length (AA)	Binding target	Accession number	Conserved Domains	Strains	Reference
Mub	4326	Human IECs, mucus	AAV43217	YSIRK_signal super family[cl04650]; MucBP super family[cl05785] x2	*Lactobacillus acidophilus* NCFM	(Altermann 2005)
Mub	3269	Mucus	AAF25576	MucBP super family[cl05785] x14	*Lactobacillus reuteri* 1063	(Roos and Jonsson 2002)
SlpA	1017	Human IECs	AAV43202	none	*Lactobacillus acidophilus* NCFM	(Altermann 2005)
SlpA	465	Human IECs, collagen, laminin and fibronectin	CAA78618	none	*Lactobacillus brevis* ATCC 8287	(Jakava-Viljanen, Åvall-Jääskeläinen et al. 2002)
Slp	437	Human IECs	AAZ99044	PBP2_NikA_DppA_OppA_like super family[cl01709]; SLAP[pfam03217]	*Lactobacillus helveticus* R0052	(Johnson-Henry, Hagen et al. 2007)
FbpA	563	Human IECs,fibronectin	AAV42987	DUF814 super family[cl05307]	*Lactobacillus acidophilus* NCFM	(Altermann 2005)
CbsA	440	Collagen I and IV, laminin, bacterial lipoteichoic acids (LTA)	AAB58734	SLAP[pfam03217]	*Lactobacillus crispatus*	(Sillanpaa, Martinez et al. 2000)

Adhesin	Length (AA)	Binding target	Accession number	Conserved Domains	Strains	Reference
EF-Tu	396	Human IECs, mucus	AAS08831	EF_Tu[cd01884]; EFTU_III[cd03707]; EFTU_II[cd03697]	*Lactobacillus johnsonii* NCC533	(Granato, Bergonzelli et al. 2004)
GroEL	543	Human IECs, mucus	AAS08453	GroEL[cd03344]	*Lactobacillus johnsonii* NCC533	(Pridmore 2004)
Msa	1010	Mucus via mannose binding	CCC78612	lectin_L-type[cd01951]; MucBP super family[cl05785] x3	*Lactobacillus plantarum* WCSF	(Gross, van der Meulen et al. 2008)
CnBP	263	Human IECs, mucus	CAA68052	PBPb[cd00134]	*Lactobacillus reuteri*	(Roos and Jonsson 2002)
MapA	263	Human Caco-2 cells, mucus	CAC05301	PBPb[cd00134]	*Lactobacillus reuteri* 104R	(Miyoshi, Okada et al. 2006)
LspA	1209	Human IECs	YP_535207	none	*Lactobacillus salivarius* UCC118	(Claesson 2006)
SpaC	895	Human mucus	CAR86339	vWFA[cd00198]; Peptidase_M14NE-CP-C_like super family[cl15700] x3	*Lactobacillus rhamnosus GG*	(Kankainen, Paulin et al. 2009)
SpaF	983	Human mucus	CAR88267	Peptidase_M14NE-CP-C_like super family[cl15700] x2	*Lactobacillus rhamnosus* GG	(von Ossowski, Reunanen et al. 2010)
Lar0958	1229	Human mucus	BAG25474.1	YSIRK_signal [TIGR01168]	*Lactobacillus reuteri* JCM 1112	Etzold, McKenzie et al., 2014

Adhesin	Length (AA)	Binding target	Accession number	Conserved Domains	Strains	Reference
BopA	593	Human Caco-2 cells	CAM97356	PBP2_Lpqw[cd08501];SBP_bac_5[pfam00496]	*Bifidobacterium bifidum* MIMBb75	(Guglielmetti, Tamagnini et al. 2009)

1.5.3 Known sugar targets of bacterial adhesins

Adhesin/ligand pairs that have been characterised so far include pathogenic species and very few for enteropathogens. Resources such as SugarBindDB (www.sugarbinddb.com) contain data on lectins from several mammals' pathogens associated with their carbohydrate ligands. As of December 2013, we found that 69 bacterial strains and 32 toxins were recorded along with 15 entries for described lectin/adhesins (Shakhsheer, Anderson et al. 2013). A closer look at the species revealed that most were pathogenic species affecting the respiratory or urinary tracts. The adhesins found on uropathogenic strains of *E. coli* (UPEC) are the best characterised fimbrial adhesins. Amongst these, the UPEC P-pili group, containing distinct fimbrial adhesins for which different variants have been identified (including class I, class II, and class III). P-fimbriae recognise globoseries of glycosphingolipids that contain the galabiose (Gal-α-(1→4)-Gal) core structure found on human P-blood group antigens (Kuehn, Heuser et al. 1992). Another online resource, the 'Pathogen adherence to carbohydrate database (PACD, http://jcggdb.jp/search/PACDB.cgi) was also investigated. The data were subsequently mined to compile a list of human specific pathogenic bacteria binding to the GIT, along with their known adhesins and target ligands. Most entries related to *E. coli* and *Helicobacter pylori* species. But most enteric *E. coli* pathogenic lectins listed were from strains affecting animals (such as adhesins CS3, K88 and K99 respectively binding to receptors in rabbit, piglet or bovine intestines). Overall 12 lectins with their corresponding carbohydrate ligand originating from only four species could be compiled. Three of the adhesins (BabA, HpaA and SabA) were affected to three strains of the prominent gastric pathobiont *Helicobacter pylori* (Table 1.11). Host immunological conditions seem to determine the status of *H. pylori* as pathogenic or commensal bacterium (Amieva and El-Omar 2008). Its genome encodes numerous outer membrane proteins including porins. Some of them have been described as adhesins and were assigned several target ligands possibly the results of strain specificity to the host mucosa. *H. pylori* BabA for example, has been attributed 11 binding fucosyl structures. This lectin seems to be involved in the docking step of the pathogen to the stomach mucosa. BabA adhesin recognises HBG antigens (such as H-type 1 and Le^b) expressed on gastric epithelial cells. Two sialylated structures were likewise recognised as ligands of a second adhesin, SabA. Upon colonisation, inflammation of the mucosa occurs resulting in sialylation of gastric glycans. The SabA adhesin mediates *H. pylori* binding to the inflamed gastric mucosa by

recognising sialyl-Lewis antigens. A third lectin listed here for this pathogen is HpaA, which is assigned two different structures disagreeing between the two databases searched. The combining of the two ligands might be a closer answer as HpaA is a flagellar lipoprotein that binds *N*-acetylneuraminyllactosamine (reviewed by Odenbreit et al. and Backert et al. (Odenbreit, Swoboda et al. 2009; Backert, Clyne et al. 2011)).

Vibrio cholerae is the cause of cholera, an acute diarrheal infection associated with contaminated water or food ingestion. Two surface chitin-binding proteins binding GlcNAc were listed in PACD that are involved in copepod exoskeleton binding but might also be involved in GIT adherence leading to human infection (Tarsi and Pruzzo 1999). Another fimbrial adhesin, the mannose-sensitive haemagglutinin (MSHA) identified as a significant factor in adherence to Caco-2 human intestinal epithelial cells was associated with two monosaccharide ligands (Glc and Man).

Enterotoxigenic *Escherichia coli* (ETEC) use surface fimbriae (colonisation factor 1(CFA/I)) to adhere to human IECs. The pili consist of a major pilin subunit (CfaB) and a tip-adhesive minor subunit CfaE (Li, Poole et al. 2009). SugarBindDB presents three records for CFA/I-CfaB binding respectively to sialic acid linked α-(2→8) and two glycolipids with a sialic acid branch. The CfaB entry is probably related to gut adherence in piglets since the neuraminic acid is Neu5Gc, a form that does not occur in humans. Finally, the K99 lectin, which can adhere to human jejunal cells (Lindahl and Wadstrom 1984) is associated with the trisaccharide ligand 3'Sialyllactose NeuAc-α-(2→3)-Gal-β-(1→4)-Glc.

Other adhesins receptors include glycosaminoglycans such as heparan sulfate and glycoproteins from the extra cellular matrix (ECM) which harbour complex glycan structures (Table 1.12). These include laminin, fibronectin and collagen, participating in the architecture of the mucosal barrier (section 1.4.2).

Apart from the mannose-binding adhesin msa from *L. plantarum*, no carbohydrate target has yet been described for commensal adhesins. Glycan microarrays have been used in this thesis to identify the glycan ligands in selected commensal species.

Table 1.11: List of characterised lectins/ligands from human enteric pathogens

Pathogen	Lectin	Carbohydrate ligand	[1]D
E.faecalis	EfaA	Fuc	P
E.faecalis	EfaA	Gal	P
EAggEC	18 kDa	Gal	S
ETEC	K99	NeuAc-α-(2→3)-Gal-β-(1→4)-Glc	S
ETEC	CFA/1	NeuAc-α-(2→8)	S
ETEC	CfaB	NeuGc-α-(1→3)[GalNAc-β-(1→4)]-Gal-β-(1→4)-Glc-β-1-Cer	S
ETEC	CfaB	GalNAc-β-(1→4)[NeuGc-α-(1→3)]-Gal-β-(1→4)-Glc-β-1-Cer	S
***H. pylori* (17875)**	BabA	Fuc-α-(1→2)-[Gal-α-(1→3)Gal-β-(1→3)]-GlcNAc-[Fuc-α-(1→4)]	S
***H. pylori* (17875)**	BabA	Fuc-α-(1→2)-[GalNAc-α-(1→3)-Gal-β-(1→3)]-Fuc-α-(1→4)-[GlcNAc]	S
***H. pylori* (17875)**	BabA	Fuc-α-(1→2)-[GalNAc-α-(1→3)-Gal-β-(1→3)]-GlcNAc	S
***H. pylori* (J99)**	BabA	Fuc-α-(1→2)-[GalNAc-α-(1→3)-Gal-β-(1→3)]-GlcNAc-[Fuc-α-(1→4)]	S
***H. pylori* (J99)**	BabA	Fuc-α-(1→2)-Gal-β-(1→3)-Fuc-α-(1→4)-[GlcNAc]	S
***H. pylori* (P466)**	BabA	Fuc-α-(1→2)-Gal-β-(1→3)-GlcNAc	S
***H. pylori* (P466)**	BabA	Gal-α-(1→3)-Gal-β-(1→3)[Fuc-α-(1→2)-]-GlcNAc[Fuc-α-(1→4)]	S
***H. pylori* (J99)**	BabA	GalNAc-α-(1→3)-Gal-β-(1→3)-[Fuc-α-(1→2)-]-Fuc-α-(1→4)-[GlcNAc]	S
***H. pylori* (J99)**	BabA	GalNAc-α-(1→3)-Gal-β-(1→3)-[Fuc-α-(1→2)-]-GlcNAc	S
***H. pylori* (J99)**	BabA	GalNAc-α-(1→3)-Gal-β-(1→3)-[Fuc-α-(1→2)-]-GlcNAc-[Fuc-α-(1→4)]	S
***H. pylori* (J99)**	BabA	Gal-α-(1→3)-Gal-β-(1→3)-[Fuc-α-(1→2)-]-Fuc-α-(1→4)[GlcNAc]	S
H. pylori	HpaA	Gal-β-(1→4)Glc(β1-1)Cer	S
H. pylori	HpaA	NeuAc-α-(2→3)-Gal-β-(1→4)	P
H. pylori	SabA	NeuAc-α-(2→3)-Gal-β-(1→4)	P
H. pylori	SabA	NeuAc-α-(2→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)Gal-β-(1→4)Glc(β1-1)Cer	P
V. cholerae	MSHA	Glc; Man	P
V. cholerae	CBP 36 and 56 KDa	GlcNAc	P

[1]Abbrevations for database sources (D) are as follows: S: SugarBindDB; P: PACD. (http://sugarbind.expasy.org/; http://jcggdb.jp/search/PACDB.cgi)

Table 1.12: Adhesins from enteric pathogens adhering to epithelial cells or ECM molecules

Species	Adhesive structure	Adhesin	[1]Ligand	Reference(s)
Campylobacter jejuni	Outer membrane protein	CadF	Fibronectin	(Monteville, Yoon et al. 2003)
Enterococcus faecalis	Surface protein	AS	IECs	(Wells, Moore et al. 2000)
	Surface protein	Ace	Collagen, laminin	(Nallapareddy, Qin et al. 2000)
	Surface protein	Acm	Collagen	(Nallapareddy, Singh et al. 2006)
	Surface protein	Scm	Collagen, fibrinogen	(Sillanpaa, Nallapareddy et al. 2008)
Streptococcus agalacticae	Surface protein	FbsA	Fibrinogen	(Schubert, Zakikhany et al. 2002)
	Surface protein	Lmb	Laminin	(Spellerberg, Rozdzinski et al. 1999)
	Surface protein	PavA	Fibronectin	(Holmes, McNab et al. 2001)
Yersinia enterocolitica	Afimbrial adhesin-TAA	YadA	Collagen, laminin	(Schulze-Koops, Burkhardt et al. 1992; Flugel, Schulze-Koops et al. 1994)
Shigella flexnerii	Outer membrane protein	OspE1/ OspE2	IECs	(Faherty, Redman et al. 2012)
Salmonella enterica	Curli	Agf	IECs	(Collinson, Clouthier et al. 1996)
	Chaperone-Usher pili	Lpf	IECs	(Collinson, Clouthier et al. 1996)
	Afimbrial adhesin-TAA	MisL	Fibronectin, IECs	(Dorsey, Laarakker et al. 2005)
	Chaperone-Usher pili	Pef	IECs	(Nicholson and Low 2000)
	Afimbrial adhesin	RatB	IECs	(Kingsley, Humphries et al. 2003)
	Afimbrial adhesin	ShdA	IECs	
	Afimbrial adhesin	SinH	IECs	
	Chaperone-Usher pili	FimH	DCs	(Muscas, Rossolini et al. 1994)
DAEC	Afimbrial adhesin	AIDA-I	IECs	(Servin 2005)
EAEC	Aggregative fimbriae	AfaE	IECs	(Lalioui, Jouve et al. 1999)

Species	Adhesive structure	Adhesin	[1]Ligand	Reference(s)
ETEC	CFA fimbriae	CfaE	IECs	(Gaastra and Svennerholm 1996; Li, Poole et al. 2009)
EPEC	Bundle-forming pili	bfpA	IECs	(Cleary, Lai et al. 2004)
	Afimbrial adhesin (Intimin)	eae	IECs	(Cleary, Lai et al. 2004)
	T3SS	espA	IECs	(Cleary, Lai et al. 2004)
EHEC	Chaperone-Usher pili	ECPa	IECs	(Rendon, Saldana et al. 2007)
	Afimbrial adhesin	Efal	IECs	(Badea, Doughty et al. 2003)
	Afimbrial adhesin (Intimin)	eae	IECs	(Fitzhenry, Pickard et al. 2002)
	T4P	HCP	Fibronectin, laminin	(Xicohtencatl-Cortes, Monteiro-Neto et al. 2009)

[1]Abbreviations are as follows: IECs, intestinal epithelial cells; TAA, trimeric autotransporter; T3SS, type 3 secretion system; T4P, type 4 pili.

1.6 Commercial interest in human gut commensals

1.6.1 Probiotics and prebiotics

Probiotic strains have been defined by the World Health Organisation as 'live micro-organisms which, when administered in adequate amounts, confer a health benefit on the host' (Schlundt and Pineiro, 2001). There is an increasing interest in research programs to explore their beneficial effects along with prebiotics, the latter being indigestible substrates for the host that stimulate the growth of commensal bacteria (e.g. fructans such as inulin or fructo-oligosaccharides). Prebiotics can also be produced by commensals. EPSs, for instance, promote the growth of other beneficial microbes from the microbiota such as bifidobacteria (Korakli, Gänzle et al. 2002).

Some criteria have been proposed for the development of probiotics for human consumption. The strains should be from human origin, safe, viable in the GIT, acid and bile stable, clinically demonstrated to be beneficial and able to adhere to the mucosa (Culligan, Hill et al. 2009). So far the scientific evidence on their beneficial aspects resides in results coming from animal models or *in vitro* studies as well as small-scale clinical trials in humans. Understanding the adhesion mechanism from probiotic strains would be

of importance since these are considered transient in the GIT or temporary residents (O'Hara and Shanahan 2007). Strains from lactobacilli and bifidobacteria, generally considered as safe (GRAS), are the most relevant to the probiotic industry. They have been the subject of numerous mucus adhesion studies and their EPSs are the most studied in commensals (Marcobal, Southwick et al. 2013).

Several diseases, including inflammatory bowel diseases, diarrheal and infectious diseases, auto-immune illnesses, heart and hepatic diseases, and metabolic disorders, could theoretically be prevented or treated using bacterial therapy by combining the action of pro- and prebiotics (synbiotic effect) (Candela, Perna et al. 2008). Synbiotic formulas could also be develop to supplement bovine milk to feed pre-term infants or babies that do not get the benefits of human breast milk (e.g. bifidogenic effect of HMOs) (Boehm, Jelinek et al. 2004). Similarly, selected pro- and prebiotics could help in enriching or limiting the alteration of the microbiota observed in ageing (Tiihonen, Ouwehand et al. 2010).

Lactobacillus rhamnosus strain 'GG' ATCC 53103 (LGG) has been the most studied probiotic species studied so far with over 500 publications listed in Pubmed (as of June 2013). LGG was isolated by Drs Goldin and Gorbach from a healthy human (feces) following certain criteria such as acid stability, adhesion to mucus and antimicrobial activity. The patented strain (Patent US4839281, New England Medical Centre Inc.) has been commercially developed since 1990 as a probiotic by the Finnish company Valio Ltd. LGG scientifically proven health benefits include the prevention and cure of acute diarrhea in children and a prophylactic role in antibiotic-associated diarrhea (Doron, Snydman et al. 2005). Other health claims relate to the treatment of relapsing *Clostridium difficile* colitis (Silva, Jacobus et al. 1987) and the prevention of atopic diseases (Kalliomaki, Salminen et al. 2001; Kalliomaki, Salminen et al. 2003).

1.6.2 Other commercial applications

In silico approaches could be used to identify useful commensal enzymes. Proteomes could be searched for glyco-related enzymes such as GTs or GHs and proteins produced in bacterial expression systems. GHs from mucin-degrading commensals such as *Akkermansia sp.* or HMO-scavenger *Bifidobacterium longum* subsp. *infantis* could be of great interest in glycobiology.

1.7 Aim of this thesis

The overall scope of this project was to advance knowledge on commensal glycobiology, especially to fulfill the present gap on the effector molecules involved in the interaction between commensal bacteria and the lining of the gut. Recent technological advances in glycomics (databases, synthesis of glycoconjugates and glycosylation-related high troughput assays) have provided new tools to study lectin-glycan interactions. Lectin and glycan microarrays were developed to facilitate this goal and tested on species from three prominent commensal genera of the human GIT (*Faecalibacterium*, *Bifidobacterium* and *Lactobacillus)*. The methods developed and used to achieve this goal are detailed in Chapter 2.

The first objective was to use an *in silico* approach to identify adhesins in representative species of commensals. We analysed the proteomes of these commensals using mathematical algorithms for the prediction of protein subcellular localisation sites. The predicted surface exposed proteins were further analysed by searching conserved domains to identify candidates that could represent adhesins with lectin functionality. A transcriptomic approach was then used to study the expression of the adhesin candidates in various conditions. The results of this work are found in Chapter 3.

The second objective was to discover and identify the glycan epitopes of selected bacterial adhesins. Labelled whole cells and isolated pili from selected strains were incubated on microarrays containing glycan structures found in the GIT. The results of this work are described in Chapter 4.

The third objective was to demonstrate that lectin microarray technology could be used to profile commensal species to provide information on cell surface glycosylation or could be used as a screening tool in search for novel EPSs with desirable properties. Lectin signatures of whole cells were first determined. EPS from these strains were isolated using two different methods. Commensal EPSs were subsequently labelled using a novel protocol and lectin binding profiles investigated. To support these profiles, monosaccharide analysis were finally carried out on isolated exopolysaccharides. The results from this work are detailed in Chapter 5.

Chapter 2: Materials and methods

2.1 *In silico* analyses

2.1.1 Analysis of human glycans from the gastro-intestinal tract

The 'Consortium for Functional Glycomics' (CFG, USA) funded by the US National Institute for Health (NIH) provides comprehensive online resources based on analytical-research on glycans. The consortium conducted profiling of glycans derived from mammalian glycoproteins and glycolipids using several analytical methods including mass spectrometry. Glycan databases available from its online gateway (http://www.functionalglycomics.org) were used to draw up a list of specific human intestinal glyco-conjugates, which could be potential targets for bacterial adhesins. Keywords, such as 'gastro-intestinal tract', 'intestine, 'colon' or 'mucosa', were entered in the glycan search entry and an inventory of glyco-conjugates obtained. Information related to their compositions was also used to define the structural groups of glycans, the families, whether glycosphingolipids, *O*-linked glycans or *N*-linked glycans. Details of glycan sub-families were likewise available (See Appendix).

2.1.2 Analysis of surfaceomes

Three reference strains, *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302, *Bifidobacterium longum* subsp. *infantis* ATCC 55813 and *Faecalibacterium prausnitzii* A2-165 were selected from the NIH-funded Human Gut Microbiome Project (http://genome.wustl.edu/projects/detail/human-gut-microbiome). Their proteomes were analysed using SurfG+, predictive software for surface exposed protein localisation. The UNIX-based tool was developed for protein localisation and identification of surface exposed loops that are of interest for vaccines development. The analysis follows a flow-scheme of logical steps using several bioinformatics packages programmed in Java (Barinov, Loux et al. 2009). The authors have validated this approach on a Gram+ bacterium, *Streptococcus pyogenes* and on lactobacilli to compare strains from the human gastro-intestinal tract (GIT) and from yogurt (dairy strains). Starting from a protein sequence in a FASTA format, the procedure follows a series of logical steps (Figures 2.1 and 2.2) to test the protein sequentially through a transmembrane helix predictor (TMMOD) (Kahsay, Gao et al. 2005), a secretion signal predictor (SignalP) (Petersen, Brunak et al.), a lipoprotein signal predictor (LipoP) (Juncker, Willenbrock et al. 2003) and a sequence alignment for protein profiles (HMMER) (http://hmmer.org). Arrangements were made with the Département de Mathématique informatique et genome, INRA, Jouy-en-Josas, France, to analyse our three potential probiotic strains: *Lactobacillus paracasei* subsp. *paracasei*, *Bifidobacterium longum* sp. and *Faecalibacterium prausnitzii* (Table 2.1). Results from SurfG+ were presented as spreadsheet tables and contained the classification of the proteins into predicted cell compartments whether secreted, found in the cytoplasm, within the cell wall, or exposed on the bacterial surface (surfaceome). The predicted

surface exposed proteins (SEPs) were then analysed using BLASTp algorithms from NCBI (http://www.ncbi.nlm.nih.gov) to find homologs to described adhesins. Sequences were also entered in the Conserved Domains Database (http://www.ncbi.nlm.nih.gov) to investigate known motifs found on adhesins such as the mucus binding domain (MucBP). Shortlisted putative adhesins were then refined further by comparison to a list of functionally characterised adhesins from the literature (Chapter 1, section 1.5.2).

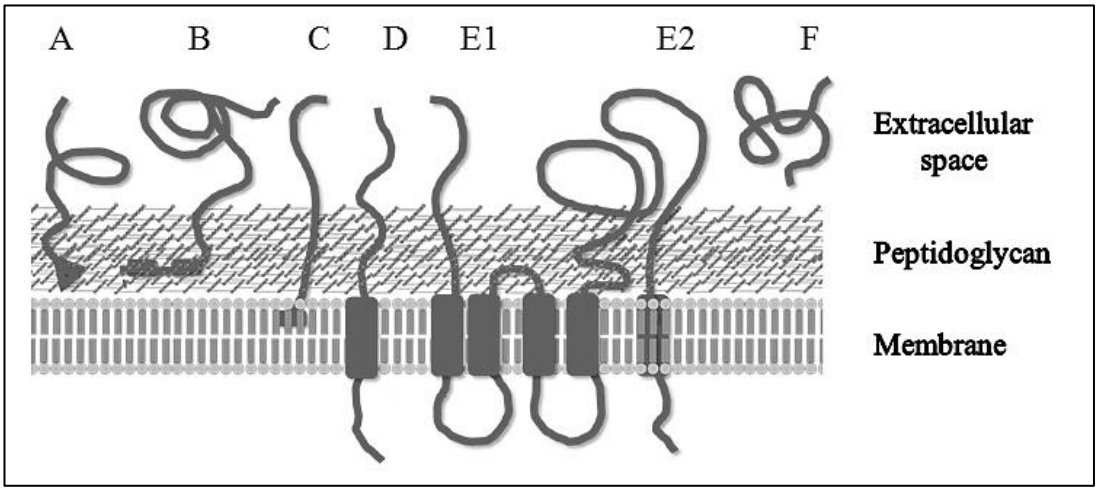

Figure 2.1: Surface exposed proteins on a Gram positive cell wall as predicted by SurfG+

A: Surface exposed proteins (SEP) covalently bound to the cell wall; B: SEP non covalently bound to the cell wall; C: membrane-bound lipoprotein; D: membrane anchored protein through N or C-terminal trans-membrane helix (TMH); E: membrane anchored through several trans-membrane helixes (TMH); E1: surface exposed N or C-terminal end; E2: surface exposed loop; F: secreted protein. Adapted from Barinov et al. (Barinov, Loux et al. 2009)

Table 2.1: List of genomes submitted to SurfG+ program

GenBank accession	Strain	Genome size (bp)	Number of proteins
NZ_ACGY00000000	*Lactobacillus paracasei* subsp. *paracasei* ATCC 25302	2,885,587	3042
NZ_ACHI00000000.1	*Bifidobacterium longum* subsp. *longum* ATCC 55813	2,372,858	2109
NZ_ACOP00000000	*Faecalibacterium prausnitzii* A2-165	3,080,849	3475

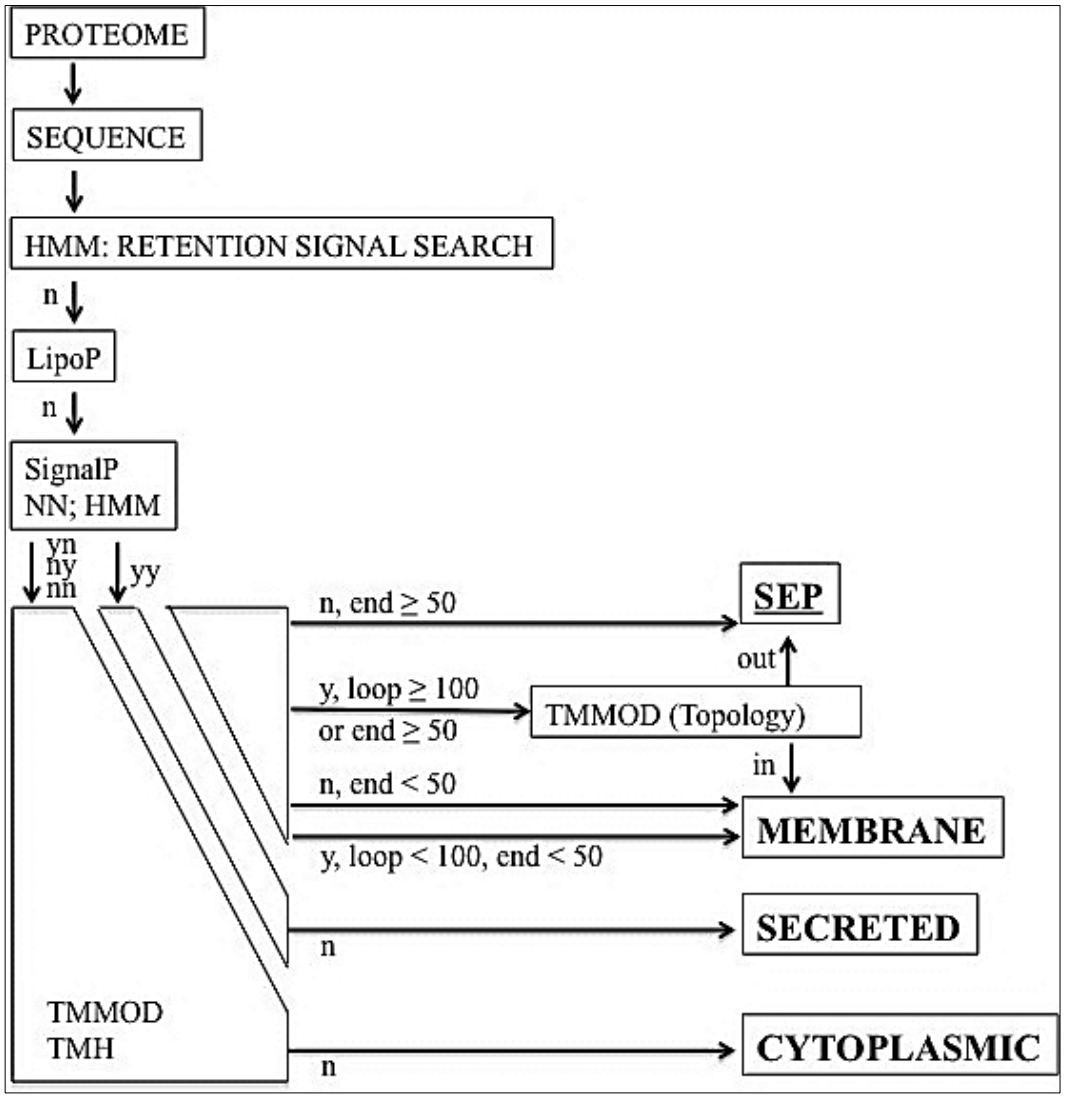

Figure 2.2: SurfG+ packages flow scheme

Protein sequences were analysed through a series of logical decisions represented by the arrows and classified into four categories (surface-exposed protein (SEP), Membrane, Secreted or Cytoplasmic). 'y' refers to a positive search for the feature and 'n' to a negative search. For SignalP, the first character represents the result of the NN method and the second character the result of the HMM method. Mature proteins were analysed using TMMOD. Commas represent the Boolean expression "AND"; 'end' and 'loop' cut off values are expressed as numbers of amino acids; 'in', 'end' or 'loop' are predicted at the cytoplasmic side of the cell membrane; 'out', 'end' or 'loop' are predicted outside the cytoplasmic membrane. Adapted from Barinov et al. (Barinov et al., 2009).

2.2 Bacterial strains

2.2.1 Type strains and clinical isolates

The bacteria used in this study were all human commensals from the gastro-intestinal tract (GIT) listed in Table 2.2. Stocks were either purchased through Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany) or Laboratorium voor Microbiologie Universiteit, Gent (BCCM/LMG Bacteria Collection, Gent, Belgium).

Table 2.2: List of strains

Species name	Identification	Human origin	Description	References
Bifidobacterium breve	UCC2003	infant stool	Gram positive	(Fanning, Hall et al. 2012)
Bifidobacterium longum* subsp. *infantis	DSM 20088 (ATTC 15697)	infant intestine	Gram positive	(Mattarelli, Bonaparte et al. 2008)
Faecalibacterium prausnitzii	DSMZ 17677 (A2-165)	fecal sample	Gram positive	(Duncan, Hold et al. 2002)
Lactobacillus paracasei* subsp. *paracasei	DSMZ 5622 (ATTC 25302)	milk	Gram positive	(Collins, Phillips et al. 1989)
Lactobacillus johnsonii	LMG 9436 (ATTC 33200)	blood	Gram positive	(Fujisawa, Benno et al. 1992)
Lactobacillus rhamnosus GG	LMG 18243 (ATCC 53103)	faeces	Gram positive	(Silva, Jacobus et al. 1987)

2.2.2 Exopolysaccharide ‘EPS’ mutants

EPS mutants, derived from *Bifidobacterium breve* UCC2003 wild type (WT), were provided by Prof. Douwe Van Sinderen, Alimentary Pharmabiotic Centre, UCC, Ireland (Table 2.3). *In silico* analysis performed by our collaborators on the *B. breve* UCC2003 genome identified an EPS operon (termed ‘eps’), with an unusual *modus operandi.* This strain is able to produce two types of EPS, the synthesis of which is controlled by either half of a bidirectional gene cluster. Thus, an alternate EPS synthesis is possible by means of reorienting its promoter (Figure 2.3). The alternative ‘inverted’ EPS producer was isolated and termed INV in this study. Mutants were also constructed by knocking out parts of the cluster by homologous recombination using a non-replicative plasmid (O'Connell Motherway, O'Driscoll et al. 2009). Briefly, DNA was inserted to disrupt the transcription of genes encoding glycosyltransferases (Bbr_0430, Bbr_0440 and Bbr_0441). The expected expression product would be different EPSs with shorter repeating units (SF440 and SF441) or no EPS expression (SF430). In this case, the insertion mutant was engineered in the monocistronic Bbr_0430 gene, which encodes the priming glycosyltransferase. The SF430 clone presented a different phenotype since it appears as ‘sediment’ in broth culture (Figure 2.4). In addition, a ‘spontaneous’ mutant presenting phenotypic resemblance to SF430 was fortuitously isolated (DEL). Genome analysis revealed it had lost the DNA region between two identical insertion sequence elements encompassing the ‘eps’ operon.

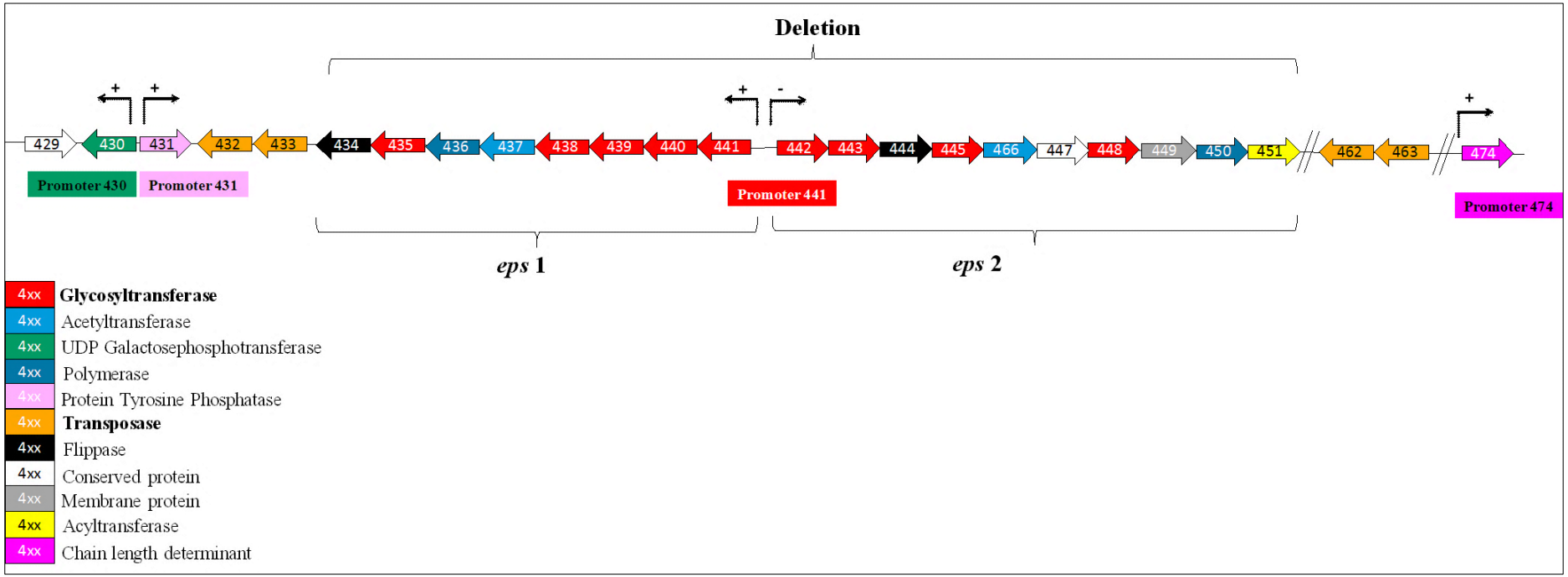

Figure 2.3: 'eps' gene cluster from *Bifidobacterium breve* UCC2003

Bidirectional gene cluster in *Bifidobacterium breve* UCC2003, enabling the strain to produce two types of EPSs: the 'wild' type EPS and, an alternate EPS (INV) synthesised by means of promoter reorientation. Each gene is represented as an arrow. The various glycosyltransferases are coloured in red.

Adapted from Fannin et al. (Fanning, Hall et al. 2012)

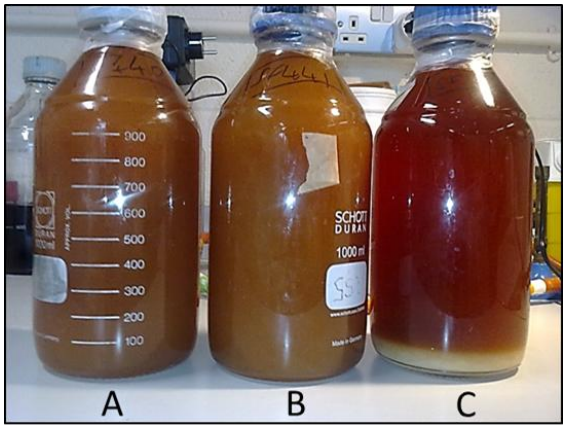

Figure 2.4: Phenotypic characteristics from *B.breve* UCC2003 mutants

1L cultures of *B. breve* UCC2003 with clone SF440 (A), SF441 (B) and showing sedimentation of cells in the non-EPS producer SF430 clone (C).

Table 2.3: List of *Bifidobacterium breve* UCC2003 'EPS' clones

Isogenic derivatives	Description	Mutation	EPS
WT	Wild type	None	Type 1 EPS
INV	Wild type mutant	Alternate transcription	Inverted EPS - type 2 EPS
DEL	Wild type mutant	Unknown (possible transposition)	Spontaneous non-EPS producer
SF440	Insertion mutant	Glycosyltransferase Bbr_0440	Type 3 EPS
SF441	Insertion mutant	Glycosyltransferase Bbr_0441	Type 4 EPS
SF430	Insertion mutant	Priming glycosyltransferase Bbr_0430	Non EPS-producer

2.2.3 *Lactococcus lactis* clones

L. lactis is a non-pathogenic strain ideal for producing surface exposed or secreted proteins due to its low extracellular proteolytic activity and it permits functional studies using either whole cell or membrane fractions. The standard host strain for nisin-inducible vectors is *Lactococcus lactis* NZ9000 [pepN::nisRnisK], based on *L. lactis* subsp. *cremoris* MG1363, which is a plasmid-free offshoot of wild-type strain NCDO712 (Gasson 1983). NZ9000 has been engineered to contain the nisin regulatory genes nis*R* and nis*K* integrated into its chromosome within the lysyl-aminopeptidase (pep*N*) gene. One nisin-inducible vector is pNZ8150, based on the pSH71 rolling circle replicon (De Vos 1987). It is a wide-ranging host vector containing the chloramphenicol resistance-encoding gene (CmR) and the nis*A* promoter, followed by the ScaI restriction enzyme site, enabling the downstream insertion of blunt end PCR fragments, exactly at the ATG translational start codon (De Ruyter, Kuipers et al. 1996; Mierau and Kleerebezem 2005). Nisin acts as an inducing peptide in the transformed NZ9000 cells via a two-component regulatory system. The bacteriocin nisin is first sensed by the histidine protein kinase (nis*K*) that autophosphorylates, then, in turn, provokes the phosphorylation of the response regulator (nis*R*). Subsequently, nis*R* acts as a transcription factor on the promoter nis*A* and results in overexpression of the protein of interest inserted in the cells (Van der Meer, Polman et al. 1993). Plasmid pNZ8150 was used for cloning the bifidobacteria pili clusters represented on Figure 2.5 (*B. bifidum* PRL2010 pili). The genetic engineering was performed in UCC, as described before, by Dr Mary O'Connell-Motherway (O'Connell Motherway, O'Driscoll et al. 2009; Turroni, Serafini et al. 2013). The nisin producing culture and the engineered clones are listed in Table 2.4.

Table 2.4: Strain used in the NICE® lactococcal expression system

***Lactococcus lactis* clones**	**Description**
Strain NZ9700	nisin producing culture
NZ9000-pNZ8150-1709-1707-5	expression *of B. bifidum* pil2
NZ9000-pNZ8150-1709-1707-19	expression *of B. bifidum* pil2
NZ9000-pNZ8150-1709-1707-47	expression *of B. bifidum* pil2
NZ9000-pNZ8150-282-283+pNZEM-284-1	expression *of B. bifidum* pil3
NZ9000-pNZ8150-282-283+pNZEM-284-2	expression *of B. bifidum* pil3
NZ9000-pNZ8150-282-283+pNZEM-284-3	expression *of B. bifidum* pil3
NZ9000-pNZ8150-282-283+pNZEM-284-4	expression *of B. bifidum* pil3
NZ9000 pNZ8150+pPTPi	control

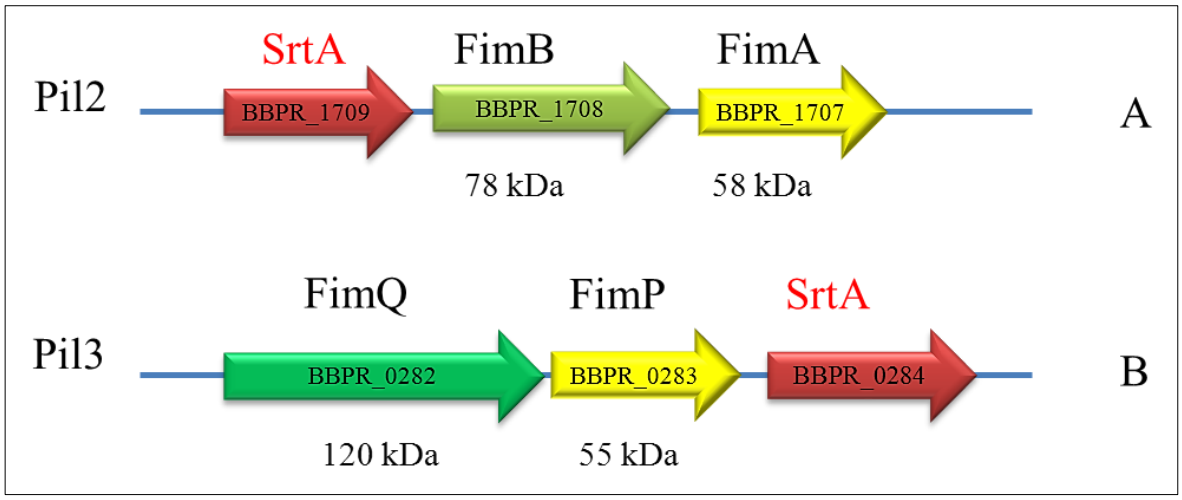

Figure 2.5: *B. bifidum* PRL2010 pili clusters

A: Schematic of *B. bifidum* PRL2010 Pil2 gene cluster featuring the sortase A (srtA), the putative major pilin (FimA) and the suspected minor pilin (FimB), NCBI gene ID and expected protein sizes are shown. B: Pil3 gene cluster featuring the sortase A (srtA), the putative major pilin (FimP) and the suspected minor pilin (FimQ), NCBI gene ID and expected molecular mass.

2.2.4 *Escherichia coli* strains

Two *E. coli* strains (Table 2.5) were used as positive controls to test the neo-glycoconjugate microarrays. They were obtained from Dr Aoife Boyd, Dept. of Microbiology, NUIG. Cells from *E. coli* ATCC 25922 express P fimbriae adhering to galabiose, whereas cells from *E. coli* ATCC 35218 produce type 1 fimbriae binding mannose.

Table 2.5: *E. coli* strains used for the neo-glycoconjugate microarrays

Strain	Origin	Description
***Escherichia coli* ATCC 35218**	Canine	Production of type 1 fimbriae - Quality control strain for bioMerieux.
***Escherichia coli* ATCC 25922**	Clinical isolate	Production of P fimbriae - Quality control strain for bioMerieux.

2.3 Culturing conditions

All media were purchased and prepared following manufacturer's instructions except Luria Bertani media (LB) (home-made) and M17 media. The latter was not supplemented with lactose but with sterile 0.5% glucose instead. Media were also supplemented when required with selective antibiotics including chloramphenicol (Sigma-Aldrich, Ireland, #C0378), tetracycline (Sigma-Aldrich, Ireland, #T7660) or erythromycin (Sigma-Aldrich, Ireland, #E5389).

2.3.1 Anaerobic culturing

Routinely, bacteria were grown on agar plates (Table 2.6) anaerobically in sealed pouches (BD GasPak™ EZ Anaerobe Systems, UK, #260683) for 48h at 37°C except *F. prausnitzii* (24h incubation). Growth curves as well as some experiments, requiring a particular growth phase, were performed in liquid culture. For this purpose, colonies were transferred to relevant broth media (Table 2.7) supplemented with 0.05% (w/v) L-cysteine (Sigma-Aldrich, Ireland, #C6852) to keep a reduced environment and adequate antibiotics added when required. Subsequently glycerol stocks were made up and stored at -80°C. Bacteria were passaged twice in broth to acclimatise the cells to their environment prior to using them in experiments.

Table 2.6: Growth conditions for the microarray 'whole cell' experiments

Strain / isogenic clone	Media	[1]Selection
B. longum* subsp. *infantis	Reinforced clostridial agar (Oxoid Ltd, UK, #CM0151)	none
F. prausnitzii	Wilkins-Chalgren agar (Oxoid Ltd, UK, #CM0619)	none
L. johnsonii	MRS agar (Oxoid Ltd, UK, #CM0361)	none
L. paracasei* subsp. *paracasei	MRS agar (Oxoid Ltd, UK, #CM0361)	none
L. rhamnosus GG	MRS agar (Oxoid Ltd, UK, #CM0361)	none
***B. breve* UCC2003**	Reinforced clostridial agar (Oxoid Ltd, UK, #CM0151)	none

Strain / isogenic clone	Media	[1]Selection
***B. breve* INV**	Reinforced clostridial agar (Oxoid Ltd, UK, #CM0151)	Cm 3µg/ml
***B. breve* SF430**	Reinforced clostridial agar (Oxoid Ltd, UK, #CM0151)	Tet 5ug/ml
***B. breve* SF440**	Reinforced clostridial agar (Oxoid Ltd, UK, #CM0151)	Tet 5ug/ml
***B. breve* SF441**	Reinforced clostridial agar (Oxoid Ltd, UK, #CM0151)	Tet 5ug/ml
***B. breve* DEL**	Reinforced clostridial agar (Oxoid Ltd, UK, #CM0151)	Cm 3µg/ml

[1]Cm: chloramphenicol , Em: erythromycin and Tet: tetracyclin.

2.3.2 Exopolysaccharides production

All commensal strains and mutants were first revived on agar plates (Table 2.2) for 48h at 37°C except *F. prausnitzii* (24h incubation). Colonies from the plate were transferred to a 20 mL syringe containing required broth media, adequate selection and incubated in static conditions for 8h at 37°C (Table 2.7). The content was then injected to a 250mL glass bottle (DURAN, Germany, #1127076) and incubated overnight under the same conditions. Finally, a 10% inoculum was transferred to a 1L glass bottle (DURAN, Germany, #1127076) and incubated for a further 48h under similar conditions.

Table 2.7: Growth conditions for the exopolysaccharide experiment

Strain / isogenic clone	Media	[1]Selection
B. longum* subsp. *infantis	MRS broth (Oxoid Ltd, UK, #CM0359)	none
F. prausnitzii	Wilkins-Chalgren broth (Oxoid Ltd, UK, #CM0643)	none
L. johnsonii	MRS broth (Oxoid Ltd, UK, #CM0359)	none
L. paracasei* subsp. *paracasei	MRS broth (Oxoid Ltd, UK, #CM0359)	none
L. rhamnosus GG	MRS broth (Oxoid Ltd, UK, #CM0359)	none
***B. breve* UCC2003**	MRS broth (Oxoid Ltd, UK, #CM0359)	none
***B. breve* INV**	MRS broth (Oxoid Ltd, UK, #CM0359)	Cm 3µg/ml
***B. breve* SF430**	MRS broth (Oxoid Ltd, UK, #CM0359)	Tet 5ug/ml

Strain / isogenic clone	Media	[1]Selection
***B. breve* SF440**	MRS broth (Oxoid Ltd, UK, #CM0359)	Tet 5ug/ml
***B. breve* SF441**	MRS broth (Oxoid Ltd, UK, #CM0359)	Tet 5ug/ml
***B. breve* DEL**	MRS broth (Oxoid Ltd, UK, #CM0359)	Cm µg/ml

[1] Cm: chloramphenicol and Tet: tetracyclin.

2.3.3 Culturing of *Escherichia coli* strains

Luria Bertani (LB) media was made by adding to 900mL of distilled water, 5g yeast extract (Sigma-Aldrich, Ireland, #Y0875), 10g tryptone (Sigma-Aldrich, Ireland, #T9410) and 10g NaCl (Sigma-Aldrich, Ireland, #S7653) then adjusting the volume with water to 1L. Agar was added (5g/L) for the preparation of LB agar plates. *Escherichia coli* strains were grown on LB agar plates overnight at 37°C. Colonies from the plate were then transferred to 50mL centrifuge tubes (Sarstedt, Ireland, #62.548.004) containing 50mL of LB broth and incubated in static condition at 37°C overnight. A 2% inoculum was then transferred to fresh LB media and bacteria grown overnight again in similar conditions prior to experiments.

2.3.4 Aerobic culturing of *Lactococcus lactis* clones

L. lactis clones were plated on M17 agar (Oxoid Ltd, UK, #CM0785) supplemented with 0.22µm sterile-filtered 0.5% glucose (Sigma-Aldrich, Ireland, #1181302) and associated selective antibiotics (Table 2.8) and grown in aerobic condition at 30˚C for 48h. A colony was picked from each plate and transferred to an eppendorf tube containing 1mL of M17 broth supplemented with 0.5% glucose plus relevant antibiotics and placed at 30˚C overnight in static conditions. The culture was then transferred to a 50mL tube containing 24mL of similar medium and grown overnight in similar conditions. Since the production of nisin is best at stationary phase, cultures from strain NZ9700, producing this antibacterial peptide, were harvested at this stage by centrifugation (8000*g*, 5min, 4˚C), the supernatant retained and sterile-filtered through 0.22µm syringe filters (Millipore, Ireland, #SLGP033NS) to eliminate any remaining cells. A 2% inoculum from all other overnight cultures was then transferred to fresh M17 media, also supplemented with 0.5% glucose plus relevant antibiotics, and grown to exponential phase in similar conditions (OD_{600} 0.5, during approximately 2h 45min). At this point, cells were induced by adding nisin to each culture (2µL/mL of media), the tubes gently swirled and returned to the 30˚C incubator

until harvest.

Table 2.8: Growth conditions of *L. lactis* strains

***Lactococcus lactis* clones**	**Media**	**[1]Selection**
Strain NZ9700	GM17	Cm 5 µg /ml
NZ9000-pNZ8150-1709-1707-5	GM17	Cm 5 µg /ml
NZ9000-pNZ8150-1709-1707-19	GM17	Cm 5 µg /ml
NZ9000-pNZ8150-1709-1707-47	GM17	Cm 5 µg /ml
NZ9000-pNZ8150-282-283+pNZEM-284-1	GM17	Cm 3 µg /ml, Em 3 µg /ml
NZ9000-pNZ8150-282-283+pNZEM-284-2	GM17	Cm 3 µg /ml, Em 3 µg /ml
NZ9000-pNZ8150-282-283+pNZEM-284-3	GM17	Cm 3 µg /ml, Em 3 µg /ml
NZ9000-pNZ8150-282-283+pNZEM-284-4	GM17	Cm 3 µg /ml, Em 3 µg /ml
NZ9000 pNZ8150+pPTPi	GM17	Cm 3 µg /ml, tet 5 µg /ml

[1]Cm: chloramphenicol, Em: erythromycin and Tet: tetracyclin.

2.4 Adhesin gene expression

2.4.1 Bacterial growth curves

Bifidobacterium longum subsp. i*nfantis* ATCC 15697 was purchased from DSMZ, Germany (#DSM20088). Reinforced Clostridial Agar (RCA, Oxoid, UK #CM0151) plates were streaked from glycerol stocks, placed in an anaerobic container with gas generating sachets (BD GasPak™ EZ Anaerobe Pouch System, UK, #260683) and incubated for 48h at 37˚C. Bacteria were subsequently grown in broth in syringes to keep the conditions anaerobic. One colony was picked and transferred to a 1mL syringe containing De Man, Rogosa, Sharpe broth (MRS, Oxoid, UK, #CM0359) and incubated overnight at 37˚C. A 5% inoculum was passaged a second time in MRS broth in a 20mL syringe to alleviate stress and incubated overnight at 37˚C. A 5% inoculum was then used to inoculate 50mL syringes containing regular MRS media, MRS supplemented with 0.22µm sterile-filtered 0.1% mucin from porcine stomach (PGM, Sigma-Aldrich, Ireland, #M1778), MRS supplemented with the salt gut osmolarity NaCl 0.3M (Sigma-Aldrich, Ireland, #S3014) or both 0.1% PGM and 0.3M NaCl. Growth was measured over time by optical density measurements (600nm) using dilutions when appropriate (Biophotometer, Eppendorf, Germany, # 6132 000.008) and pH monitored at the same time points (Symphony, VWR, Ireland, #89231-664).

Lactobacillus paracasei subsp. *paracasei* ATCC 25302 was purchased from DSMZ, Germany (DSM5622). Bacteria were grown as described above.

Faecalibacterium prusnitzii A2-165 was purchased from DSMZ, Germany (DSM17677). Bacteria were grown similarly to *B. longum* subsp. *infantis* but were grown on Wilkins-Chalgren anaerobe agar plates (WCA, Oxoid, UK, #CM0619) or Wilkins-Chalgren anaerobe broth (WC, Oxoid, UK, #CM0643) and incubated for 24h at 37˚C.

2.4.2 RNA isolation

The RNA extraction procedure was adapted from Karatzas et al. (Karatzas, Brennan et al. 2010). Three biological replicates of bacteria were plated, grown in broth as described above then harvested at lag phase or log phase. Following initial growth curves, the lag and log phases were determined to be respectively T=2h and T=6h for *Faecalibacterium prausnitzii*, T=3h and T=7h for *Bifidobacterium longum* subsp. i*nfantis* and T=3h and T=8h for *Lactobacillus paracasei* subsp. *paracasei.* Samples (20mL) were immersed in 50ml tubes containing 6mL of a solution composed of 5% phenol (v/v) and 95% ethanol (v/v) to stop RNA synthesis and to prevent the degradation of the synthesised RNA. Tubes were incubated on ice for 15min and subsequently centrifuged (5000*g*, 10min, 4˚C then 9000*g*, 3min, 4˚C). The supernatants were discarded and pellets immediately stored at -80˚C until extraction. Mechanical disruption of the cells was performed using a 3min homogenisation cycle in a bead beater (Mini Beadbeater™, Biospec products, USA, #607) repeated twice with the samples left on ice for 5min between cycles to prevent overheating and RNA degradation. Total RNA isolations were carried out using a commercial kit following the manufacturer's protocol (RNEasy Mini Kit, Qiagen, UK, #74104) including an on-column DNase I step (RNase-Free DNase Set, Qiagen, UK, #79254). The DNase step was repeated if required following the reverse transcription quality control step (RT negative).

2.4.3 RNA quality control

Total RNA from each biological replicate for the five conditions investigated was quantified and the 260/280 ratio estimated using the NanoDrop (ThermoScientific, Belgium, #NanoDrop2000). The RNA quality was assessed using the integration number (R.I.N) determined by a micro-fluidics based platform (2100 Bioanalyser, Agilent, Ireland, #G2943CA) using a specific assay (RNA6000 nano kit, Agilent Technologies, Ireland, #5067-1511) and following the manufacturer's protocol. This RNA6000 nano-assay consists of a chip that provides a quick assessment of the purity, concentration and integrity of up to 12 RNA samples. Several chips were used to check the R.I.N of all samples. When required (i.e. R.I.N<7.0), samples were re-run on a new chip. If a sample did not satisfy the threshold a second time, RNA isolation was repeated.

2.4.4 Reverse transcription

Reverse transcription was carried out using total RNA to generate first strand cDNA template for each sample and also by pooling all RNA from each strain to test gene-specific primers for the real time PCR (qPCR). The reverse transcription was performed with a commercial formulation providing enhanced cDNA synthesis efficiency (SuperScript VILO, Life Technologies, USA, #117574-250). A master mix was prepared as per manufacturer's protocol with 2µl RNA (corresponding to 1.5 to 2µg). A control for the carry-over of genomic DNA ('RT negative') was generated by using the same mix without the enzyme. Tubes were then incubated at 25°C for 10 minutes, then at 42°C for 60 minutes. The reverse transcription reaction was terminated by a last step at 85°C for 5 minutes. Synthesised cDNA products were stored at -20°C until use (qPCR).

2.4.5 Primers design

Oligonucleotides were designed to candidate genes using Primer3Plus software (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) with the advanced qPCR settings and amplicon size set between 100bp and 150bp (Tables 2.9-2.11).

Table 2.9: *Lactobacillus paracasei* subsp. *paracasei* qPCR primer sets

NCBI accession	Name	[1] Oligos
ZP_03965231.1	Elongation factor Tu	LEFTU_F GCGCATCAATATGTGCGTAG
		LEFTU_R TGCTGCTCCAGAAGAAAAGG
ZP_03963800.1	GROEL chaperonine	LGROEL_F CAAGGACAACACCACGATTG
		LGROEL_R TTTCACGGTCGAAGTCACTG
ZP_03965124.1	Fibronectin binding protein A	LFBPA_F AATGACACGTAGGGGATTGC
		LFBPA_R TGCGTTACGACTCAACCAAG
ZP_03964327	Collagen binding protein	LCnBP_F GCTCTTTAAACCGGTTGTGG
		LCnBP_R CACTGTTTCGCTCAATCTGG
ZP_03963300.1	Fimbriae subunit	Lpara5075_F TCTCGGGTTTAATGGCACTC
		Lpara5075_R GATCAGTCGCACGATCAATG
ZP_03963571.1	Adhesion exoprotein	Lpara28086_F ATTCCGACGCTGACAATACC
		Lpara28086_R CCTGAAAGCCCAATGTTGTC
ZP_03963572.1	Adhesion exoprotein	Lpara31701_F TGAAGCGTCAGGTCTCAATG
		Lpara31701_R CCGTCAACAACAATGTCCTG
ZP_03963620.1	Outer membrane protein	Lpara8786_F TCAACCACTGCAAAGTCCAC
		Lpara8786_R ATCACTGGCGAATTCCTGAC
ZP_03964000.1	Collagen adhesion protein	Lpara8692_f AGTAGCGTTGAATGGGATGG
		Lpara8692_R ATACCGGAACCCCAATAAGG
ZP_03964057.1	Cna B domain protein	Lpara48083_F TTGAGACGACTGCACCAAAG
		Lpara48083_R CGTGAGTGCATTGGTATTGG
ZP_03964058.1	Cell wall surface anchor family protein	Lpara50135_F CAAGAAGTAAAGGCGCCAAC
		Lpara50135_R TCATCGCTGGCAACATAGAC
ZP_03963680.1	Possible outer membrane protein	Lpara2797_F ACTGGATTACCGCGAACAAG
		Lpara2797_R CAATCGGATCAACCCAGTTC
ZP_03963682.1	Pilus subunit protein	Lpara5677_F ACCTGTCGATTCAAGGGATG
		Lpara5677_R ATCTCAAAGTCCGTGGTTGG
AF182724	16S ribosomal RNA	LAF182724_F ATACATAGCTGGCCGGCGGC
		LAF182724_R CCCACTTCGCTCGCCGCTAC

[1] 'F' refers to the forward (sense) primer and 'R' the reverse (anti-sense) primer

Table 2.10: *Bifidobacterium longum* subsp. *infantis* qPCR primer sets

NCBI accession	Name	[1]Oligos
YP_002323370.1	Elongation factor Tu	BEFTU_F CGGTGATCATGTTCTTCACG
		BEFTU_R GCGGTATCACCATCAACATC
YP_002322176.1	GROEL chaperonine	BGROEL_F ATGCCAAGAAGGTCATCGTC
		BGROEL_R CAGCTTCTCACGGTCGTAATC
YP_002321787.1	LPXTG-motif cell wall anchor domain protein	BLSPA_F CCCAGATATGGGATCAATCG
		BLSPA_R CACCACTTTCACCAAGATGG
YP_002322242.1	Extracellular solute-binding protein, family 3	BCnBP_F ACGATTTTGGTCAGCAGGTC
		BCnBP_R CAATCTTGATGTCCGCATCC
YP_002323851.1	Extracellular solute-binding protein, family 5	BBOPA_F TTGGACTTGCCGTTGTACTG
		BBOPA_R CCTTCCAGAACGGTGAAATC
YP_002323915.1	G5 domain protein	Blon2813700_F CTCCACACAAACCATCATCG
		Blon2813700__R GTTTCGGATTTCAGCGTCTC
YP_002322905.1	Narrowly conserved hypothetical protein	Blon1603301_F ATCCCGGTACCGATTTATGG
		Blon1603301_R GCGAAATTCAGGGACGTATC
YP_002322015.1	LPXTG-motif cell wall anchor domain protein	Blon596214_F TGGTGTTGGTGATGGTGAAC
		Blon596214_R GCCTGAAGATCGAGAAATGG
YP_002322725.1	Allergen V5, possible large adhesin	Blon1451659_F AGAAGGCCAGTTACGAGAAGG
		Blon1451659_R ACCCCTTGTTGACGATGTTC
YP_002323586.1	CHAP domain containing protein	Blon2373318_F GTTGATCTTCTCGGCAATGG
		Blon2373318_R CAAGGGTTCTCATTCCGTTC
YP_002322168.1	Von Willebrand factor	Blon790303_F CGGCAATGTCTTTGAGGTTC
		Blon790303_R ATTCCAATACCACGGACCAC
YP_002322323.1	Peptidoglycan-binding LysM	Blon4886_F CGTGATGACTGGTTCAATGG
		Blon4886_R TAATCCTCCCACGCACATTC
YP_002323430.1	Hypothetical protein Blon_1985	Blon1985_F CACGTACTGGGTTTTCATCG
		Blon1985_R CGGTCACGATGGAATATGTG
BlonR0026	16S rRNA	BR0026_F AATCCGCTGGCAACACGGGG
		BR0026_R GGGCGGGTTCACAGGTGGTG

[1] 'F' refers to the forward (sense) primer and 'R' the reverse (anti-sense) primer

Table 2.11: *Faecalibacterium prausnitzii* qPCR primers sets

NCBI accession	Name	[1]Oligos
ZP_05614002	Elongation factor Tu	FEFTU_F TAGTGACGAGCATCGGTCTG
		FEFTU_R CATGGACTATGCCAACATCG
ZP_05614692	Fibronectin binding protein	FGROL_F TGATACCCTGGCAAATACCG
		FGROL_R AGCCATAGTTGGGCTTGTTG
ZP_05616399	Nickel ABC transporter, periplasmic nickel-binding protein	FBOPA_F TGACCTTGACCGTGATCTTG
		FBOPA_R CTATGTGCTGACGGATTTCG
ZP_05616027	Chaperonin GroL	FFBPA_F TGAAGAAGTTCTGGGCGTTC
		FFBPA_R AGGGAATGAAGAGCATCACC
ZP_05614429.1	Amino acid ABC transporter, periplasmic amino acid-binding protein	FCnBP_F CTTTGACATGGTCTGCAACG
		FCnBP_R TCTTGATCTCGTCGTTGTCC
ZP_05613528.1	Putative outer membrane autotransporter barrel	Fpraus28666_F TTGCGAAGATTCGTTCTGG
		Fpraus28666_R TGAAACAGGACCCTTTCTGG
ZP_05613970.1	Putative cell surface protein	Fpraus12278_F TCTTCAATCCGGATCACTCC
		Fpraus12278_R ACTGGTACAGGCGGAAATTG
ZP_05614164.1	Putative lipoprotein	Fpraus10789_F TGACATCGATAGCGACGAAC
		Fpraus10789_R TCATTGTCTGCGTTGGCTAC
ZP_05614788.1	Putative LysM domain protein	Fpraus22516_F CAGCCTGCGCATCTATTATG
		Fpraus22516_R ATGGCAGTAATTCCCTCTGC
ZP_05614977.1	Putative lipoprotein	Fpraus37846_F TGGACGGCGATATGAGTATG
		Fpraus37846_R TGTTGCCATCCAGGTCTTTC
ZP_05615565.1	Protein-export membrane protein	Fpraus21095_F AAGGCGAAGAACACGAAGTG
		Fpraus21095_R CGTGACCATCGTTATTGTGG
ZP_05615730.1	The GLUG motif protein	Fpraus29194_F GCAATGTGACCGTGAACATC
		Fpraus29194_R TTGAAAACCGTCTCCTCACC
ZP_05616094.1	Putative S-layer y domain protein	Fpraus18355_F GCAGCCGAAATTATCGTACC
		Fpraus18355_R AGATGGCAGACAATGTGCAG
ZP_05616166.1	Pilin	Fpraus22746_F AAATCAGGACTGCACGGAAG
		Fpraus22746_R TTGCAGAAGCAGGCACATAC

NCBI accession	Name	[1]Oligos
ZP_05616171.1	Pilin isopeptide linkage domain protein-like protein	Fpraus26651_F TGTACTATCTGCGCGTTTCC
ZP_05616171.1		Fpraus26651_R TGTTGAAGGTGATGCTGCTC
ZP_05616293.1	Putative fasciclin domain protein	Fpraus2792_F AACTGACCATCGGCGAATAC
		Fpraus2792_R GCGTGGATCAAATGATAGGC
ZP_05616615.1	Putative autotransporter/adhesin	Fpraus12467_F GATGGTGTTTGTGGTTGTGG
		Fpraus12467_R TAACTCCGACAATGGCACAG
AJ270469.2	16S rRNA	FAJ270469_F TGGGCTCAACCCATAAACTG
		FAJ270469_R TCCCGATATCTACGCATTCC

[1] 'F' refers to the forward (sense) primer and 'R' the reverse (anti-sense) primer

2.4.6 Real Time PCR

Real-time PCR was used to determine the relative expression at mid-log phase of the adhesin candidate genes for each condition using the lag phase as calibrator and the rRNA gene (16S) as reference gene. The conditions explored were various culture media (regular broth media, media supplemented with 0.1% mucin (PGM), media supplemented with NaCl 0.3M (gut osmolarity) or media supplemented with both mucin and NaCl). Real time PCR runs were performed in triplicate on three independent biological samples with the reference ribosomal gene on each plate. The two step-qPCR method was carried out using a SYBR Green® dye mix (QuantiTect SYBR® Green PCR Kit, Qiagen, UK, #204143) in a qPCR thermal cycler with FAM/SYBR® Green (492nm-516nm) and ROX/Texas Red (585nm-610nm) filters (Mx3000p, Stratagene, Agilent Technologies, Ireland, #401403). PCR reactions were slightly modified from manufacturer's protocol. They consisted of a 20µL total reaction volume containing 10µL of 2X Quantitect SYBR green PCR master mix, 1µL of 10µM sense primer, 1µL of 10µM anti-sense primer, 3µL of dH_2O and 5µL of cDNA template (~10ng/µL). The thermal profile was 95˚C for 15min, followed by 40 cycles of 94 ˚C for 15s, 60 ˚C for 30s, 72 ˚C for 30s. DNA melting curves were carried out at the end of each run using a ramping rate of 1 ˚C /min over a temperature range of 60 ˚C to 95 ˚C. Specificity of the amplicon was confirmed by the presence of a single peak. Ct values were normalised to the reference dye (ROX) to take into account pipetting errors and collected at the same threshold number to compare each plate. Data were then normalised to the housekeeping gene (16S rRNA) and fold changes between conditions

calculated following the 'delta delta Ct' method for relative quantification of gene expression (Livak and Schmittgen 2001) using the lag phase as the calibrator. Statistical tests were subsequently carried out in the form of analysis of variance (ANOVA) and post hoc Tukey test, between the four conditions (regular media, media supplemented with PGM, NaCl or both) for the whole set of candidates or between conditions for each putative adhesin.

2.5 Ultrastructure visualisation

2.5.1 Scanning electron microscopy

Bacteria were grown on relevant agar plates for 48h or in broth for 16h as previously detailed (section 2.3), washed twice with sterile TBS 1x (5000*g*, 2min, 4°C), and adjusted to an OD_{600} of 1 (approx. 10^8 cfu/ml) in TBS 1x. The cells were fixed with 2.5% gluteraldehyde (Sigma-Aldrich, Ireland, #G5882) overnight, followed by serial dehydration through graded ethanol solutions (Sigma-Aldrich, Ireland, #E7023): 30%, 50%, 70%, 80%, 90% ethanol with each step corresponding to a 10min incubation period. Pellets were then resuspended in 100% ethanol and placed on round coverslips (Sarstedt, Ireland, #83.1840.002) in a 12-well tissue culture plate and air-dried. Samples were subsequently treated twice with a drying agent, 100% hexamethyl disilizane for 15 min (HMDS, Sigma-Aldrich, Ireland, #379212) and allowed to air-dry for at least 2h in a fume hood. Samples were then fixed to metal stubs using carbon adhesive tabs (Agar Scientific, UK, #G3348N), gold sputter coated and viewed with the scanning electron microscope (Hitachi, Japan, S-4700).

2.5.2 Transmission electron microscopy

Bacteria were harvested at required time-points by centrifugation, washed twice in sterile TBS 1x (5000*g*, 2min, 4°C) and adjusted to OD_{600} 1 in TBS 1x (approx. 10^8 cfu/ml). A formvar carbon grid (Agar Scientific, UK, #S162) was placed shiny side towards a drop of 10µl of cells and left 1 minute to mount. Excess liquid was subsequently removed using silk paper and the grid mounted on a drop of 10µl negative staining solutions, either with phosphotungstic acid from 0.1% to 1% (PTA, Sigma-Aldrich, Ireland, #79690), Alcian blue 1% (Sigma-Aldrich, Ireland, #05500) or ammonium molybdate 1% (Sigma-Aldrich, Ireland, #A1343). Staining was carried out for 1min, excess liquid was then removed using

silk paper and the grid air-dried prior to visualisation. Samples were viewed and photographed using a transmission electron microscope (Hitachi, Japan, H-7500).

2.6 Isolation of exopolysaccharides (EPS)

The EPS isolation was adapted from Tallon et al. and Grant et al (Tallon, Bressollier et al. 2003; Grant 2011). Briefly, cells were first separated from the growth media. Then, either the cell pellet was treated with a chelating agent (EDTA) or the media was ultrafiltered through a 100kDa cut-off membrane. Proteins were then removed by TCA precipitation and the EPS recovered by ethanol precipitation. The preparations were treated with proteases and nucleases as described in more detail below (Figure 2.6). An MRS media sample was used as a negative control and processed identically to the bacterial samples throughout the EPS extraction procedure.

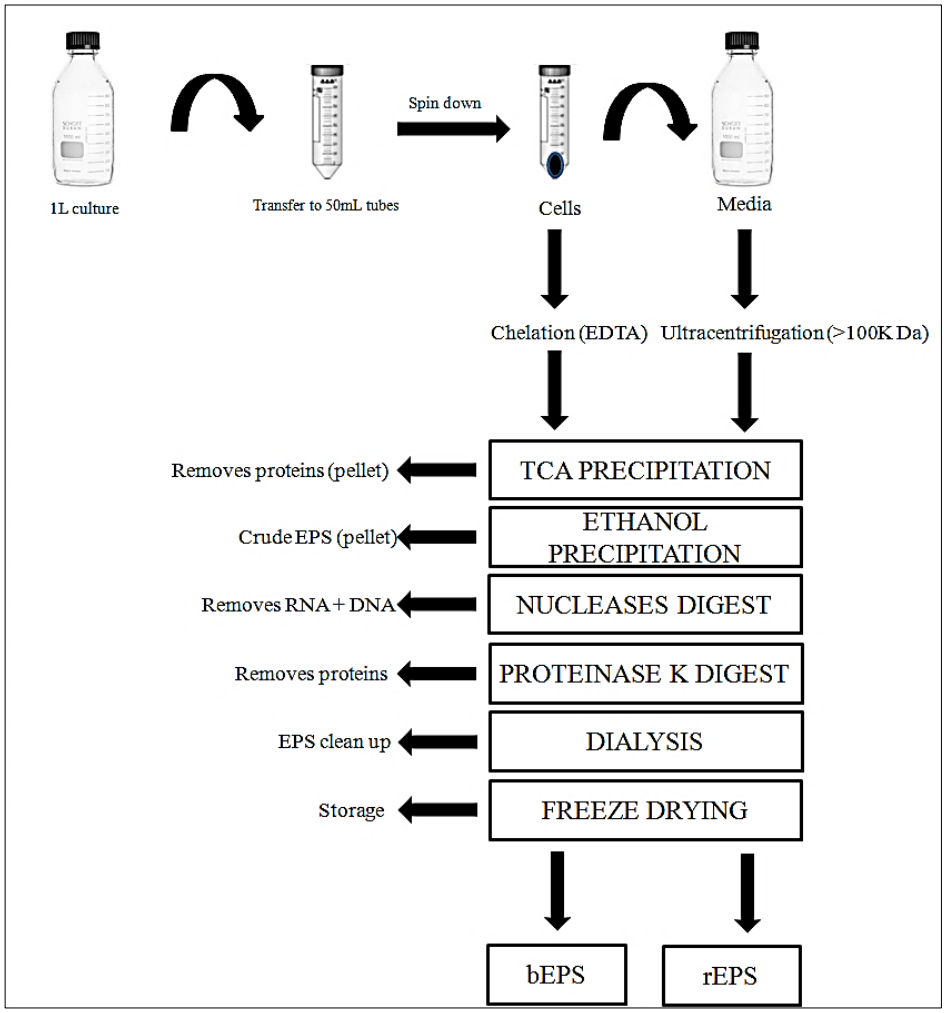

Figure 2.6: Schematic of EPS isolation procedures

Representation of the two methods used to extract cell-bound EPS (bEPS) by chelation of cells and released EPS (rEPS) by ultrafiltration of the culture media.

2.6.1 Cell-bound EPS isolation (bEPS)

One litre of bacterial culture was transferred to 50mL tubes that were centrifuged at 15000*g* (10min, 4 °C). The cell pellets were kept on ice, washed with 20mL TBS 1x, and centrifuged again (15000*g*, 10min, 4°C). Cell pellets were weighed and resuspended in 50mL EDTA 0.05M (Sigma-Aldrich, Ireland, #E6758) then incubated on ice for 4h on a rocking platform. Samples were centrifuged (8000*g*, 20 min, 4°C) to remove the cells. The supernatant containing the cell-bound EPS (bEPS) was precipitated by adding 20% (v/v) trichloroacetic acid (TCA, Sigma-Aldrich, Ireland, #T6399) and the mixture incubated overnight at 4°C. Samples were then centrifuged (8000*g*, 20 min, 4°C) to remove proteins. EPSs were recovered through precipitation of the supernatant by addition of two volumes of ice-cold ethanol (Fisher Scientific, Ireland, #10048291) followed by an overnight incubation at -20°C. The crude bEPS pellet was collected by centrifugation (8000*g*, 20 min, 4°C), resuspended in 2mL distilled water (dH_2O) before treatment with nucleases. The digestion procedure consisted of an overnight incubation in a water bath set at 37°C with RNase A (Sigma-Aldrich, Ireland, #R6513) and DNase I (Sigma-Aldrich, Ireland, #D7691), both at a final concentration of 0.1mg/mL. Following this step, proteinase K was added to a final concentration of 0.2mg/mL and the mixture incubated for three days in a water bath set at 37°C. The bEPS subsequently underwent dialysis (Pierce, UK, Slide-A-Lyzer Dialysis Cassettes, 3.5K MWCO, #66330) for two days against 5L of dH_2O at 4°C, with changes three times daily. The dialysed bEPS was then frozen at -80°C and freeze-dried. To remove possible contamination from glycolipids, the EPS was resuspended in 500µL of dH_2O and homogenised with 10mL of a solution of a 2:1 (v/v) chloroform/methanol (Sigma-Aldrich, Ireland, #C2432 and #322415) as described in the Folch extraction procedure (Folch 1957). After dispersion, the whole mixture was shaken for 20min in an orbital shaker at room temperature (250rpm). The homogenate was then centrifuged (3000*g*, 20min) and the upper liquid phase containing the polar molecules recovered. The bEPS was dialysed a second time as described above, frozen at -80°C and freeze-dried before quantification and quality control.

2.6.2 Media-released EPS isolation (rEPS)

The media, free from the cells (1L) and containing the released EPS (rEPS), was ultrafiltered using a tangential filtration flow device consisting of a polyethersulfone membrane with a molecular weight cut off of 100kDa (Millipore, France, Pellicon XL50 Biomax #PXB100C50) connected to a peristaltic pump at 32rpm (Watson-Marlow, UK, 101UR, #JW101/UR32). The process was stopped when 900mL of permeate to 100mL retained media was reached (in approximately 10h). The Pellicon cassette was purged with 10mL of TBS 1x and the purged volume was pooled to the retentate. The system was regenerated by filtering 1L of 0.1N NaOH (Sigma-Aldrich, Ireland, #221465) and rinsed with 10mL of TBS 1x prior to treating a new sample. A 20% v/v TCA was subsequently added to the retentate, the mixture incubated overnight at 4°C, then centrifuged (8000*g*, 20 min, 4°C) to remove proteins. The rEPSs were precipitated and treated as detailed above with bEPSs.

2.6.3 EPS quantification

2.6.3.1 Phenol sulphuric assay

The phenol sulphuric assay to quantify carbohydrates was adapted from Cuesta *et al.* (Cuesta, Suarez et al. 2003). Standards from 0 to 140pmoles of haptenic sugars including glucose (Glc, Sigma-Aldrich, Ireland, #G7375), galactose (Gal, Sigma-Aldrich, Ireland, #G0750), fucose (Fuc, Sigma-Aldrich, Ireland, #F8150), mannose (Man, Sigma-Aldrich, Ireland, #M6020) and rhamnose (Rha, Sigma-Aldrich, Ireland, #R3875) were tested first with this method prior to the EPS samples. Briefly, 150µL of sulphuric acid (Sigma-Aldrich, Ireland, #339741) was added rapidly to 50µL EPS samples or sugar standards in a micro-titre plate followed by 30µL of a 5% (v/v) phenol solution (Sigma-Aldrich, Ireland, #P1037). The plate was incubated at 90°C for 5min in a PCR block and a further 5min at room temperature. Absorbance was measured at 490nm on a plate reader (Spectra Max, Molecular probes).

2.6.3.2 Monsigny assay for the quantification of polysaccharides

The assay to quantify neutral carbohydrates was adapted from Monsigny et al. (Monsigny, Petit et al. 1988). A volume of 20µL EPS samples was added to a micro-titre plate followed by 20µL of 0.6% (w/v) of freshly made up resorcinol solution (Sigma-Aldrich, Ireland, #398047). The plate was slowly shaken for 2min before adding 100µL of a 75%

sulphuric acid solution (Sigma-Aldrich, Ireland, #339741), mixed by pipetting up and down and incubated in an oven for 1h at 80ºC. The plate was subsequently cooled down at room temperature for 5 min and centrifuged before measuring the absorbance at 450nm on a plate reader (Spectra Max, Molecular probes). To quantify EPSs, solutions of dextran and glucose from 0 to 150pmoles were used as standards. Regarding the purified labelled EPS, standards were prepared from each activated EPS sample from 0 to 2.5g/L.

2.6.3.3 Quality control

The EPSs isolated using the two described procedures were checked using gradient polyacrylamide gels for electrophoresis (SDS-PAGE). Due to the large molecular weights expected from EPS, samples were run on four SDS-PAGE 3-8% mini gels (NuPAGE® Novex®, Invitrogen, USA, #EA03752BOX) in Tris-acetate-SDS buffer (Invitrogen, USA, #LA0041). A quantity, estimated by colorimetric assays of 200nmoles of EPS samples, was prepared with loading buffer (4x NuPAGE® LDS, Invitrogen, USA, #NP0008) and the gel run at 150V for 45min. Subsequently, the gels were stained to detect the presence of protein contaminants with coomassie blue or silver stain, the presence of charged carbohydrates with alcian blue or neutral polysaccharides using the periodic acid-Schiff staining method. Albumin (2µg) and lipopolysaccharides (LPS, 2µg) were used respectively as protein and carbohydrate controls on the gels. One SDS-PAGE was added directly to 20mL of the coomassie solution (Mybio, UK, #ISBT) as per manufacturer's protocol and left for 3h prior to rinsing twice in distilled water and subsequent imaging. A second SDS-PAGE was washed in ultrapure water for 5min then fixed in a 30% ethanol (Sigma-Aldrich, Ireland, #E7023), 10% acetic acid (Sigma-Aldrich, Ireland, #320099) solution for 15min and silver stained following the manufacturer's procedure (Pierce silver staining kit, Thermo Scientific, UK, #24597). A third gel was stained with alcian blue. For this process, two washing solutions were made up. The first washing solution (WS1) consisted of 25% ethanol and 10% acetic acid (v/v) diluted in dH_2O; the second washing solution (WS2) contained 15% ethanol and 5% acetic acid (v/v) in dH_2O. The alcian blue staining solution was prepared by mixing extensively 0.125% (w/v) alcian blue (Sigma-Aldrich, Ireland, #05500) in the first washing solution. The gel was first washed three times in WS1 for 5min, 10min then 15min at 50°C with gentle agitation. It was then transferred into the alcian blue staining solution for 15min at 50°C. The gel was then washed again three times in WS 1 for 1min, 4min, 5min and then twice in WS2 for 2min and 4min at 50°C. A fourth SDS-PAGE gel was stained with a mixture of periodic acid

and Schiff's reagent. After being washed twice for 1min in dH_2O to eliminate most of the SDS, the gel was fixed for 1h in a 25mL solution of 30% ethanol (v/v) and 5% acetic acid (v/v). The gel was subsequently washed twice with 25ml of a 10% ethanol solution (v/v) for 5min, then twice in dH_2O for 15min. The carbohydrates were subsequently oxidised to aldehydes by soaking the gel in 25mL of a 1% (w/v) periodic acid solution for 1h (Sigma-Aldrich, Ireland, #P7875) at room temperature with gentle agitation. The excess periodic acid was removed from the gel by three washes for 5min in 25ml of a 3% acetic acid solution (v/v) and subsequently stained in 20mL of Schiff's reagent during 1h (Sigma-Aldrich, Ireland, #84655). Excess Schiff reagent was washed off in several successive rinses in dH_2O prior to imaging. Gels were photographed using a trans-illuminator (UVI Pro, VWR, UK, #286-755-110).

2.7 Surface exposed proteins isolation

Bacteria were grown for 48h at 37˚C on MRS agar plates (Oxoid, UK, #CM0361) in anaerobic conditions (GasPak™ EZ Anaerobe Pouch System, BD, UK, # 260683). Cells were then transferred from plate to MRS broth (Oxoid, UK, #CM0359) and grown twice overnight at 37˚C in static conditions to acclimatise the cells. A 5% inoculum was transferred to fresh MRS medium in 50mL syringes for a further overnight at 37˚C. Cells were centrifuged (12000*g*, 4min, 4˚C) and supernatant discarded. Pellets were washed once with PBS 1x (Sigma-Aldrich, Ireland, #P4497) and centrifuged using the same conditions. A solution of 1M LiCl (Sigma-Aldrich, Ireland, #L7026) was prepared in PBS 1x and 10mL added to resuspend the bacterial pellets. Cells were placed on ice on a rocker for 4h before centrifugation (12000*g*, 4min, 4˚C). Supernatants were subsequently retained, filtered through a 0.22µm membrane (Millipore, Ireland, #SLGP033NS) and digested for 3h with nucleases: DNase I (Sigma-Aldrich, Ireland, #D5025) and RNase A (Sigma-Aldrich, Ireland, #R4642), in a water bath set at 37˚C. The protein mixtures were dialysed overnight against 2L dH_2O at 4˚C (SnakeSkin Dialysis Tubing, 3.5K MWCO, ThermoScientific, USA, #8842). The surface exposed proteins (SEPs) were freeze-dried and stored at -20 ˚C.

2.8 Pili isolation

2.8.1 *Faecalibacterium prausnitzii*

This procedure was modified from previously published physical methods (McMichael and Ou 1979; Hanson, Hempel et al. 1988). *Faecalibacterium prausnitzii* cells were grown overnight in Wilkins-Chalgren anaerobe broth at 37˚C in static condition (Oxoid, UK, #CM0643). A 2% inoculum was then transferred to fresh medium and grown in similar conditions for a second overnight to acclimatise the bacteria. Cells were transferred to 50mL tubes and harvested by centrifugation (8000*g*, 10min, 4˚C), supernatants discarded and pellets kept on ice. Cells were resuspended in 40mL of a chilled solution of chloramphenicol (34µg/mL) using a plastic rod to gently mix pellets. Tubes were centrifuged (10000*g*, 5min, 4˚C) and supernatants discarded, being careful to retain any hazy material near the pellets. These were resuspended in 20mL of the same chilled solution using a plastic rod before centrifugation (10000*g*, 5min, 4˚C). Supernatants were discarded again and pellets resuspended in 20mL of chilled 1x 'TBSm' buffer. The 'TBSm' solution is a low salt modified version of the TBS buffer, the 10x concentration consists of 20mM Tris-HCl (Sigma-Aldrich, Ireland, #252859), 100mM NaCl (Sigma-Aldrich, Ireland, #S3014), 1mM $CaCl_2$ (Sigma-Aldrich, Ireland, #C1016) and 1mM $MgCl_2$ (Sigma-Aldrich, Ireland, #M8266), brought to pH7.2 with HCl. Fimbriae were sheared on ice (17000rpm, 10min) using a high-performance disperser homogeniser (Ultra Turrax®, IKA, Germany). Sheared cells were centrifuged (7000*g*, 20 min, 4˚C) and the supernatant transferred to a 20mL syringe with a 0.22µm filter (Millipore, Ireland, #SLGP033NS) to remove any remaining cells. The filtered solution was aliquoted in 1.5mL tubes and stored at -20˚C.

2.8.2 All other Gram positive bacteria

Two procedures were used in the case of Gram positive bacteria and summarised in Figure 2.7.

2.8.2.1 Whole cell enzymatic treatment

This protocol was adapted from the direct enzymatic method of pili isolation for affinity purification (Reunanen, von Ossowski et al. 2012). Bacteria were grown for 48h on agar plates with relevant supplements and conditions (as described in section 2.3). Cells were transferred to broth and cultured twice overnight in relevant conditions to acclimatise the

cells. An inoculum was transferred to fresh medium in a 1L glass bottle (DURAN, Germany, #1127076) and grown to log phase. Cells were harvested at mid-log phase by centrifugation in 50mL tubes (8000*g*, 10min, 4˚C), the pellets washed with PBS, resuspended with 100mL of protoplastic buffer (50mM Tris-HCl, 5mM $MgCl_2$; 5mM $CaCl_2$; 20% sucrose; 10mg/ml lysozyme; 450U/ml mutanolysin, pH7.5) and incubated at 37°C for 3h. The suspensions were centrifuged (5000*g*, 10min, 4˚C) to remove cellular debris and supernatants containing the proteins enzymatically released from the cell wall were collected for use.

2.8.2.2 Fractionation method

Strains were routinely grown for 48h on agar plates as described (section 2.3). Cells were transferred to broth and cultured twice overnight in relevant conditions to acclimatise the cells. An inoculum was then transferred to fresh medium in 50mL centrifuge tubes or syringes. Bacteria were cultured to mid-log phase prior to harvest by centrifugation in 50mL tubes (8000*g*, 10min, 4˚C). Pellets were retained, washed twice with cold PBS (5000*g*, 10min, 4˚C), supernatants discarded and pellets kept on ice all through the rest of the procedure. Cold PBS (1mL) was added to resuspend the cells and the homogenates transferred to sterile 2mL tubes containing 500µL of glass beads (Sigma-Aldrich, Ireland, #G8772). After each time-point, cells were placed in a cell disrupter (MiniBeadBeater, BioSpec, USA, #607) for 2min, repeated three times, with a 5min pause on ice between runs. Glass beads and unbroken cells were removed by slow centrifugation (1000g, 5min, 4˚C). The supernatants were transferred to fresh tubes and 0.5mL cold PBS added before high speed centrifugation (20000*g*, 25min, 4˚C). Supernatants containing the cytosolic fractions were transferred to new tubes and immediately stored at -20˚C. Pellets containing the cell wall fractions were resuspended with 100µL of digest buffer [50mM Tris-HCl, pH8.0; 5mM $MgCl_2$, 5mM $CaCl_2$; 10mg/mL lysozyme (Sigma-Aldrich, Ireland, #L6876), 150U/mL mutanolysin (Sigma-Aldrich, Ireland, #M9901)], incubated in a water bath at 37˚C for 3h and immediately stored at -20˚C until use.

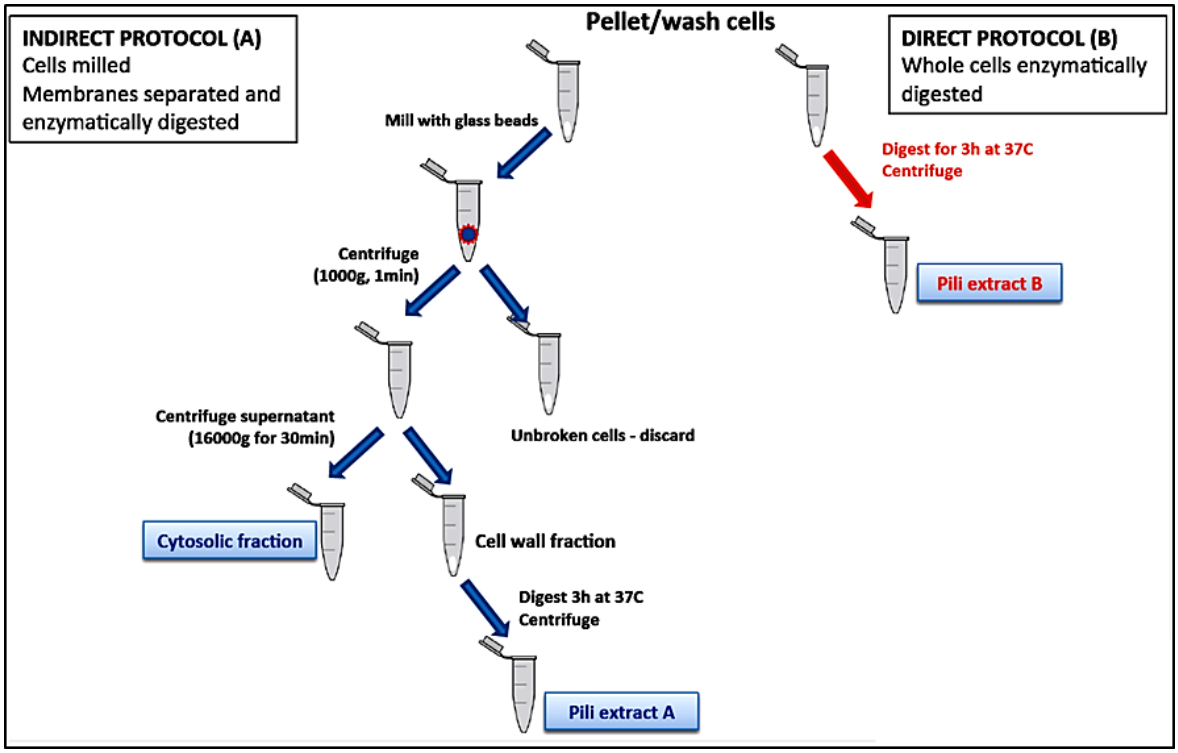

Figure 2.7: Pili isolation procedures for Gram positive bacteria

Direct and indirect enzymatic isolation of pili modelled on *L. lactis* clones.

2.8.3 Nisin-controlled gene expression system in *Lactococcus lactis*

2.8.3.1 Nisin-producing culture

The production of the antibacterial peptide nisin is best at stationary phase, therefore, the producing culture strain NZ9700, was grown as detailed previously (section 2.3.4). The overnight culture was centrifuged (8000*g*, 5min, 4˚C), supernatant containing nisin retained and filtered through 0.22µm (Millipore, Ireland, #SLGP033NS) to eliminate remaining cells. Aliquots were kept on ice for immediate use or else frozen at -20 ˚C.

2.8.3.2 Expression time-course experiment

All *L. lactis* clones were grown to mid-log phase (OD_{600} 0.5) in static conditions (as detailed in section 2.3.4). At this point (T0), one tube from each strain was kept as the 'Time 0: no-induction' control and immediately harvested by centrifugation (5000*g*, 10min, 4˚C). All others were induced by adding nisin to each culture (2µL/mL of media, section 2.8.3.1), tubes gently swirled and returned to the 30˚C incubator until harvest. Cells were harvested by centrifugation (5000*g*, 10min, 4˚C) at various time-points post-induction i.e. 60min (T=1h), 120min (T=2h) or 180min (T=3h), supernatants discarded and pellets re-suspended in 25mL PBS. Bacteria were centrifuged (5000*g*, 10min, 4˚C), supernatants

discarded and pellets kept on ice all through the rest of the procedure described below. For each time-point including T0, an aliquot of cells was kept, washed in PBS twice and fixed using 4% paraformaldehyde (Fischer Scientific, Ireland, #AC41678-5000) for TEM imaging.

2.8.2.3 Fractionation method

The fractionation procedure for pili isolation is similar to the one described in section 2.8.2

2.8.3.4 Direct enzymatic digest of whole cells

L. lactis clones were cultured as described (section 2.3.4). A 2% inoculum was transferred to fresh broth in a 1L glass bottle (DURAN, Germany, #1127076). Clones were grown to log phase (OD_{600} of 0.5) at 30°C in 1L of M17 broth supplemented with 0.5% glucose and relevant antibiotics at which point they were induced with nisin (2µL/mL). Cells were harvested after a 3h incubation period by centrifugation in 50mL centrifuge tubes (8000*g*, 10min, 4˚C), pellets washed with PBS and resuspended with 100mL of protoplastic buffer as described previously (section 2.8.2.1). The suspensions were centrifuged (5000*g*, 10min, 4˚C) to remove cellular debris. Supernatants containing pili, enzymatically released from the cell walls, were collected, aliquoted and stored immediately at -20 ˚C.

2.9 Fluorescent labelling of whole bacteria

Labelling of bacterial DNA was adapted from Hsu and Mahal (Hsu and Mahal 2006). Since the permeability of the dye is strain-specific, several optimisations steps were undertaken. First of all, the optimal dye uptake for each species of anaerobic bacteria was determined. Secondly, the washing procedure to remove the free dye prior to incubation on the microarray was optimised. Finally, titration slides were performed with several densities of bacteria to optimise signal to noise ratio.

2.9.1 Commensal strains and culturing

Bacteria from glycerol stocks were grown on agar plates as described (section 2.3). Colonies were re-suspended in 1.5mL sterile 1x TBSm, washed three times (5000*g*, 2min, 4°C) and pellets re-suspended in 1.5mL sterile 1xTBSm.

2.9.2 SYTO82 staining optimisation steps

The optical density (OD) of bacteria was measured on a biophotometer at 600nm (Eppendorf, Germany, #6131000.012) and a working volume of bacteria of 2mL at OD_{600} 2 was prepared. Bacterial DNA was stained using a cell-permeant, orange-fluorescent cyanine dye (SYTO82, Invitrogen, USA, #S11363). The choice of dye was based on its excitation/emission spectra (541nm/560nm) that matched our scanner laser but also on the fact that SYTO82 dye binds nucleic acids as opposed to the membrane of the cell, which would affect the binding results. To optimise species-specific dye uptake, SYTO82 dye was added to the cells with a final concentration ranging from 0 to 100 µM. Optical densities were then adjusted to the working concentration (OD_{600} 2) and fluorescence measured by loading 200µL of stained bacteria in triplicates in a black plate (NUNC, UK, #237105). The relative fluorescence values were measured on a plate reader (Molecular Devices, USA, Spectramax M5e) at excitation/emission wavelengths of 541/560nm along with the OD_{600} reading to calculate the dye uptake. The optimal concentration of dye to be used was then chosen based on the highest fluorescence peak.

Bacteria were stained according to species-specific dye concentration and subsequently placed in an incubator for 1h at 37°C (Sanyo, Japan, #MIR220U) with gentle shaking (60rpm). To avoid cross-reactivity with the lectins, free dye molecules were washed off using 1x TBSm-T buffer consisting of 1xTBSm supplemented with 0.05% (v/v) Tween 20 (Sigma-Aldrich, Ireland, #P1379). The number of washing steps post-staining was optimised and seven washes proved necessary to get optimal signal to background ratio. Titration lectin slides were performed to determine adequate concentration of cells per sub-array on the microarray slide. Following staining, optical densities were checked and adjusted to ensure bacterial concentration reached the working concentration (OD_{600} of 2). Subarrays were incubated with different volumes of stained bacteria in 1x TBSm-T (20 to 60µL out of a total volume of 70µL to cover the lifterslip). Volumes were chosen on the basis of the optimal signal to background ratio below feature saturation (RFU 65000 on the microarray scanner).

The optimal titres were as follows: 5µM for *B. infantis* and *F. prausnitzii*: 10µM for *Lactobacillus johnsonii* and LGG, and 20µM for *L. paracasei* (Figure 2.8). Dye titration was also carried out on *Bifidobacterium breve* UCC2003 wild type (WT) and isogenic EPS

mutants. Based on the uptake values, the optimal SYTO®82 dye concentrations chosen were 10µM for all *B. breve* isogenic clones.

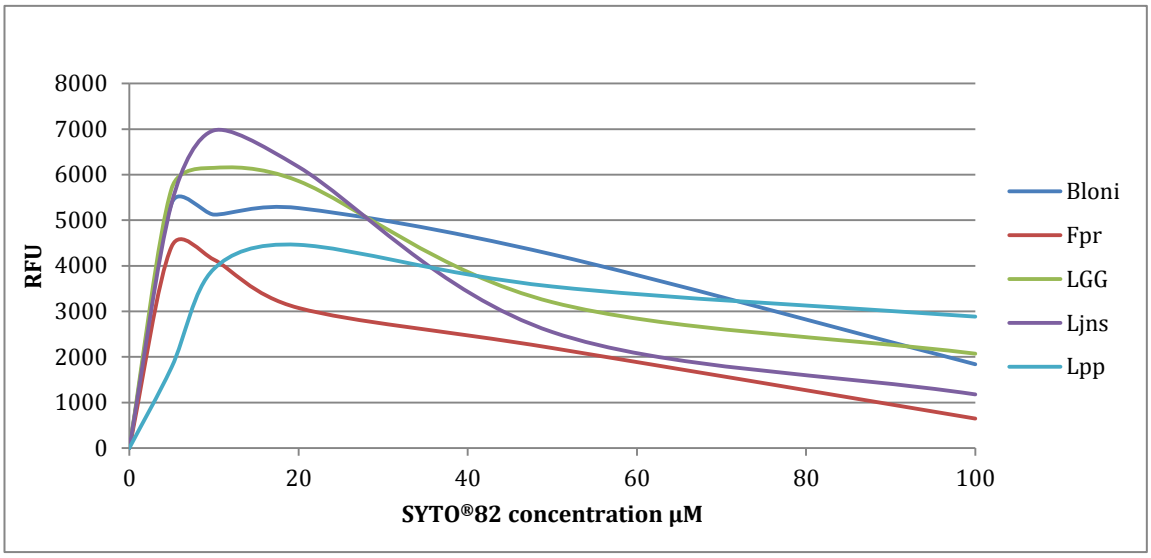

Figure 2.8: Dye uptake by five commensal strains

Relative fluorescence values (RFU) across SYTO®82 dye concentrations for *B. longum* subsp. *infantis* (Bloni); *F. prausnitzii* (Fpr)*; L. rhamnosus* GG (LGG)*; L. johnsonii* (Ljns)*; L. paracasei* subsp. *paracasei* (Lpp).

2.10 Fluorescent labelling of exopolysaccharides

The labelling procedure of EPS was based on the optimisation of dextran, a commercial glucan from *Leuconostoc sp.*, as described in Le Berre *et al.*, unpublished.

2.10.1 Activation of aldehyde groups

A mass of 10mg of EPS was weighed on a fine balance prior to re-suspension in 1mL of 50mM sodium metaperiodate (Sigma-Aldrich, Ireland, #311448). The activation of the aldehyde groups was conducted for 12h at 4°C in the dark. The activated EPS was dialysed against dH_2O (Pierce, UK, Slide-A-Lyzer Dialysis Cassettes, 3.5kDa MWCO, 3mL #66330) for 24h at 4°C in the dark, with two changes of water, prior to freezing at -80 °C and freeze drying.

2.10.2 Labelling of EPS

The EPS was re-suspended in 300µL of dH_2O. Half of this volume (150µL) was conjugated with 125µg of a fluorescent dye (AlexaFluor® hydrazide AF555, Invitrogen, USA, A20501MP) corresponding to a (w/w) 40:1 EPS/dye ratio. The mixture was

incubated for 2h at room temperature before storing at 4°C protected from the light until purification.

2.10.3 Purification of 'AF555-EPS' conjugate

The fluorescent dye-EPS conjugate ('AF555-EPS') was purified by polyacrylamide gel filtration (BioGel P6 polyacrylamide resin, BioRad, Germany, #150-4134) with 40-90µm bead size and a fractionation range of 1000-6000 Da in a 10cm column. Twenty fractions of 1mL were collected following the addition of elution buffer (1x TBSm) to the resin.

2.10.4 Quantification of labelled EPS

Absorbance of each eluted fraction was checked at OD_{555} nm with 1xTBSm used as baseline on the Nano Drop (ThermoScientific, Belgium, #NanoDrop2000). Monsigny assays (detailed in section 2.6.3.2) were then carried out on all fractions, using standards made of activated EPS from 0 to 2.5 mg/mL. A fraction was called as positive when an Alexa peak (555nm) matched a glycan peak from the Monsigny assay (450nm). For all species studied, single high peaks were observed suggesting one main type of EPS isolated. The positive fractions of EPS from each strain were tested on microarrays for optimal signals as mentioned (section 2.9.2).

2.10.5 Labelling of surface exposed proteins and pili

Surface exposed proteins, pili isolates or antibody mixtures were conjugated in the dark with fluorescent dyes, either an Alexa Fluor® AF647 NHS (excitation/emission of 651/672nm) or else an Alexa Fluor® AF555 NHS ester (excitation/emission of 555/565nm) as detailed in the manufacturer's protocol with some modifications (Life Technologies, USA, respectively #20187 or #20186). Briefly, proteins were first quantified using a bicinchoninic acid assay (BCA) with BSA standards from 0 to 2.5mg/mL (BCA Protein Assay Kit, Thermo Scientific, USA, #23227) and diluted to 2mg/mL with a volume of 500µL TBSm. One dye vial (1mg) was re-suspended in 200µL of fresh 1M sodium bicarbonate solution at pH9.5 and used for four labelling reactions. A volume of 450µL of each sample was combined to 50µL of the fluorescent reagent in a glass vial to obtain a 0.1M sodium bicarbonate solution and gently mixed by flicking. Vials were protected from light by covering with tin foil and placed at 4˚C overnight. Labelled proteins were purified using gel filtration with either P30 (Life Technologies, USA, #20187), P6 resin (Bio-Gel P-6DG, BioRad, Germany, #150-0738) or Sephadex G50 (Illustra ProbeQuant, GE

Healthcare, USA, #28-9034-08) resins depending on molecular mass. Subsequently, samples were concentrated using adapted filters of 10kDa cut-off for SEPs (Amicon 10KDa, Millipore, Ireland, #UFC501096) and 300kDa cut-off for pili mixtures (Vivaspin 300KDa, Sartorius, France, # VS0151).

2.10.6 HRP conjugation of glycoproteins

The procedure was adapted and modified from the method of Hudson (Hudson L 1989; Rojas, Ascencio et al. 2002; Wang, Wei et al. 2008) with purified porcine gastric mucin (PGM), asialofetuin (ASF) and fetuin (FET). Briefly, 2mL of a solution of 4mg/mL horseradish peroxidase (HRP, Sigma-Aldrich, Ireland, #P8375) was aliquoted in glass tubes and activated using 400µL of freshly made 0.1M sodium metaperiodate (Sigma-Aldrich, Ireland, #S1878) for 1h at room temperature by gentle rocking. The mixtures were then dialysed at 4˚C for 48h in dialysis cassettes (Slide-A-Lyzer Dialysis Cassettes, 3.5kDa MWCO, 0.5mL, Thermo Scientific, #66333) against 1L 0.001M sodium acetate buffer, pH4.5 (Sigma-Aldrich, Ireland, #S2889) changed twice daily. One millilitre of the oxidised HRP solution was transferred to clean glass tubes containing 20µL of 0.1M sodium bicarbonate pH9.5. To this mixture, 1mL of a solution of either 4mg/mL of purified porcine gastric mucin (PGM, kindly provided by Prof Steve Carrington, UCD), asialofetuin (ASF, Sigma-Aldrich, Ireland, #A4781) or fetuin (FET, Sigma-Aldrich, Ireland, #F2379) was added and placed on a rocker for 2h at room temperature for the conjugation to take place. Freshly prepared sodium borohydride buffer was made up at a concentration of 4mg/mL with deionised water (Sigma-Aldrich, Ireland, #480886) and 100 µL added to each conjugate to reduce any free HRP. The mixtures were then dialysed in dialysis cassettes (Slide-A-Lyzer Dialysis Cassettes, 3.5K MWCO, 0.5mL, ThermoScientific, #66333) at 4˚C for 48h against 1L 0.1M borate buffer, pH7.4 (Sigma-Aldrich, Ireland, #B6768) changed twice daily. The HRP-labelled conjugates were then filtered using 100kDa MWCO membranes for HRP-PGM conjugates (Amicon 100kDa, Millipore, Ireland, #UFC910008) or 50kDa for HRP-FET and HRP-ASF conjugates (Amicon 50kDa, Millipore, Ireland, #UFC905008) to eliminate the free HRP (44kDa). The retentates were then aliquoted (100µL) in sterile microtubes containing 100µL of a sterile 80% glycerol solution (Sigma-Aldrich, Ireland, #G5516) and stored at -20˚C until use.

2.11 Protein detection

2.11.1 SDS and native polyacrylamide gels

Protein separation was carried out in non-reducing conditions at 150V, 50mA using pre-cast polyacrylamide gels for 40min to 1h. All gradient polyacrylamide gels for electrophoresis (PAGE) were bought in, loaded and run with relevant buffers as per manufacturers' indications. Native 3-8% PAGE were purchased through MyBio (Runblue native gel, 3-20%, 12 Well, Expedeon, Ireland, #BCN32012); sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) were purchased from Invitrogen (NuPAGE® Novex®, 3-8% Tris-Acetate protein gels, Life Technologies, USA, # EA03752BOX or NuPAGE® Novex®, 4-12% Bis-Tris protein gels, MOPS buffer, Life Technologies, USA, #NP0322BOX). Gels were stained using silver staining (Pierce Silver stain, ThermoScientific, USA, #24612) according to manufacturer's procedure and visualised on a gel documentation system (UVITEC, UK, #UVIdoc). Replicates of gels were transferred to PVDF membrane for immunoblotting purposes.

2.11.2 Transfer to PVDF membrane

Proteins were transferred to sequencing grade polyvinylene diflouride membrane (PVDF) with 0.45µm pores (Immobilon-P Membrane, Millipore, Ireland, #IPVH00010), previously washed twice in 100% methanol (Sigma-Aldrich, Ireland, #459836), using a semi-dry transfer apparatus (Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell, BioRad, Germany, #170-3940) at 10V, 1.5mA/cm^2 for 2h. The dry PVDF membrane was cut to the size of the gel and soaked into the anode solution, consisting of 0.3M Tris (Tris(hydroxymethyl)aminomethane, Sigma-Aldrich, Ireland, #252859), 0.1M glycine (Sigma-Aldrich, Ireland, #G8898) and 0.0375% (w/v) SDS (Sigma-Aldrich, Ireland, #L6026). A sheet of thick blotting paper (Thick Blot Paper, BioRad, Germany, #170-3932) was also soaked in the anode buffer and carefully applied on the anode to avoid air bubbles, following by the PVDF membrane, on top of which was placed the polyacrylamide gel. A sheet of thick blotting paper (Thick Blot Paper, BioRad, Germany, #170-3932) was soaked in a cathodic solution made of 0.3M aminocaproic acid (Sigma-Aldrich, Ireland, #A2504), 0.03M Tris (Sigma-Aldrich, Ireland, #252859) and 0.0375% SDS (Sigma-Aldrich, Ireland, #L6026) and placed on the gel. Air bubbles were expelled from the 'sandwich' by rolling over it a 15mL tube. The anode was then placed on the

sandwich for transferring proteins. Following transfer, PVDF membranes were rinsed (protein side up) for 5min in TBSm-T prior to blocking overnight at 4˚C with 1% BSA (Sigma-Aldrich, Ireland, #A2153) made up in TBSm-T. Prior to staining, the blocked membrane was washed twice in TBSm-T for 5min on a rocker at room temperature and kept wet.

2.11.3 HRP-glycoprotein staining

The western blots were incubated with the relevant HRP conjugates (200µL aliquot) diluted in 10mL TBSm-T, for 2h, at room temperature, with gentle shaking. Subsequently, the blots were washed three times in TBSm-T for 5min and once in TBSm before proceeding with the chemiluminescent assay (section 2.11.5).

2.11.4 Serum purification and immuno-detection

Anti-pilin sera (listed in Table 2.12) were provided by Dr Mary O'Connell-Motherway, UCC. These originated from rabbits immunised with crude extracts of the cytosolic fractions of clones engineered to express pilin subunits (anti-FimA, anti-FimP, anti-FimQ from B. *bifidum* PRL2010) in an *E.* coli expression system. Immunoglobulins IgG from these sera were purified by small-scale affinity purification with protein A resin following a spin purification procedure as per manufacturer's instructions (NAb spin kit, 1mL for antibody purification, ThermoScientific, USA, #89978). Each eluted fraction (four in total) were stored at 4˚C, their relative absorbance measured at 280nm (NanoDrop 2000, ThermoScientific, USA, #NanoDrop2000) and the major fraction used for western blotting or fluorescent labelling. The PVDF membranes were incubated with either un-purified sera or purified IgG diluted 1/1000 in TBSm-T for 1h, at room temperature with gentle shaking. Subsequently, the blots underwent three washes in TBSm-T for 5min each. Polyclonal HRP-conjugated swine anti-rabbit IgG antibodies (DakoCytomation, Denmark, #P0217) were then diluted at 1/1000 and incubated on the membranes for 1h, at room temperature, with gentle shaking. The blots were washed again three times in TBSm-T for 5min and rinsed in TBSm before the chemiluminescent assay.

Table 2.12: List of antibodies used in the study

Date produced	ID	Pilin	Size of immunogen	Target
26/07/2012	IE12005	BBPR_1707	56 kDa	*B. bifidum* FimA
26/07/2012	IE12004	BBPR_283	51 kDa	*B. bifidum* FimP
26/07/2012	IE12003	BBPR_282	121 kDa	*B. bifidum* FimQ

2.11.5 Chemiluminescent assay

The chemiluminescent reaction follows the oxidation of luminol to 3-aminophthalate by the HRP enzyme. The commercial reagents used contain luminol and modified phenols to amplify the reaction for enhanced chemiluminescence (Western Bright ECL kit, Advansta, Ireland, #K-12045-D20). Solutions were prepared as per manufacturer's instructions and blots immersed in the ECL solution for 30sec to 2min depending on the intensity required. Blots were placed in an image station set at 428nm ladder (Kodak 4000mm, USA, #8642985) and images processed to compare detected bands against the protein ladder (Page Ruler Plus, ThermoScientific, #66333).

2.12 Lectin microarrays

The whole procedure relating to the lectin microarrays was based on the method developed in the Glycoscience group, NUI, Galway (Gerlach, Krüger et al. 2013).

2.12.1 Lectin microarray printing

Lectins were spotted on special glass slides (Nexterion® slides H, Schott, USA, #1070936), which were coated with a polymer that helps to preserve the native tri-dimensional structure of molecules, therefore maintaining their conformation and functionality. Slides stored at -20°C were thawed at room temperature for 1h prior to printing. Each lectin probe was prepared at a concentration of 0.5mg/mL in a 1mM solution of 1x PBS, pH 7.4 containing their appropriate sugar storing solutions (1mM of lactose, fucose, galactose, mannose or *N*-acetyl-D-glucosamine). Probes were centrifuged to remove insoluble materials (4000*g*, 2min, 20°C) and checked for any precipitation, denoting a deteriorated lectin. According to the printing map, 50µL of each probe were pipetted to a 384-well plate (Molecular Probes, USA, #GenetixX6001). The plate was centrifuged to remove air bubbles (150g, 2min, 20°C) and placed in the microarray spotter along with new glass slides. A volume of 1nL of each probe was printed according to the manufacturer's manual using a piezo-driven non-contact and low volume dispenser with a 90µm uncoated glass nozzle at 62% humidity (Scienion AG, Germany, #SciFLEXARRAYER S3). Six replicates of 52 features were spotted, corresponding to 42 lectins along with controls such as dH_2O, 1x PBS and empty spaces, totalling 416 spots per subarray (Table 2.13). The slide comprised eight subarrays in total (Figure 2.9). Printed microarrays were incubated in a humidity chamber overnight to enable conjugation, and

the remaining functional groups were capped by incubating the slide for 1h with a freshly prepared blocking solution of 100mM ethanolamine (Sigma-Aldrich, Ireland, #411000) in 50mM sodium borate, pH 8.0 (Sigma-Aldrich, Ireland, #S9640). Slides were washed three times in 1x PBS-T buffer composed of 1x PBS and 0.05% Tween-20 (Sigma-Aldrich, Ireland, #P1379). They were subsequently dried by centrifugation (150g, 5min) and stored at 4°C with desiccant until use (within one month). Quality controls were carried out with two serum glycoproteins of known glycan compositions, fetuin (Sigma-Aldrich, Ireland, #F2379) and asialofetuin (Sigma-Aldrich, Ireland, #A4781), labelled with Alexa Fluor dye AF555 (Invitrogen, USA, #A20174) to assess the presence of all features on each print batch.

2.12.2 Lectin microarray incubation

The lectin microarrays were removed from 4°C storage and left 15min at room temperature to acclimatise prior to incubation. Following the titration slide results and the optimal signal determination (section 2.9.2), samples were diluted to a volume of 70µL with 1x TBSm-T. Fluorescently labelled glycoproteins AF555-fetuin and AF555-asialofetuin were also diluted to 1µg/mL from stocks and used as experimental and slide to slide controls. A 'turtle' coverslip (8-well gasket slide) comprised of eight defined areas or subarrays (Agilent Technologies, USA, #G2534-60015) was placed in a gasket holder. Two subarray wells were filled with the control glycoproteins and 70µL of samples added to the six free wells (see Figure. 2.8 for slide layout). The microarray was then laid onto the coverslip and the assembled cassette incubated with gentle rotation (1h, 23°C) in a hybridisation oven (Agilent Technologies, USA, #G2545A). Following incubation, the slide was washed using a microarray automatic wash station (Implen, Germany, # Advawash), set with the following parameters: 1x TBSm-T for two minutes repeated six times, then 1x TBSm rinse for two minutes.

Table 2.13: Lectin microarray composition list

[1]ID	Abbreviation	Origin	Species	Common name	[2]Major Ligand(s)	Vendor
1	AIA, Jacalin	Plant	*Artocarpus integrifolia*	Jack fruit lectin	Gal (sialylation independent)	EY Labs
2	RPbAI	Plant	*Robinia pseudoacacia*	Black locust lectin	Gal, GalNAc	EY Labs
3	PA-I	Bacteria	*Pseudomonas aeruginosa*	Pseudomonas lectin	Gal, Gal derivatives	Sigma Aldrich
4	SNA-II	Plant	*Sambucus nigra*	Sambucus lectin-II	Gal/GalNAc	EY Labs
5	SJA	Plant	*Sophora japonica*	Pagoda tree lectin	βGalNAc	EY Labs
6	DBA	Plant	*Dolichos biflorus*	Horse gram lectin	GalNAc	EY Labs
7	APP	Plant	*Aegopodium podagraria*	Ground elder lectin	GalNAc	EY Labs
8	SBA	Plant	*Glycine max*	Soy bean lectin	GalNAc	EY Labs
9	VVA-B4	Plant	*Vicia villosa*	Hairy vetch lectin	GalNAc	EY Labs
10	BPA	Plant	*Bauhinia purpurea*	Camels foot tree lectin	GalNAc/Gal	EY Labs
11	WFA	Plant	*Wisteria floribunda*	Japanese wisteria lectin	GalNAc/Sulfated GalNAc	EY Labs
12	ACA	Plant	*Amaranthus caudatus*	Amaranthin	Sialylated/Gal-β-(1→3)-GalNAc	Vector Labs
13	ABL	Fungi	*Agaricus bisporus*	Edible mushroom lectin	Gal-β-(1→3)-GalNAc, GlcNAc	EY Labs
14	PNA	Plant	*Arachis hypogaea*	Peanut lectin	Gal-β-(1→3)-GalNAc	EY Labs
15	GSL-II	Plant	*Griffonia simplicifolia*	Griffonia lectin-II	GlcNAc	EY Labs
16	sWGA	Plant	*Triticum vulgaris*	Succinyl WGA	GlcNAc	EY Labs
17	DSA	Plant	*Datura stramonium*	Jimson weed lectin	GlcNAc	EY Labs
18	STA	Plant	*Solanum tuberosum*	Potato lectin	GlcNAc oligomers	EY Labs
19	LEL	Plant	*Lycopersicum eculentum*	Tomato lectin	GlcNAc-β-(1→4)-GlcNAc	EY Labs
20	NPA	Plant	*Narcissus pseudonarcissus*	Daffodil lectin	Man-α-(1→6)-	EY Labs
21	GNA	Plant	*Galanthus nivalis*	Snowdrop lectin	Man-α-(1→3)-	EY Labs
22	HHA	Plant	*Hippeastrum hybrid*	Amaryllis agglutinin	Man-α-(1→3)-Man-α-(1→6)-	EY Labs

[1]ID	Abbreviation	Origin	Species	Common name	[2]Major Ligand(s)	Vendor
23	ConA	Plant	*Canavalia ensiformis*	Jack bean lectin	Man	EY Labs
24	Lch-B	Plant	*Lens culinaris*	Lentil isolectin B	Man, fucose dependent	EY Labs
25	PSA	Plant	*Pisum sativum*	Pea lectin	Man, fucose dependent	EY Labs
26	WGA	Plant	*Triticum vulgaris*	Wheat germ agglutinin	GlcNAc/Neu5Ac	EY Labs
27	MAA	Plant	*Maackia amurensis*	Maackia agglutinin	Neu5Ac-α-(2→3)-Gal	EY Labs
28	SNA-I	Plant	*Sambucus nigra*	Sambucus lectin-I	Neu5Ac-α-(2→6)-Gal/GalNAc	EY Labs
29	PHA-L	Plant	*Phaseolus vulgaris*	Kidney bean leuko-agglutinin	Tri/tetra-antennary β-Gal/Gal-β-(1→4)-GlcNAc	EY Labs
30	PCA	Plant	*Phaseolus coccineus*	Scarlet runner bean lectin	GlcNAc in complex oligomers	Sigma Aldrich
31	PHA-E	Plant	*Phaseolus vulgaris*	Kidney bean erythro-agglutinin	Biantennary, bisecting GlcNAc,β-Gal/Gal-β-(1→4)-GlcNAc	EY Labs
32	RCA-I/120	Plant	*Ricinus communis*	Castor bean lectin I	Gal-β-(1→4)-GlcNAc	Vector Labs
33	CPA	Plant	*Cicer arietinum*	Chickpea lectin	Complex glycopeptides	EY Labs
34	CAA	Plant	*Caragana arborescens*	Pea tree lectin	Gal-β-(1→4)-GlcNAc	EY Labs
35	ECA	Plant	*Erythrina cristagalli*	Coral tree lectin	Gal-β-(1→4)-GlcNAc oligomers	EY Labs
36	AAL	Fungi	*Aleuria aurantia*	Orange peel fungus lectin	α-Fuc-(1→6)	Vector Labs
37	LTA	Plant	*Lotus tetragonolobus*	Lotus lectin	α-Fuc-(1→3)	EY Labs
38	UEA-I	Plant	*Ulex europaeus*	Gorse lectin-I	α-Fuc-(1→2)	EY Labs
39	EEA	Plant	*Euonymous europaeus*	Spindle tree lectin	α-Gal	EY Labs
40	GSL-I-B4	Plant	*Griffonia simplicifolia*	Griffonia lectin-I	α-Gal	EY Labs
41	MPA	Plant	*Maclura pomifera*	Osage orange lectin	α-Gal	EY Labs
42	VRA	Plant	*Vigna radiata*	Mung bean lectin	α-Gal	EY Labs
43	MOA	Fungi	*Marasmius oreades*	Fairy ring mushroom lectin	α-Gal	EY Labs

[1]Lectin printing ID

[2]Abbreviations are as follows, Gal, galactose; GalNAc, *N*-acetylgalactosamine; Glc, glucose; GlcNAc, *N*-acetylglucosamine; Man, mannose; Fuc, fucose; Neu5Ac: sialic acid

The microarray slide was centrifuged to dry (350g, 5min, 23°C). Three biological and two technical replicates of each bacterial species were incubated on the microarray for the 'whole cell' experiment. The lectin arrays consisted in three technical replicates of one labelling reaction in the case of EPS profiling.

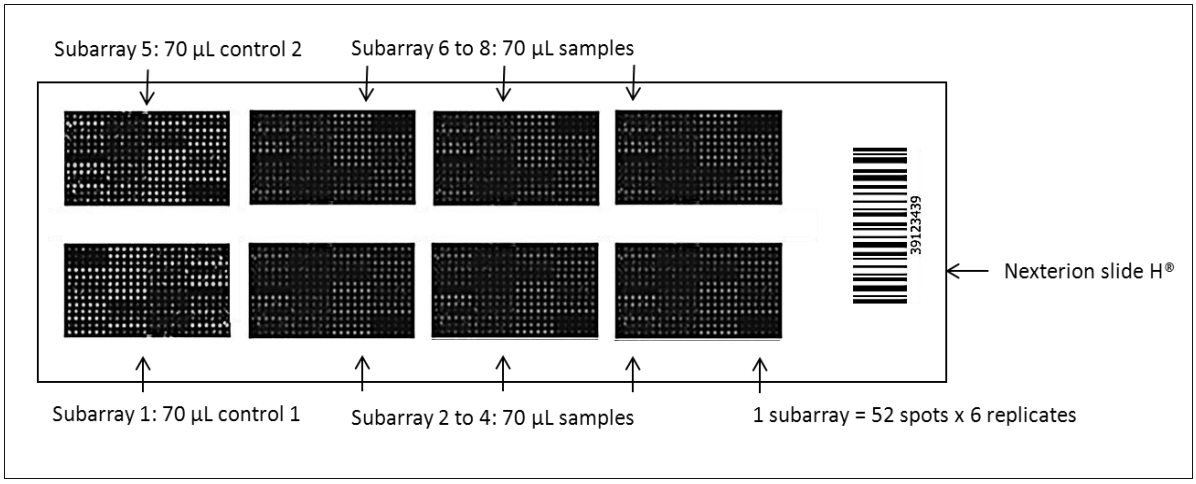

Figure 2.9: Microarray slide layout

Schematic of a typical microarray slide featuring eight subarrays, two of these (subarrays 1 and 5) were used for controls and the remaining six for samples.

2.12.3 Inhibition experiments

Co-incubation of labelled whole cells or EPS samples with simple sugars was carried out to check for sugar-mediated binding to lectins. Briefly, subarrays containing each a final concentration of 100mM of haptenic sugars (Fuc, Gal, Glc, GlcNAc, Man or Rha) were co-incubated directly on the coverslip with the samples and processed as the neat slides.

2.12.4 Scanning of the microarray slide

The slide was placed in a microarray scanner (Perkin Elmer, USA, #900-3011537000 or else Agilent Technologies, USA, #G2505-60502) and scanned using the Cy3/AF555 laser channel (with an emission of 543nm). Spot intensities were exported as image files (*.tif) into the signal extraction software.

2.12.5 Signal extraction

The signal intensities were converted using extraction software (GenePix v6.1.0.4, Molecular Devices, UK, #5004757). The microarray file (*.gal) was fitted onto the image file using adaptive diameter circular alignment (70–130%) based on the 230µm feature diameter. Data were expressed as medians of the spot intensities minus the background (F543 median-B543). No normalisation procedures were performed at this stage. Signal values were exported to spreadsheet software (Microsoft Excel 2010, Microsoft, USA, # 11320797128).

2.12.6 Data analysis

Data intensities across technical and biological replicate subarrays were normalised to the per-subarray median total intensity mean. Signals for each entity (i.e. lectin) were then averaged for replicates and plotted on a chart for easy visualisation of binding patterns (a radar plot or a histogram). Lectin signatures of serum glycoproteins (AF555-labelled fetuin and asialofetuin) used as pattern controls were examined from slide to slide to validate the experiment. Signals from each biological replicate were normalised by scaling all signals to the global mean and exported to MS Excel for calculation of probe means, standard deviations and for graphical purposes.

2.12.7 Statistical analysis

Statistical analysis was carried out using microarray analysis software (Genespring GX version 11.5, Agilent, USA, #G3784AA). Signal intensities were transferred to the software and submitted to scale normalisation to minimise non-biological variability across arrays. Heat maps representing the dataset and hierarchical clustering of entities were performed for each individual strain or EPS using Euclidian distances. The data were also sorted into conditions (strains or type of EPS). Means between conditions were compared using one-way analysis of variance (ANOVA) with Benjamini-Hochberg false discovery rate (FDR) of 5% followed by Tukey post hoc tests to determine which features were responsible for the significant differences obtained ($P<0.05$). Student *t*-tests were similarly carried out to compare bound and released EPS within each strain. Binding fold change between conditions were calculated and significant binders selected above a cut-off of 2-folds. Statistically significant data were presented in a table along with the FDR-corrected *P*-value for comparison purposes between strains or EPS types.

2.13 Monosaccharides analyses

2.13.1 Acid hydrolysis

The equivalent of 100µg of the desalted EPS was resuspended in 4M trifluoroacetic acid (TFA) and hydrolysed in an incubator at 100°C for 16h (Biermann 1988). The hydrolysate, on cooling down, was evaporated to complete dryness using a concentrator (Eppendorf, Germany, #5301). The procedure was repeated three times using high grade 18.2MΩ water to remove all TFA. Dried hydrolysates were resuspended in 250µL of 18.2MΩ water, filtered through 3kDa MWCO spin columns (Amicon, Sigma-Aldrich, Ireland, #Z677094) and evaporated to dryness again before storage at -20°C until analysis.

2.13.2 Monosaccharides analysis using HPAEC_PAD

For monosaccharide analysis, dried hydrolysates were dissolved in 100µL of 18.2MΩ water and the released sugars were identified by high-pressure anion exchange chromatography-pulse amperometric detection (HPAEC_PAD) on an ICS3000 system (Dionex, CA, #289185). The equivalent of 1µg of each hydrolysate was injected onto a carbohydrate analytical column (Dionex CarboPac PA20, Thermo Scientific, USA, #060142) preceded by guard and amino-trap columns to minimise interference from peptides, and separated with isocratic elution by 18mM NaOH at a flow rate of 0.35mL/min over 18.5min at 25°C. As controls, 18.2MΩ water and 50µg of dextran (Sigma-Aldrich, Ireland, #D4772) were used together with an internal standard with each sample (100 pmoles of 2-deoxy galactose). In addition, a standard set of monosaccharides (fucose, rhamnose, galactosamine, glucosamine, galactose, glucose and mannose) was injected prior to each sample to monitor retention drift over the course of the run. A standard curve was also calculated using the aforementioned standards (Mix7) from 10 to 100pmoles. Dilutions of samples were carried out when necessary in order to keep within the range of standards peaks. Response factors (Rf) were calculated for each standard using the formula: Rf = (Area of the standard residue in Mix7*internal standard concentration in Mix 7 in µg/ml)/(Sugar concentration in µg/ml in Mix7*internal standard area in Mix 7).

2.13.3 Monosaccharides analysis using HPLC

MRS media from Oxoid contains 20g/L of glucose but it also contains unknown polysaccharides in the yeast extract (4g/L) and traces of glycogen in the ‘Lab lemco’ beef

extract (8g/L). To check if these components were interfering with the rEPS isolation procedure, MRS broth was prepared in the same way as all the rEPS samples described previously (section 2.6.2). Briefly, 1L of MRS broth was autoclaved and processed through tangential flow filtration with a 100kDa cut-off ultrafiltration device. The permeate was treated with TCA and the polysaccharides recovered by precipitation with ethanol. They were treated with nucleases and proteinase K before being extensively dialysed and freeze-dried. The dry material was hydrolysed as described previously (section 2.13.1) before being labelled with anthranilamide (2-aminobenzamide, Sigma-Aldrich, Ireland, #A89804) following a method adapted from Bigge et *al*. (Bigge, Patel et al. 1995). As a comparison and to check the extent of contamination by MRS polysaccharides in rEPS samples, the rEPS fraction from *Lactobacillus rhamnosus* GG was also labelled. Briefly, 1mL of 0.35M labelling solution was prepared by dissolving 47.6mg of anthranilamide in 1mL of a solution of 30% acetic acid diluted in DMSO in an amber tube. This solution was then added to a new amber tube containing 62.84mg of sodium cyanoborohydride (NaCNBH4, Sigma-Aldrich, Ireland, #156159) to make up a 1M final solution of the reducing NaCNBH4. The solution was mixed until complete dissolution, protected from light with tin foil and used within an hour. A volume of 60µL of the labelling reagent was added to the dried sample (100µg) in an amber tube and incubated with gentle agitation (65°C, 2.5h, 200 rpm). The labelled sugars were purified using cartridges (Glyko S cartridges, Prozyme, USA, #GC210) according to the manufacturer's instructions. Samples were subsequently stored at -20˚C protected from the light until separation by high performance liquid chromatography (HPLC system, Waters, France, #Alliance).

2.13.4 ^{1}H-NMR

^{1}H-NMR was used on LGG bEPS to verify the presence of rhamnose and the absence of rhamnose in the sample. Dry material from the sample was deuterised and prepared by Dr Tinaig Le Costaouec and run on the NMR instrument in the Department of Chemistry, NUIG.

2.14 Neo-glycoconjugate microarrays

2.14.1 Printing of slides

Neo-glycoconjugate (NGC) microarrays were printed and blocked 'in-house' in the Glycoscience group as described previously (Kilcoyne, Gerlach et al. 2012; O'Boyle, Houeix et al. 2013). In brief, BSA and HSA were used as molecular scaffolds for presentation of the linker-bound glycan and conjugation to the microarray slide surface. The glycans were bound to three atom spacers unless specified. Probes comprising glycoproteins, commercial or 'in-house' NGCs and linker controls were prepared at a concentration of 1mg/mL based on weight (glycoproteins) or following BCA assay. A maximum of 52 probes were printed per subarray at 1mg/mL in PBS (10mM sodium phosphate, 140mM NaCl, 30mM KCl, 2mM potassium phosphate, pH7.4). The total number of NGCs and glycoproteins prepared was 76 and was therefore divided between twinned A and B slides (Tables 2.14 and 2.15) with 20 features redundant between platforms. Probes were cooled to 10°C for the duration of the print cycle. The 52 features were printed in replicates of six spots per subarray with a total of eight subarrays (13 x 24 spots each) on a special glass slide as mentioned in section 2.12. Microarrays were then transferred to a slide chamber and suspended above a saturated salt water solution for 8h to allow completion of linking reactions. Slides were subsequently blocked for 1h, in 50mM sodium borate, pH8.0 and 100mM ethanolamine added after pH adjustment for capping unconjugated matrix. Following blocking, microarrays were washed three times for 5min in PBS containing 0.05% Tween 20 (PBS-T) and once in PBS for five more minutes. Slides were centrifuged (450g, 5min, 15°C) to dry and stored in a suitable holder sealed in a bag with desiccant at 4°C until use.

2.14.2 Incubation

A titration slide was initially carried out to determine signal to noise ratio and feature overload. Samples, consisting in labelled whole-cells (OD_{600} 2 stock) or labelled-proteins (10µg) were diluted to a final volume of 70µL with TBSm-T. Each lectin control solution (lectins AIA, MAA, UEA-I or WFA at 1µg/mL) or sample solution were dispensed in relevant wells of an 8-well gasket slide placed in a hybridisation cassette (Figure 2.9). The glycan array was placed on top of the gasket slide and the pressure plate added to close the cassette. Microarrays were incubated at 23°C for 1h with gentle end-end rotation (Agilent

Technologies, USA, #G2545A). Microarray chambers were opened in a square basin containing TBSm-T and transferred immediately to the Advawash station (Implen, Germany, # Advawash) with a similar program (section 2.12.2) and centrifuged (350g, 5 min, 23°C) to dry. Inhibition experiments were carried out as described in section 2.12.3

2.14.3 Scanning, data extraction and analyses

Dry slides were transferred to a microarray scanner and signals extracted as in section 2.13.5. For each twinned microarray platform, data were analysed as described previously (section 2.12.6). The signals corresponding to the 20 common probes between slides A and B were summed, then averaged per platform and the global mean calculated between A and B. Subsequently, a correction factor was applied to each platform to combine the unique features from A and B. The data corresponding to the resulting 76 glycans were transferred to spreadsheet software for graphical purpose (MS Excel). Microarray data were then transferred to GX statistical software for strain to strain comparison as detailed in section 2.12.7.

Table 2.14: List of features printed on the neo-glycoconjugate microarray platform A

[1]ID	Name of feature	[2]Source	Cat. no.	[3]Comments	Abbreviation	Description/Structure
1	Fetuin	Sigma	F2379		Fetuin	Fetuin
2	Asialofetuin	Sigma	A4781		ASF	ASF
3	PBS	MK			PBS	PBS
4	Ovalbumin	Sigma	A5503		Ov	Ovalbumin
5	RNAse B	Sigma	R1153		RB	RNAseB
6	Transferrin	Sigma	T8158		Xferrin	Transferrin
7	4AP-HAS	MK			4APHSA	4-aminopyridine-HAS
8	α-Crystallin from bovine lens	Sigma	C4163		a-C	α-crystalline (*Bos taurus*)
9	Mana1,3(Mana1,6)Man-BSA	Dextra	NGP1336	3 atom sp	M3BSA	Man-α-(1→3)-[Man-α-(1→6)]-Man-BSA
10	GlcNAc-BSA	Dextra	NGP1101	14 atom sp	GlcNAcBSA	GlcNAc-Sp14-NH2(Lys)-BSA
11	LacNAc-BSA	Dextra	NGP0201	3 atom sp	LacNAcBSA	Gal-β-(1→4)-GlcNAc-Sp3-BSA
12	3'SialylLacNAc-BSA	Dextra	NGP0301	3 atom sp	3SLNBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc-APD-HSA
13	3´-Sialyllactose-APD-HSA, 0.5 mg	IsoSep	60/67	3 atom sp	3SLacHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-Glc-APD-HSA
14	6´-Sialyllactose-APD-HSA, 0.5 mg	IsoSep	60/93	3 atom sp	6SLacHSA	Neu5Ac-α-(2→6)-Gal-β-(1→4)-Glc-APD-HSA
15	2'Fucosyllactose-BSA	Dextra	NGP0307	3 atom sp	2FLBSA	Fuc-α-(1→2)-Gal-β-(1→4)-Glc-Sp3-BSA
16	3'Sialyl-3-fucosyllactose-BSA	Dextra	NGP0405	3 atom sp	3SFLBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)[Fuc-α-(1→3)]-Glc-Sp3-BSA
17	H Type II-APE-BSA, 0.5 mg	IsoSep	60/54	3 atom sp	H2BSA	Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc-β1-APE-BSA
18	Blood Group A-BSA	Dextra	NGP6305	6 atom sp	BGABSA	GalNAc-α-(1→3)-[Fuc-α-(1→2)]-Gal
19	Blood Group B-BSA	Dextra	NGP6323	6 atom sp	BGBBSA	Gal-α-(1→3)-[Fuc-α-(1→2)]-Gal
20	Gala1,3Galb1,4GlcNAc-HSA	Dextra	NGP2334	3 atom sp	GGGNHSA	Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc-HSA
21	Gala1,3Gal-BSA	Dextra	NGP0203	3 atom sp	Ga3GBSA	Gal-α-(1→3)-Gal-Sp3-BSA

[1]ID	Name of feature	[2]Source	Cat. no.	[3]Comments	Abbreviation	Description/Structure
22	Galb1,4GalBSA, **1 mg**	Dextra	NGP0204	3 atom sp	Gb4GBSA	Gal-β-(1→4)-Gal-Sp3-BSA
23	Gala1,2GalBSA, **1 mg**	Dextra	NGP0202	3 atom sp	Ga2GBSA	Gal-α-(1→2)-Gal-Sp3-BSA
24	4AP-BSA	MK			4APBSA	4-aminopyridine-BSA
25	Lacto-*N*-fucopentaose I-BSA	Dextra	NGP0503	3 atom sp	LNFPIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-BSA
26	Lacto-*N*-fucopentaose II-BSA	Dextra	NGP0501	3 atom sp	LNFPIIBSA	Fuc-α-(1→3)Gal-β-(1→3)-GlcNAcβ-(1→3)-Gal-β-(1→4)-Glc-BSA
27	Lacto-*N*-fucopentaose III-BSA	Dextra	NGP0502	3 atom sp	LNFPIIIBSA	Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-BSA
28	Lacto-*N*-difucohexaose I-BSA	Dextra	NGP0601	3 atom sp	LNDHIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-Sp3-BSA
29	LNDI-BSA/ Lewis b-BSA	IsoSep	60/04	3 atom sp	LebBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-BSA
30	Lewis x-BSA	Dextra	NGP0302	3 atom sp	LexBSA	Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
31	Di-Lex-APE-BSA, 0.5 mg	IsoSep	61/64	3 atom sp	DiLexBSA	Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-BSA
32	Di-Lewisx-APE-HSA, 0.5 mg	IsoSep	61/59	3 atom sp	DiLexHSA	Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β(1→4)[Fuc-α-(1→3)]GlcNAc-β1-O-APE-HSA
33	Tri-Lex-APE-HSA, 0.5 mg	IsoSep	61/56	3 atom sp	3LexHSA	Gal-β(1→4)-[Fuc-α-(1→3)]GlcNAc-β(1→3)Gal-β(1→4)-[Fuc-α-(1→3)]-GlcNAc-β(1→3)Gal-β(1→4)-[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-HSA

[1]ID	Name of feature	[2]Source	Cat. no.	[3]Comments	Abbreviation	Description/Structure
34	3'Sialyl Lewis x-BSA	Dextra	NGP0403	3 atom sp	3SLexBSA3	Neu5Ac-α(2-3)-Gal-β(1→4)-[Fuc-α(1→3)]-GlcNAc-Sp3-BSA
35	3'Sialyl Lewis x-BSA	Dextra	NGP1403	14 atom sp	SLexBSA14	Neu5Ac-α-(2→3)-Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-Sp3-BSA
36	6-Sulfo Lewis x-BSA	Dextra	NGP0603	3 atom sp	6SuLexBSA	(SO4)-6Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
37	6-Sulfo Lewis a-BSA	Dextra	NGP0604	3 atom sp	6SuLeaBSA	(SO4)-6Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-Sp3-BSA
38	3-Sulfo Lewis a-BSA	Dextra	NGP0304	3 atom sp	3SuLeaBSA	(SO4)-3Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-Sp3-BSA
39	3-Sulfo Lewis x-BSA	Dextra	NGP0303	3 atom sp	3SuLexBSA	(SO4)-3Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
40	Difucosyl-para-lacto-N-hexaose-APD-HSA, 0.5 mg (Lea/Lex)	IsoSep	61/57	3 atom sp	DFPLNHHSA	Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-β-(1→3)-Gal-β(1→4)-Glc-APD-HSA
41	Lewis a-BSA	Dextra	NGP0303	3 atom sp	LeaBSA	Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-Sp3-BSA
42	Lewis y-tetrasaccharide-APE-HSA, 0.5 mg	IsoSep	60/95	3 atom sp	LeyHSA	Fuc-α-(1→2)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-HSA
43	Tri-fucosyl-Ley-heptasaccharide-APE-HSA, 0.5 mg	IsoSep	61/63	3 atom sp	3FLeyHSA	Fuc-α-(1→2)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-HSA
44	Lacto-N-neotetraose-APD-HSA, 0.5 mg	IsoSep	60/72	3 atom sp	LNnTHSA	Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA
45	Lacto-N-tetraose-APD-HSA, 0.5 mg	IsoSep	60/97	3 atom sp	LNTHSA	Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA

[1]ID	Name of feature	[2]Source	Cat. no.	[3]Comments	Abbreviation	Description/Structure
46	Sialyl-LNF V-APD-HSA, 0.5 mg	IsoSep	61/61	3 atom sp	SLNFVHSA	Fuc-α-(1→2)-Gal-β-(1→3)-[NeuAc-α(2-6)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA
47	Monofucosyl, monosialyllacto-N-neohexaose-APD-HSA, 0.5 mg	IsoSep	61/62	3 atom sp	MMLNnHHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc-β-(1→3)-[Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→6)]-Gal-β-(1→4)Glc-APD-HSA
48	Sialyl-LNnT-penta-APD-HSA, 0.5 mg	IsoSep	61/68	3 atom sp	SLNnTHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-1→4)-Glc-APD-HSA
49	GM1-pentasaccharide-APD-HSA, 0.5 mg	IsoSep	61/69	3 atom sp	GM1HSA	Gal-β-(1→3)-GalNAc-β-(1→4)-[Neu5Ac-α-(2-3)]-Gal-β-(1→4)-Glc-APD-HSA
50	Asialo-GM1-tetrasaccharide-APD-HSA, 0.5 mg	IsoSep	60/96	3 atom sp	aGM1HSA	Gal-β-(1→3)-GalNAc-β-(1→4)-Gal-β-(1→4)-Glc-APD-HSA
51	Globo-N-tetraose-APD-HSA, 0.5 mg	IsoSep	60/99	3 atom sp	GlobNTHSA	GalNAc-β-(1→3)-Gal-α-(1→4)-Gal-β-(1→4)-Glc-APD-HSA
52	Globotriose-APD-HSA, 0.5 mg	IsoSep	60/90	3 atom sp	GlobTHSA	Gal-α-(1→4)-Gal-β-(1→4)-Glc-β1-APE-HSA

[1]Identification number (ID), name; [2]Source: commercial or in-house conjugation (MK); [3]'sp' refers to spacer.

Table 2.15: List of features printed on the neo-glycoconjugate microarray platform B

[1]ID	Name of feature	[2]Source	Cat. no.	[3]Comments	Abbreviation	Description/Structure
53	Fetuin	Sigma	F2379		Fetuin	Fetuin
54	Invertase	Sigma	I4504	grade VII	Inv	Invertase
55	PBS				PBS	PBS
56	Ovalbumin	Sigma	A5503		Ovalbumin	Ovalbumin
57	Fibrinogen	Sigma	F4129		Fibrin	Fibrinogen
58	alpha-1-antitrypsin	Sigma	A6388		A1AT	alpha-1-antitrypsin
59	4AP-HSA-	MK			4APHSA	4AP-HSA-
60	α-Crystallin from bovine lens	Sigma	C4163		a-C	a-Crystallin
61	Mana1,3(Mana1,6)Man-BSA	Dextra	NGP1336	3 atom sp	M3BSA	Man-α-(1→3)-[Man-α-(1→6)]-Man-BSA
62	GlcNAc-BSA	Dextra	NGP1101	14 atom sp	GlcNAcBSA	GlcNAc-Sp14-NH2(Lys)-BSA
63	Ceruloplasmin	Sigma	C-3007	human, type III	Cerulo	Ceruloplasmin
64	alpha-1-acid glycoprotein	Sigma	G9885		AGP	α-1-acid glycoprotein
65	3´-Sialyllactose-APD-HSA, 0.5 mg	IsoSep	60/67	3 atom sp	3SLacHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-Glc-APD-HSA
66	6´-Sialyllactose-APD-HSA, 0.5 mg	IsoSep	60/93	3 atom sp	6SLacHSA	Neu5Ac-α-(2→6)-Gal-β-(1→4)-Glc-APD-HSA
67	LacNAc-a-4AP-BSA	MK		3 atom sp	LacNAcaBSA	Gal-β-(1→4)-GlcNAc-α-4AP-BSA
68	LacNAc-b-4AP-BSA	MK		3 atom sp	LacNAcb4APBSA	Gal-β-(1→4)-GlcNAc-β-4AP-BSA
69	H Type II-APE-BSA, 0.5 mg	IsoSep	60/54	3 atom sp	H2BSA	Fuc-α-(1→2)-Gal-β(1→4)-GlcNAc-β1-APE-BSA
70	H-Type 2-APE-HSA, 0.5 mg	IsoSep	60/89	3 atom sp	H2HSA	Fuc-α-(1→2)-Gal-β(1→4)-GlcNAc-β1-APE-HSA
71	Ovomucoid	Sigma	T-9253		Ovomuc	Ovomucoid

[1]ID	Name of feature	[2]Source	Cat. no.	[3]Comments	Abbreviation	Description/Structure
72	Gala1,3Galb1,4GlcNAc-HSA	Dextra	NGP2334	3 atom sp	GGGNHSA	Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc-HSA
73	Gala1,3Gal-BSA	Dextra	NGP0203	3 atom sp	Ga3GBSA	Gal-α-(1→3)-Gal-Sp3-BSA
74	L-Rhamnose-Sp14-BSA	Dextra	NGP1106	3 atom sp	RhaBSA	L-Rha-Sp14
75	Gal-a-ITC-BSA	MK		3 atom sp	XGalaBSA	Gal-α-ITC
76	4AP-BSA	MK			4APBSA	4AP-BSA
77	Lacto-*N*-fucopentaose I-BSA	Dextra	NGP0503	3 atom sp	LNFPIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-BSA
78	Man-a-ITC-BSA	MK		3 atom sp	XManaBSA	Man-α-ITC-BSA
79	Lac-b-4AP-BSA	MK		3 atom sp	XLacbBSA	Lac-β-4AP-BSA
80	Man-b-4AP-BSA	MK		3 atom sp	XManbBSA	Man-β-4AP-BSA
81	LNDI-BSA/ Lewis b-BSA	IsoSep AB	60/04	3 atom sp	LebBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-BSA
82	Lewis x-BSA	Dextra	NGP0302	3 atom sp	LexBSA	Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-Sp3-BSA
83	Gal-b-ITC-BSA	MK			XGalbBSA	Gal-β-ITC-BSA
84	Xyl-b-4AP-BSA	MK			XylbBSA	Xyl-β-4AP-BSA
85	Xyl-a-4AP-BSA	MK			XylaBSA	Xyl-α-4AP-BSA
86	3'Sialyl Lewis x-BSA	Dextra	NGP0403	3 atom sp	3SLexBSA3	Neu5Ac-α-(2→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
87	Glc-b-4AP-BSA	MK			XGlcbBSA	Glc-β-4AP-BSA
88	Fuc-a-4AP-BSA	MK			FucaBSA	Fuc-α-4AP-BSA
89	6-Sulfo Lewis a-BSA	Dextra	NGP0604	3 atom sp	6SuLeaBSA	(SO4)-6-Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-Sp3-BSA
90	Fuc-b-4AP-BSA	MK			FucbBSA	Fuc-β-4AP-BSA
91	Glc-b-ITC-BSA	MK			GlcbITCBSA	Glc-β-ITC-BSA
92	Gal-b-4AP-BSA	MK			Galb4APBSA	Gal-β-4AP-BSA
93	Neu5Gc-4AP-BSA	MK			Neu5GcBSA	Neu5Gc-4AP-BSA

[1]ID	Name of feature	[2]Source	Cat. no.	[3]Comments	Abbreviation	Description/Structure
94	Lewis y-tetrasaccharide-APE-HSA, 0.5 mg	IsoSep	60/95	3 atom sp	LeyHSA	Fuc-α-(1→2)-Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-β1-O-APE-HSA
95	PBS	MK			PBS	PBS
96	PBS	MK			PBS	PBS
97	Lacto-N-tetraose-APD-HSA, 0.5 mg	IsoSep	60/97	3 atom sp	LNTHSA	Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA
98	PBS	MK			PBS	PBS
99	PBS	MK			PBS	PBS
100	PBS	MK			PBS	PBS
101	Collagen type IV	Sigma	C8374		CollagenIV	Collagen type IV
102	Globotriose-HSA	Dextra	NGP2340		D-GlobTHSA	Gal-α-(1→4)-Gal-β-(1→4)-Glc-Sp3-BSA
103	Globo-N-tetraose-APD-HSA, 0.5 mg	IsoSep	60/99	3 atom sp	GlobNTHSA	GalNAc-β-(1→3)-Gal-α-(1→4)-Gal-β-(1→4)-Glc-APD-HSA
104	Globotriose-APD-HSA, 0.5 mg	IsoSep	60/90	3 atom sp	GlobTHSA	Gal-α(1→4)-Gal-β-(1→4)-Glc-β1-APE-HSA

[1]Identification number (ID), name

[2]Source: commercial or in-house conjugation (MK)

[3]'sp' refers to spacer

2.15 Mucin microarrays

Mucin extraction and purification were performed in Prof. Stephen Carrington's laboratory, Veterinary Science Centre, University College Dublin, as detailed previously (Alemka, Whelan et al. 2010). These procedures were licensed by the Department of Health and Children, Ireland, in accordance with the Cruelty to Animals Act (Ireland 1897), the European Community Directive 86/609/EC and were sanctioned by the Animals Research Ethics Committee, University College Dublin. In brief, mucus were scraped from different sections of the gastro-intestinal tract (GIT) of 8 animal species including *Bos taurus* (cow), *Cervus elaphus* (deer), *Equus ferus caballus* (horse), *Ovis aries* (sheep), *Sus scrofa domesticus* (pig), *Mus musculus* (mice), *Rattus norvegicus* (rat) and *Gallus gallus domesticus* (chicken). In addition, mucins produced by two human cell lines derived from colonic carcinoma were purified. Human gastric MUC5AC glycoproteins were obtained from HT29-MTX-E12 cells, which produce an adherent mucus layer. Clones were grown on Transwell filters for 21 days, the mucus layer removed from the cells and treated with *N*-acetylcysteine as previously described (Alemka, Clyne et al. 2010). LS174T cells, on the other hand, secrete mainly human intestinal MUC2 mucin (Tom, Rutzky et al. 1976). Supernatants from tissue culture were therefore collected and purified once the cells reached 60 to 80% confluence as detailed previously (Behrens, Stenberg et al. 2001). The mucin array consisted altogether of 48 GIT mucins and glycoproteins (Table 2.16). The printing was optimised for each mucin and executed by Dr Michelle Kilcoyne, Glycoscience group, NUI, Galway as previously reported (Kilcoyne, Gerlach et al. 2012). Incubation, scanning, data extraction and analysis procedures for labelled whole cells or labelled proteins were performed as detailed in sections 2.12.2 to 2.12.7.

Table 2.16: List of glycoproteins printed on the mucin microarray

Print ID	Sample ID	Species/Description	Mucins' origin/Glycoprotein
1	6	Equine	Stomach
2	10	Ovine	Abomasum antrum
3	11	E12	MUC5AC
4	12	Ovine	Descending colon
5	13	Ovine	Ileum
6	18	Ovine	Spiral colon
7	71	Chicken	Proximal small intestine
8	35	Ovine	Jejunum
9	36	Ovine	Duodenum
10	37	Porcine	Gastric mucin
11	41	Chicken	Large intestine
12	52	Equine	Duodenum
13	55	Deer	Jejunum
14	56	Deer	Large intestine
15	57	Bovine	Abomasum
16	58	Bovine	Duodenum
17	59	Equine	Small intestine
18	60	Equine	Left ventral colon
19	61	Bovine	Spiral colon
20	62	Deer	Duodenum
21	65	Equine	Right ventral colon
22	66	Equine	Dorsal colon
23	67	Deer	Abomasum
24	70	Chicken	Cecum
25	204	LS174T	MUC2
26	85	Porcine	Descending colon
27	85b	Porcine	Descending colon
28	86	Porcine	Jejunum
29	87	Porcine	Spiral colon
30	87b	Porcine	Spiral colon

Print ID	Sample ID	Species/Description	Mucins' origin/Glycoprotein
31	102	Porcine	Stomach
32	103	Porcine	Cecum
33	105	Mouse	Large intestine
34	106	Mouse	Cecum
35	107	Mouse	Stomach
36	121	Mouse	Small intestine
37	139	Rat	Ileum
38	140	Rat	Duodenum
39	141	Rat	Cecum
40	146	Rat	Stomach
41	147	Rat	Colon
42	ASF	Fetal calf serum	Asialofetuin
43	RB	Bovine pancreas	Rnase B
44	Fetuin	Fetal calf serum	Fetuin
45	Xferrin	Human	Transferrin
46	Ovomuc	Chicken egg white	Ovomucoid
47	PBS	PBS	PBS
48	PBST	PBS + 0.025% Tween 20	PBS + 0.025% Tween 21
49	PEW1	Pigeon	Pigeon egg white

Chapter 3: *In silico* identification of putative adhesins and experimental evaluation

Key points:

- Computer programs were used to predict the surface-exposed proteins (SEPs) of selected commensal proteomes
- The SEPs were further analysed by searching conserved domains and by sequence homology to proteins involved in adhesion
- Candidates were identified as potential adhesins for the selected commensal species from *Faecalibacterium*, *Bifidobacterium* and *Lactobacillus* genera
- Four putative pilus clusters were also identified for *L. paracasei* subsp. *paracasei*
- Expression analysis using real-Time PCR revealed five candidate genes of interest of *L. paracasei* subsp. *paracasei* to be up-regulated in media supplemented with mucin and salt

3.1 Introduction

The human microbiome is often referred to as the 'forgotten organ' or as our 'other' genome (O'Hara 2007). Commensal microorganisms composing the intestinal microbiota appear to be key factors in host homeostasis as atrophy of the microbial 'organ' correlates with inflammation in the gut (Grangette 2005; Packey and Sartor 2009; Littman and Pamer 2011; Robles Alonso and Guarner 2013). Commensal bacteria are, therefore, the subject of increasing research interest aimed at understanding their beneficial effects on the host and exploring their commercial value as probiotic strains. According to the adopted definition by the World Health Organisation, probiotics are 'live microorganisms which when administered in adequate amounts confer a health benefit on the host'. Amongst criteria considered for selection of probiotic bacteria are their capacities to adhere to the intestinal epithelial cells or the mucosa. Adhesion is believed to be essential for a strain to exert beneficial effects such as immunomodulation and pathogen exclusion on its host (Servin and Coconnier 2003; Galdeano, de Moreno de LeBlanc et al. 2007).

Adherence mechanisms in bacteria have to date been studied in pathogenic species and remain poorly understood in commensals. Attachment is usually initiated by a docking step corresponding to a reversible attachment of the bacteria to the surface. Following this step, anchoring or a more permanent attachment can occur (Dunne 2002). Bacteria usually attach to surfaces using hair-like ultra-structures called pili or fimbriae that reach out from the cells and contain a special protein called adhesin. Not all adhesins have a glycan target but most display a lectin-like function (Linke and Goldman 2011). The structures of pili differ between bacteria. The fimbriae in Gram negative and Gram positive species have been reviewed extensively and are summarised in Figure 3.1. In Gram negative bacteria (Figure 3.1 A), the prototypical pilus as described for *E. coli* type 1 fimbria is assembled through the chaperone-usher system and corresponds to the non-covalent attachment of a major pilin forming the pilus shaft (Fim A) and minor pilin units (FimF, FimG, FimH) constituting the fibrillum including a tip adhesin with a lectin function (FimH binding to mannose) (Knight, Berglund et al. 2000). In Gram positive bacteria such as *Corynebacterium diphteriae* (Figure 3.1 B), a special enzyme called sortase catalyses the assembly of pilin subunits into pili, and the anchoring of pili to the cell wall (Marraffini, Dedent et al. 2006; Mishra, Das et al. 2007). The pilin-specific sortase (SrtC) cleave the sorting signal of the secreted pilin and forms peptide bonds between pilin subunits. This

results in the covalent attachment of major subunits (such as Spa A) forming the pili backbone and a minor pilin (Spa B) decorating the shaft and presenting a tip adhesin (Spa C). The assembled structure is then anchored to the PG layer by a housekeeping sortase (SrtA) (Proft and Baker 2009; Kline, Dodson et al. 2010; Hendrickx, Budzik et al. 2011).

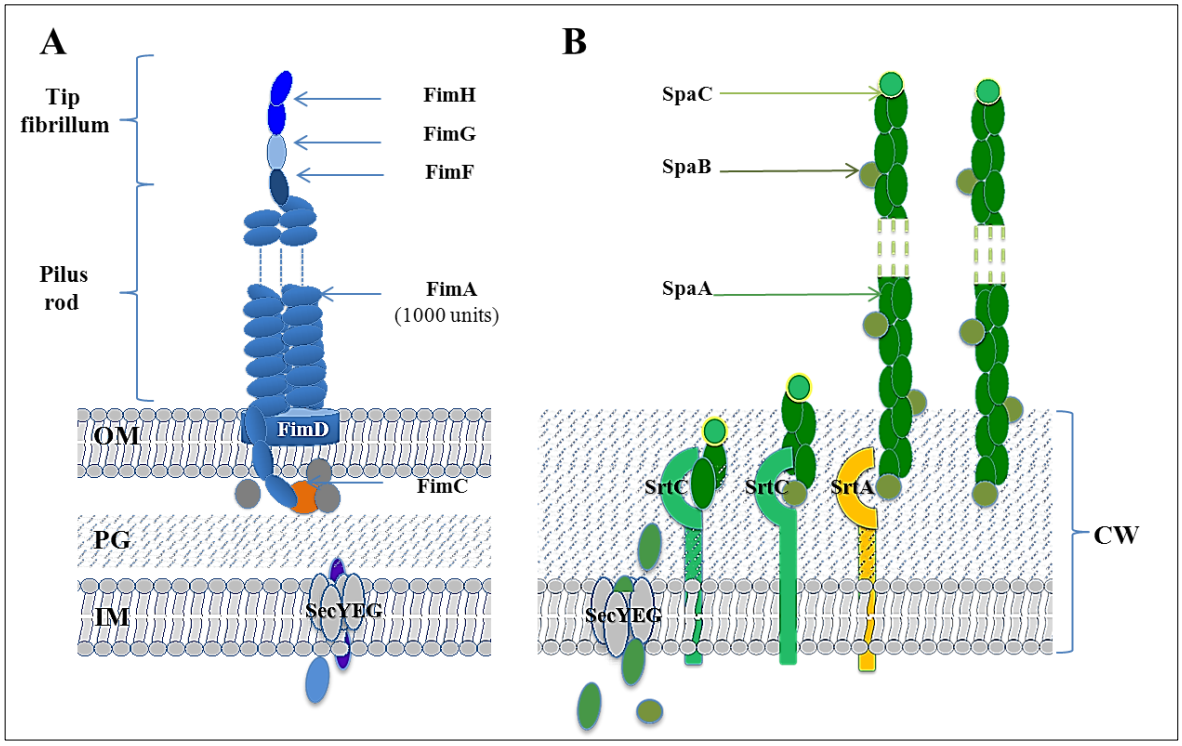

Figure 3.1: Pili encountered in bacteria

(A) Typical chaperone-usher (CU) pili of Gram negative bacteria for *E. coli*, featuring the chaperone (FimC), the usher (FimD), the major backbone pili (FimA) building the pilus rod, the minor pilin subunits (FimF, FimG and the tip mannose-adhesin FimH) forming the fibrillum.
(B) Typical sortase-dependent pili in Gram positive bacteria as described for *C. diphteriae*, *A. naeslundii* and *S. pyogenes*, featuring the pilin-specific sortase (SrtC) that covalently links the tip adhesin SpaC, the minor pilin SpaB and the major backbone unit SpaA. Termination of the assembly involves the covalent attachment of the structure to the cell wall by the house-keeping sortase (SrtA). Abbreviations are as follows: translocon (secYEG); outer membrane (OM); inner membrane (IM); peptidoglycan layer (PG); cell wall (CW).
Adapted from Scott and Zahner 2006; Mishra, Das et al. 2007; Kline, Dodson et al. 2010; Hendrickx, Budzik et al. 2011; Korea, Ghigo et al. 2011; Busch and Waksman 2012.

Adhesion does not always involve pili and many pathogens are known to express cell surface proteins to adhere closely to the epithelium or to extracellular matrix proteins (ECM) such as collagen, fibrinogen, fibronectin or laminin. These microbial surface components recognising adhesive matrix molecules (MSCRAMMs) have been identified in numerous species and display an 'IgG-like' fold domain (Patti, Allen et al. 1994;

Navarre and Schneewind 1999). In Gram negative bacteria, afimbrial adhesins are on the outer membrane and assembled through several types of secretion systems (Gerlach and Hensel 2007). In Gram positive bacteria, which display an outermost thick layer of peptidoglycan, cell wall proteins involved in adhesion include transmembrane proteins, lipoproteins and peptidoglycan anchored molecules. The latter can be covalently fixed through sortase-dependent LPXTG signals or non-covalently attached through lysin motifs (LysM), repeats of the dipeptide Glycine-Tryptophan ('GW' modules) and S-layer proteins (Desvaux, Dumas et al. 2006).

Recent studies have implicated various surface-exposed proteins from commensal bacteria with adhesion to epithelial cells (Buck, Altermann et al. 2005), ECM proteins (Styriak and Nemcova 2003) and mucins (Rojas, Ascencio et al. 2002). A number of proteins binding to mucus have recently been identified in lactobacilli that show lectin-like interactions. These include sortase-dependent pilin subunits such as SpaB and SpaC from the probiotic *L. rhamnosus* GG (von Ossowski, Reunanen et al. 2010). However, most correspond to afimbrial large surface proteins with highly repetitive motifs involved in mucus adhesion such as MUB or MucBP domains (MacKenzie, Tailford et al. 2009). Moreover, several proteins from this genus were detected with a MucBP domain (Van Tassell and Miller 2011). No such domains have been identified yet in bifidobacteria. In this genus, a single outer surface lipoprotein from *B. bifidum* MIMBb75 (BopA) has been characterised as an adhesin, mediating attachment to intestinal cells (Guglielmetti, Tamagnini et al. 2009). Recently, O'Connell et al. discovered the type IVb tight adherence pili (Tad) as a conserved locus in all bifidobacterial genomes sequenced to date, suggesting a common mechanism for colonisation of the GIT (O'Connell Motherway, Zomer et al. 2011). Very little is known about the specific ligands of microbial adhesins in the GIT. Out of the 16 commensal adhesins described to date, only one sugar ligand has been functionally identified in a sortase-dependent surface protein, termed the mannose-specific adhesin (msa) of *L. plantarum*, which adheres to mucus via mannose (Pretzer 2005). Conserved domains in the sequence of msa (*L. plantarum* CCC78612) include a Lectin_L-type motif (cd01951) as well as three MucBP domains (cl05785).

The host epithelium is covered by a layer of mucus composed of various mucin glycoproteins. These glycoproteins are heavily glycosylated with about half of their mass composed of *O*-glycans that present many attachment sites for the microbiota (Dekker,

Rossen et al. 2002; Byrd and Bresalier 2004; Perez-Vilar 2007). The Consortium for Functional Glycomics (CFG) conducted mass spectrometric profiling of glycans derived from mammalian glycoproteins and glycolipids (http://www.functionalglycomics.org/). The online portal was interrogated to draw up a list of specific human intestinal glycoconjugates using keywords such as 'gastro-intestinal tract', 'intestine', 'colon' or 'mucosa'. Altogether, 95 human GIT glycans were compiled (Appendix 1), the two largest glycans being a complex tetra-antennary *N*-linked glycan of 14 residues and a core 3-derived *O*-linked structure of 12 monosaccharides. The majority of the structures (65%) were associated with *O*-linked glycans, especially the core 3 sub-family with the structure GlcNAc-β-(1→3)-GalNAc-α-Ser/Thr. Of the *N*-linked family (25 % of all structures), 80% were complex structures. The remaining (10%) was made up of glycolipids.

Bioinformatics tools have been developed to curate the vast amount of data generated by genomic, proteomic and glycomic research and this approach could help in our understanding of the cross-talk between bacteria and host. Predictive computer modelling of bacterial adhesion to mucosal surfaces has been employed in infection studies (Lee, Lim et al. 2000; de Jong, Vissers et al. 2007; Yousef and Espinosa-Urgel 2007). Programs for the prediction of surface-exposed proteins such as SurfG+ or SPAAN, which permit the rapid screening of whole proteomes, have been developed (Sachdeva, Kumar et al. 2005; Barinov, Loux et al. 2009). The predictive localisations rely on well-established algorithms (hidden Markov pattern recognition models) that detect specific sequences such as transmembrane helices (TMH) (characteristic of membrane proteins), signal sequences (secreted proteins or lipoproteins) and retention signals (linked to the cell wall) (Eddy 1998; Juncker, Willenbrock et al. 2003; Bendtsen, Nielsen et al. 2004; Boekhorst, de Been et al. 2005; Gardy, Laird et al. 2005; Kahsay, Gao et al. 2005). Java program SurfG+ was designed specifically to identify proteins that protrude from the cell wall as a mean of identifying potential vaccine candidates. It was validated on Gram positive *Streptococcus pyogenes.* SurfG+ was also used to compare lactobacilli strains from the human GIT or from yoghurt (Barinov 2009; Barinov, Loux et al. 2009). Barinov *et al*. reported that commensal lactobacilli contained a higher percentage of surface exposed proteins than dairy strains. Moreover, half of the LPXTG anchored proteins from commensal lactobacilli from *L. acidophilus, L. johnsonii and L. gasseri* contained mucus-binding domains (Barinov 2009).

Here, we describe an *in silico* approach to identification of adhesins involved in host-microbe interactions. Commensal species representing three genera of commensals, the genomes of which have been sequenced - *Bifidobacterium longum sp., Lactobacillus paracasei* subsp. *paracasei* and *Faecalibacterium prausnitzii* - were first investigated using the SurfG+ flow scheme to identify putative adhesins based on their predicted location (surface-exposed). Secondly, the literature was mined to list the commensal adhesins known to date and to align these sequences with the species of interest to find homologous proteins. Thirdly, candidate genes of interest, coding for putative adhesins, were selected and examined experimentally by qPCR.

3.2 *In silico* analysis of the proteome from commensal species

Bacterial surface-exposed proteins were investigated using the SurfG+ program to analyse three commensal species: *Faecalibacterium prausnitzii* A2-165 (abbreviated as *F. prausnitzii*), *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302 (abbreviated as *L. paracasei*) and *Bifidobacterium longum* subsp. *longum* ATCC 55813 (abbreviated as *B. longum*) (Chapter 2 Table 2.1). Results from SurfG+ were presented as spreadsheet tables and included the classification of the proteins into predicted cell compartments (Table 3.1).

The relative number of proteins of each type was similar between all three species. The majority of the proteins analysed were either cytoplasmic (72% to 76%) or membrane-bound (13% to 16%). A small proportion was potentially secreted (2% to 4%). The predicted surface exposed proteins (SEPs) represented approximately 8% of the total proteome. The total number of SEPs was 212 for *F. prausnitzii*, 162 for *B. longum* subsp. *longum* and 213 for *L. paracasei* subsp. *paracasei* (Table 3.1).

Once the surface exposed localisation was established, it was possible to determine from the analysis the topology of the identified SEPs in the three strains and to define various types such as lipoproteins, loops, proteins exposing an amino terminal (Nt), protruding with a carboxy terminal (Ct) and possessing or not cell wall retention signals. The highest proportion of lipoproteins (E, Figure 3.2) was found in *F. prausnitzii* in comparison to the two other Gram positive species. In the latter, a higher number of proteins containing a retention signal was noticed (R, Figure 3.2). In addition, a higher level of loops was found in the *Lactobacillus* strain.

Table 3.1: SurfG+ prediction of protein localisation

	[2]Fpr		Blon		Lpp	
Localisation	[1]Number	[1]%	Number	%	Number	%
Cytoplasm	2209	76.3	1476	75.4	2022	72.5
Membrane	372	12.9	273	13.9	454	16.2
[1]SEP	212	7.3	162	8.2	213	7.6
Secreted	101	3.5	45	2.3	100	3.5
Total predicted	2894	100.0	1956	100.0	2789	100.0

[1]Figures indicate the number of proteins in each category and corresponding percentages of the total number of proteins per strain;
[2]Abbreviations are as follows, SEP: surface exposed protein; Blon: *B. longum* subsp. *longum*; Fpr: *F. prausnitzii*; Lpp: *L. paracasei* subsp. *paracasei.*

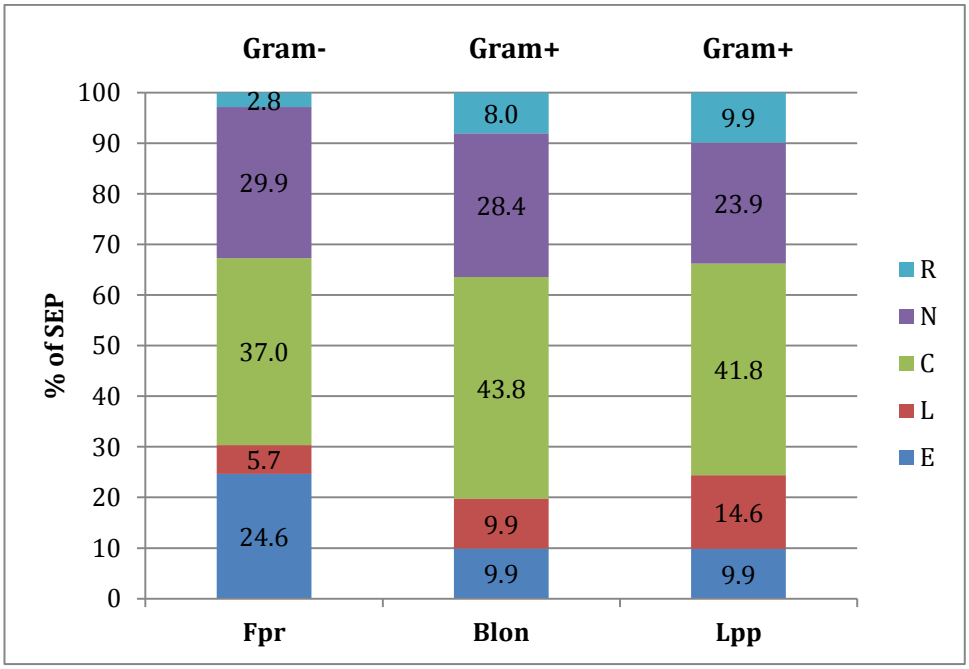

Figure 3.2: SEPs topology as predicted by SurfG+ software

Percentages of SEPs per strain for each structural type; R: proteins containing a retention signal; N: transmembrane helix with surface exposed N terminal end; C: transmembrane helix with surface exposed C-terminal end; L: surface exposed loop; E: Lipoprotein.
Refer to Table 3.1 for abbreviations.

Another interesting aspect of the analysis is the chain length of the predicted surface exposed proteins. Above all, 90% of the SEPs contain less than 1000 amino acids (AA) and the vast majority comprises between 251 and 500 AA for the three species tested. *B. longum* presents the highest number of large proteins, more than twice that of the other strains (18 SEPs above 1000 AA as opposed to 7 SEPs for *L. paracasei* and six for *F. prausnitzii* (Figure 3.3).

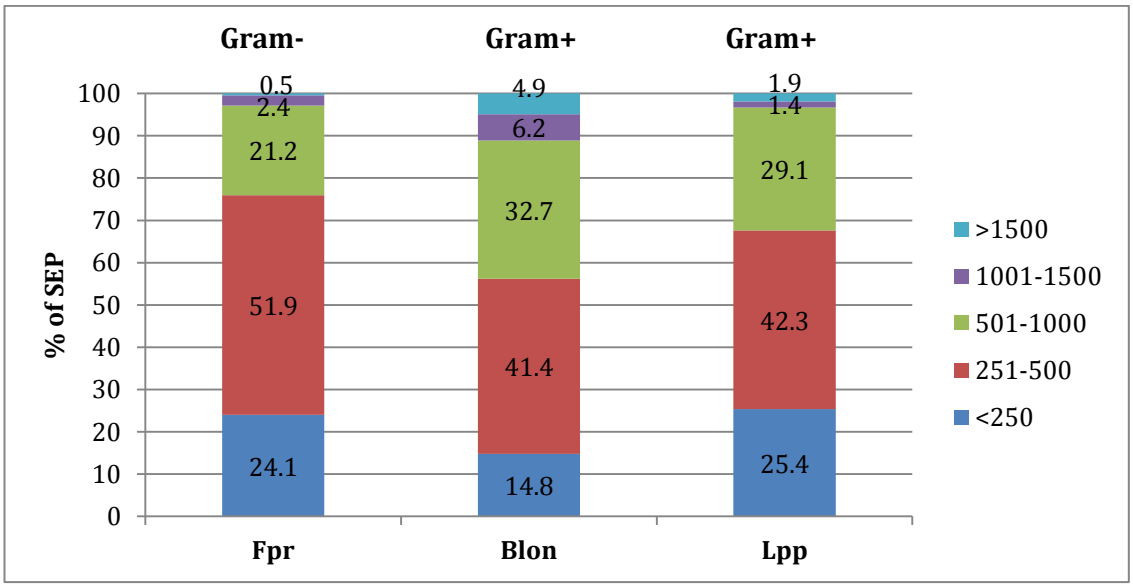

Figure 3.3: SEPs size as predicted by SurfG+ software

Percentages of proteins in each size group based on the chain length (AA) in each strain. Refer to Table 3.1 for abbreviations.

3.3 Selection of putative adhesin candidates

The predicted surface exposed protein sequences obtained from SurfG+ were first entered in the NCBI BLASTp program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify matches with non-redundant sequenced bacterial proteins. Secondly, they were investigated for conserved motifs found in known adhesins using the Conserved Domains Database. These included various signals described in pilin subunits and motifs found in bacterial proteins binding mucus or ECM. Thirdly, alignment tools were used to shortlist selected proteins. In addition, the commensal proteomes were searched for homologous sequences to the functionally characterised adhesins from commensal bacteria (Chapter 1, section 1.5.2). A list of putative candidate proteins was finally generated for each strain (Tables 3.2 to 3.8), namely *F. prausnitzii*, *L. paracasei* and *B. longum*. The latter was known as

Bifidobacterium longum subsp. *infantis* ATCC 55813 in the Human Microbiome project at the start of this study before being re-classified as subspecies *longum*. Since the strain available for experimentation in the laboratory was *B. longum* subsp. *infantis* ATCC 15697, the latter was also investigated for putative adhesins.

3.3.1 *Faecalibacterium prausnitzii* A2-165

Three proteins that presented the highest homology to the functionally characterised commensal adhesin list (Chapter 1, section 1.5.2) based on sequence analysis were selected. These were the translation factor EF_Tu (71% conserved), the fibronectin-binding protein FbpA (32%) and the chaperonin GroEL (64%). Most molecules of interest were predicted to be surface exposed by SurfG+ (Table 3.2). A putative lipoprotein (ZP_05614164) was included in the candidate list due to its homology to a group of proteins containing an immunoglobulin Ig-like domain that is found in molecules involved in Gram negative cell adhesion, such as intimins or invasins (Tsai, Yen et al. 2010). One SEP also presented a conserved fasciclin domain which is also involved in cell adherence (ZP_05616293) (Moody and Williamson 2013). Moreover, two sequences that presented S-layer protein characteristics (ZP_05615730 and ZP_05616094) as well as a pilin (isopeptide linkage domain-like protein) were included in the candidate list. However, the best candidates seem to be the two largest proteins (adhesin/autotransporter ZP_05616615 of 1005 AA and pilin ZP_05616166 of 1032 AA). The autotransporter (ZP_05616615) has homology to a filamentous hemagglutinin from *Burkholderia sp.* and an adhesin from *Yersinia pseudotuberculosis*, while the protein ZP_05616166 is homologous to the minor pilin Cpa, which is located at the tip of the sortase-dependent Type III pilus of *Streptococcus pyogenes* (Quigley, Zähner et al. 2009). Interestingly, a sortase predicted in the secreted proteins compartment (sequence EEU95448) was detected in *F. prausnitzii* with a sortase class B domain (Sortase_B_2[cd05826]). Other cell compartments were also explored in the case of *F. prausnitzii*. Two interesting sequences were found located in the cytoplasm (a putative fimbrial protein FimC, ZP_05615807) and in the membrane (type I pilus protein CsuB, ZP_05615656). Both FimC and CsuB correspond to major assembly proteins (chaperones) of the typical chaperone-usher type 1 pili from Gram negatives (Figure 3.1 A) (Knight, Berglund et al. 2000; Tomaras, Flagler et al. 2008).

Table 3.2: Putative adhesins for *F. prausnitzii* A2-165 following SurfG+ predictive software analysis

[1]Accession	[2]AA	Name	[3]CDD	[4]Potential adhesin homologue
ZP_05614002	401	translation elongation factor Tu	EF_Tu [cd01884]; EFTU_II[cd03697]; EFTU_III[cd03707]; PRK00049[PRK00049]	EF-Tu (71%, AAS08831, *Lactobacillus johnsonii* NCC 533)
ZP_05614692	546	fibronectin-binding protein	DUF814 super family [cl05307]; FbpA [pfam05833]	FBPA (32%, AAV42987, *Lactobacillus acidophilus* NCFM)
ZP_05616027	585	chaperonin GroL	GroEL [cd03344]; groEL[PRK12849]	GROEL (64%, AAS08453, *Lactobacillus johnsonii* NCC 533)
ZP_05614164	413	putative lipoprotein	Periplasmic_Binding_Protein_Type_1 super family [cl10011]	ZP_07631870 Ig domain-containing protein [*Clostridium cellulovorans* 743B]
ZP_05615730	431	the GLUG motif protein		YP_001394473 surface-layer protein [*Clostridium kluyveri* DSM 555]
ZP_05616094	727	putative S-layer domain protein	SLH super family [cl02857]	ZP_07635694 S-layer domain protein [*Ruminococcaceae bacterium* D16]
ZP_05616166	1032	pilin	Flg_new super family [cl09714]	ABM66469 Cpa [*Streptococcus pyogenes*]
ZP_05616171	259	pilin isopeptide linkage domain protein-like protein	strep_pil_rpt super family [cl14038]	EFF17027 pilin isopeptide linkage domain protein [*Ruminococcus albus* 8]
ZP_05616293	363	putative fasciclin domain protein		ACV81612 periplasmic binding protein/LacI transcriptional regulator [*Nakamurella multipartita* DSM 44233]

[1]Accession	[2]AA	Name	[3]CDD	[4]Potential adhesin homologue
ZP_05616615	1005	putative autotransporter/adhesin		YP_003908198 filamentous hemagglutinin family outer membrane protein [*Burkholderia* sp. CCGE1003]
ZP_05615807	91	putative fimbrial assembly protein FimC	TraX super family[cl05434]	cytoplasm
ZP_05615656	114	type I pilus protein CsuB		membrane

[1]Candidate list of putative adhesins for *F. prausnitzii* A2-165 with accession number in NCBI database
[2]Amino acid chain length (AA)
[3]Conserved domains present in the sequence (CDD)
[4]Homologous sequences in GenBank following BLASTp and Blink to known commensals adhesins

3.3.2 *Bifidobacterium longum* subspecies

3.3.2.1 *B. longum* subsp. *longum* ATCC 55813

As for *F. prausnitzii,* some proteins presented relatively high homology to the functionally characterised commensal adhesins (Table 3.3). These were homologous to the translation factor EF_Tu (EEI81120, 67% identical) and the chaperonin GroEL (EEI80571, 59%). The only adhesin functionally characterised so far in bifidobacteria is BopA, a lipoprotein from *B. bifidum* MIMBb75 involved in adhesion to intestinal epithelial cells (Guglielmetti, Tamagnini et al. 2009). In *B. longum* subsp. *longum,* the best homology obtained with BopA was for an ABC transporter (EEI79775, 44%), which was selected. Three proteins with PG surface anchors were also included. The three sequences were found to be homologous to a secretion protein (ZP_03976134), an S-layer protein (ZP_03976356) and to virulence factor-related protein 'mVin' (ZP_03977128) found in several bifidobacteria. Another sequence (ZP_03977104) was also included as it presented an anchor domain, a fimbrial motif found on major pilins and classified as a lectin-like domain (isopeptide formation D2 domain, TIGR04226, see Table 1.7), a collagen binding domain (CnaB, pfam05738) and as it was homologous to an adhesion protein from *Bacillus cereus.* Three final sequences selected presented homology to fimbrial proteins: ZP_03977101, homologous to a fimbrial protein from *B. longum*, ZP_03976131, with a CnaB domain analogous to the major pilin FimA from *B. dentium* Bd1, and ZP_03976132 homologous to an internalin from *Listeria monocytogenes.*

As well as adhesins, the predicted SEPs listed by SurfG+ for *Bifidobacterium longum* subsp. *longum,* also contained four sortases. Three of these show homology to pilin-specific sortases (C sortase family) and thus might be involved in the assembly of three sortase-dependent pili in this strain (Table 3.4) (Hendrickx, Budzik et al. 2011).

Table 3.3: Putative adhesins for *B. longum* subsp. *longum* ATCC 55813 following SurfG+ predictive software analysis

[1] Accession	[2]AA	Name	[3]CDD	[4]Potential adhesin homologue
EEI81120	399	elongation factor Tu	EF_Tu [cd01884]; EFTU_II [cd03697]; EFTU_III [cd03707]; PRK00049[PRK00049]	EF-Tu (67%, AAS08831, *Lactobacillus johnsonii* NCC 533)
EEI80571	541	chaperonin GroEL	GroEL [cd03344]; groEL[PRK12849]	GROEL (59%, AAS08453, *Lactobacillus johnsonii* NCC 533)
EEI79775	546	ABC transporter, substrate-binding protein, family 5	SBP_bac_5[pfam00496]; PBP2_NikA_DppA_OppA_like [cd00995]	BOPA (44%, ADE67063, *Bifidobacterium bifidum*)
ZP_03976134	1011	cell wall surface anchor family protein	SCP_bacterial [cd05379]	secretion protein [*Bifidobacterium longum*] [WP_012577602]
ZP_03976356	785	S-layer domain protein	[COG1361] S-layer domain	S-layer protein [*Bifidobacterium longum* DJO10A; YP_001954835.1]
ZP_03977128	687	conserved hypothetical protein	MATE_MurJlike [cd13123]; mviN [TIGR01695]	virulence factor MVIN family protein [*Bifidobacterium breve* UCC2003; YP_007555704]
ZP_03977104	537	conserved hypothetical protein	RrgB_K2N_iso_D2 [TIGR04226]; Cna_B [pfam05738]; Gram_pos_anchor [pfam00746]	collagen adhesion protein [*Bacillus cereus* AH621] [EEK73966.1]
ZP_03977101	2573	exoprotein involved in heme utilization or adhesion		fimbrial protein [*Bifidobacterium longum*] [WP_011068613.1]
ZP_03976131	571	conserved hypothetical protein	Cna_B [pfam05738]	FimA fimbrial subunit-like cell surface protein [*Bifidobacterium dentium* Bd1; YP_003359706.1]

[1] Accession	[2]AA	Name	[3]CDD	[4]Potential adhesin homologue
ZP_03976132	2327	conserved hypothetical protein	Flg_new [pfam09479] 2 repeats	internalin [*Listeria monocytogenes* str. 4b H7858][ZP_05266073.1]

[1] Candidate list of putative adhesins with accession number in NCBI database
[2] Amino acid chain length (AA)
[3] Conserved domains present in the sequence (CDD)
[4] Homologous sequences in GenBank following BLASTp and Blink to known commensals adhesins

Table 3.4: Putative sortases of *B. longum* subsp. *longum* ATCC 55813 predicted by SurfG+

NCBI accession	Name	Chain length (AA)	[1]Topology	[2]CDD
ZP_03975680	Sortase	377	L	Sortase_C_3[cd05827]
ZP_03976133	Sortase family protein	325	C	Sortase_C_3[cd05827]
ZP_03977102	Sortase family protein	328	N	Sortase_C_3[cd05827]
ZP_03977489	Sortase	412	L	Sortase_D_5[cd05830]

[1]Predicted topology in SurfG+ with L: surface exposed loop; C: transmembrane helix with surface exposed C-terminal end; N: transmembrane helix with surface exposed amino-terminus
[2]Conserved domain in the sequence (CDD, NCBI database)

3.3.2.2 *B. longum* subsp. *infantis* ATCC 15697

Bifidobacterium longum subsp. *infantis* ATCC 15697 (*B. infantis* Reuter 1963), contains a larger proteome than the *B. longum* subsp. *longum* strain ATCC 55813 but a plasmid present in the latter is missing in *B. infantis*. Because of significant homologies to commensal adhesins, four proteins were selected including the translation factor EF_Tu (66% identical to the known adhesin), the chaperonin GroEL (59%), together with two solute binding proteins (YP_002322242 and YP_002323851) homologous to respectively FbpA (38%) and to BopA (44%) (Table 3.5).

Furthermore, proteins presenting non-covalent attachment domains such as LysM (YP_002322323) or S-layer protein (YP_002323430) were considered.

A large protein (YP_002322015 of 1480 AA) containing an 'LPXTG' motif was homologous to a glucan-binding surface anchor molecule of *Streptococcus suis*. In streptococci, these proteins are lectins that are specific for dextran and are involved in aggregation and virulence (Okamoto-Shibayama, Sato et al. 2006).

A protein (YP_002322168) with the von Willebrand factor motif (vWFA; [cd00198]) was also retained as some lectins such as the tip adhesin SpaC from the sortase-dependent pilus of *L. rhamnosus* GG contain this domain.

Finally, three other sequences were of interest. These were a homologue of the adhesion exoprotein of strain ATCC 55813 (YP_002322905), a CHAP domain containing protein (YP_002323586) homologous to a surface antigen from *B. longum* DJ010A and the allergen V5 (YP_002322725) which is homologous to a large adhesin from *Haemophilus somnus.*

No homologous sequences were found to the list of sortases from strain ATCC 55813 obtained following SurfG+ analysis. Twelve sequences could be identified in the genome of *B. infantis* following a search in NCBI database; these are not necessarily involved in pilus assembly. The genome of strain ATCC 15697, revealed an overabundance of the Family 1 of solute binding proteins (F1SBPs) compared to other bifidobacterial genomes (Garrido, Kim et al. 2011). Yet, Garrido et al. noticed the absence of sortase-dependent pili. F1SBPs are ATP-binding cassette (ABC) transporters associated with the import of oligosaccharides that can be involved in adhesion to human intestinal cells *in vitro* for this strain (Garrido, Kim et al. 2011).

However, a type IVb tight adherence pilus (Tad) has been described recently as conserved in all bifidobacteria (O'Connell Motherway, Zomer et al. 2011). The *Tad* locus is composed of genes coding for integral membrane proteins involved in the pilus assembly, namely *Tad*Z, *Tad*B, *Tad*C and the associated ATPase *Tad*A. *Tad*Z seems to be involved in the localisation of pili to the poles of bacteria (Mishra, Das et al. 2007; Xu, Christen et al. 2012). The *Tad* pilus is comprised of polymers of pilin subunits, with the fimbrial low molecular weight protein *flp* (6.5 kDa) being the major pilin (Inoue, Tanimoto et al. 1998). The shaft is decorated by the minor pilins *Tad*E and *Tad*F that might be the adhesins. The secreted pre-pilins contain a hydrophobic leader peptide, which is cleaved by the *Tad*V peptidase and processed for assembly. Alignments were therefore performed using *Bifidobacterium breve* UCC2003 *Tad* locus to determine the possible cluster in *B. infantis* ATCC 15697 (Figure 3.4, Table 3.6). Some encoded proteins were not defined in the NCBI database, such as TadB and TadC, but both sequences contained the type II secretion system domain (T2SF) found in proteins involved in the assembly of type IV pili. Similarly, *Tad*Z homolog contained the septum site-determining protein domain (minD) found in *B. breve* TadZ. TadE was identified by homology between *B. breve* gene BBR_0136 and *B. infantis* gene Blon_0155, both coding for a short pilin of 95 AA. The protein homologous to the TadF pilin from *B. breve* was referred to as 'TadE family protein' in *B. infantis* and was slightly larger (128AA compared to 126 AA in *B. breve*).

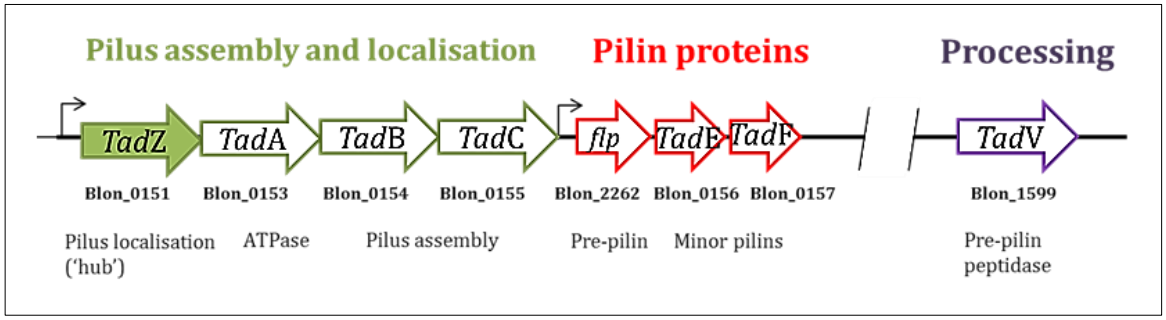

Figure 3.4: Proposed *Tad* locus from *B. longum* subsp. *infantis* ATCC 15697

Schematic based on the conserved tad locus in the genus *Bifidobacterium* based on *B. breve* UCC 2003 sequences (O'Connell Motherway *et al.*, 2011). Each arrow represents an open reading frame (ORF) with the *B. longum* subsp. *infantis* gene identification given below the arrow and the generic gene name given within the arrow. The functions of the encoded proteins are indicated below the gene name.

Table 3.5: Putative adhesins for *B. longum* subsp. *infantis* ATCC 15697 following SurfG+ predictive software analysis

[1] Accession	[2]AA	Name	[3]CDD	[4]Potential adhesin homologue
YP_002323370	399	elongation factor Tu	EF_Tu [cd01884]; EFTU_II [cd03697]; EFTU_III [cd03707]; PRK00049[PRK00049]	EF-Tu (66%, AAS08831, *Lactobacillus johnsonii* NCC 533)
YP_002322176	541	GROEL chaperonine	GroEL [cd03344]; groEL[PRK12849]	GROEL (59%, AAS08453, *Lactobacillus johnsonii* NCC 533)
YP_002322242	270	extracellular solute-binding protein, family 3	SBP_bac_3 [pfam00497]; PBPb [cd00134]	FbpA (38%, AAV42987, *Lactobacillus acidophilus* NCFM)
YP_002323851	577	extracellular solute-binding protein, family 5	PBP2_Lpqw [cd08501]; SBP_bac_5 [pfam00496]; DdpA [COG0747]	ADE67063 (44% BopA)
YP_002322323	119	peptidoglycan-binding	LysM [cd00118]	LysM-like protein [YP_001954112; *B. longum* DJO10A]
YP_002323430	388	hypothetical protein Blon_1985	COG1361 [COG1361]	S-layer protein [YP_001954835; *B. longum* DJO10A]
YP_002322015	1480	LPXTG-motif cell wall anchor domain protein	LPXTG_anchor super family [cl11747]; TIGR04228, adhesin isopeptide-forming domain	glucan-binding surface-anchored protein [YP_003028232, *Streptococcus suis* BM407]
YP_002322905	373	narrowly conserved hypothetical protein	DUF2974 [pfam11187]	adhesion exoprotein [ZP_03977104, *B.longum* subsp. *longum* ATCC55813]
YP_002323586	318	CHAP domain containing protein	NLPC_P60 [cl11438]	Surface antigen protein [YP_001955727, *B. longum* DJO10A]

[1] Accession	[2]AA	Name	[3]CDD	[4]Potential adhesin homologue
YP_002322725	973	Allergen V5/Tpx-1 family protein, possible large adhesin	SCP CAP [pfam00188]	large adhesin [YP_719270, *Haemophilus somnus* 129PT]

[1] Candidate list of putative adhesins accession number in NCBI database
[2] Amino acid chain length (AA)
[3] Conserved domains present in the sequence (CDD)
[4] Homologous sequences in GenBank following BLASTp and Blink to known commensals adhesins

Table 3.6: *Tad* locus in *Bifidobacterium longum* subsp. *infantis* ATCC 15697

[1]Gene			[2]Protein - *B. infantis*			
Name	***B. breve***	***B. infantis***	**Accession**	**AA**	**Name**	**CDD**
***Tad*A**	BBR_0133	Blon_0152	YP_002321659	355	type II secretion protein E	VirB11 Type IV secretory pathway [COG0630]; RecA-like_NTPases super family[cl17233]
***Tad*B**	BBR_0134	Blon_0153	YP_002321660	217	hypothetical protein	T2SF [pfam00482]
***Tad*C**	BBR_0135	Blon_0154	YP_002321661	200	hypothetical protein	T2SF [pfam00482]
***Tad*E**	BBR_0136	Blon_0155	YP_002321663	95	hypothetical protein	DUF4244 [pfam14029]
***Tad*F**	BBR_0137	Blon_0156	YP_002321664	128	*Tad*E family protein	*Tad*E [pfam07811]
flp	BBR_0138	Blon_0157	YP_002323698	125	*Tad*E-like protein	helicase/secretion neighborhood *tad*E_like_DECH [TIGR03816];
***Tad*Z**	BBR_0132	Blon_0151	YP_002321658	316	hypothetical protein	MinD [cd02036]
***Tad*V**	BBR_0901	Blon_1599	YP_002323056	157	prepilin type IV	

[1]Homologous genes of the *Tad* locus between *B. breve* UCC2003 and *B. infantis*
[2] Corresponding protein in *B. infantis* with accession number (NCBI), amino-acid length (AA), name and conserved domains (CDD)

3.3.3 *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302

Most of the commensal adhesins have so far been characterised in lactobacilli. However, none as yet have been described in *L. paracasei*. Adhesin candidates for the latter are listed in Table 3.7. Amongst them, four proteins were homologous to functionally characterised commensal adhesins. The protein with accession number ZP_03965231 presented high homology to the translation factor EF_Tu (87% identical), ZP_03963800 to the chaperonin GroEL (84% identity), ZP_03965124 to the fibronectin-binding protein A (FbpA, 47%) and ZP_03964327 to the collagen-binding proteins (CnBP/MapA, 57%), all of which are known adhesins in lactobacilli.

In addition, six SEPs contained collagen-binding domains (CollagenBindB [cl14619]). These included the largest protein (ZP_03964000, 2240 AA), which possesses nine repeats of collagen-binding motif and which is homologous to a collagen adhesion protein from *Lactobacillus rhamnosus* LSM2-1. Two sequences (ZP_03963620 and ZP_03964057) showed high similarity to the collagen-binding proteins from *L. plantarum* (ACT 98655) and *Enterococcus faecium*. Another SEP (ZP_03964058) aligned with the major pilin PilA, also from *E. faecium.*

Finally, several sequences were homologous to proteins from a genetically related strain *Lactobacillus rhamnosus* GG (LGG), also a member of the *L. case*i group. Two large proteins of respectively 1039 AA (ZP_03963571) and 1173 AA (ZP_03963572), shared common sequences with two adhesion extracellular proteins from LGG (exoproteins CAR88818 and WP_003662334). The genome of LGG presents sortase-dependent pili clusters. In the predicted SEPs list from SurfG+, two sequences with CollagenbindB domains [cl14619] (ZP_03963682 and ZP_03963680) were homologous to two proteins encoded by the SpaFED pilus cluster of LGG. These were respectively homologous to the major pilin SpaD (YP_003172116) and the adhesin SpaF (BAI42808). A fimbrial subunit (ZP_03963300) possessing an LPXTG anchor, was also presenting high homology to SpaA, the major pilin encoded by the *spaCBA* pilus cluster in LGG (CAR86337). The predicted SEPs also contain three sortases (listed in Table 3.8). One sequence is homologous to a sortase from *L. casei* (srtA, WP_003588431) and two show homology to class C sortases i.e. pilin-specific sortases (srtC, Sortase_C_3 [cd05827]). Altogether, this suggests the presence of at least two sortase-dependent pili clusters in *L. paracasei*.

Table 3.7: Putative adhesins for *L. paracasei* subsp. *paracasei* ATCC 25302 following SurfG+ predictive software analysis

[1]Accession	[2]AA	Name	[3]CDD	[4]Potential adhesin homologue
ZP_03965231	396	elongation factor Tu	EF_Tu [cd01884]; EFTU_II [cd03697]; EFTU_III [cd03707]; PRK00049[PRK00049]	EF-Tu (87%, AAS08831, *Lactobacillus johnsonii* NCC 533)
ZP_03963800	544	GROEL chaperonine	GroEL [cd03344]; groEL[PRK12849]	GROEL (84%, AAS08453, *Lactobacillus johnsonii* NCC 533)
ZP_03965124	574	fibronectin binding protein A	DUF814 super family[cl05307]; FbpA[pfam05833]	FBPA (47%, AAV42987, *Lactobacillus acidophilus* NCFM)
ZP_03964327	270	collagen binding protein	PBPb [cd00134]	CnBP/MapA (57%, CAA68052, CnBP *L. reuteri*)
ZP_03963300 (EEI69168)	334	fimbriae subunit	LPXTG_anchor super family [cl11747]	pilus specific protein, major backbone protein, SpaA [CAR86337, *L. rhamnosus GG*]
ZP_03963571 (EEI68918)	1069	adhesion exoprotein	KxYKxGKxW super family [cl14016]; LRR_5[pfam13306]: 6 repeats; Big_3 super family[cl06524]; Big_3 super family[cl06524]	adhesion exoprotein [CAR88818, *L. rhamnosus* GG]
ZP_03963572 (EEI68919)	1173	adhesin	LPXTG_anchor [TIGR01167]	adhesin [WP_014570244, *L. rhamnosus GG*]
ZP_03963620 (EEI68842)	611	outer membrane protein	CollagenBindB super family [cl14619]; Collagen_bind super family[cl05349]; COG4932 [COG4932]	collagen binding protein [ACT98655, *L. plantarum*]
ZP_03964000 (EEI68395)	2240	collagen adhesion protein	CollagenBindB super family [cl14619]: 9 repeats; COG4932 [COG4932]	collagen adhesion protein [EEN79876, *L. rhamnosus* LMS2-1]

ZP_03964057 (EEI68452)	655	Cna B domain protein	CollagenBindB super family [cl14619]: 2 repeats	collagen adhesion protein [ZP_05674116, *Enterococcus faecium*]
ZP_03964058 (EEI68453)	683	cell wall surface anchor family protein	CnaB pfam[05738]: 2 repeats; RrgB_K2N_iso_D2 [TIGR04226]	putative pilus subunit protein PilA [ZP_06676416, *E. faecium*]
ZP_03963680 (EEI68682)	931	possible outer membrane protein	CollagenBindB super family [cl14619]: 2 repeats	pilus specific protein, ancillary protein involved in adhesion, SpaF [BAI42808, *L. rhamnosus* GG]
ZP_03963682 (EEI68684)	519	pilus subunit protein	CnaB pfam[05738]; RrgB_K2N_iso_D2 [TIGR04226]; COG4932 [COG4932]	pilus specific protein, SpaD, [YP_003172116, *L. rhamnosus* GG]

[1] Candidate list of putative adhesins accession number in NCBI database

[2] Amino acid chain length (AA)

[3] Conserved domains present in the sequence (CDD)

[4] Homologous sequences in GenBank following BLASTp and Blink to known commensals adhesins

Table 3.8: Putative sortases of *L. paracasei* subsp. *paracasei* ATCC 25302 predicted by SurfG+

NCBI accession	Name	Chain length (AA)	[1]Topology	[2]CDD
ZP_03963301	Sortase	348	C	Sortase_C_3[cd05827]
ZP_03963898	Sortase	235	C	Sortase_A_1[cd06165]
ZP_03964059	Possible sortase	278	L	Sortase_C_3[cd05827]

[1]Predicted topology in SurfG+ with L: surface exposed loop; C: transmembrane helix with surface exposed C-terminal end;
[2]Conserved domain in the sequence (CDD, NCBI database)

Based on conserved domains, sequence alignments and genome-walking from the predicted adhesins in *L. paracasei*, three putative clusters encoding sortase-dependent pili were identified. Each of these contained one pilin-specific sortase (SrtC).

One cluster (Figure 3.5 A) included a protein (termed Pil1, with sequence number EEI68452) containing two repeats of collagen-binding domain (CnaB [pfam 05738]). These repeated motifs are found in major backbone pilin subunits, forming the stalk. A second pilin (termed Pil2, EEI68453) contained a fimbrial isopeptide formation D2 domain (RrgB_K2N_iso_D2 super family [cl18869]), which is found in streptococcal adhesins with lectin-like properties (Marchler-Bauer, Zheng et al. 2013).

In addition, two clusters corresponded to sortase-mediated pilus assembly (Spa) fimbriae (*SpaCBA* in Figure 3.5 B and SpaFED in Figure 3.5 C). The prototypical Spa pili were first described for *Corynebacterium diphtheriae* and *Streptococcus pyogenes*. In these pathogens, the Spa pilus consists of a major backbone pilin (SpaA or SpaD) that forms the shaft and two minor pilin subunits, one at the base bound to the cell wall (SpaB, SpaE) and a larger one at the tip of the polymer (SpaC and SpaF), which is responsible for the adhesion (Ton-That, Marraffini et al. 2004; Rogers, Das et al. 2011).

The gene cluster that might correspond to the *SpaCBA* operon in *L. paracasei* (Figure 3.5 B) consists of four pilin subunits, two of which present homology to the tip adhesin SpaC from LGG. The largest pilin with sequence number EEI69165 (termed here SpaC_vWF, 65kDa), contains adhesin-like conserved domains such as the von Willebrand Factor motif found in the mannose-specific adhesin (msa) from *L. plantarum* (vWFA [cd00198]), as

well as in SpaC from LGG; a collagenBindB domain ([cl4619]) and a metal ion-dependent adhesion site (MIDAS), suggesting this protein could display a lectin-like function (Gross, van der Meulen et al. 2008; Marchler-Bauer, Lu et al. 2011). The *SpaCBA* cluster from *L. paracasei* harbours two genes that might be involved in the regulation of the transcription. A DNA recombinase (resolvase, EEI69164) is found directly upstream of the sortase**.** Resolvases are replication origin components located upstream of pili clusters in virulence plasmids such as the EAF plasmid found in EPEC and ETEC strains producing bundle forming pili (Sohel, Puente et al. 1993). A regulatory protein (ArpU, EEI68878) containing a sigma factor (domain accession [TIGR01637]) needed for initiation of RNA synthesis is also found upstream of the pili cluster.

The proposed *SpaFED* operon (Figure 3.5 C) contains three pilin subunits (namely SpaD, SpaE and SpaF) including a possible large adhesin (SpaF, EEI68682, 102kDa) and might also contain several regulatory elements**.** These could include a transcriptional regulator, the universal stress protein (usp, protein accession number EEI68681); a motif required for the binding of DNA, the DNA-binding helix turn helix protein (DBHTP, EEI68680), which is involved in pili phase variation in *Neisseria sp.* (Arvidson and So 1995); a tyrosine phosphatase found in signalling pathways (TPase, EEI68679); and a protein involved in the control of gene expression, the iron-sulphur cluster-binding protein (QueG, EEI68878). The latter is identical to the epoxyqueuosine reductase from a different strain of *L. paracasei* (protein accession number EPC74700) that contains solenoid protein domain repeats (HEAT) and a QueG domain involved in tRNA maturation.

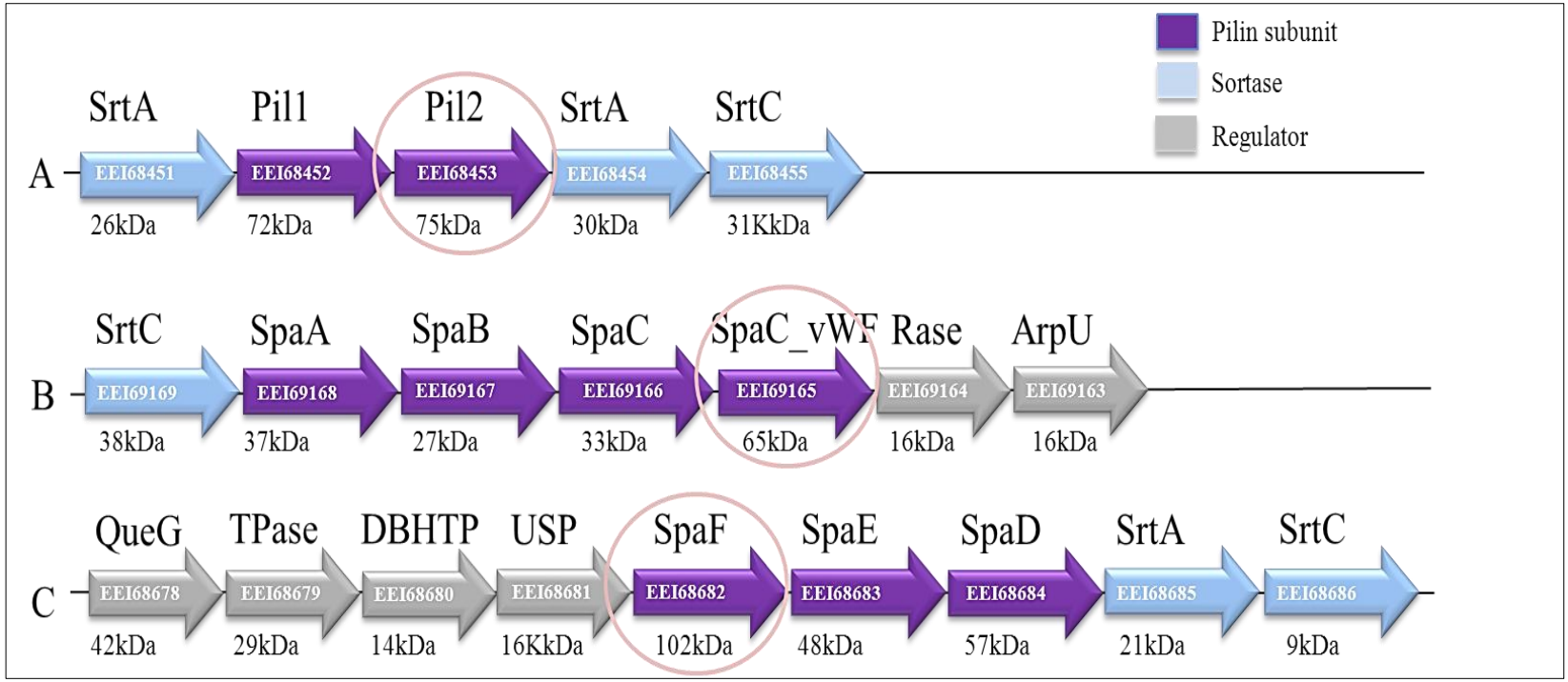

Figure 3.5: Putative pilus clusters in *Lactobacillus paracasei* subsp. *paracasei* ATTC 25302

Each letter represents a putative cluster of genes involved in pilus formation. Each arrow corresponds to a gene with the protein accession number given inside (NCBI database). A blue arrow corresponds to a gene encoding a sortase (srtA and srtC), a purple arrow to a pilin subunit (Pil1, Pil2, SpaA, SpaB, SpaC, SpaC_vWF, SpaD, SpaE, SpaF), a pink circle to a putative adhesin (Pil2, SpaC_vWF, SpaF) and a grey arrow to a putative transcriptional regulator (USP: universal stress protein; DBHTP: DNA bind helix turn helix protein; TPase: tyrosine phosphatase; QueG: epoxyqueueosine reductase-like protein; Rase: resolvase; ArpU: ArpU regulatory protein). Estimated molecular mass of the proteins are given in kDa below the arrows.

3.4 Preliminary experimental evaluation of putative adhesins

Following *in silico* analysis, *L. paracasei*, *B. infantis* and *F. prausnitzii* were cultured, RNA extracted from the bacteria and qPCR carried out on selected adhesin genes to check for *in vitro* expression (Chapter 2, section 2.4).

3.4.1 Quality control of cultured commensal strains

Commensal bacteria were cultured initially on agar plates then sub-cultured in relevant broth media. Primers based on the 16S ribosomal genes were designed specifically for each strain and a generic set corresponding to *E. coli* was included since it is the main bacteria used in the laboratory and therefore a major potential contaminant. Strains were tested for cross-contamination and the presence of *E. coli* by Polymerase Chain Reaction (PCR). Electrophoresis on agarose gel was used to visualise PCR products (Figure 3.6). These were purified and sequenced in both direction using gene specific primers. Sequences were aligned using CLUSTALW and the consensus sequence (full length) checked for homology to known proteins using the BlastN program (NCBI suite). All four sequences hit the 16S ribosomal gene from the relevant species confirming the cultured bacteria were the expected strains and devoid of contaminants.

3.4.2 Determination of optimal growth conditions for expression study

The three commensals were grown in various media conditions in an attempt to simulate the *in vivo* growth environment of the bacteria (Chapter 2, section 2.4). Strains were grown anaerobically with gentle rocking to mimic the peristaltic movements of the gut and relevant media were supplemented with mucin, with salt or both. Growth of the micro-organisms was measured by reading the optical density at 600nm for each media condition to determine the time frame corresponding to the log phase. The pH was also monitored during growth as the increasing acidity of the media correlates with the production of butyric, lactic and acetic acids, which are the main end products of the glucose fermentation in the three relevant species.

Neither PGM, nor salt seemed to affect the kinetics of the growth of any of the species listed (Figure 3.7).

F. prausnitzii (Figure 3.7 A and D) presented the fastest growth in broth reaching the stationary phase within 10h. This was also the case on agar plates where colonies were seen following an overnight incubation as opposed to 48h for the Gram positive strains.

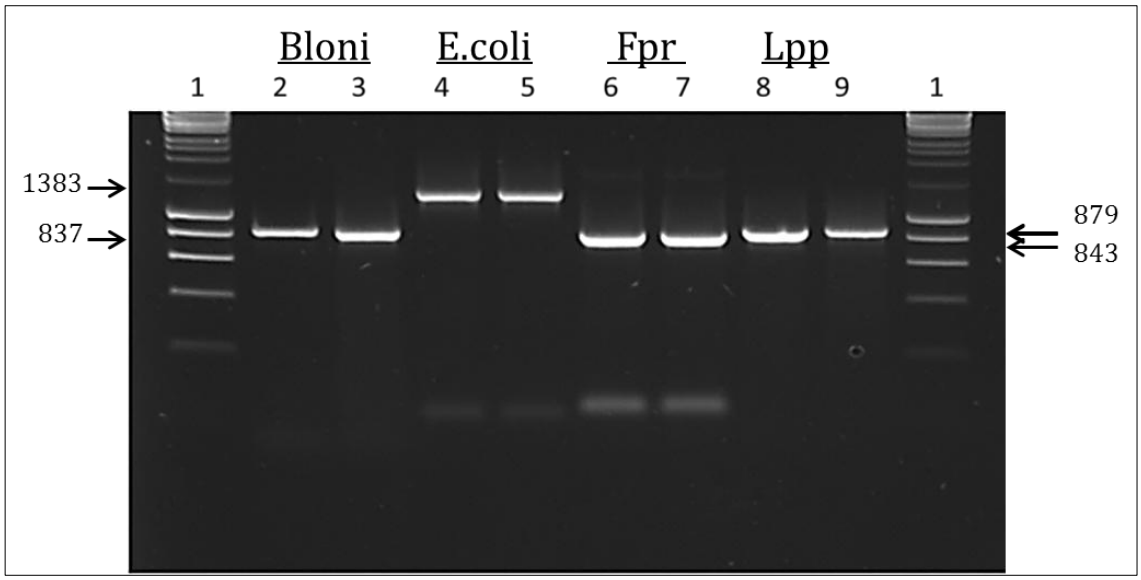

Figure 3.6: Quality control by PCR

Agarose gel (1.5 %) of PCR products stained by Sybr®Safe. Amplification was carried out with species-specific 16S rRNA primers.
Lane 1 : molecular ladder (Hyperladder II, Bioline); lane 2 : *Bifidobacterium longum* subsp. *infantis* colony grown on RCM agar plate; lane 3 : *B. infantis* colony grown in RCM broth; lane 4 : *Escherichia coli* colony from LB agar plate; lane 5 : *E. coli* colony grown in LB broth.; lane 6 : *Faecalibacterium prausnitzii* colony from WC agar plate; lane 7 : *F. prausnitzii* colony grown in WC broth.; lane 8 : *Lactobacillus paracasei* subsp. *paracasei* colony from MRS agar plate; lane 9 : *L. paracasei* subsp. *paracasei* colony grown in MRS broth. Amplicon sizes are expressed in DNA base pairs (left and right arrows).

The log phase for *F. prausnitzii* was estimated to be from 3h to 8h, when the acidity reached pH5 (Figure 3.7 A and D). Once the stationary phase was reached, the pH of the media gradually increased reaching pH6 (~16h), corresponding to a possible change of metabolism.

The log phase for *B. infantis* was longer (Figure 3.7 B and E), between 4h and 12h post-inoculation, with mid-log phase at approximately 7-8h. A more acidic environment was produced compared to *F. prausnitzii*, reaching pH levels below 4 at the end of the log phase.

L. paracasei (Figure 3.7 C and F) was the slowest growing species with log growth estimated to be between 5h and 16h post inoculation and mid-log phase approximated at 10h. Similarly to *B. infantis*, the media pH levels were below pH4 at the end of the log phase.

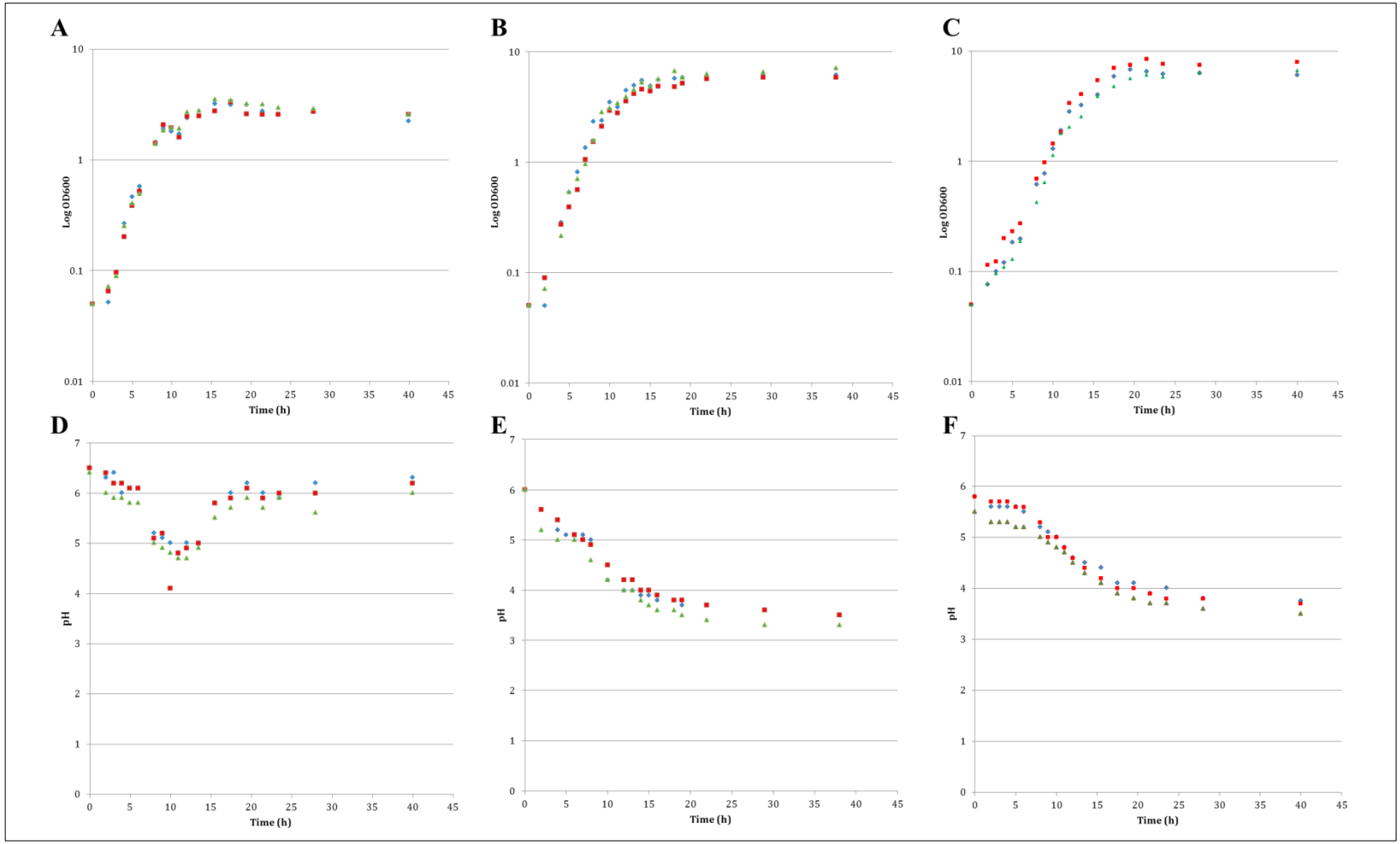

Figure 3.7: Kinetics of growth of selected commensals

Growth of bacteria and associated pH measurements in regular media (◆), media supplemented with 0.1% PGM (■) or with salt (0.3M NaCl, ▲); A, D: *Faecalibacterium prausnitzii*; B, E: *Bifidobacterium longum* subsp. *infantis*; C, F: *L. paracasei* subsp. *paracasei*.

3.4.3 Expression analysis by qPCR

Following growth curve analysis for each strain, total RNA was isolated from cells harvested as follows: cells in mid-log phase (i) in regular broth, (ii) in broth supplemented with mucin, (iii) in broth supplemented with salt, and (iv) in media supplemented with both mucin and salt. RNA was also isolated from cells in lag phase in regular broth (as the control). Samples were quantified and quality assessed as measured by the RNA integrity number (R.I.N.) (Table 3.9).

Table 3.9: RNA quality control

	Bloni			Fpr			Lpp		
Sample[1]	**Yield[2]**	**260/280[3]**	**RIN[4]**	**yield**	**260/280**	**RIN**	**yield**	**260/280**	**RIN**
1	**38.05**	**2.14**	**10**	**5.84**	**2.23**	**8.9**	**8.79**	**2.2**	**10**
2	**33.61**	**2.17**	**10**	**4.35**	**2.21**	**9.5**	**7.10**	**2.09**	**10**
3	**21.72**	**2.14**	**10**	**5.47**	**2.16**	**9.3**	**12.47**	**2.15**	**10**
4	46.74	2.16	10	26.99	2.16	9.7	11.40	2.15	10
5	64.25	2.17	10	28.70	2.16	9.7	11.23	2.12	10
6	63.20	2.16	10	29.91	2.14	8.8	14.29	2.14	10
7	54.19	2.15	10	18.052	2.18	7.4	22.30	2.14	10
8	35.38	2.15	9.9	18.49	2.19	10	16.27	2.14	10
9	48.01	2.16	10	10.33	2.2	9.2	19.21	2.14	9.9
10	45.47	2.15	10	26.02	2.16	9.4	12.88	2.11	10
11	25.45	2.13	10	11.31	2.21	10	13.94	2.09	10
12	49.33	2.16	10	28.60	2.16	9.6	8.44	2.19	7.4
13	48.91	2.16	10	28.29	2.15	9.5	4.96	2.04	10
14	77.82	2.17	10	18.59	2.19	9.5	10.16	2.12	10
15	77.28	2.16	10	37.12	2.17	10	10.81	2.13	10

[1]Samples 1-3 correspond to the controls (lag phase, in regular broth), 4-6 to regular media (log phase), 7-9 broth supplemented with PGM 0.1% (log phase), 10-12 media supplemented with 0.3M NaCl (log phase) and 13-15 broth supplemented with both PGM and salt (log phase).
[2]Yields of RNA preparations (in µg)
[3]260/280 ratio quality control
[4]RNA integrity number (RIN)
Abbreviations are as follows *Bifidobacterium longum* subsp. *infantis* (Bloni), *Faecalibacterium prausnitzii* (Fpr), *Lactobacillus paracasei* subsp. *paracasei* (Lpp)

Primers were designed for all candidate genes for use in qPCR to determine the relative expression of the candidate adhesin genes using the 16S rRNA gene as a housekeeping reference. To check the primers sets, all samples from each strain were pooled, amplified with the housekeeping gene and PCR products checked on agarose gels. No amplification was obtained for any candidate genes for *F. prausnitzii* except for the 16S rRNA. Only two PCR products (for 16S rRNA and the EF-Tu) were identified for *B. infantis*. On the other hand, all fifteen primer sets from *L. paracasei* resulted in amplification of products of the appropriate size. Real-time PCR amplification of individual *L. paracasei* samples was then performed using all primer pairs for the potential adhesins. Fold changes were calculated by the delta-delta Ct method (Livak and Schmittgen 2001) using the lag phase sample as the calibrator. Statistical analysis by ANOVA and Tukey *post hoc* test identified significant up-regulation of candidate genes from lag phase to log phase in regular MRS media ($P<0.01$), MRS supplemented with mucin (MRS+M, $P<0.05$) and MRS supplemented with both salt and mucin (MRS+MN, $P<0.001$) (Fig. 3.8, Table 3.10).

Multiple comparisons of candidates by Two-Way ANOVA identified five genes up-regulated significantly ($P<0.05$) in MRS+MN media supplemented with both salt and mucin (MRS+MN). One of these (LEFTU) corresponded to the translation factor EF-Tu, displaying moonlighting functions in bacteria and shown to be surface-exposed and involved in adhesion to intestinal cells, fibronectin or mucins in lactobacilli strains such as *L. johnsonii* and *L. plantarum* (Granato, Bergonzelli et al. 2004; Ramiah, van Reenen et al. 2007; Nishiyama, Ochiai et al. 2013). It was recently reported that EF-Tu binds the carbohydrate moieties of sulfomucins in *L. reuteri* (Nishiyama, Ochiai et al. 2013). The four other significantly up-regulated candidate genes all shared homology to known pilins and adhesins described in other strains including LGG (Figure 3.9). These were gene L2797 encoding a possible outer membrane protein and homologous to the minor pilin SpaF (protein sequence EEI68682), gene L5677 encoding a pilus subunit homologous to the major pilin SpaD (protein sequence EEI68684), gene L5075 encoding a fimbriae subunit showing homology to another major pili, SpaA (protein sequence EEI69168), and gene L48083 encoding a CnaB domain protein homologous to a collagen adhesion protein (protein sequence EEI68452).

Table 3.10: Expression of candidate adhesins from *L. paracasei* subsp. *paracasei*

	[1]Fold change							
[2]Gene ID	MRS	sd	MRS+M	sd	MRS+N	sd	MRS+MN	sd
LGROEL2	**1.71**	0.79	**1.25**	0.29	**0.53**	0.38	**1.85**	1.46
LFBPA	**1.25**	0.84	**0.91**	0.31	**1.49**	1.12	**1.16**	0.23
LCnBP2	**1.63**	0.48	**1.63**	0.3	**1.28**	0.7	**1.19**	0.23
LEFTU	**1.03**	0.31	**1.17**	0.03	**1.22**	0.05	**1.59**	0.33
L2797	**1.02**	0.34	**0.86**	0.07	**1.11**	0.49	**2.54**	1.02
L28086	**2.08**	2.23	**1.22**	0.34	**2.35**	1.94	**1.58**	1.53
L31701	**1.50**	0.55	**1.04**	0.25	**1.26**	0.42	**1.42**	0.26
L48083	**1.16**	0.21	**1.61**	0.64	**1.19**	0.52	**2.51**	1.9
L50135	**1.06**	0.13	**1.11**	0.16	**1.14**	0.39	**1.71**	0.71
L5677	**1.72**	1.4	**0.94**	0.2	**1.09**	0.88	**2.61**	1.71
L5075	**1.43**	0.26	**1.42**	0.27	**1.10**	0.2	**2.58**	0.89
L8692	**1.40**	0.73	**0.69**	0.41	**1.04**	1.29	**1.50**	1.29
L8786	**1.26**	0.5	**1.20**	0.37	**0.93**	0.18	**1.75**	0.65
Mean	**1.40**	0.67	**1.16**	0.28	**1.21**	0.66	**1.85**	0.94

[1]Fold change in expression of candidate genes normalised with 16S house-keeping gene and with lag phase used as the calibrator, between regular MRS media, MRS supplemented with PGM 0.1% (MRS+M), MRS supplemented with NaCl 0.3M (MRS+N) or MRS with both additives (MRS+MN). Significant data are coloured in red.

[2]Gene ID for *L. paracasei* subsp. *paracasei* as detailed in Chapter 2, section 2.4.5

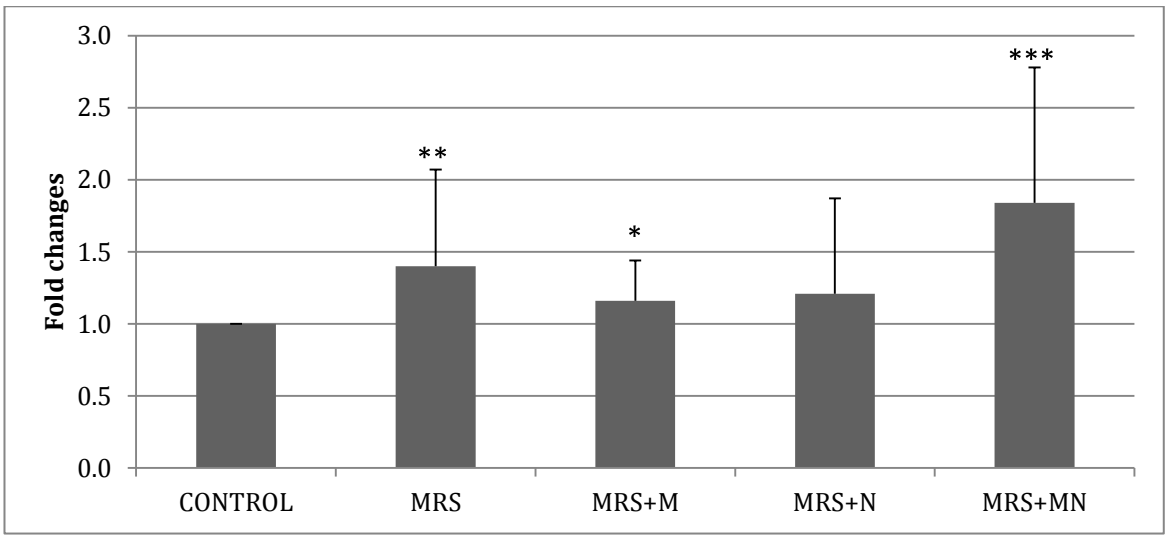

Figure 3.8: Average expression of all adhesin candidates from *L. paracasei* subsp. *paracasei* in different media

Fold changes of expression of all *L. paracasei* subsp *paracasei* adhesin candidates in the five different media conditions (control: lag phase; regular: MRS broth; PGM: MRS supplemented with mucin; NaCl: MRS supplemented with salt; P+N: MRS supplemented with mucin and salt; *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$)

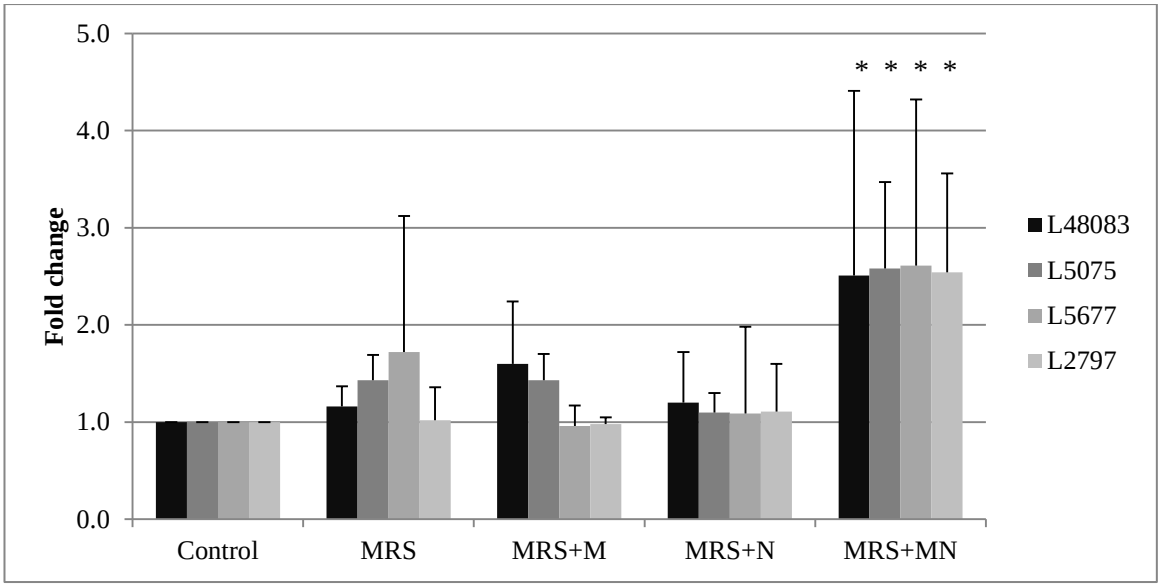

Figure 3.9: Expression of pilin candidates from *L. paracasei* subsp. *paracasei* in different media

Control: MRS, lag phase; MRS: MRS, log phase; MRS+M: MRS supplemented with mucin; MRS+N: MRS supplemented with salt; MRS+MN: MRS supplemented with mucin and salt. L48083, collagen adhesin homologue; L57075, SpaA homologue; L5677, SpaD homologue; L2797, SpaF homologue)

3.5 Discussion

Predictive computer modelling of bacterial adhesion, which has been used in vaccine development, is a potentially useful tool to select surface-exposed proteins (Barinov, Loux et al. 2009). The proteomes of representative commensal species were analysed using SurfG+, software for the prediction of protein subcellular localisation. The results from SurfG+ classified proteins into predicted surface exposed (SEP), secreted, cytoplasmic or membrane molecules. Details of the putative topology of the SEP were likewise given: whether predicted to present a loop, an exposed amino terminal (Nt) or an exposed carboxy terminal (Ct), and whether the SEP was a lipoprotein or contained a retention signal. In summary, the number of SEPs was similar between the three strains investigated, namely *F. prausnitzii*, *L. paracasei* and *B. longum*, and the majority contained less than 500 AA. Nevertheless, the topology differed, especially between Gram positive strains and *F. prausnitzii.* For instance, in the latter, 25% of the SEPs contained a lipoprotein signal but less than 3 % had retention motifs, whereas proportions were 10% regarding both in Gram positive. The abundance of lipoproteins in *F. prausnitzii* classified as SEPs could be expected due to the presence of the two membranes. This could also be an over-estimate resulting from the program's original design based on the properties of Gram positive bacteria. Moreover, in Gram negative *E. coli*, the described lipoproteins are not surface exposed but face the periplasm (Bos, Robert et al. 2007). However, lipoproteins from both Gram negative and Gram positive species have been shown to be involved in bacterial virulence (Kovacs-Simon, Titball et al. (2011). In some Gram negative bacteria such as *Borrelia sp.*, lipoproteins protrude through the outer layer of the outer membrane (OM) and are being investigated as vaccine candidates (Kovacs-Simon, Titball et al. 2011). In *B. burgdorferi* differential expression of lipoproteins has also been linked to the transition to life inside the mammalian host (Caimano, Iyer et al. 2007).

3.5.1 Putative adhesins of *F. prausnitzii*

Duncan et al. reported *F. prausnitzii* to stain as Gram-negative bacteria but to be phylogenetically related to the Gram-positive bacteria of the *Clostridium leptum* group (Firmicutes) following 16S rRNA sequencing (Duncan, Hold et al. 2002). Since the predictive SurfG+ program targeted Gram positive bacteria, proteins predicted to locate to other cell compartments were explored in this case. However, all selected candidates were ultimately predicted to be surface-exposed. These included a putative lipoprotein

containing an immunoglobulin Ig-like domain found in molecules such as intimin or invasin involved in adhesion of Gram negative cells (Tsai, Yen et al. 2010). Four proteins were homologous to pilin proteins from the typical chaperone-usher (CU) type 1 pilus found in Gram negative bacteria. Two sequences located by SurfG+ in the membrane and cytoplasm compartments actually corresponded to two chaperones responsible for type 1 pili assembly (FimC and CsuB). This suggests that at least two CU pili are present in this strain. In addition, two proteins corresponded to fimbriae subunits including a pilin presenting some homology to Cpa, the tip adhesin of the Gram positive strain *S. pyogenes.* Interestingly, three sortase sequences are present in the genome of *F. prausnitzii*, including one predicted by SurfG+ as SEP. There is therefore circumstantial evidence of the presence of a sortase-dependent-like pilus in *F. prausnitzii* and the covalent assembly of the Cpa-like pilin. In addition to pili subunits, afimbrial adhesins seem to occur in *F. prausnitzii.* A large surface autotransporter (1005 AA) with a hemagglutinin domain (pfam05860) presented homology to a filamentous lectin from *Burkholderia* sp. and an adhesin from *Yersinia pseudotuberculosis*. Filamentous hemagglutinins are secretory proteins that have been reported to adhere to epithelial cells in the respiratory tract (Kajava, Cheng et al. 2001; Hull, Donovan et al. 2002).

The potential of the *in silico* approach was then evaluated by analysis of expression of the selected adhesin genes under different experimental conditions. This experimental validation was unsuccessful for *F. prausnitzii*, despite high quality RNA and the visualisation of fimbriae using TEM. Amplification of the housekeeping gene 16S rRNA was detectable but no amplicons were obtained with any of the selected sets of primers related to the adhesin candidates. Inhibitory effects of PCR amplification can be due to numerous factors including cell constituents or high GC contents in the DNA (56.4% GC in *F. prausnitzii*) (Wilson 1997) and were not investigated further in this case. This could also suggest the design of the primers was inadequate. Even though it seems unlikely that all sequences were wrong, sequence inaccuracy can occur in the NCBI database and is not unusual in bacteria (Richardson and Watson 2012). *F. prausnitzii* is actually the only species in its genus and the sequenced genome of clinical isolate A2-165 is still in the process of verification.

3.5.2 Putative adhesins of *B. longum* subspecies

Several proteins were homologous to fimbrial subunits or contained motifs found in pilins

such as the RrgB_K2N_iso_D2 domain. This motif is involved in the cross-linking of pilin subunits and has been found in fimbrial proteins with lectin-like function and involved in adhesion to host structures (Spraggon, Koesema et al. 2010). SEPs with such domains include a conserved protein from *B. longum* subsp. *longum* homologous to a collagen adhesion protein from *Bacillus cereus.* Another protein from the same strain was homologous to a major pilin (FimA) from *B. dentium*. Interestingly, a genomic neighbour coding for a larger SEP (2327 AA) contained leucine-rich repeats similar to the ones found in *Listeria sp.* and presented some homology to *Listeria monocytogenes* internalin-F. The latter is an adhesin involved in the invasion of enteric cells and the subsequent dissemination of the pathogen from cell to cell and across the intestinal and placental barriers (Bonazzi, Lecuit et al. 2009). It is tempting to speculate on a similar system for bifidobacterial species residing in humans such as the subspecies from *B. longum.* Recent reports of bifidobacteria and lactobacilli strains found in human milk suggest the possible translocation of these strains through the intestinal barrier and towards the mammary glands. Some studies propose this might be carried out by macrophages and breast-feeding could represent a way to colonise the new-born with maternal probiotic strains (Martín, Langa et al. 2004; Sinkiewicz and Nordström 2005; Martín, Jiménez et al. 2009). Interestingly, lactobacilli and bifidobacteria have been detected in meconium, umbilical cord and amniotic fluid suggesting that invasion of the intestinal or the placental barriers might actually start *in utero* (Satokari, Gronroos et al. 2009; Funkhouser and Bordenstein 2013).

Four sortases were predicted by SurfG+ as SEP in *B. longum* subsp. *longum*. Following alignment using blastp, three of these showed homology to sortases belonging to class C (SrtC), which corresponds to pilin specific trans-peptidases (Hendrickx, Budzik et al. 2011). Altogether, this suggests the presence of three pili cluster for this strain similarly to *B. bifidum* PRL2010 (Turroni, Serafini et al. 2013). Interestingly, the strain on which the experimental validation was attempted, *B. longum* subsp. *infantis* is devoid of sortase-dependent pili but instead contains a larger proteome than *B. longum* subsp. *longum* including glycoenzymes and a high number of solute binding proteins. These have been suggested to play a role in adhesion of *B. infantis* to the infant host (Garrido, Kim et al. 2011). Another factor that might be important for this strain in the adhesion to infants is the high yield of exopolysaccharides produced (Chapter 5, section 5.2.2), which has been associated with increased adherence of several bifidobacteria including *B. longum* species

(Abbad Andaloussi, Talbaoui et al. 1995; Alp, Aslim et al. 2010). TEM photographs (not shown) corroborated results described elsewhere of the non-piliation of *B. infantis*, at least, *in vitro* (Foroni, Serafini et al. 2011). Nevertheless, four candidates in *B. infantis* presented putative adhesin functions. One conserved protein was found to be homologous to the collagen adhesion exoprotein of strain ATCC 55813 mentioned above. A second protein containing a CHAP domain was similar to a surface antigen from *B. longum* DJ010A. An allergen V5 was also homologous to a large adhesin from *Haemophilus somnus*. Furthermore, a large protein was homologous to a streptococcal lectin, namely the glucan-binding surface anchor molecule of *S. suis*, which is involved in virulence and aggregation (Ma, Lassiter et al. 1996; Okamoto-Shibayama, Sato et al. 2006). Recently, O'Connell Motherway et al. discovered a novel type of pili in *B. breve* UCC2003, a type IVb tight adherence pilus (Tad), which is conserved in all bifidobacteria (O'Connell Motherway, Zomer et al. 2011). Alignments were thus performed using *B. breve* to determine the possible *Tad* operon cluster in *B. infantis* and the corresponding pilin subunits of what seems to be the only pilus in this strain.

In *B. infantis*, validation by qPCR was only successful using primer sets corresponding to housekeeping genes and the elongation factor EF-Tu. None of the genes encoding the proposed adhesins showed any amplification. As revealed by transcriptomic analysis in other bifidobacterial species, fimbriae are not necessarily produced in laboratory as opposed to *in vivo* conditions (Foroni, Serafini et al. 2011; O'Connell Motherway, Zomer et al. 2011). However, supplementing the growth media with host-derived molecules such as lactose or mucin can result in an up-regulation of pilin genes in some strains, suggesting that pilin expression is an adaptation to a specific ecological niche. This is the case for instance in *B. adolescentis* ATCC 15703, which when grown in the presence of lactose, fructo-oligosaccharide (FOS) or GlcNAc displayed an increase in the transcription of pilin genes FimP, FimQ and the sortase gene (srtA) (Foroni, Serafini et al. 2011; Turroni, Serafini et al. 2013).

3.5.3 Putative adhesins of *L. paracasei*

Microbial surface components recognising adhesive matrix molecules (MSCRAMMs) are cell surface adhesins that interact with the host extracellular matrix (ECM). For instance, several bacterial fibronectin-binding proteins have been reported in the last decade such as FbpA in *Lactobacillus acidophilus* NCFM (Buck, Altermann et al. 2005). Fibronectin, a multidomain glycosylated protein, seems to play the role of a bridge between the bacterial adhesins and the integrins of the epithelial cells (Joh, Wann et al. 1999). In our investigation, a high number of candidates for *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302 contained the collagen-binding domain (CollagenBindB, cl4619) that mediates bacterial adherence to collagen. The glycosylation of this ECM glycoprotein is particular and affects differently the various sub-types of collagen (28 discovered so far) as well as their functions. Post-translational modification can add *O*-linked galactose or the *O*-linked galactosyl-glucose disaccharide to hydroxy-lysyl residues of the protein backbone. Additionally, some types of collagen contain *N*-linked glycans, however, these structures have not yet been resolved (Spiro 1967; Prockop 1995; Adamczyk, Struwe et al. 2013). One of the cell surface proteins with nine repeats of the CollagenBindB domain is a large protein of 2240 AA homologous to a collagen adhesion protein from *Lactobacillus rhamnosus* LMS2-1. *L. rhamnosus* is also a member of the *L. casei* group of lactobacilli and several surface exposed proteins from *L. paracasei* presented homology to the well-studied probiotic *L. rhamnosus* GG (LGG). Two proteins (EEI68918 and EEI68919 of respectively 1039 AA and 1173 AA) resemble an adhesin from LGG (CAR88818). Six leucine-rich repeats (LRR) are found on EEI68918. Interestingly, this specific type of conserved domains is found in BspA, an adhesin from *Bacteroides sp.* (Sharma, Sojar et al. 1998) and also in BspA–like cell surface antigens from *Trichomonas sp.* involved in the binding of those pathogens to fibronectin and oral mucosa (Noel, Diaz et al. 2010).

In LGG, two sortase-dependent pili clusters (*SpaCBA* and *SpaFED*) have been described (Kankainen, Paulin et al. 2009). In the SEP list predicted for *L. paracasei*, two sequences (EEI68684 and EEI68682) were homologous to two proteins encoded by the SpaFED pilus cluster of LGG. These correspond respectively to the major pilin SpaD (LGG02370) and the ancillary pilin SpaF (BAI42808). The latter was reported as the human mucus-binding adhesin of the *SpaFED* pilus (von Ossowski, Reunanen et al. 2010). In addition, a fimbriae subunit (ZP_03963300/EEI69168) possessing an LPXTG anchor, also presented high homology to SpaA, the major pilin encoded by the *SpaCBA* pilus cluster in LGG

(CAR86337). The presence of at least two sortase-dependent 'Spa' pili operons is further suggested by the presence of two pili specific sortases (srtC) in the genome of *L. paracasei*. These homologies have been consolidated into three putative gene clusters, encoding sortase-dependent pili in *L. paracasei* that have not been described elsewhere (Figure 3.5). Based on conserved domains, sequence alignments and genome-walking, the adhesin was predicted amidst the various proteins. All three clusters contain at least one pilin-specific sortase and two of them correspond to sortase-mediated pilus assembly 'Spa' fimbriae (termed *SpaCBA* and *SpaFED*). In Gram positive bacteria, the genes for pilin subunits and the specific sortase are usually found clustered but the housekeeping sortase that attaches the whole structure to the cell wall is often found elsewhere in the genome (Scott and Zahner 2006; Hendrickx, Budzik et al. 2011). Although some strains producing sortase-dependent pili only possess the one housekeeping transpeptidase, namely srtA (Mandlik, Swierczynski et al. 2008), remarkably, the three clusters in *L. paracasei* contain several sortases of both housekeeping and pili-specific classes (SrtA and SrtC respectively in Figure 3.5B and C). This could serve as an adaptation trait to synthesize the fimbriae efficiently and rapidly following the detection of external stimuli.

All selected genes for *L. paracasei* were upregulated at log phase following qPCR. Amongst these, four encoded pilin subunits were significantly up-regulated in the culturing media that mimicked the gut environment (i.e. in the presence of mucin and salt). Interestingly, the up-regulated pilin genes were part of the proposed pili operons, namely coding for the major pilin EEI68453 (gene L48083), SpaA (gene L5075), SpaD (gene L5677) and SpaF (gene L2797), and, corresponding respectively to clusters A, B and C (Figure 3.10). Furthermore, appendages were observed on TEM at mid-log phase in similar growing conditions in that strain (Chapter 4, section 4.4). Another gene, coding for the moonlighting protein EF-Tu, was significantly up-regulated in similar culturing conditions. This gene was selected based on homology to a characterised adhesin from *L. johnsonii* NCC533 (AAS08831). EF-Tu is a cytoplasmic protein that was detected on the cell surface in *L. johnsonii* NCC533 and shown to be implicated in adhesion to human intestinal cells and mucins (Granato, Bergonzelli et al. 2004). EF-Tu, as well as two other homologs of mucin binding adhesins Mub and MapA, were up-regulated in *L. plantarum* 423 when grown in the presence of 0.01% mucin (Ramiah, van Reenen et al. 2007).

Altogether, this suggests that *L. paracasei* is able to adapt to specific niches in the GIT by switching on various pili and adhesins. The prototypical *SpaCBA* and *SpaFED* pili were

first described for *Corynebacterium diphtheriae* and *Streptococcus pyogenes*. In these pathogens, the Spa pilus consists of a major pilin (SpaA or SpaD) that forms the shaft and two minor pilins, one at the base tethered to the cell wall (SpaB, SpaE) and a larger one at the tip of the polymer (SpaC and SpaF), which is responsible for the adhesion (Ton-That, Marraffini et al. 2004; Rogers, Das et al. 2011). Pili in Gram positive bacteria are composed of two to three different subunits consisting of one major unit and one or two minor pilins. So far no structural backbone pilin (major subunit) has been shown to display adhesin characteristics (Kline, Fälker et al. 2009; Proft and Baker 2009; Hendrickx, Budzik et al. 2011). Surprisingly, the putative *SpaCBA* operon in *L. paracasei* (Figure 3.5C) consists of four pilin genes, two of which presenting homology to SpaC. The largest pilin 'SpaC_vWF' (with sequence number EEI69165, 65kDa) contains adhesin-like conserved domains such as the von Willebrand Factor motif found in the mannose-specific adhesin of *L. plantarum*, collagenBindB domain and a metal ion-dependent adhesion site (MIDAS), suggesting a lectin-like function (Gross, van der Meulen et al. 2008; Marchler-Bauer, Lu et al. 2011). The proposed adhesins in the two Spa pili clusters (*SpaCBA* and *SpaFED*), respectively SpaC and SpaF, have been described in LGG recently (von Ossowski, Reunanen et al. 2010; Rintahaka, Yu et al. 2014) and revealed binding to human mucus in addition to another minor pilin from the *SpaCBA* operon, spaB. A trimeric model of the fimbriae encoded by the SpaCBA cluster was suggested based on TEM using immunogold-staining of antibodies raised against the various pilins from LGG (von Ossowski, Reunanen et al. 2010; Rogers, Das et al. 2011). This supported a very different model compared to the *C. diphtheriae* prototype, with a backbone formed by the major subunit SpaA as well as the presence of the minor pilin SpaB in between SpaA subunits and the decoration of the shaft by the adhesin SpaC (Reunanen, von Ossowski et al. 2012).

Similarly to the *Tad* pilus genes from *B. breve* UCC2003 that are only transcribed in *in vivo* conditions, *SpaFED* pilin genes were not transcribed *in vitro* in the case of LGG, possibly due to the lack of an environmental stimuli (O'Connell Motherway, Zomer et al. 2011; Reunanen, von Ossowski et al. 2012). Stress-related conditions such as temperature, iron deficiency, high salt concentration or acidity in the culture media have been found to increase the adherence of gut pathogens such as *Clostridium difficile* (Waligora, Barc et al. 1999). The adherence of some lactobacilli to Caco2 cells is also increased at low pH (Greene and Klaenhammer 1994). Interestingly, the proposed *SpaFED* operon in *L. paracasei* harbours a universal stress protein (usp, protein accession number EEI68681,

Figure 3.5 C). Bacteria sense and adapt to their surrounding environment via quorum sensing. This communication system is intra- or interspecific and functions by means of small molecules known as autoinducers that regulate gene expression simultaneously across the bacterial population as its density increases (Miller and Bassler 2001; Kaper and Sperandio 2005). Autoinducers are involved in the virulence of pathogenic species such as *Vibrio cholerae* and *Pseudomonas aeruginosa* by favouring biofilm formation and cell aggregation (Zhu, Miller et al. 2002; Shiner, Rumbaugh et al. 2005). Functions of universal stress proteins (usp) are largely unknown. However, usp in *E. coli* have been linked to oxidative stress defense and also cell adhesion (uspG) (Nachin, Nannmark et al. 2005). Moreover, the expression of usp genes is upregulated by quorum sensing in some species of *Burkholderia* (Kim, Goo et al. 2012). The usp gene in *L. paracasei* might thus be part of the regulatory system of the Spa operon for that strain and may function as a 'switch' for the piliation. Moreover, three proteins involved in gene regulation are found directly upstream of the usp gene. These correspond to a transcriptional regulator (DNA-binding helix turn helix protein, EEI68680), a tyrosine phosphatase (EEI68679) and an iron sulfur-binding protein (EEI68878), which is identical to a regulatory protein (QueG, EPC74700) (Agris 2004; Sutak, Lesuisse et al. 2008).

Phase variation of piliation involving transcriptional elements has been described in *E. coli* for type 1 and P fimbriae formation (Eisenstein 1981; Nou, Skinner et al. 1993). The molecular 'switch' of the *fim* operon, responsible for the production of type 1 fimbriae, consists of an invertible DNA element containing the promoter. Recombinases (such as FimB and FimE) act to invert that element and enable the transcription or the silencing of the operon upon the detection of a temperature change (Wolf and Arkin 2002; Holden, Totsika et al. 2006). Remarkably, in the SpaCBA cluster from *L. paracasei*, a recombinase is found directly upstream of the sortase (resolvase, EEI69164) as well as a regulatory protein (ArpU, EEI68878) containing a possible sigma factor (TIGR01637). These factors have been associated with virulence in a number of species by acting to redirect transcription initiation (Kazmierczak, Wiedmann et al. 2005).

Chapter 4: Glycan ligands of selected commensal adhesins

Key points:

- Neo-glycoconjugate microarrays were used to identify ligands of commensal adhesins found in species from the human GIT, originating from the genera *Faecalibacterium, Bifidobacterium* and *Lactobacillus*
- This approach was validated using two *E. coli* strains producing two different types of fimbriae with two tip adhesins binding two different glycans (Type 1 fimbriae-FimH binding mannose and P fimbriae-PapG binding galabiose)
- Glycan ligands were identified using whole cells as well as isolated pili for the selected strains (*F. prausnitzii*, *B. bifidum* PRL2010 and *L. paracasei* subsp. *paracasei*)
- Glyco-profiling was also carried out using microarrays containing natural GIT mucins to identify specific niches colonised by the selected strains

4.1 Introduction

Metagenomic analysis has revealed that the microbiota from a human being consists of about 1000 bacterial strains, 80% of these being host specific and 20% being conserved amongst healthy individuals. This phylogenetic core consists of 67 dominant species from the genera of *Faecalibacterium, Ruminococcus, Eubacterium, Dorea, Bacteroides, Alistipes* and *Bifidobacterium* (Ley, 2008; Tap, 2009; Lepage, Gut 2013). Bifidobacteria are amongst the first colonisers of the human infant and constitute a dominant group of microorganisms in the early microbiota, coinciding with the development of the new born immune system (Turroni, Ribbera et al. 2008). The gut contains the highest numbers of immune cells of all organs and commensals have been shown to be involved in the development of the adaptive immune system (Bailey, 2001; Montilla, review 2004). There is evidence that commensal bacteria interact with their eukaryotic host by expressing adhesive molecules on their cell surface that bind to specific host cell receptors or soluble macromolecules (see Chapter 1). These enable the colonisation of specific niches in the gastro-intestinal tract (GIT) and produce effector molecules involved in symbiosis (Marco, Pavan et al. 2006; O'Flaherty and Klaenhammer 2010).

Studies to date on bacterial adhesion and immunogenic molecules in the gut have mainly focussed on human pathogens, as potential candidates for vaccines, rather than with commensals (Neutra and Kozlowski 2006; Carvalho, Ching Ching et al. 2012). Although a highly specific system has been reported between the intestinal cell receptors (mostly glycoproteins acting as pattern recognition receptors, PRR) and adhesins of pathogenic bacteria (mostly glycan-binding proteins), evidence is only beginning to emerge that similar mechanisms exist for attachments of commensal strains and their host cells (Blackburn, Husband et al. 2009; Kline, Fälker et al. 2009; Korea, Ghigo et al. 2011). The primary mechanism of bacterial adhesion typically involves multimeric proteinaceous appendages, such as fimbriae or pili, which reach out from the cell wall bearing specific subunits termed adhesins (Chapter 1, section 1.4). It has been demonstrated recently that the genome of the model commensal strain *Lactobacillus rhamnosus* GG contains Sortase-dependent pilus assembly gene clusters, namely SpaCBA and SpaFED gene clusters, encoding two sortase-dependent pili with SpaC and SpaF likely coding for the adhesins. SpaCBA pilin subunits have been visualised on LGG cells using immunogold electron microscopy and western blotting, however, no SpaFED pili has been expressed under *in*

vitro conditions (Reunanen, von Ossowski et al. 2012). Functional characterisation using a SpaC knockout strain showed the involvement of the SpaCBA cluster in adhesion to human intestinal cells, biofilm formation and immunomodulation (Kankainen, Paulin et al. 2009; Lebeer, Vanderleyden et al. 2010; von Ossowski, Reunanen et al. 2010).

In bifidobacteria (Figure 4.1), experimental evidence of the existence of sortase-dependent and Type IV pili was reported only very recently (Foroni, Serafini et al. 2011; O'Connell Motherway, Zomer et al. 2011). The genome of *Bifidobacterium bifidum* PRL2010 harbours three Spa gene clusters, transcribed as polycistronic mRNA. Transcriptomic studies revealed commonly expressed extracellular proteins, under *in vitro* and *in vivo* conditions, including two of the three sortase-dependent pili (*pil2* and *pil*3), which are differentially expressed according to the carbohydrate source present in the culturing medium (Foroni, Serafini et al. 2011; Serafini, Strati et al. 2013). The predicted pili comprise in both case, a major backbone pilin (FimA or FimP) and a minor subunit possibly corresponding to the adhesin (FimB and FimQ). Because bifidobacteria are fastidious in their culturing requirements, particularly strain *B. bifidum* PRL2010, the two Spa pili gene clusters from *B. bifidum* PRL2010 (*pil*2 and *pil*3) were cloned into a bacterial expression system using *Lactococcus lactis* (NICE®) (Mierau and Kleerebezem 2005). Expression of the coding sequences of these *B. bifidum* PRL2010 appendages in non-piliated *Lactococcus lactis* increased adherence to human enterocytes through binding to ECM proteins and bacterial aggregation. Furthermore, piliated *L. lactis* cells induced a higher TNFα response in mice, suggesting that bifidobacterial sortase-dependent pili present immunomodulation functions (Turroni, Serafini et al. 2013).

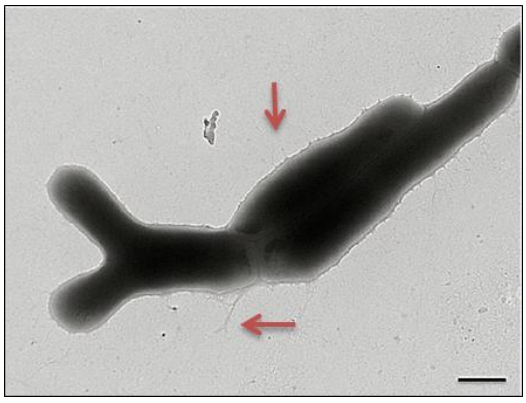

Figure 4.1: Piliated *Bifidobacterium breve* UCC2003

TEM photograph of PTA-stained *B. breve* UCC2003 cells grown at mid-exponential phase in broth media and expressing sortase-dependent pili (red arrows). Scale bar represents 500 nm.

The first crystal structure of an adhesin was described for the type 1 fimbria tip adhesin (FimH) from uropathogenic *Escherichia coli* strains (UPEC) (Clarke, Haigh et al. 2003; Niemann, Schubert et al. 2004). *E. coli*, a rod-shaped Gram negative bacterium is the most studied prokaryote. Its versatile physiology enables the bacterium to multiply in various ecological niches by adapting its enzymatic machinery (Estrada-Garcia, Hodges et al. 2013). *E. coli* is well suited to the intestinal environment due to its ability to utilise low concentration of nutrients. It is able to utilise glucose as the sole carbon source and can grow in both anaerobic and aerobic environments. Therefore, *E. coli* has been used widely as a faecal pollution marker ('coliform count') as well as a tool for biotechnology companies or scientific research (especially in genetic engineering and in the production of recombinant proteins). Some *E. coli* strains are pathogenic causing food-poisoning, gastroenteritis, meningitis and urinary diseases (Kaper, Nataro et al. 2004). More than 700 antigenic types (termed serotypes) have been described so far, based on the molecules they display on their outer membrane: polysaccharidic ('O'), flagellar ('H') or capsular ('K') antigens (Johnson, Orskov et al. 1994). The virulence of the diseases caused by *E. coli* serotypes is determined by the expression of several adhesins, invasins and toxins (see Chapter 1, Table 1.9). The persistence of *E. coli* in the GIT likely involves adhesins binding to mannose and galabiose-containing receptors from the intestinal mucosa (Wold, Thorssen et al. 1988). Urinary tract infections are essentially the result of colonisation of the bladder by enteric *E. coli* from faeces. *E. coli* serotypes have the ability *in vivo* to express multiple fimbriae, the best studied being pyelonephritis-associated pilus (PAP or P fimbria) and type I fimbria. Of interest to our study, both of their adhesins are amongst the few that have been characterised regarding target glycans. Both are Chaperone-Usher (CU) pili structures, for which assembly has been extensively reviewed (Waksman and Hultgren 2009; Busch and Waksman 2012; Wurpel, Beatson et al. 2013). Their biosynthesis typically involves a molecular chaperone, located in the periplasm, which mediates proper folding of fimbrial subunits, and an outer membrane 'usher' that assembles these subunits to form the pili. The expression of the two fimbriae is governed by two different complex transcriptional regulation systems (Hull, Donovan et al. 2002; Melican, Sandoval et al. 2011).

The P fimbria is encoded by a methylation-dependent operon (*pap* operon) composed of a cluster of eleven genes. The shaft consists of a non-covalent polymerised backbone pilin (PapA) and a large tip fibrillum composed of an assembly of a major pilin (PapE) and

minor pilin subunits (PapF and PapK). The tip adhesin (PapG) binds in a lectin-manner to the P-blood group antigen that is present on human erythrocytes as well as in various tissues. The minimal structure for adhesion is the disaccharide Gal-α-(1→4)-Gal-β (galabiose), which is present on glycosphingolipids from epithelial cells (Svenson, Källenius et al. 1983; Jacobson 1986). The expression of the *pap* operon is regulated by a complex epigenetic phase-variation system resulting in differential piliation of *E. coli* cells. Briefly, regulatory proteins such as the Leucine responsive regulatory protein (Lrp), bind to DNA sequences to block the methylation at particular sites, leading to the control of pilin gene expression (Nou, Skinner et al. 1993; Casadesús and D'Ari 2002). Under this phase variation control mechanism, individual cells are either piliated (phase 'ON') or non piliated (phase 'OFF'). Moreover, the 'ON' to 'OFF' switch rate is 100-fold higher than the 'OFF' to 'ON' rate causing a mixed, but mostly unpiliated, population (Blyn, Braaten et al. 1990).

The type 1 fimbriae of *E. coli* mediate attachment to a different glycan ligand - mannose-containing glycoproteins - through the FimH adhesin. These fimbriae are conserved amongst the Enterobacteriacea and are expressed in the majority of both commensal and pathogenic strains of *E. coli* (Dhakal, Kulesus et al. 2008; Blackburn, Husband et al. 2009). type 1 fimbriae are encoded by the *fim* cluster, a polycistronic operon comprising nine genes expressing structural and regulatory proteins. The fimbria consists of a 'piling up' of the major pilin (FimA), forming the shaft, and the fibrillum made from the minor pilins (FimF, FimG) and the tip-adhesin (FimH), which contains the mannose-binding lectin. The periplasmic chaperone (FimC) and the outer-membrane usher (FimD) are responsible for the assembly of the structure on the outer membrane (Knight, Berglund et al. 2000). The transcription of the complex is controlled by an invertible DNA element (the *fim* 'switch', FimS) that contains the promoter responsible for the expression of the fimbrial structural genes. When the promoter is positioned in the right orientation ('ON'), it allows the transcription of the *fim* genes (FimACDFGH), the subsequent expression of the *fim* subunits and thus the formation of the pilus. On the contrary, the expression gets silenced when the promoter is in the 'OFF' direction. The system is regulated through recombinases (FimB and FimE), which catalyse site-specific recombination, and the ratio of the two enzymes (FimB/FimE) regulates the phase variation switching. A high ratio of FimB/FimE favours the 'ON' phase whereas a lower ratio switches off expression. Furthermore, this

regulation is under the influence of 20 auxiliary proteins that respond to environmental triggers such as pH and osmolarity (Schwan 2011).

Recent studies have given some insights into the mechanisms involved in the colonisation of the kidney by *E. coli* (Sebaihia, Wren et al. 2006; Wiles, Kulesus et al. 2008). Uropathogenic *E. coli* (UPEC) strains seemingly adhere first to the tubular epithelium by expressing P fimbriae (early attachment), followed by the production of type 1 pili. The latter provide a supplementary mechanism of adherence. In addition, the colonisation can be enhanced by the subsequent aggregation of the cells and the formation of a biofilm via FimH binding to specific mannosyl-glycoproteins found in urine (Wright, Seed et al. 2007).

Unlike PapG and FimH, the vast majority of adhesins have not been characterised in terms of their specific ligand or glycotope. Recent advances in glycomics technology such as glycoconjugate synthesis and glycan microarrays have provided new tools to study lectin-glycan interactions. In this study, we have used high-throughput glycan microarrays technology to profile the bacterial surface lectome from selected commensal strains. We demonstrate as a proof of concept the potential of neoglycoconjugate arrays to detect adhesin ligands using two reference strains of *E. coli* expressing the two different types of pili mentioned above. We then use the same approach to identify potential adhesion targets of three prominent human commensal strains, *Faecalibacterium prausnitzii* (abbreviated as *F. prausnitzii), Bifidobacterium bifidum* PRL2010 (abbreviated as *B. bifidum*) and *Lactobacillus paracasei* subsp. *paracasei* (abbreviated as *L. paracasei*).

4.2 Confirmation of glycan ligands of type 1 and P fimbriae from *E. coli* by NGC microarrays analysis

Reference *E. coli* strains, namely the type 1 fimbriae-producing strain (ATCC 35218) and the P fimbriae-producing strain (ATCC 25922), which are known to bind respectively mannose via FimH and galabiose via PapG, were used to confirm the ability of our NGC microarrays to detect adhesin targets.

4.2.1 Growth and pili visualisation

The two *E. coli* strains showed similar growth patterns in LB medium although the type 1 fimbriae-producing strain had a slightly shorter doubling time compared to the P fimbriae-producing strain. Mid-exponential phase (OD_{600}~0.5 to 0.6) was determined as 2.5h and stationary phase was reached by both strains at ~5h (OD_{600}~1). The presence of pili was checked by transmission electronic microscopy (TEM) in the mid-exponential phase (OD_{600} 0.6; T=2.5h) with negative staining and revealed appendages on both strains (Figure 4.2). This mid-exponential phase time point was subsequently used for all the glyco-array experiments.

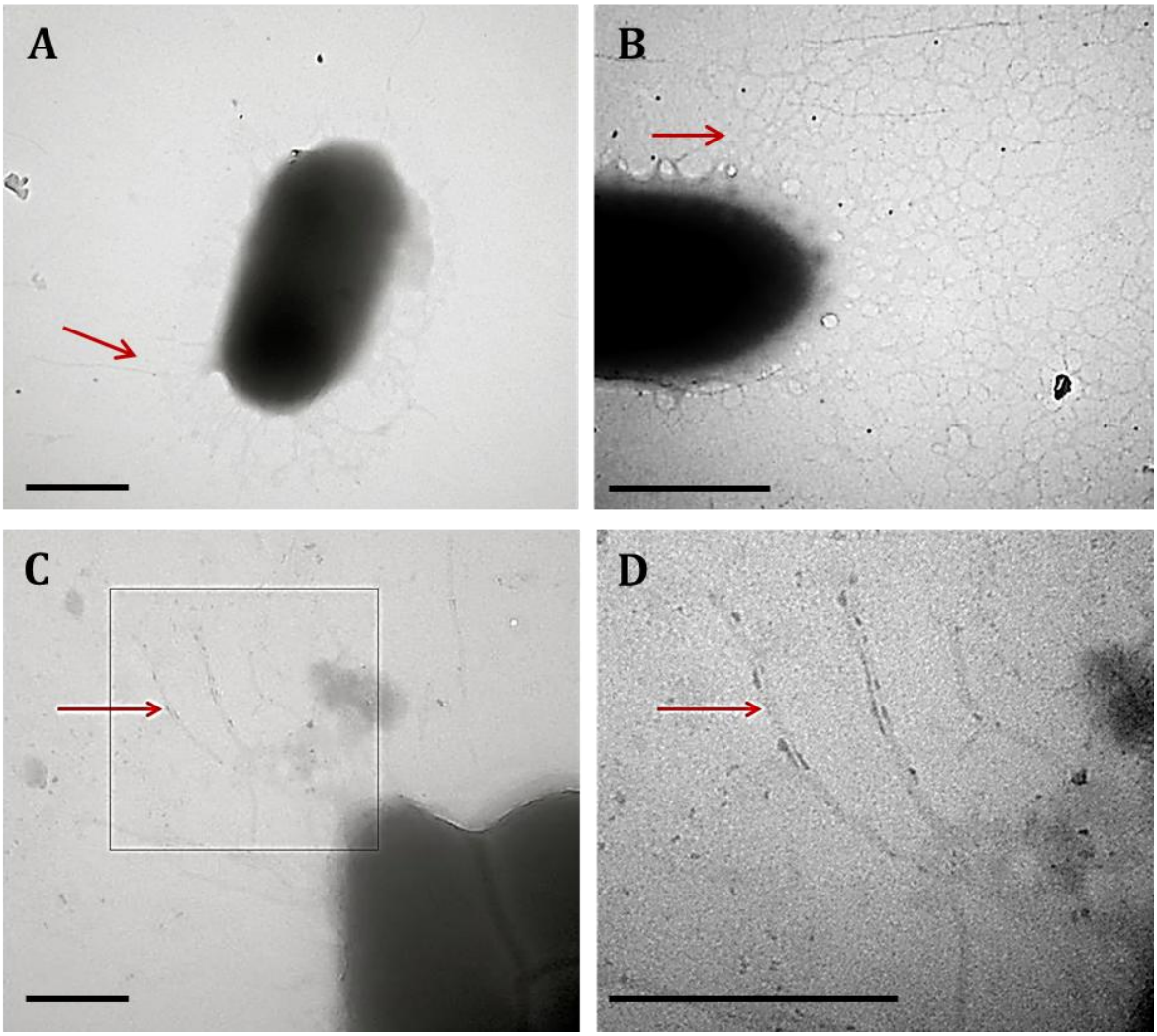

Figure 4.2: TEM photographs of *E. coli* piliated strains

Cells were harvested at mid-exponential phase (OD_{600}~0.5) and negatively stained with PTA. Scale bar represents 500nm. A red arrow points at a fimbria or pilus
A: Appendages surrounding type 1 fimbriae-producing strain (ATCC35218). B: Magnification of A. C: Appendages surrounding P fimbriae -producing strain (ATCC25922). D: Magnification of C.

4.2.2 Neoglycoconjugate microarray analysis of *E. coli* strains

Cells, harvested at mid-exponential phase, were stained using a DNA-binding fluorescent dye (SYTO®82) and incubated on the glycan arrays. Data are presented as histograms corresponding to combined normalised RFU values from two different microarray platforms and totalling 76 non-redundant glycans representing motifs found in the human gut. Overall, both strains bound strongly to monosaccharides glucose (Glucβ) and galactose (Galβ) that are linked to the microarray by the isothiocyanate linker (Figure 4.3).

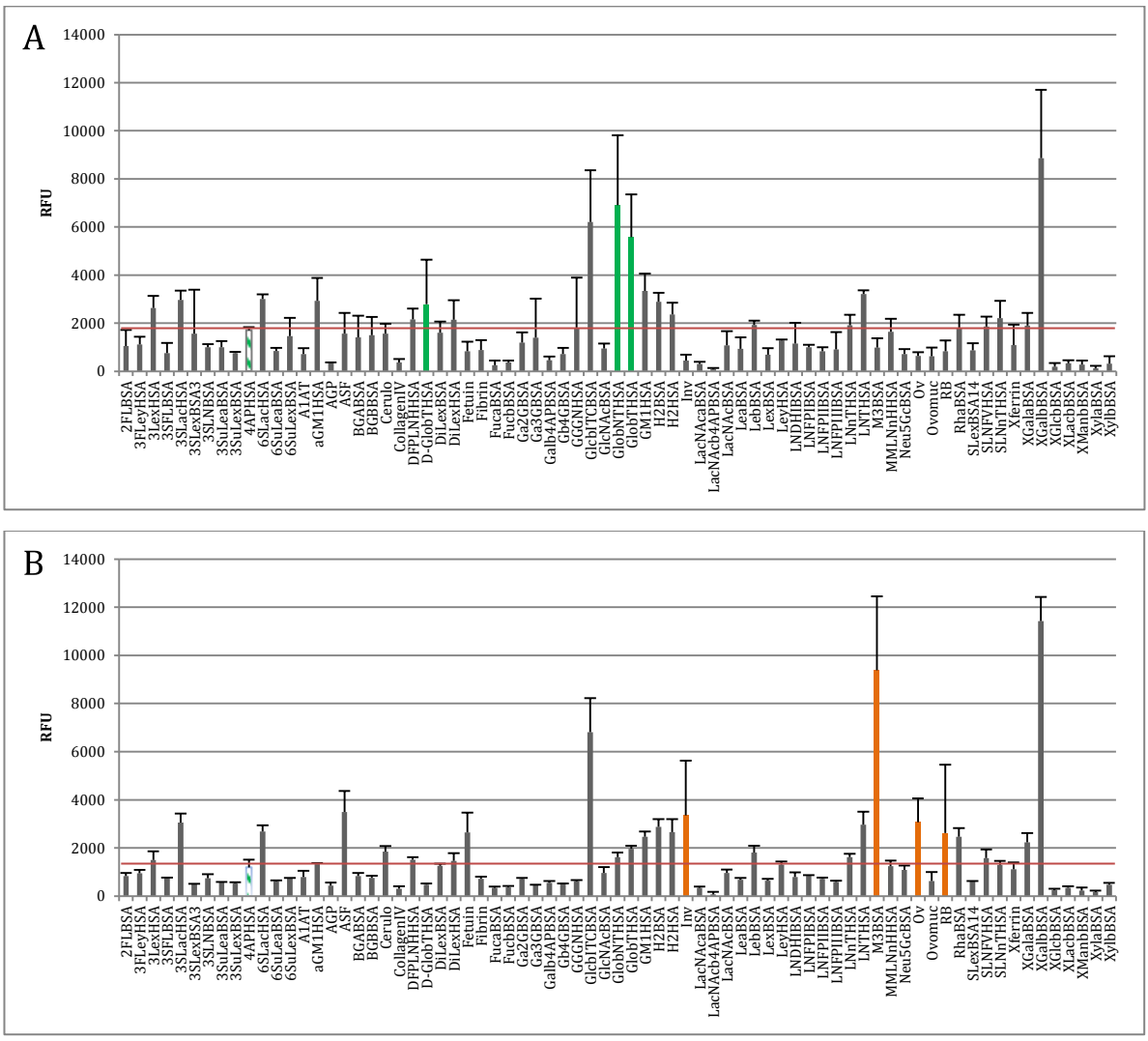

Figure 4.3: Binding of PapG- and FimH- producing *E.coli* strains on NGC arrays

Normalised signals for NGC microarrays using labelled whole cells. All values are averages from three microarrays and error bars represent the standard deviation of the mean. Horizontal red bar denotes the binding threshold signal corresponding to the linker 4APHSA value. A: *E. coli* strain ATCC 25922 producing P fimbriae. Glycans containing galabiose are represented in green. B: *E. coli* strain ATCC 35218 producing type 1 fimbriae. High mannose-containing glycans are highlighted in orange. Abbreviations of NGCs are found in Table 2.14 and Table 2.15.

Differences in binding were noted between the two strains. The P fimbriae- producing strain (ATCC 25922) interacted strongly to the glycosphingolipids globotriose of structure Gal-α-(1→4)-Gal-β-(1→4)-Glc-β1- (represented by two NGCs, D-GlobTHSA and GlobTHSA) and globo-*N*-tetraose (GlobNTHSA, GalNAc-β-(1→3)-Gal-α-(1→4)-Gal-β-(1→4)-Glc-β1-) (highlighted in green in Figure 4.3 A). In contrast, strong signals were given by the type 1 fimbriae-producing strain (ATCC 35218) with the mannosyl trisaccharide (M3BSA, Man-α-(1→3)-[Man-α-(1→6)]-Man). Three glycoproteins, ribonuclease B (RB), invertase (Inv) and ovalbumin (Ov), all containing high mannose *N*-glycans also bound this strain but with lower intensities (highlighted in orange in Figure 4.3 B).

Statistical comparison of differential binding between the two strains using the Student *t*-test ($P< 0.05$, FDR 5%) identified four glycans. The P fimbriae-producing strain (ATCC 25922) showed significantly higher binding to glycosphingolipids globotriose (GlobTHSA, $P<0.001$) and globotetraose (GlobNTHSA, $P<0.004$) with respectively 2.7- and 4-fold increase compared to the other strain. These two glycolipids contain galabiose (Gal-α-(1→4)-Gal-β). Inhibition experiments using 100 mM of galactose resulted in a potent decrease of the signal (84%) in both cases. The trisaccharide M3BSA ($P< 0.002$) and ovalbumin ($P< 0.004$) presented an increase in binding for the Type 1 fimbriae-producing strain (ATCC 35218) with respectively 9.8- and 4.8-fold increase compared to strain ATCC 25922. Competition experiments with D-mannose (100mM) resulted in inhibition of binding to both mannosyl trisaccharide M3BSA (93%) and ovalbumin (74%).

4.2.3 Discussion

The NGC microarrays developed in-house were produced as two platforms each of 52 features totalling 76 non-redundant glycans, and were previously assessed using a panel of lectins (Kilcoyne, Gerlach et al. 2012). The experimental procedure was optimised with bacterial cells from the incubation step to the analysis of data. As a proof of concept, two strains of *E. coli*, which produce *in vitro* two types of pili with known carbohydrate ligands were studied. These were type 1- and P fimbriae -producing strains binding respectively mannosyl structures and galabiose-containing glycans.

To avoid interfering with cell surface proteins, bacteria were labelled using a nucleic acid fluorescent dye. Binding of the cells to the glycan arrays were then optimised by titration to determine optimal signal to noise ratio. Low amounts of cells did not give acceptable signals whereas high amounts led to aggregation and high background. Biases originating

from the various steps from printing to scanning were removed by normalisation. These involved corrective factors, based on the total signal means intra-slide (between subarrays) and inter-slide (between the two platforms) to combine the 76 unique features. Comparison between strains was then achieved by scale-normalisation using microarray statistical software.

The piliation of commonly used *E. coli* strains is regulated by complex mechanisms (phase variation) such as promoter inversion (e.g. type 1 fimbriae) or DNA methylation (e.g. P fimbriae) (Clegg, Wilson et al. 2011). A survey of the literature to assess the optimal conditions for the *in vitro* expression of bacterial pili revealed diverse methodologies between researchers. While some groups use bacteria in stationary phase in agar (Wang, Hsu et al. 2012) or broth media (Korhonen, Nurmiaho et al. 1980), others utilise cells in various stages of the exponential phase (Müller, Åberg et al. 2009) or after several sub-culturing steps in broth media (Turner, Chen et al. 2005). In our case, to mimic the environment of the gastro-intestinal tract, strains were grown at 37˚C in static and anaerobic conditions. Cells were recovered from glycerol stocks on agar plates and sub-cultured twice in broth. Bacteria were harvested at mid-exponential phase and both *E. coli* strains showed fimbriae-like appendages surrounding the cells when visualised by TEM. This time-point and methodology was subsequently chosen for the glycan microarray experiments. Binding threshold between lectin and glycans on microarray is generally considered as five times the fluorescent background value (Kilcoyne, Gerlach et al. 2012). The 4-aminophenyl linker (4AP), which is being used for 75% of the NGC, was always above the latter and could be considered as hydrophobic interactions. Therefore, in this study, the binding threshold was defined by the signal of the feature 4APHSA.

Overall, the expected glycans were bound by both type 1 and P fimbriae-producing strains and specificity was supported by inhibition experiments with free monosaccharides. Statistical comparison between strains using a highly stringent test (false discovery rate of 5 %) resulted in the predicted outcomes. The mannosyl trisaccharide (M3BSA, Man-α-(1→3)[Man-α-(1→6)]-Man) and ovalbumin, a glycoprotein containing high mannose *N*-glycans including Man-α-(1→3)[Man-α-(1→6)]-Man, were significantly bound by the type 1 fimbriae-producing strain (ATCC 35218). In contrast, P fimbriae-producing cells (ATCC 25922) showed significant higher affinity to the galabiose (Gal-α-(1→4)-Gal-β)-

containing glycans featured on the array (globotriose and globo-N-tetraose). These results gave us confidence to carry out the experiments with other strains.

4.3 Identification of adhesion targets of *Faecalibacterium prausnitzii*

4.3.1 Analysis of *F. prausnitzii* NGC arrays

Hair-like structures were clearly visible by TEM on *F. prausnitzii* cells grown either in broth media (Figure 4.4 A) or on cells grown on agar plates (Figure. 4.4 A and B). Appendages at the apex of aggregated cells when magnified resembled rakes with a stem of ~200nm in length (Figure 4.4 C).

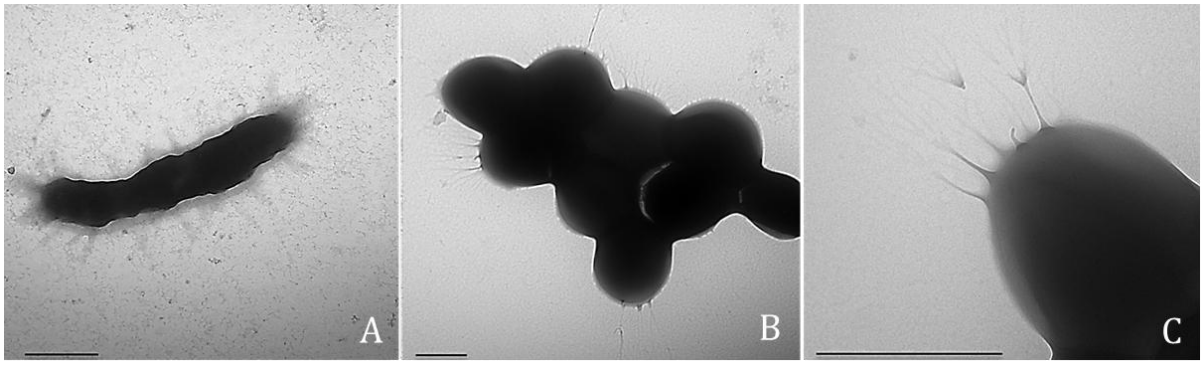

Figure 4.4: TEM photographs of *F. prausnitzii*

A: PTA-stained *F. prausnitzii* cells grown in broth media. B: PTA-stained *F. prausnitzii* cells grown on agar plates. C: Magnified appendages at the apex of a cell grown on agar plates. Scale bar corresponds to 500 nm.

Cells grown on agar plates were labelled with SYTO®82, incubated on microarrays and the NGC profile analysed (Figure 4.5A). A considerable number of features were bound by the Gram-negative strain above the threshold, defined as the binding to the 4-aminophenyl linker (4APHSA). However, differences in intensity between these NGC were marked and 20 high binders were identified out of a total of 76 unique glycan motifs (Figure 4.5 B).

Strongest signals were obtained with some of the glycoproteins printed on the microarray, such as α-crystallin (a-C, containing *O*-GlcNAc with phosphorylation), fibrinogen (Fibrin, containing *N*-glycans) and ovomucoid (Ovomuc, containing 4 *N*-glycans). Asialofetuin (ASF, containing 2 *N*- and 3 *O*- glycans) was also a strong binder as were monosaccharides β-glucose (GlcbITC), α-mannose (XManaBSA) and rhamnose (RhaBSA). A group of

human milk oligosaccharides were also strongly bound. These were 3'- and 6'-sialyllactose (3SLacHSA, Neu5Ac-α-(2-3)-Gal-β(1→4)-Glc; 6SLacHSA, Neu5Ac-α-(2→6)-Gal-β-(1→4)-Glc), lacto-*N*-tetraose (LNTHSA, Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc), lacto-*N*-neotetraose (LNnTHSA Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc), sialyllacto-*N*-fucopentaose (SLNVHSA, Fuc-α-(1→2)-Gal-β-(1→3)-[NeuAc-α-(2→6)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc). Binding was also high to the blood group H-type 2 trisaccharide (present as two NGCs: H2BSA and H2HSA of structure Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc) and oligosaccharides presenting Lewis antigens such as the Le^b hexasaccharide (LebBSA, Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc) and the dimeric Le^a/ Le^x octasaccharide difucosyl-*para*-lacto-*N*-hexaose (DFPLNHHSA, Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc). Four glycans found on glycolipids were also included amongst the high binders. These were the globosides globotetraose (GlobNTHSA, GalNAc-β-(1→3)-Gal-α-(1→4)-Gal-β-(1→4)-Glc) and globotriose (GlobHSA, Gal-α-(1→4)-Gal-β-(1→4)-Glc), and the gangliosides GM1 (GM1HSA, Gal-β-(1→3)-GalNAc-β-(1→4)-[Neu5Ac-α-(2→3)]-Gal-β-(1→4)-Glc) and the asialo-GM1 (aGM1HSA, Gal-β-(1→3)-GalNAc-β-(1→4)-Gal-β-(1→4)-Glc).

Pili were then isolated by a physical method from stationary phase cultures of *F. prausnitzii*, labelled and incubated on NGC arrays (as described in Chapter 2, section 2.8.1). Fewer features above threshold were bound by the isolated pili compared to whole cells (Figure 4.5C). Highest binders included monosaccharide glycoconjugates and some glycoproteins. These were α-mannose (XManaBSA), β-galactose (XGalbBSA), and β-glucose (GlcbITC). NGCs α-mannose and β-glucose were also two of the strongest binders from the whole cell profile. Asialofetuin (ASF, containing 2 *N*- and 3 *O*- linked glycans) as well as α-crystallin (a-C, *O*-GlcNAc with phosphorylation) were common strong binders between cells and isolated pili. In addition, RNaseB (RB, high mannose *N*-linked glycan $Man_{5\text{-}9}GlcNAc$), ceruloplasmin (Cerulo, 7 *N*-linked glycans) and ovalbumin (Ov, high mannose and hybrid *N*-linked glycans) interacted strongly compared to all other NGCs.

To sum up, binders were identified for *F. prausnitzii* cells. Most NGCs were bound by the strain, which could be explained by the presence of various adhesins on its cell surface. In contrast, isolated pili bound to very few NGCs but strongly to α-mannose glycoconjugate

and RNase B, suggesting a tropism towards mannosyl structures by the corresponding adhesin(s).

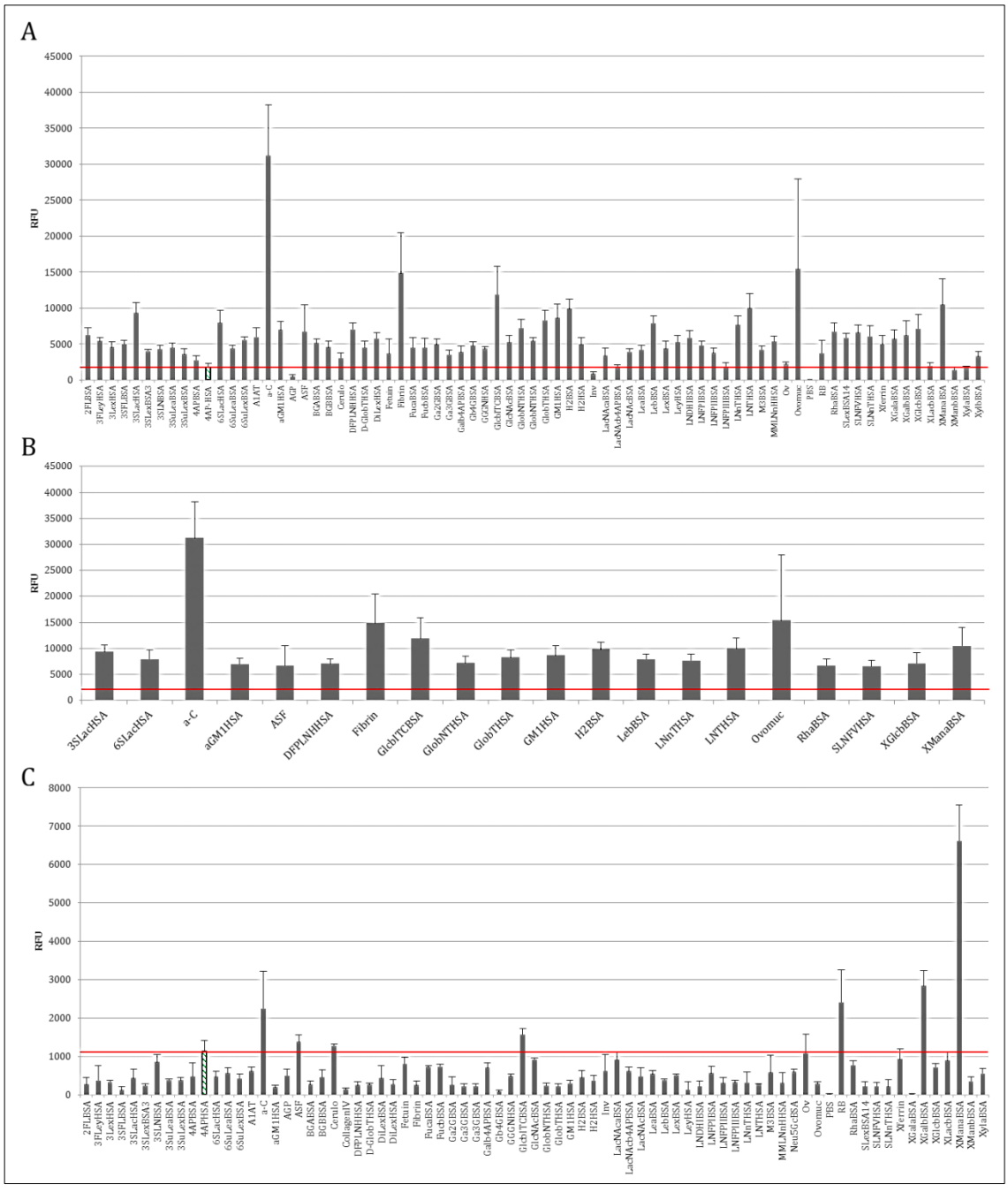

Figure 4.5: Binding of *F. prausnitzii* to NGC microarrays

A: Normalised signals for NGC microarrays using labelled whole cells
B: Top 20 glycans bound using labelled *F. prausnitzii* whole cells
C: Normalised signals for NGC microarrays using labelled *F. prausnitzii* pili extract
All signals are averages from three microarrays. Error bars represent the standard deviation
Horizontal red bar denotes the binding threshold signal corresponding to the linker 4APHSA value.
Abbreviations of NGCs are found in Table 2.14 and Table 2.15.

4.3.2 Analysis of *F. prausnitzii* on mucin arrays

Whole cells and isolated pili from *F. prausnitzii* were incubated on a natural GIT mucin microarray platform to further investigate binding specificities. Labelled whole cells gave weak signals and a high background but binding to eight mucins and six glycoproteins was above the threshold (defined as five times the background signal value, Figure 4.6A). The eight mucins were all intestinal mucins from various species and included, in order of signal intensity, rat duodenum, mouse large intestine, chicken proximal small intestine, equine small intestine, human MUC2, porcine cecum, porcine jejunum and ovine spiral colon. The glycoproteins bound were RNAse B, asialofetuin, pigeon egg white (PEW), ovomucoid, transferrin and fetuin.

The labelled pili gave lower background on the microarray (Figure 4.6B). Signals were stronger with mucins and bound all 14 features mentioned above. The highest binders amongst mucins were colonic mucins, namely the human MUC2 secreted by the LS174T cell line, followed by the rat duodenum mucin. Strong binding was also found with asialofetuin, fetuin and pigeon egg white all containing complex *N*-linked glycans.

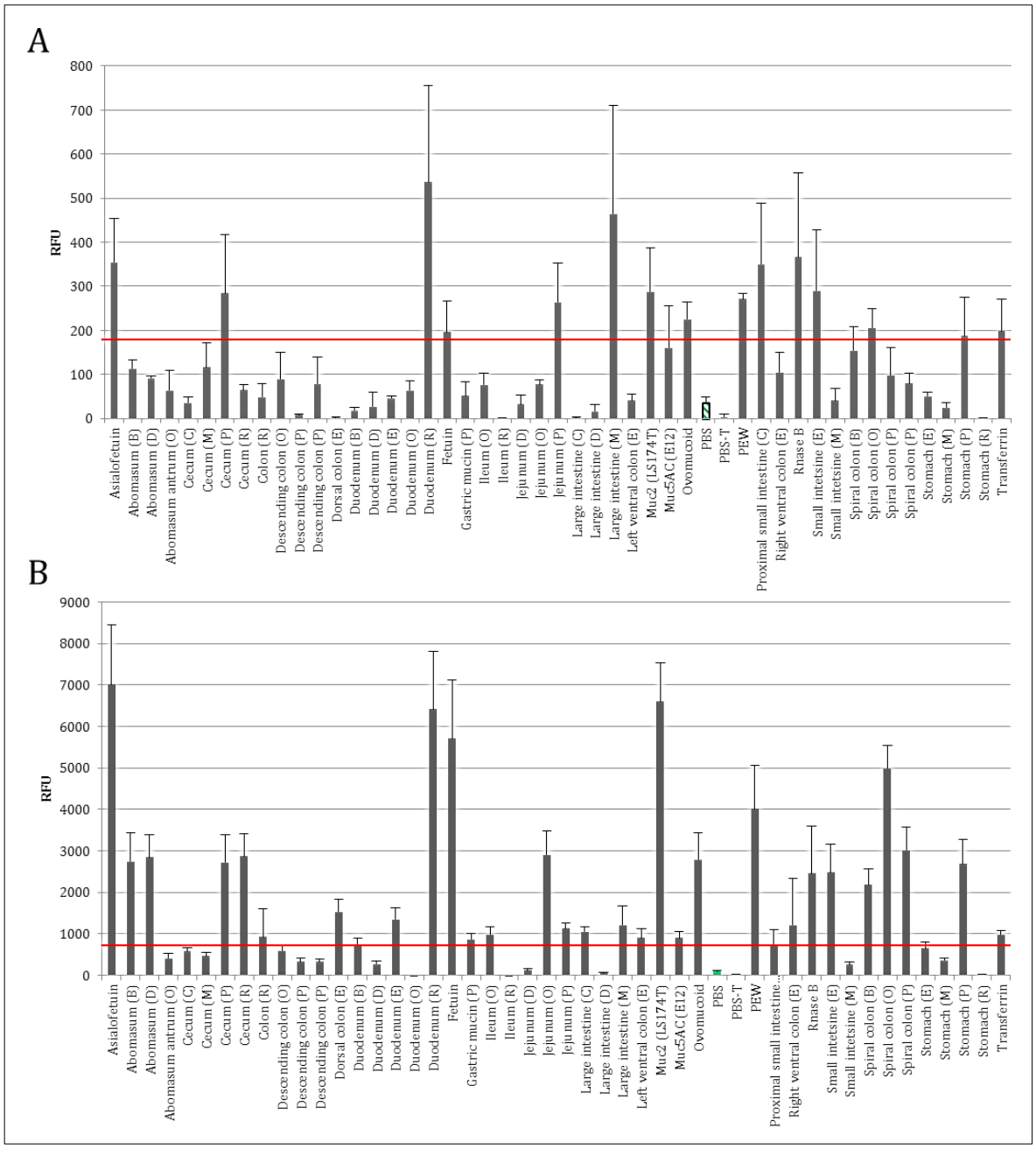

Figure 4.6: Binding of *F. prausnitzii* to mucin arrays

A: Normalised signals for NGC microarrays using labelled whole cells

B: Normalised signals for NGC microarrays using labelled *F. prausnitzii* pili extract

All signals are averages from three microarrays. Error bars represent the standard deviation of the mean. Horizontal red bar denotes the binding threshold signal corresponding to five times the background value. Abbreviations are as follows: bovine (B), cervine (D), equine (E), galline (C), murine-mouse (M), porcine (P), murine-rat (R), human (E12, LS174T).

4.3.3 Western blotting of pili isolates with asialofetuin

Due to the strong binding of both *F. prausnitzii* whole cells and pili to glycoproteins such as asialofetuin, pili extracts were separated by PAGE and western blotting carried out with peroxidase-labelled glycoproteins.

Extracts displayed several high molecular weight bands on native PAGE with bands ~300 kDa and 350 kDa but especially strong at 150 kDa (Figure 4.7 A). These could correspond to fimbriae of various lengths or large adhesins. When processed in denaturing conditions a high number of protein bands were visible (Figure 4.7 B). They might correspond to pilin subunits under various degree of polymerisation. A small degree of contamination by cytoplasmic proteins cannot be excluded from dead cells or following the various centrifugation steps of the pili isolation procedure (detailed in Chapter 2). A high background was noticed compared to coomassie when loading similar amounts and staining the SDS-PAGE with silver, suggesting the presence of polysaccharides and possible glycosylation of fimbriae (Figure 4.7C).

Since asialofetuin was one of the strongest binder following the microarray analysis, the SDS-PAGE was blotted and reacted with in house-prepared HRP-labelled asialofetuin (Figure 4.7D). The strongest band was visible ~114 kDa, which could correspond to the putative adhesins found *in silico* (ZP_05616166 of 114 kDa or ZP_05616615 of 110 kDa, see Chapter 3, Table 3.2). A doublet (~ 60-65 kDa) was also strong, as well as a low molecular weight band (~ 27 kDa) that could correlate with pilin ZP_05616171 (Chapter 3, Table 3.2). Inhibition of the signals with a sugar cocktail (consisting of 100mM Gal, Man and GlcNAc) was not evident after 1h incubation and limited to a band ~32kDa (not shown). Binding to asialofetuin might thus result from electrostatic interactions. A fluorescent probe such as 8-anilino-1-naphthalenesulfonic acid (ANS) that binds with high affinity to hydrophobic surfaces of proteins could help to confirm this result (Celej, Dassie et al. 2005).

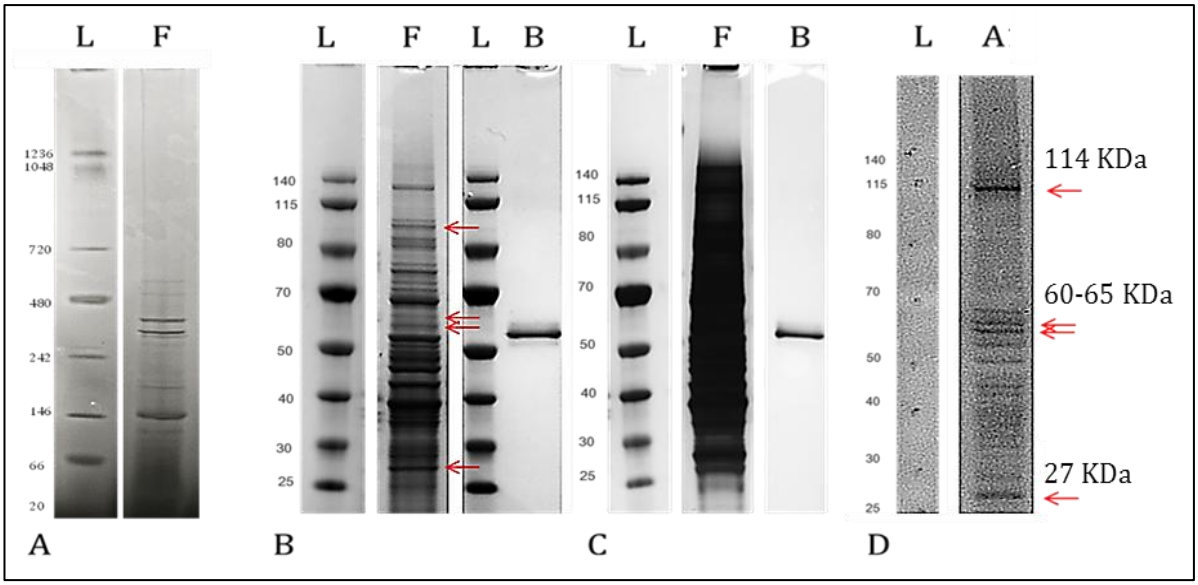

Figure 4.7: Binding of *F. prausnitzii* pili to asialofetuin

A: Pili extract on 3-8% native PAGE (coomassie blue staining). B: Pili extract on 4-12% SDS-PAGE (coomassie blue staining). C: Pili extract on 4-12% SDS-PAGE (silver staining). D: Western blots using HRP-labelled asialofetuin. Abbreviations are as follows: L, molecular weight ladder (kDa); F: pili extract (50 µg); B, BSA (2 µg, ~64KDa); A, HRP-labelled asialofetuin Red arrows represent putative pilin subunits

4.3.4 Discussion

Faecalibacterium prausnitzii, a prominent commensal bacterium in adults, is considered a biomarker of intestinal homeostasis (Dore and Corthier 2010; Fujimoto, Imaeda et al. 2013; Galecka, Szachta et al. 2013; Miquel, Martin et al. 2013). To our knowledge, the adhesion mechanism of *F. prausnitzii* to host mucus has not yet been addressed. Difficulties in obtaining mono-associated animals with *F. prausnitzii* to study its adhesion have been reported and could explain this (Wrzosek, Miquel et al. 2013). According to Wrzosek et al., *F. prausnitzii* colonisation of the GIT happens late after birth, when the oxido-reduction potential in the gut has been reduced following the establishment of other commensal strains such as *Bacteroides thetaiotaomicron*. The binding of *F. prausnitzii* was thus investigated on glycan arrays and several binders were identified. These might represent ligands to several surface exposed proteins including two large adhesins discovered *in silico* (Table 3.2, Chapter 3). The strongest binder, which was not bound by any other commensals studied, was fibrinogen. Several microbial surface components recognising adhesive matrix molecules (MSCRAMM) that binds to fibrinogen have been identified (Patti, Allen et al. 1994; Schubert, Zakikhany et al. 2002; Niemann, Schubert et al. 2004). The glycosylation of fibrinogen consists of *N*-linked glycans, mainly biantennary

digalactosylated monosialylated structures and biantennary digalactosylated disialylated structures (Adamczyk, Struwe et al. 2013).

This interaction is of interest and might in part enlighten the anti-inflammatory role of *F. prausnitzii* described by Sokol et al. (Sokol, Pigneur et al. 2008; Sokol, Seksik et al. 2009). Fibrinogen, an acute phase glycoprotein, is involved in the coagulation process and polymerises into fibrin to form the blood clot. It is also considered a key regulator of inflammation and disease. Several studies have shown that extravascular fibrinogen that deposits in tissue exerts a pro-inflammatory function through the binding of various molecules such as the integrin Mac-1 (Adams, Schachtrup et al. 2007; Steinbrecher, Horowitz et al. 2010; Raghu and Flick 2011; Davalos and Akassoglou 2012; Adamczyk, Struwe et al. 2013). The latter is expressed on leukocytes from the innate immune system (granulocytes and monocytes), and Raghu et al. demonstrated that disrupting binding with fibrinogen reduced local inflammation (Raghu and Flick 2011).

It has also been shown that plasma fibrinogen induces the synthesis of particular lectins (selectins) on platelets through interactions with integrins (Schwertz, Zimmerman et al. 2009; Yang, Lang et al. 2009). Selectins are constitutively expressed on leukocytes (L-selectins) and are up-regulated in patients suffering from Crohn's disease (Pooley, Ghosh et al. 1995). In addition to their physiological role, selectins display signalling functions following their binding to various glycoproteins, which results in T-cells recruitment and homing (Ley and Kansas 2004). One of the described ligands, common to all selectins, is the P-selectin glycoprotein ligand 1 (PSGL-1). PSGL-1 is a dimeric mucin of yet poorly described glycosylation, consisting of branched *O*-glycans including sialyl Le^x (Carlow, Gossens et al. 2009; Lo, Antonopoulos et al. 2013), which is also involved in homeostasis and immunity (Zarbock, Ley et al. 2011). Commensal bacteria such as *Bacteroides intermedius* and *B. gingivalis* bind to fibrin, a network of fibrinogen and possess potent fibrinogen-degrading activity (Lantz, Allen et al. 1990; Lantz, Allen et al. 1991; Houston, Blakely et al. 2010). It is thus tempting to speculate that the binding to fibrinogen by *F. prausnitzii* cells and its degradation would attenuate the inflammatory response. Interactions of the surface proteins from *F. prausnitzii* cells with fibrinogen could prevent or 'compete' with the binding of leukocytes (through Mac-1), resulting in downregulation of the expression of selectins. Therefore, a reduced interaction of the selectins with their ligands such as PSGL-1 would attenuate the inflammatory cascade and reduce the local inflammation.

F. prausnitzii was found by TEM to express hair-like appendages. When incubated on the same glycan array, both isolated pili from *F. prausnitzii* and the whole cells bound strongly to α-crystallin (containing *O*-linked GlcNAc) and α-linked mannose. Pili extracts bound preferentially glycoproteins displaying hybrid *N*-linked glycans or high mannosyl structures including ceruloplasmin (Harazono, Kawasaki et al. 2006). Pili and whole cells shared common strongest binders on mucin arrays such as asialofetuin and the rat duodenum mucin. Natural immobilised mucins on an array format are considered a better representation of the glycan environment and as sensed by the bacteria *in vivo* (Kilcoyne, Gerlach et al. 2012). Mucins would also be the first substrate the bacteria would adhere to when colonising a particular niche in the adult gut. An important metabolic effort by the bacterial cell is necessary for the synthesis of multimeric appendages such as pili (Müller, Åberg et al. 2009) and as for the Type-1 fimbriae, regulation by phase variation could be a selective advantage for a particular strain with phases 'ON' for initial colonisation to mucins. An 'OFF' phase could then follow with the close adherence to its host using other surface exposed molecules such as MSCRAMM.

4.4 Identification of adhesion targets of *L. paracasei* subsp. *paracasei*

4.4.1 NGC profiling of whole cells of three lactobacilli

Following expression experiments under various conditions, up-regulation of four pilin genes were noticed by qPCR when culturing *L. paracasei* in broth media supplemented with mucin (0.1 %) and salt (0.3 M) (see Chapter 3). *L. paracasei* were thus grown in similar conditions prior to incubation on NGC microarrays. The binding profile of *L. paracasei* was compared to two other *Lactobacillus* commensals, namely *L. johnsonii* and *L. rhamnosus* GG (LGG), both known to produce pili *in vitro*. All species were grown to mid-exponential phase and the presence of appendages confirmed by TEM (red arrows on Figure 4.8). These had three different morphologies, with thick appendages essentially observed at the poles of LGG cells, thin and curly on *L. paracasei* and ciliated surrounding *L. johnsonii.*

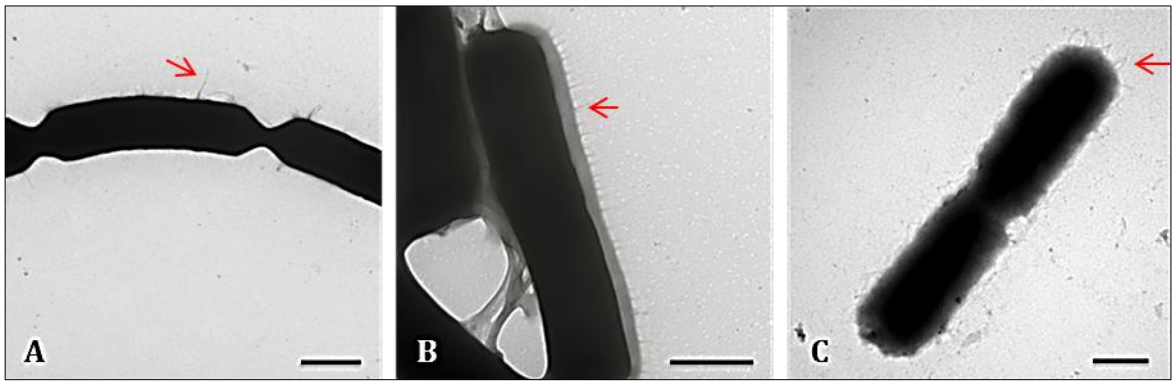

Figure 4.8: Lactobacilli pili

TEM photographs from lactobacilli cells grown in broth media to mid-exponential phase
Scale (black bar) corresponds to 500 nm. Red arrows point to possible pili
A: LGG. B: *L. johnsonii*. C: *L. paracasei*

When whole cells of the three lactobacilli were incubated on NGC microarrays, glycan-binding profiles for each species were unique (Figure 4.9). LGG was considerably different from the two other strains, showing strong binding to asialofetuin, fetuin, lactose-NGC and βGal-NGC.

Signals from the three species above the cut off value (i.e. 4APHSA, red arrow, Figure 4.9) were analysed statistically.

ANOVA followed by Tukey post hoc test revealed differential binding to 30 NGC between the two casei group members LGG and *L. paracasei* (Figure 4.10 A); 10 NGC between the acidophilus group member *L. johnsonii* and *L. paracasei* (Figure 4.10 B); 16 NGC between LGG and *L. johnsonii* (Figure 4.10 C). These results suggest that even though LGG and *L. paracasei* are phylogenetically related, both being members of the casei group, they adhere to different glycan ligandsglycan ligands and therefore might occupy specific niches in the gastro-intestinal tract.

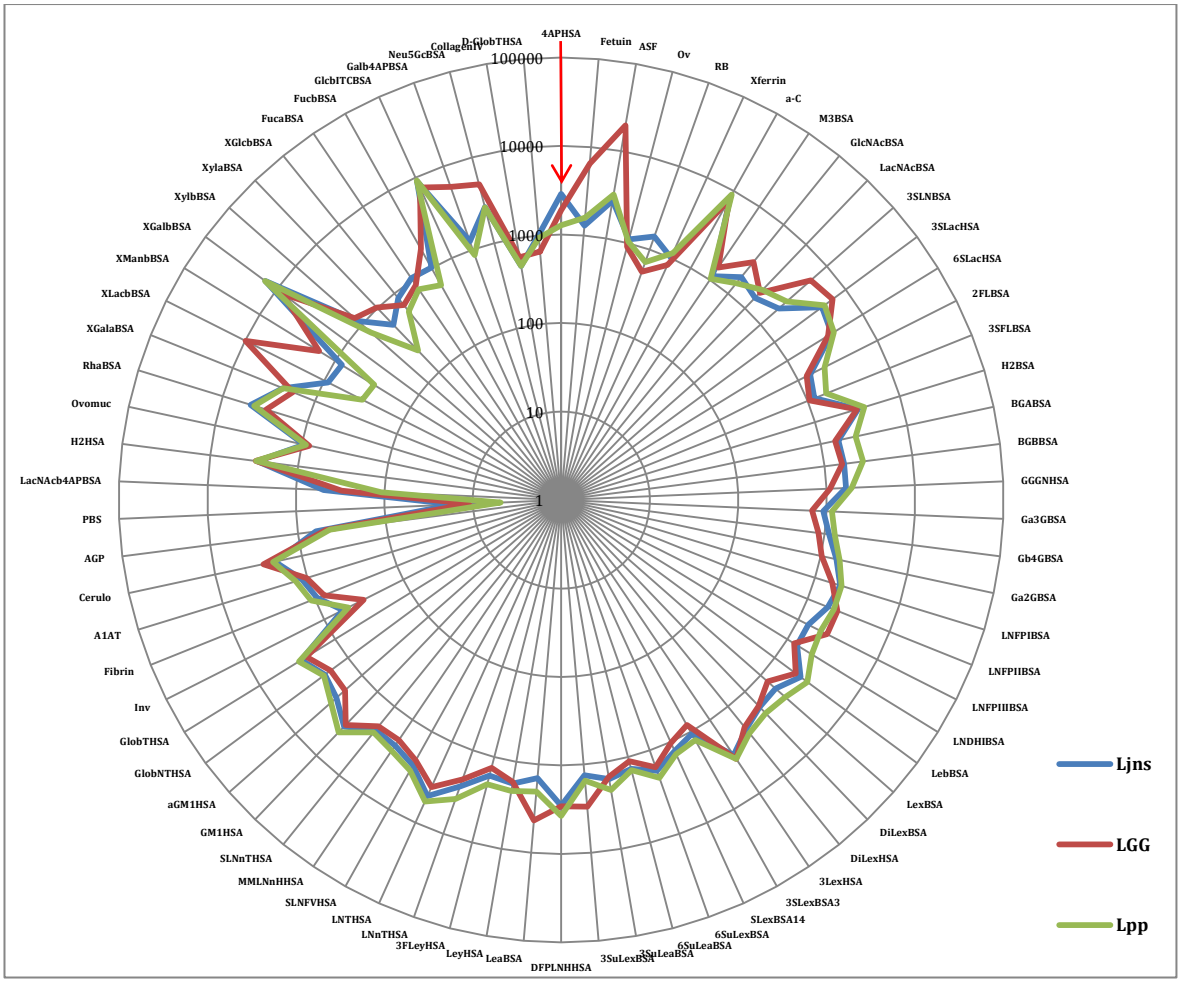

Figure 4.9: Binding of three lactobacilli species to NGC microarray

Radar plot of normalised signals from NGC microarrays analysis using whole cells from *Lactobacillus rhamnosus* GG (LGG), *L. johnsonii* (Ljns) and *L. paracasei* subsp. *paracasei* (Lpp). Red arrows points out the cut off value used in subsequent analyses. Abbreviations of NGCs are found in Table 2.14 and Table 2.15.

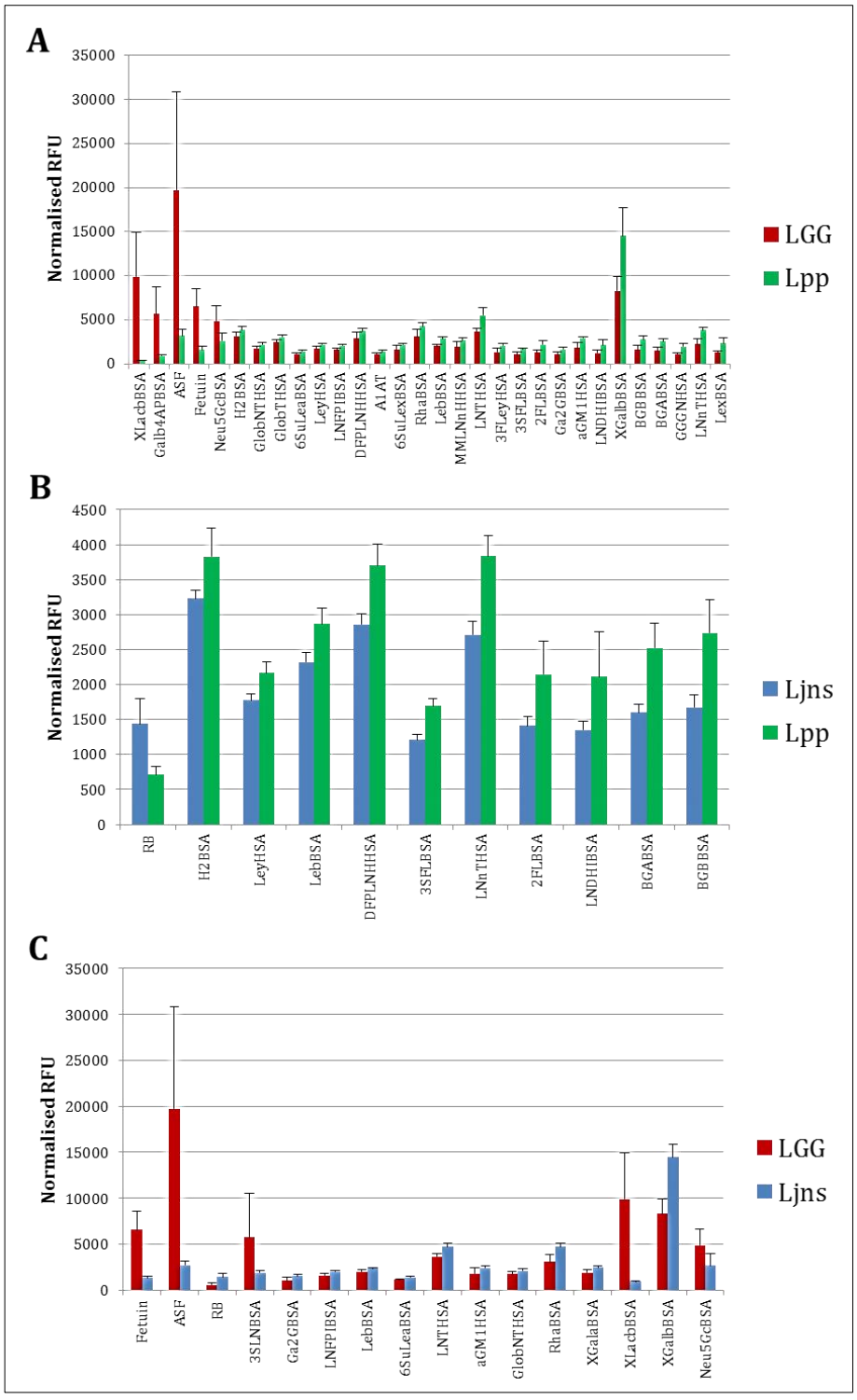

Figure 4.10: Differential binding of three lactobacilli to NGC microarray

NGCs showing significant differences in binding when each pair of lactobacilli species were compared. All signals are averages from three microarrays following scale normalisation. Error bars represent the standard deviation of the mean. A: Pairwise comparison between *Lactobacillus rhamnosus* GG (LGG) and *L. paracasei* subsp. *paracasei* (Lpp). B: Pairwise comparison between *L. johnsonii* (Ljns) and *L. paracasei* subsp. *paracasei* (Lpp). C: Pairwise comparison between LGG and *L. johnsonii* (Ljns). Abbreviations of NGCs are found in Table 2.14 and Table 2.15.

Highest binders for LGG were fetuin (glycoprotein containing three complex *N*-glycans and five *O*-glycans), asialofetuin (ASF), lactose-NGC (XLacBSA) as well as sialic acid (Neu5GCBSA). In contrast, *L. johnsonii* bound more strongly to RNase B (RB, containing high mannose *N*- glycans: $Man_{5-9}GlcNAc$) and *L. paracasei* favoured glycans containing fucose (15/30 NGCs compared to LGG; 9/10 NGCs compared to *L. johnsonii*).

Of the ten common NGC that differed between *L. paracasei* and the other two strains, seven structures contained α-(1→2)-linked fucose residues such as the ones found in the two blood group trisaccharides (BGABSA and BGBBSA, Table 4.1). The human milk core oligosaccharide lacto-*N*-neotetraose (LNnTHSA, Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc) was the only NGC not fucosylated that is also favoured by *L. paracasei*. Lacto-*N*-neotetraose contains the type 2 chain lactosamine structure (Gal-β-(1→4)-GlcNAc-β) that is found in half of the significant NGC (5/10). The lactosyl structure (Gal-β-(1→4)-Glc) is also a common feature found in 6/10 NGC.

Inhibition experiments using lactose resulted in the diminution of the signals corresponding to all ten NGC. Highest inhibitions were obtained with the Lewis antigen octasaccharide difucosyl-*para*-lacto-*N*-hexaose (50 %) and the human milk tetrasaccharide lacto-*N*-neotetraose (48 %).

Table 4.1: NGC bound differentially by *L. paracasei* subsp. *paracasei* whole cells

NGC	Structure	[1]%Inhib
H2BSA	Fuc-α-(1→2)Gal-β-(1→4)-GlcNAc-β1-	26.03
LeyHSA	Fuc-α-(1→2)-Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-β1-O-	31.23
LebBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-	33.71
DFPLNHHSA	Gal-β-(1→3)-[Fuc-α-(1→4)]GlcNAc-β-(1→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-	50.18
3SFLBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-[Fuc-α(1→3)]-Glc-	42.65
LNnTHSA	Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-	48.54
2FLBSA	Fuc-α-(1→2)-Gal-β-(1→4)-Glc-	42.77
LNDHIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-	42.33
BGABSA	GalNAc-α-(1→3)-[Fuc-α-(1→2)]-Gal-	41.98
BGBBSA	Gal-α(1→3)-[Fuc-α-(1→2)]-Gal-	41.77

[1] Inhibition by 100 mM lactose

4.4.2 Glyco-profiling of isolated pili from *L. paracasei* subsp. *paracasei*

To detect the glycan ligands of the potential adhesins present on *L. paracasei* pili (Chapter 3), these were isolated using an enzymatic procedure, labelled and incubated on NGC microarrays (Figure 4.11). In total, the pili extracts bound to 17 NGC above the threshold mark, corresponding to 14 individual glycans. Strongest interactions were obtained with NGC Glucose-β (GlcbITCBSA), Galactose-α/β (XGalaBSA, XGalbBSA) and Mannose-α (XManaBSA). These high interactions could, in part, be explained by the presentation of the NGC on the platform. These four NGC were monosaccharides conjugated in house with PITC that has been found to give higher signals when compared to the 4AP linker (Kilcoyne, Gerlach et al. 2012).

A very strong binder was also the blood group H-type 2 trisaccharide (present as two glycans: H2BSA and H2HSA of structure Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc). In addition, pili bound two glycans with the same high intensity: the human milk oligosaccharide 3' sialyllactosamine (3SLNBSA, Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc) as well as the type 2 chain disaccharide unit lactosamine (represented on the array by three NGC: LacNAcaBSA, LacNAcb4APBSA and LacNAcBSA, Gal-β-(1→4)-GlcNAc-α/β-). The αGal epitope (GGGHSA, Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc), also containing the type 2 chain disaccharide unit, also bound strongly. Ceruloplasmin (Cerulo), which contains 7 *N*-glycans including fucosylated, sialylated and terminal LacNAc structures, was the only glycoprotein that showed good binding to pili extracts. All the strong binders were inhibited by unlabelled pili (Table 4.2) or when challenged with haptenic sugars.

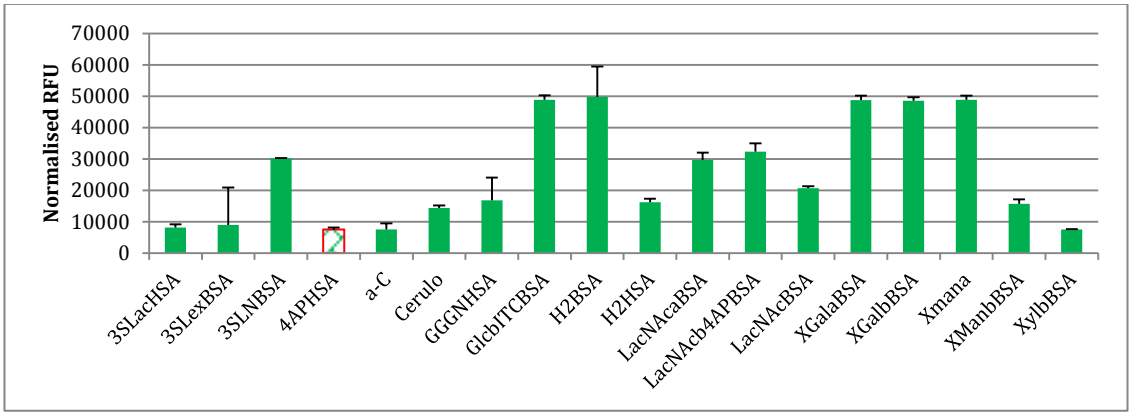

Figure 4.11: NGC bound by *L. paracasei* subsp. *paracasei* pili

Normalised signals from three NGC microarrays using labelled pili from *L. paracasei* subsp. *paracasei.* Only NGC above threshold (linker 4APHSA, highlighted in red) are shown.

Table 4.2: Competition experiments with *L. paracasei* subsp. *paracasei* pili on NGC microarrays

NGC	Structure	[1]Inh Pili	[2]Sugar	[3]Inh Hapt
4APHSA	4AP linker	nd	nd	nd
XGalaBSA	Gal-α	nd	Gal	0
XGalbBSA	Gal-β	95.26	Gal	21.3
Xmana	Man-α	95.23	Man	51.77
LacNAcBSA	Gal-β-(1→4)-GlcNAc	92.56	Lac	0
GlcbITCBSA	Glc-β	90.86	Lac	0
LacNAcb4APBSA	Gal-β-(1→4)-GlcNAc-β	68.26	Lac	25.88
3SLNBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc	60.34	GlcNAc	33.34
GGGNHSA	Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc	57.97	Lac	27.17
H2BSA	Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc-β1	55.91	Fuc	45.35
LacNAcaBSA	Gal-β-(1→4)-GlcNAc-α	53.97	Lac	18.93
H2HSA	Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc-β1	40.54	Fuc	30.01
Cerulo	7 *N*-glycans	20.41	Fuc	36.38
3SLacHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-Glc	17.67	Lac	53.97
a-C	*O*-GlcNAc with phosphorylation	12.59	GlcNAc	0
3SLexBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]GlcNAc	0	Fuc	13.4
XManbBSA	Man-β	0	Man	24.77
XylbBSA	Xyl-β	0	nd	nd

[1]Percentage signal inhibition when co-incubated with 10 µg pili
[2]Haptenic sugars used
[3]Percentage signal inhibition when co-incubated with 100mM sugars
nd: not determined

4.4.3 Mucin profiling of *L. paracasei* subsp. *paracasei*

As for *F. prausnitzii, c*ells and pili extracts from *L. paracasei* were labelled and incubated on mucins microarray (Figure 4.12). Of the top ten binders, seven glycoproteins were common between the two experiments with predominance of colonic mucins including the human MUC2 secreted by the LS174T cell line. Whole cells bound strongly to the bovine spiral colonic mucin. Highest signals for the isolated pili were the cecal mucins from pig and rat. Altogether, this suggests that *L. paracasei* predominant niche is the lower gastro-intestinal tract (colon) and that the (porcine) cecal mucins contain the glycan ligands for the extracted pili produced by *L. paracasei*.

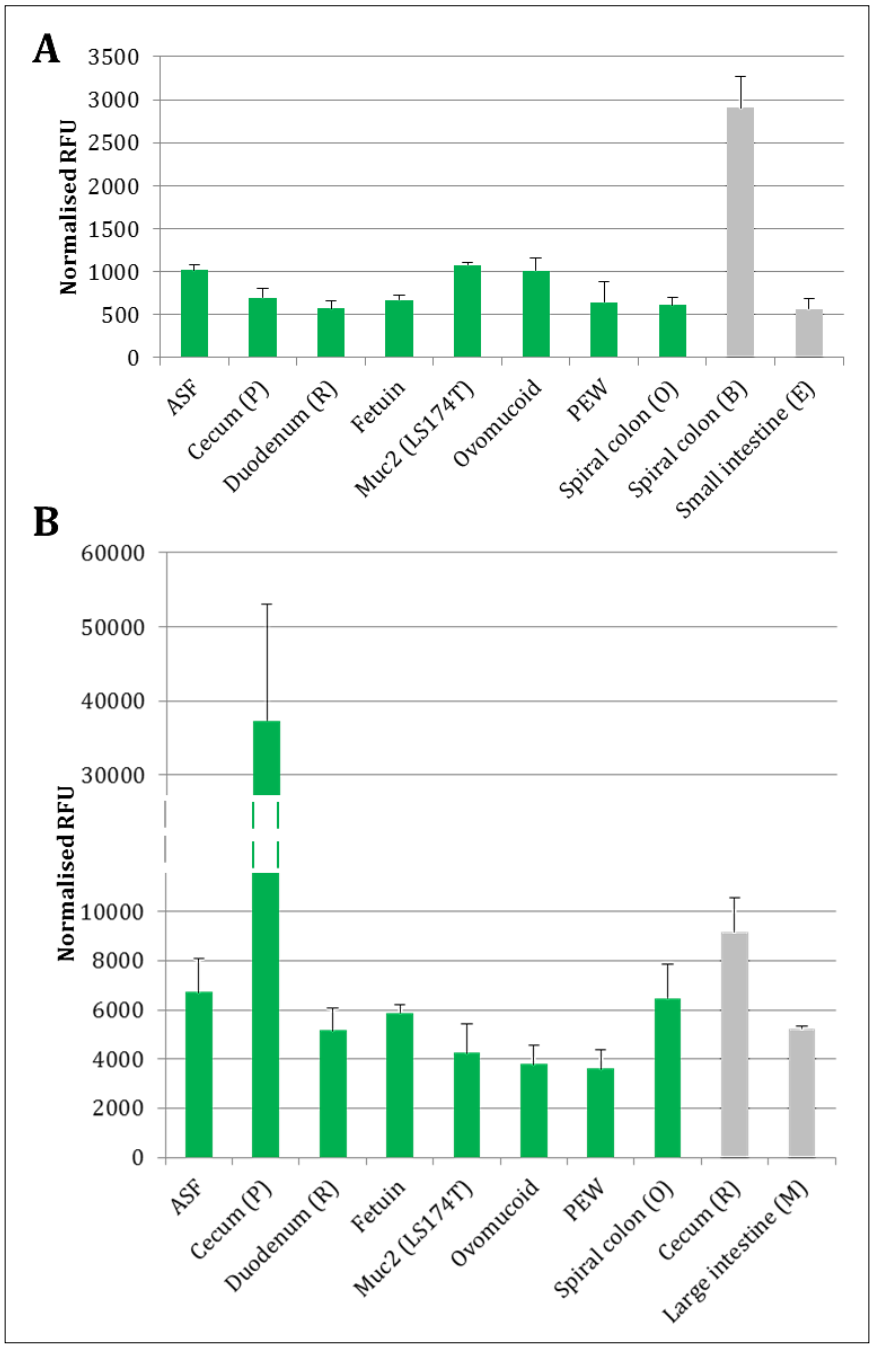

Figure 4.12: Top ten mucins bound by *L. paracasei* subsp. *paracasei* whole cells and pili extract

A: Normalised signals from microarrays using whole cells. B: Normalised signals from microarrays using isolated pili following the enzymatic digest described by *Reunanen* et al. (Reunanen, von Ossowski et al. 2012). Common mucins/glycoproteins between A and B are coloured in green. Abbreviations are as follows: asialofetuin (ASF), pigeon egg white (PEW), bovine (B), equine (E), murine-mouse (M), porcine (P), murine-rat (R), human (LS174T).

4.4.4 Western blotting of surface exposed proteins with HRP-labelled asialofetuin

Several proteins on the surface of *L. paracasei* cells could be involved in adhesion and could explain the differences in binding of whole cells compared to pili extracts on the microarrays. Surface-exposed proteins (SEP) from the bacteria were isolated using lithium chloride to investigate their putative lectin functionalities. SDS-PAGE under native conditions of the SEP preparation gave a multitude of bands ranging from ~ 20 to ~ 150 kDa (Figure 4.13 A). Immunoblotting using HRP-labelled asialofetuin highlighted a strong band at ~ 60 kDa and three fainter bands at ~ 64 kDa, ~ 42 kDa and 35 kDa. Competition with a cocktail of monosaccharides present in asialofetuin resulted in inhibition of the ~42 kDa band after 1h incubation. Inhibition of the other bands was observed following an overnight incubation.

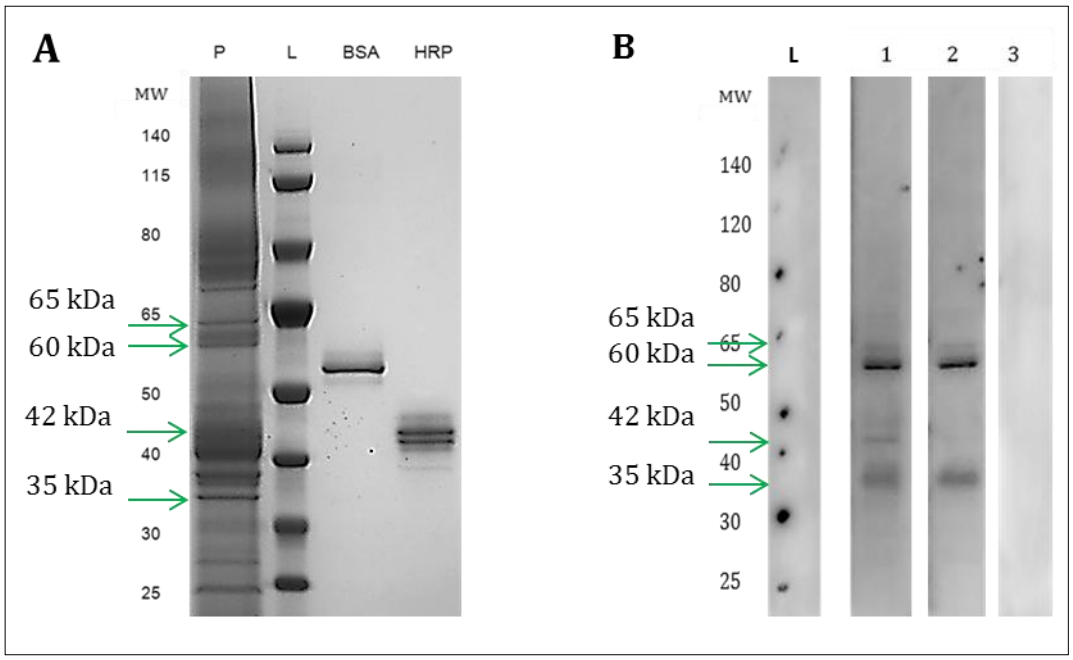

Figure 4.13: SDS-PAGE and Western Blot of surface exposed proteins from *L. paracasei* subsp. *paracasei*

A: Coomassie-stained SDS PAGE (4-12%) under native conditions with P: surface-exposed proteins (SEP) from *L. paracasei* subsp. *paracasei* (50 µg); L: molecular weight ladder (MW in kDa); BSA and HRP controls (2 µg,). B: Western blot (WB) using HRP-labelled asialofetuin (HRP-ASF) incubated with HRP-ASF for 1h (lane 1); WB incubated for 1h with HRP-ASF and competing sugar cocktail of 100mM galactose, mannose and GlcNAc (lane 2); WB incubated for 16h with HRP-ASF and competing sugar cocktail (lane 3).

These results suggest lectin functionalities of the SEP aforementioned. The larger band of 64 kDa might correspond to the collagen adhesion protein identified by *in silico* analysis (ZP_03964057.1, 65.5 kDa). The ~ 42 kDa band could correspond to the moonlighting protein EF-Tu (43 kDa), which was characterised as a cell surface adhesin in various *Lactobacilli* strains (Granato, Bergonzelli et al. 2004; Ramiah, van Reenen et al. 2007). The gene expression of the collagen adhesion protein and EF-Tu (respectively L48083 and LEFTU, Chapter 3, Table 3.10) were both up-regulated when *L. paracasei* was grown in the presence of mucin and salt. The two other bands could correspond to the putative adhesins of a sortase-dependent pilus cluster identified in Chapter 3, and which contain lectin-like characteristics (Chapter 3, Figure 3.5). These were Spa_C (EEI69166, 33 kDa) and SpaC_vWF (EEI69165, 60 kDa) that could match the ~ 35 kDa and the ~ 60 kDa bands, respectively. Interestingly, a SpaA-like protein was also up-regulated in the presence of mucin and salt. In LGG, SpaA and SpaC constitute respectively the major pilin and the adhesin of the SpaCBA.

4.4.5 Discussion

Lactobacillus paracasei subsp. *paracasei* exerts beneficial properties on its host by producing anti-microbial agents, modulating host immune functions and promoting the mucosal barrier integrity (Marzotto, Maffeis et al. 2006; Gudina, Teixeira et al. 2010; Bendali, Madi et al. 2011; Zhang, Wang et al. 2013). An important aspect for probiotic strains is the ability to adhere to and be retained in the mucosa. In lactobacilli, several adhesion molecules have been recently identified following *in silico* analysis and adhesins have been functionally characterised in 10 species (Chapter 1) (Boekhorst, de Been et al. 2005; Buck, Altermann et al. 2005; Mobili, Serradell Mde et al. 2009; van Passel, Kant et al. ; Van Tassell and Miller 2011; von Ossowski, Satokari et al. 2011) (Chapter 1). However, other than mannose-specific lectin from *L. plantarum*, no defined sugar ligand has been identified in any of the lactobacilli adhesins.

Results from our glyco-arrays following comparison to two pili-producing *Lactobacillus* strains suggest *L. paracasei* cells favour blood group and fucosylated glycans with α-(1→2) fucose linkages such as Le^b and Le^y antigens. This was supported by the strong binding to the bovine spiral colonic mucin, which showed the highest proportion of α-(1→2) linked fucose across various mammalian mucins, as identified by the interaction with UEA-I lectin and monosaccharide analysis (Kilcoyne, Gerlach et al. 2012).

Interestingly, recent findings have linked the secretor phenotype, that is to say the fucosyltransferase 2 (FUT2) polymorphism, to the composition of the microbiota (Rausch, Rehman et al. 2011; Wacklin, Tuimala et al. 2014). Previous studies using surface plasmon resonance have shown that ABO histo-blood group antigens are receptors of lactobacilli strains such as *L. brevis*, suggesting a role for these probiotic strains in pathogen exclusion by competitive binding (Uchida, Kinoshita et al. 2006).

Isolated pili also showed a strong interaction to α-(1→2) fucose with the binding of the H-type 2 trisaccharide, which was inhibited by fucose. Fucosylated oligosaccharides have been reported to be ligands for lectins from pathogenic species such as *Burkholderia sp.* and *Pseudomonas sp.*(Imberty, Wimmerova et al. 2004; Audfray, Claudinon et al. 2012; O'Boyle, Houeix et al. 2013). A panel of fucosylated ligands was defined for BambL, the fucose-binding lectin from *Burkholderia ambifaria*, following glycan arrays (Audfray, Claudinon et al. 2012). Strongest binding was found with the H-type 2 trisaccharide and Le^y antigen, suggesting the bacteria binds in preference humans of O-blood group types and secretor positive status. *In silico* molecular docking of BambL to fucosylated oligosaccharides also suggested H-type 2 as a favoured ligand (Topin, Arnaud et al. 2013). In addition, BambL displayed the highest affinity to the blood group A and H-type 2 oligosaccharides following affinity measurements by titration calorimetry (Topin, Arnaud et al. 2013).

Glycan arrays using pili extracts also showed strong binding to lactosyl glycans such as lactosamine. In addition, western blotting of surface-exposed proteins from *L. paracasei* with asialofetuin revealed four putative lectin candidates. Based on their molecular weights, these could correspond to the moonlighting protein elongation factor (EF-Tu), a collagen adhesion protein and the two Spa_C-like adhesins. Both EF-Tu and SpaC display adhesion properties to epithelial cells and mucus in various *Lactobacilli* (Granato, Bergonzelli et al. 2004; Kankainen, Paulin et al. 2009) (Chapters 1 and 3). Hyland et al. reported lactosamine to be the ligand for the bundle-forming pili from enteropathogenic strain *E. coli* EPEC (Hyland, Sun et al. 2008). Galectins form a class of animal lectins that all bind lactosamine but with refined specificities towards substituted glycans (Kasai and Hirabayashi 1996; Sato, St-Pierre et al. 2009). Some galectins such as galectin-3 are also able to bind to blood group A and B antigens (Kasai and Hirabayashi 1996). Moreover, recent studies reported on galectin-like proteins found in pathogens and parasites (Davicino, Elicabe et al. 2011). Interestingly, galectin-3 has a unique structure with a

collagen-like domain, which enables polymerisation of the protein (Romer and L. 2011). Such domain has been found in adhesin candidates of *L. paracasei* following *in silico* analysis. Furthermore, galectins are able to bind to ECM glycoproteins and are considered key molecules in immuno-regulation by their binding to leukocytes. It is therefore possible that the pili extracts of *L. paracasei* contain a single adhesin with a broader sugar specificity i.e. fucosylated and lactosyl glycans. Gene expression following *in silico* analysis resulted in the up-regulation in the presence of mucin and salt of five adhesin candidates, including the SpaD-like and SpaF-like proteins (Chapter 3). A SpaFED pilus cluster has been reported in LGG, with SpaD corresponding to the major pilin subunit and SpaF the adhesin that binds mucus (von Ossowski, Reunanen et al. 2010). We could speculate that the Spa_F adhesin from *L. paracasei* displays galectin-like functionalities recognising lactosamine, 3' sialyllactosamine and the H-type 2 trisaccharide. In addition, the binding of pili extracts to porcine cecum suggests this mucin contains these glycans and that *L. paracasei* cells colonise or reside in this lower GIT niche. The strain grows on fructans as the sole source of carbon *in vitro*, including long-chains of inulin, implying a residence in the colon, where simple sugars are scarce (Makras, Van Acker et al. 2005).

More likely, our extracts contain more than one type of fimbriae. In support of this hypothesis, four pilus clusters were identified in the proteome from *L. paracasei* (Chapter 3). Moreover, in addition to SpaD and SpaF, a SpaA-like pilin gene was up-regulated in the presence of mucin and salt. In LGG, SpaA is the major pilin of the SpaCBA cluster and SpaC constitutes the adhesin binding mucus (von Ossowski, Satokari et al. 2011). The related casei member *L. paracasei* might therefore synthesise different types of pilus to colonise various niches. The bacterium is indeed found in the infant's gut from the age of one but has been detected in the human GIT from the mouth (dental plaque) to the large colon (Maukonen, Matto et al. 2008).

4.5 Determining the target molecules of sortase-dependent pili from *Bifidobacterium bifidum* PRL2010

Lactococcus lactis clones engineered using the NICE® system to produce pili was provided by collaborators (see details in Chapter 2). These included the non-piliated strain *L. lactis* NZ9000 and a panel of sortase-dependen pili-producing clones. Two gene clusters from *B. bifidum* PRL2010, containing each two pilin subunits and a sortase, were cloned into plasmid PNZ8150. Clusters *Pil2* and *Pil3* consisted of respectively major pilin subunits FimA and FimP, and putative adhesins FimB and FimQ. A negative control, i.e. *L. lactis* transformed with an empty plasmid, was also received.

4.5.1 Optimisation of pilus expression in *L. lactis*

Plasmid PNZ8150 used in the NICE® lactococcal expression system contains a nisin-inducible promoter. The recommended induction time for protein expression is 90 minutes following the addition of nisin (Mierau and Kleerebezem 2005). Initially, a time-course experiment was carried out 1h, 2h or 3h post-nisin induction to detect the best expression time of pilus formation in our hands (Figure 4.14).

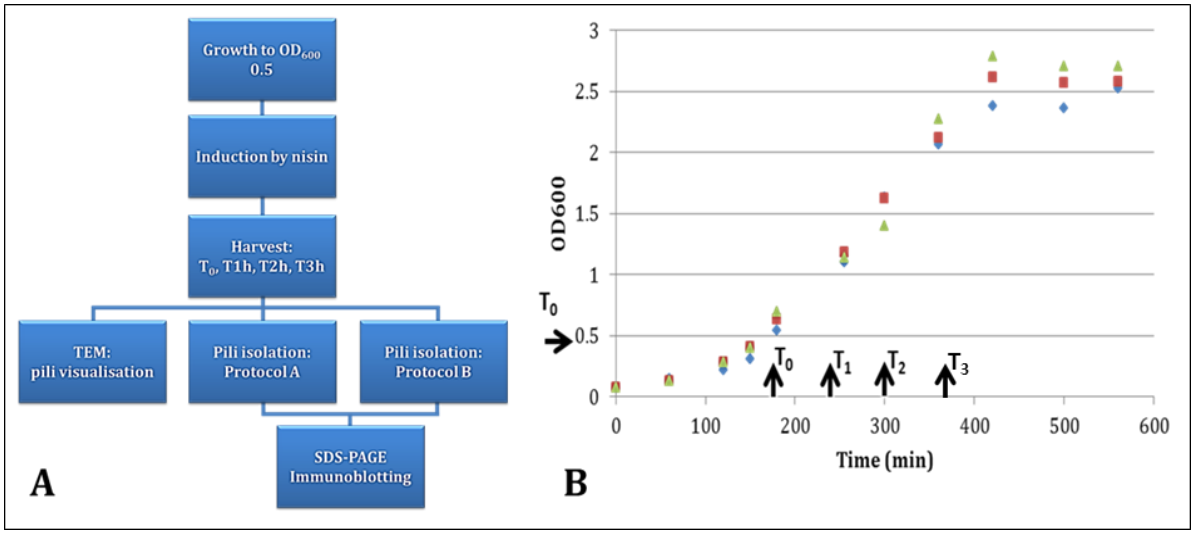

Figure 4.14: Optimisation of the expression of bifidobacterial pili in *L. lactis* clones

A: Schematic of the optimisation procedure for the expression of bifidobacterial pili
B: Growth curves of nisin-induced *L. lactis* clones
Further details of the pili isolation can be found in Chapter 2

The expression of pili was determined by TEM before and after induction with nisin (Figure 4.15). No appendages were visible before induction but they were visible on *L. lactis* clones containing pili clusters especially at 3h post-induction (T_3) compared to the non-piliated negative control. Aggregation of pili-producing clones was also observed and increasing overtime.

Appendages were isolated using two different procedures.

The first protocol involved the treatment of whole cells with muramidases mutanolysin and lysozyme, which degrade the bacterial peptidoglycan layer by hydrolysing the β(1→4) glycosidic bonds between *N*-acetylmuramic acid and *N*-acetylglucosamine (protocol A).

In the second protocol (protocol B), the bacteria were initially subjected to a milling step to break open the cells. The membrane fraction was then separated from the cytosol by centrifugation and subsequently treated with the two muramidases (see Chapter 2 for details).

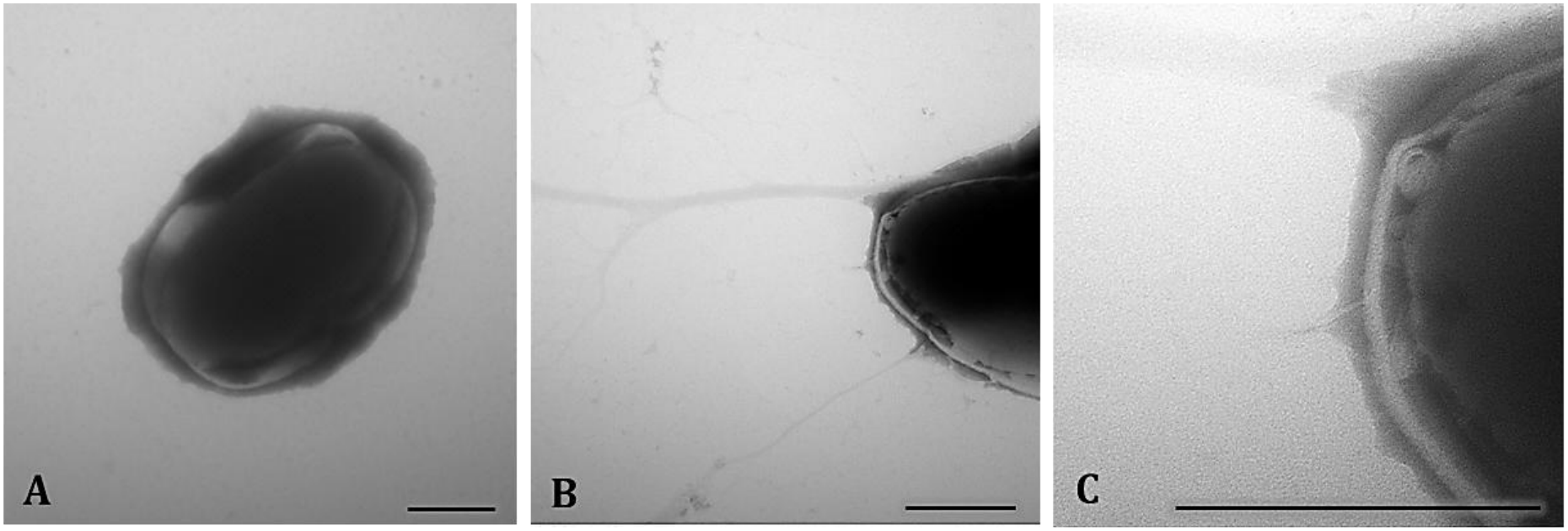

Figure 4.15: Visualisation of pilus formation by TEM post-nisin induction (T=3h)

A: Non-piliated *L. lactis* strain NZ9000 at T3h post-nisin induction
B: Transformant containing the sortase-dependant pili cluster *Pil*2 from *B. bifidum* PRL2010 at T3h post-nisin induction
C: Magnification of B
Scale bar represents 500nm

Enzymatically released proteins were quantified and separated by SDS-PAGE (3-8%) before western blotting (Figure 4.16). The pilin genes present in the two sortase-dependent pili clusters *Pil*2 and *Pil*3 from *B. bifidum* had been previously expressed in *E. coli* by our collaborators. Crude cytosolic extracts from *E. coli* clones expressing the pilin subunits FimA (56.35 kDa), FimP (51.4 kDa) and FimQ (121 kDa) were subsequently used to immunise rabbits. We were provided with the three rabbit sera in order to detect the presence of the various pilin subunits expressed in *L. lactis*. Anti-FimA serum was the only serum that gave a clear signal on immuno-blots (Figure 4.16). The pilin FimA monomer (56-58KDa) was detected in the membrane fractions and increasingly with time from 1 to 3h, where dimer and polymers are also being detected using both protocols. A 'ladder'-like effect, corresponding to the polymerisation of FimA and FimB monomers was visible at T_3 and especially using protocol A. Unspecific bands were present in the negative control but did not correspond to FimA. The optimal conditions of expression determined were thus 3h post-induction using the direct enzymatic protocol.

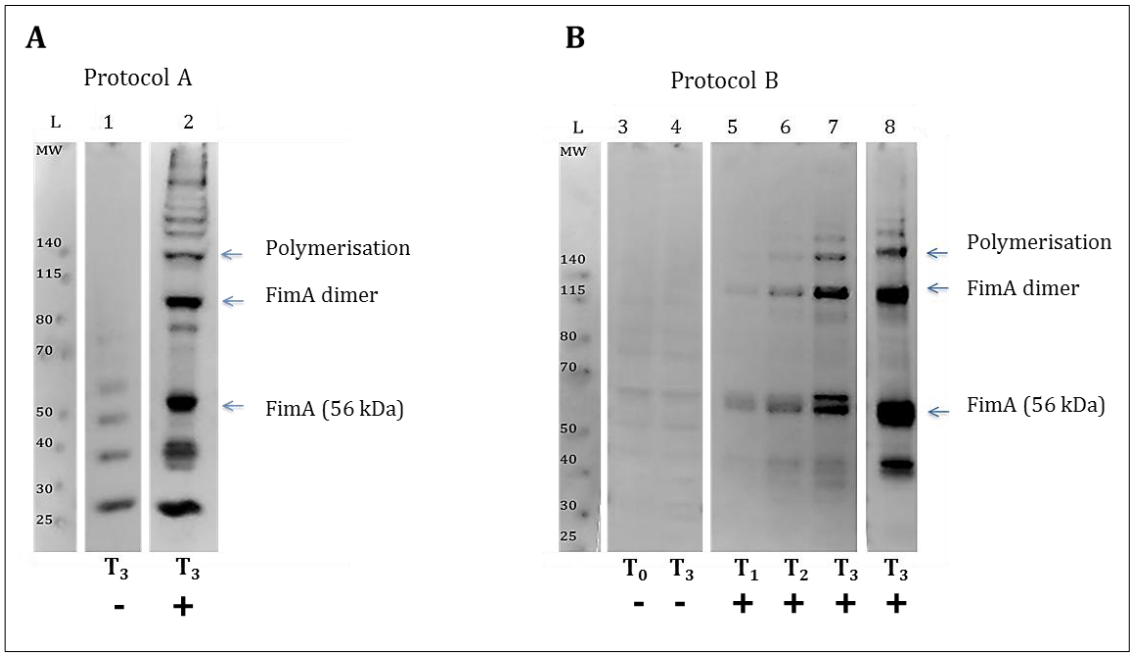

Figure 4.16: Immunoblotting of surface protein isolates of *Pil*2 expressing clones

A: Surface protein isolates following the direct enzymatic method (protocol A) from clones at 3h post-induction (T_3) of *L. lactis* negative control (lane 1, 25 µg) and clone expressing *Pil*2 (lane 2, 25 µg).

B: Surface protein isolates using the indirect procedure (protocol B) from *L. lactis* negative control before induction T_0 (lane 3, 10 µg), T_3 (lane 4, 10 µg) compared to clone expressing *Pil*2 at T_1 (lane 5, 10 µg), T_2 (lane 6, 10 µg) and T_3 (lane 7, 10 µg; lane 8, 25 µg). L denotes molecular weight ladder from 25 to 140 kDa. Blue arrows point out the possible FimA monomer and FimA or/and B subunits polymerisation.

4.5.2 Determining glycan-binding activity of isolated pili

High molecular weight protein fractions (above 300 kDa) isolated from *L. lactis* negative control, and clones expressing *Pil*2 or *Pil*3 were retained by filtration and labelled using a fluorescent dye prior to incubation on a neo-glycoconjugate (NGC) microarray. Time points corresponding to the optimal pili expression (i.e. 3h post-induction) were compared to pre-induction time points (T_0) for *Pil*2 and *Pil*3. Eighteen features corresponding to 16 glycans bound stronger to the *Pil*2-expressing clones than the negative control (Figure 4.17). Highest signals were obtained when binding the blood group H-type 2 trisaccharide (H2BSA and H2HSA, of structure Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc-R), the human milk oligosaccharide 3'-sialyllactosamine (3SLNBSA, Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc), the glycoprotein invertase (Inv, containing 14 high mannose *N*-linked glycans) and the α-Gal epitope (GGGHSA, Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc). In terms of binding fold change following induction, highest increases were seen when binding to β-mannose (XManbBSA), lactosamine (LacNAcaBSA and LacNAcb4BSA, Gal-β-(1→4)-GlcNAc-α/β), α-xylose (xylaBSA), invertase (Inv), fibrinogen (Fibrin) and the α-gal epitope (GGGHSA).

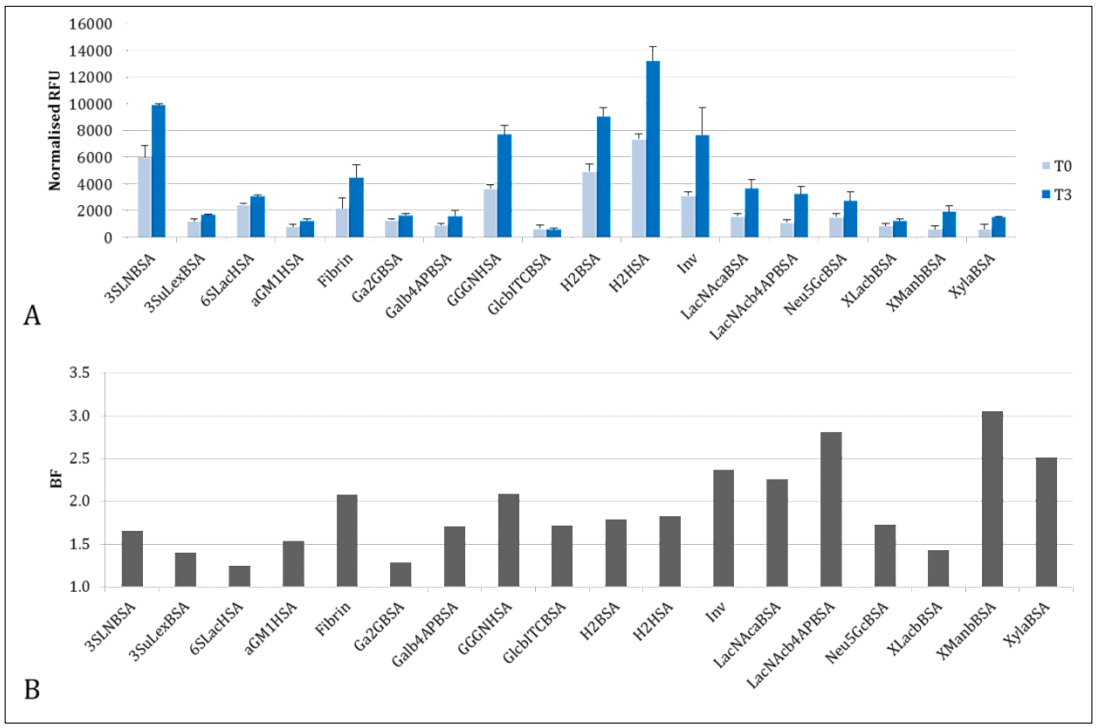

Figure 4.17: Significant glycans bound by *Pil*2 following nisin induction

A: normalised signals from three microarrays using *Pil*2 labelled pili at T_0 (pre-induction) and T_3 (post-induction). B: binding fold change between T_0 (pre-induction) and T_3 (post-induction) for 18 significant glycans.

Competition experiment with relevant sugars (Table 4.3) resulted in the inhibition of signals for most bound glycans (11/16) including the ones for which highest signals were obtained such as the blood group H-type 2 trisaccharide (above 50%).

Table 4.3: Competition experiment on *Pil*2 with haptenic sugars

Glycan	Structure	[1]*P*	[2]Hapten	[3]Inh
3SLNBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc	0.005	GlcNAc	23.3
3SuLexBSA	(SO4)-Gal-β-(1→4)-[Fuc-α-(1→3)]GlcNAc	0.029	Fuc	44.5
6SLacHSA	Neu5Ac-α-(2→6)-Gal-β-(1→4)-Glc	0.034	Gal	0
aGM1HSA	Gal-β-(1→3)-GalNAc-β-(1→4)-Gal-β-(1→4)-Glc	0.029	Gal	51.8
Fibrinogen	1-12 *N*-glycans (mainly A2G2S1, A2G2S2)	0.018	GlcNAc	0
Ga2GBSA	Gal-α-(1→2)-Gal	0.034	Gal	0
Galβ4APBSA	Gal-β-(1→4)-Gal	0.020	Gal	14.1
GGGNHSA	Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc	0.001	Gal	73.3
GlcβITCBSA	Glc-β-	0.021	Glc	0
H2BSA	Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc-β1	0.001	Fuc	59.2
H2HSA	Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc-β1	0.002	Fuc	66.4
Invertase	14 *N*-glycans (mainly high mannose)	0.003	Man	41.6
LacNAcαBSA	Gal-β-(1→4)GlcNAc-α	0.002	GlcNAc	73.1
LacNAcβ4APBSA	Gal-β-(1→4)GlcNAc-β	0.001	GlcNAc	82.8
Neu5GcBSA	Neu5Gc-	0.029	nd	nd
XLacβBSA	Lac-β-	0.044	Gal	56.4
XManβBSA	Man-β-	0.005	Man	0
XylaBSA	Xyl-α-	0.018	nd	nd

[1]*P* values of significant glycans

[2]Haptenic sugars

[3]Percentage of signal inhibition

Abbreviations are as follows: nd, not determined; A2G2S1, biantennary digalactosylated monosialylated structures; A2G2S2, biantennary digalactosylated disialylated structures

Following successful detection of *Pil*2 subunit polymerisation by immunoblotting, G class immunoglobulins (IgG) were purified from rabbit serum and a 'sandwich' binding assay was carried out on the microarray to amplify the detection of a potential ligand. The assay consisted of a secondary fluorescent anti-IgG antibody binding the primary anti-FimA polyclonal IgG detecting the isolated *Pil*2 bound to the immobilised NGC. The secondary antibody was initially incubated on NGC arrays on its own to check for non-specific binding and resulted in acceptable low level of noise (e.g. best binders consisted of linkers 4APBSA and 4APHSA with RFU values below 1800). Analysis of the incubated 'sandwich' including isolated proteins from pre-induction time point (T_0) or post-induction (T_3) were successfully carried out (Figure 4.18).

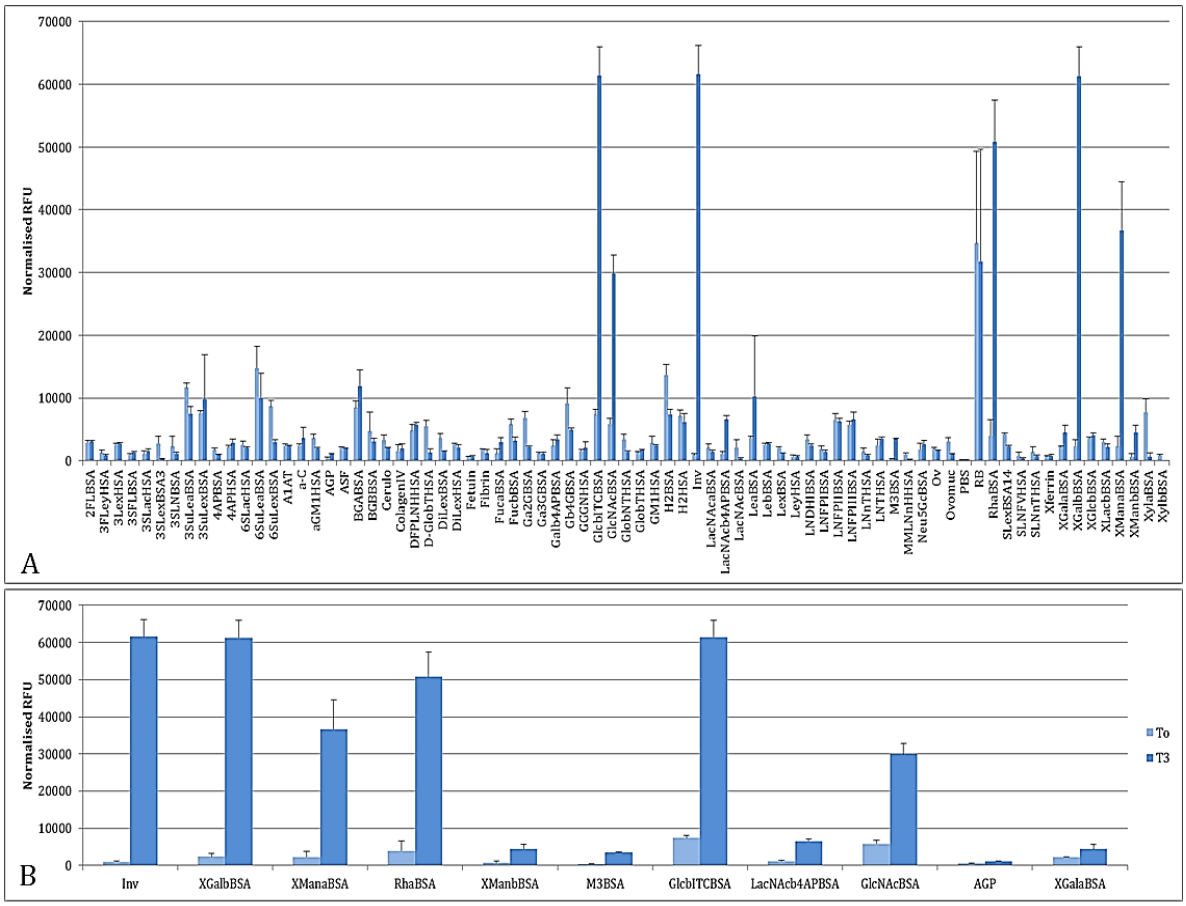

Figure 4.18: Putative FimA glycan ligands

Signals from three microarrayss following a sandwich assay on NGCs
A: Combination *Pil*2 pili pre- (T_0) and post-induction ($T_{3)}$, rabbit anti-FimA IgG and labelled Goat-anti-IgG. B: Significant binding events following *T* test (P<0.05 with FDR 5%).

Statistical comparison between time points identified 11 glycans that bound more strongly after 3h induction. Highest signals corresponded to the glycoprotein invertase, and five monosaccharides: glucose (GlcbITCBSA, β-linked to ITC linker), galactose (XGalbBSA, β-linked to ITC linker), rhamnose (RhaBSA), mannose (XManaBSA, α-linked to ITC linker) and GlcNAc (GlcNAcBSA). Other significant binders presenting lower signals included two monosaccharides: α-linked galactose (XGalaBSA) and β-linked mannose (XManbBSA); the mannosyl trisaccharide (M3BSA, Man-α-(1→3)[Man-(α1-6)]-Man); lactosamine and the alpha-1-acid glycoprotein (AGP, containing complex *N*-glycans). Most of these binding events reflected binding using labelled pili directly on the microarray. However, binding fold increase from T_0 to T_3 obtained when using labelled pili with three glycans i.e. the blood group H-type 2 trisaccharide (H2BSA and H2HSA), the human milk oligosaccharide 3'-sialyllactosamine (3SLNBSA) and the α-gal epitope (GGGHSA) were not seen using the 'sandwich' assay, implying that their binding was due to other surface exposed proteins of high molecular weight in the fraction. These results strongly suggest that *Pil*2 has a preference for glycans containing α-mannosyl structures but that it also binds strongly to monosaccharides such as rhamnose and GlcNAc.

The *Pil*3 isolate from the 3h post-induction time point (T_3) was also compared to the pre-induction time point (T_0) for (Figure 4.19). Ten glycans bound preferentially at T_3 with fold changes T_3/ T_0 from ~1.5- to 4-fold. The minimal structures involved were two trisaccharides presenting Lewis antigens (sulfated Le^a: 3SuLeaBSA, of structure (SO4)-3-Gal-β-(1→3)-[Fuc-α-(1→4)]GlcNAc; sulfated Le^x: 6SuLexBSA, (SO4)-6-Gal-β-(1→4)-[Fuc-α(1→3)]GlcNAc). Two larger oligosaccharides with Lewis antigens were also bound (Le^b hexasaccharide: LNDHI, Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc and Le^x pentasaccharide: LNFPIIIBSA, Gal-β-(1→4)-[Fuc-α-(1→3)]GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc); and six glycoproteins α-1-antitrypsin (A1AT, containing mono-fucosylated complex *N*-glycans), α-crystallin (a-C, *O*-GlcNAc residues), α-1-acid glycoprotein (AGP, complex *N*-glycans with high branching), fetuin (3 complex *N*-glycans; 5 *O*-glycans), ovalbumin (Ov, hybrid and high mannose *N*-glycans) and transferrin (Xferrin, complex *N*-glycans).

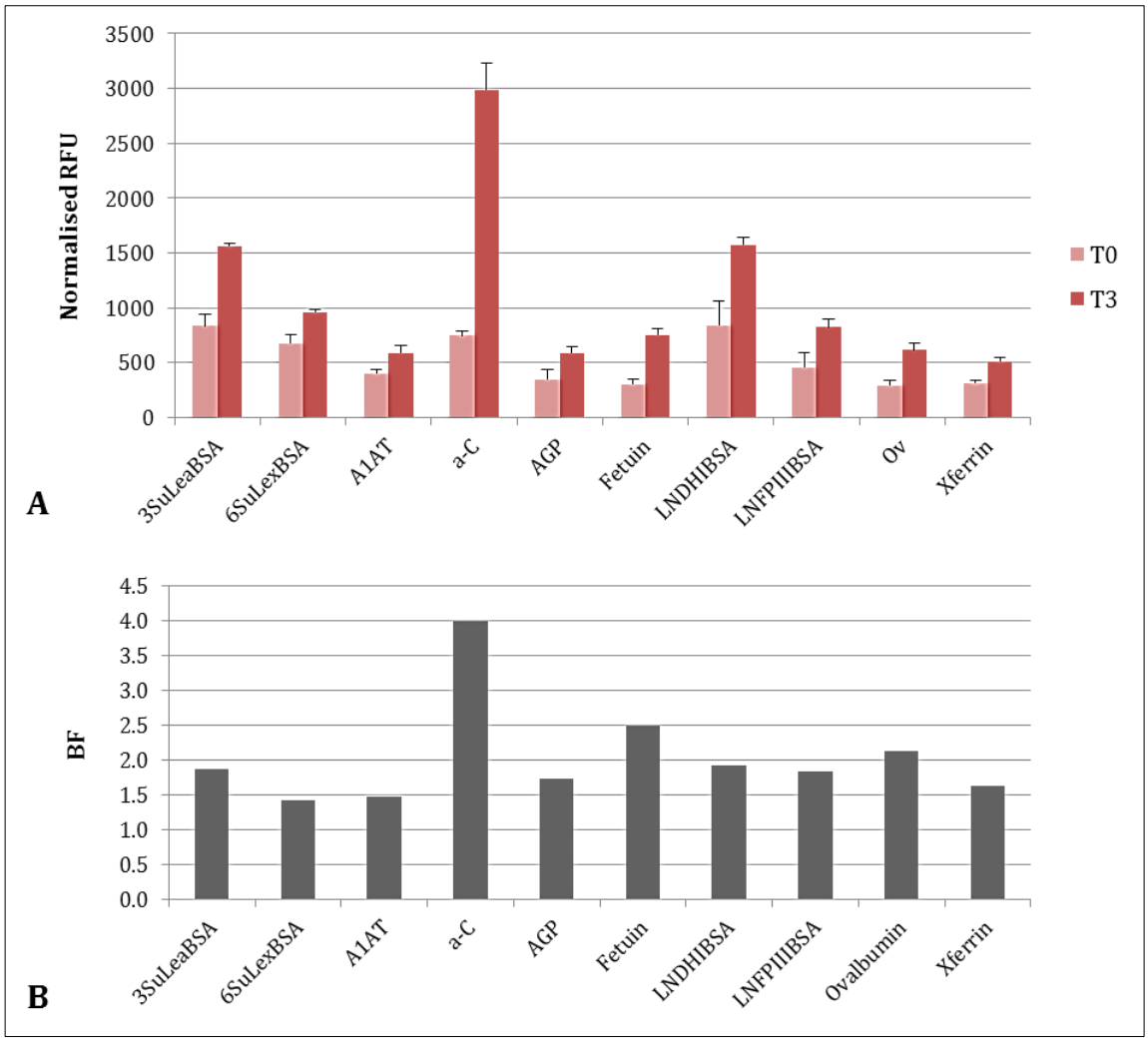

Figure 4.19: Significant glycans bound by *Pil*3 following nisin induction

A: Normalised signals from three microarrays using *Pil*3 labelled pili at T_0 (pre-induction) and T_3 (post-induction). B: Fold change T_3/ T_0 (BF) following *T* test (P<0.05 with FDR 5%) between T_0 and T_3

Competition studies with monosaccharides were carried out to confirm *Pil*3 specificities (Table 4.4). Signal inhibitions were greatest with GlcNAc on all six glycoproteins (27 to 86 %). However, no (or poor) inhibition was obtained on Lewis antigens with fucose, suggesting terminal fucose is not the ligand for *Pil*3. Larger structures could be involved with the presence of lacto-*N*-biose and fucose branching as suggested by the highest interactions observed.

Table 4.4: Competition experiment on *Pil*3 with haptenic sugars

Glycan	Structure	[1]*P*	[2]Hapten	[3]Inh
3SuLeaBSA	(SO4)-3-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc	0.005	Fucose	11.9
6SuLexBSA	(SO4)-6-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc	0.026	Fucose	6.1
A1AT	Mono-fucosylated bi-, tri-, and tetra-antennary complex *N*-glycans	0.024	GlcNAc	40.3
a-C	*O*-GlcNAc and phosphorylated	0.000	GlcNAc	86.6
AGP	Mainly triantennary complex *N*-glycans, also high branching (polylactosamine)	0.033	GlcNAc	73.4
Fetuin	3 complex *N*-glycans; 5 *O*-glycans	0.005	GlcNAc	73.1
LNDHIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc	0.022	Fucose	0
LNFPIIIBSA	Gal-β-(1→4)-[Fuc-α(1→3)]GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc	0.032	Fucose	2.2
Ovalbumin	Hybrid bi- to penta-antennary structures and high-mannose [$(Man)_5(GlcNAc)_2$ and $(Man)_6(GlcNAc)_2$] *N*-glycans	0.005	GlcNAc	37.4
Xferrin	Bi- and tri-antennary complex *N*-glycans with polylactosamine	0.008	GlcNAc	27.3

[1]*P* values of significant glycans
[2]Haptenic sugars
[3]Percentage of signal inhibition

4.5.3 Differential binding of *Pil*2 and *Pil*3 to GIT mucins

Labelled pili were also incubated on a microarray containing 40 mucins isolated from several species including two human mucins (Figure 4.20). This platform might contain glycan ligands for *B. bifidum* PRL2010 and might present the glycans as they are *in vivo*. Strongest interactions were obtained with the binding of *Pil*2 to mucins from porcine cecum and the binding of *Pil*3 to mucins from rat duodenum. Both pili also bound equally various glycoproteins and colonic mucins including the intestinal MUC2 produced by human cell line LS174T.

A comparative analysis was carried out using the two pili-expressing lactococcal clones. Significant differences were seen with ten glycoproteins differentially bound by the two pili (Table 4.5). *Pil*2 showed greater binding to three cecal mucins (from pork, rat and chicken) as well as to the mucin from the deer large intestine, whereas *Pil*3 preferably bound equine and rat colonic mucins, three others originating from the stomach of bovine, porcine and including the human MUC5AC produced by E12 cells. There was also a

significant binding by *Pil*3 to pigeon egg white printed on the platform. As supported by the NGC microarray, this mucin array analysis suggests that *Pil*2 and *Pil*3 bind to different glycan ligands suggesting *B. bifidum* PRL2010 might display an adhesion strategy by expressing various pili in response to various environments.

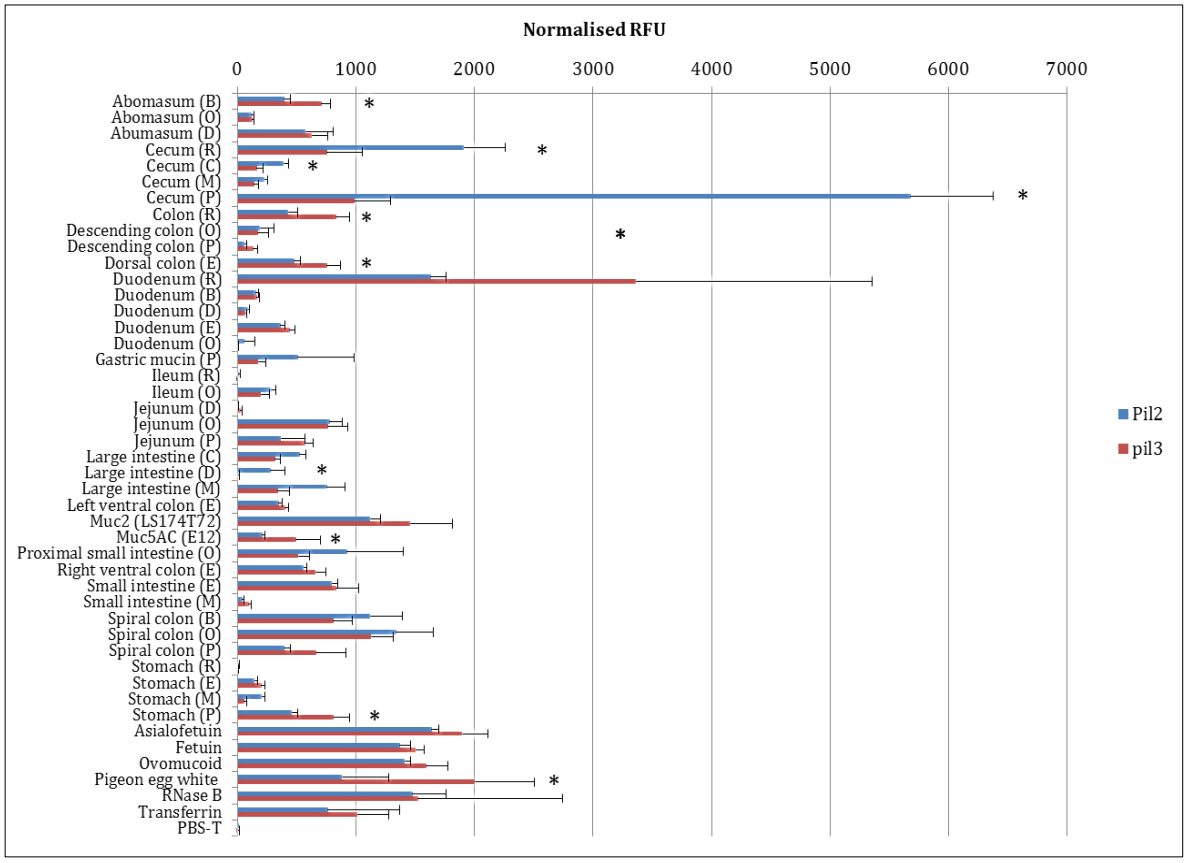

Figure 4.20: Differential binding to mucins by bifidobacterial pili *Pil*2 and *Pil*3

Normalised signals from three microarrays. Significant binding differences between *Pil*2 and *Pil*3 following *T* test ($P<0.05$ with FDR 5%) are represented by *. Abbreviations are as follows: bovine (B), galline (C), cervine (D), equine (E), murine-mouse (M), ovine (O), porcine (P), murine-rat (R), human (E12, LS174T)

Table 4.5: Differential glycan binding to glycoproteins by *Pil*2 and *Pil*3 on mucin arrays

[1]Glycoprotein	[2]*Pil*2 T_3	sd	[3]*Pil*3 T_3	sd	[4]*P*	[5]BF
Large intestine (D)	280.8	121.55	6.03	13.4	1.79E-02	48.21
Cecum (P)	5679.91	694.09	987.86	304.04	7.59E-03	5.07
Cecum (R)	1911.26	348.85	757.48	299.4	4.91E-02	2.26
Cecum (C)	391.85	40.95	166.79	46.88	2.14E-02	2.06
Dorsal colon (E)	479.21	56.21	755.43	114.02	2.03E-02	-1.83
Stomach (P)	461.25	49.35	809.44	137.92	1.90E-02	-2.04
Abomasum (B)	395.65	54.54	708.77	73.97	1.79E-02	-2.1
Colon (R)	431.18	74.38	832.06	111.28	1.79E-02	-2.26
Pigeon egg white	882.33	392.1	2005.07	499.05	2.03E-02	-2.43
MUC5AC (E12)	210.52	18.72	495.05	207.69	2.11E-02	-2.6

[1]Abbreviations as detailed in Figure 4.20

[2]Scale-normalised RFU values from three microarrays for *L. lactis* clones expressing *Pil*2 at T_3, with standard deviation of the mean (sd)

[3]Scale-normalised RFU values from three microarrays for *L. lactis* clones expressing *Pil*3 at T_3, with standard deviation of the mean (sd)

[4]*P* values of significant glycoproteins following *T* test ($P<0.05$ with FDR5%)

[5]Binding fold change (*Pil2*/*Pil3*)

4.5.4 Discussion

Prototypical sortase-dependent pili in Gram positive bacteria such as the type 1 and 2 fimbriae described for *Actinomyces oris* are made of two structural proteins FimP/FimQ and FimA/FimB respectively. FimP and FimA constitute the major pilin subunits and FimQ and FimB are the tip adhesins (Mishra, Das et al. 2007; Proft and Baker 2009). Due to the fastidious nature of *Bifidobacterium bifidum* PRL2010 in culture, two Spa gene clusters, *Pil*2 and *Pil*3, coding each for two pilin subunits respectively FimA/FimB and FimP/FimQ, as well as their associated sortase were cloned in a lactococcal vector and transformed into non-piliated *L. lactis*. Cells successfully produced pili after nisin induction as previously reported for similar clones (Turroni, Serafini et al. 2013). The surface appendages were visualised by TEM in this study and the presence of a multimeric structure supported by the high molecular weight 'ladder' effect seen on immunoblotting using an anti-pilin antibody (anti-FimA). Turroni et al. reported previously that *B. bifidum* PRL2010 adhere to intestinal cells (Turroni, Serafini et al. 2013). However, human glycan receptors have not been described yet for this species and for the genus of *Bifidobacterium* in general.

Following comparison between induced and non-induced cells, significant binders were identified for both pili. Differential binding was observed between the two Spa pili on two glyco-platforms, with *Pil3* binding to a broader range of ligands both on NGC and mucin arrays. Similar findings were reported by adhesion assay using ECM proteins, suggesting that *Pil*3 binds to a wider spectrum of glycan ligands (Turroni, Serafini et al. 2013).

A sandwich assay consisting of NGC microarrays incubated with *Pil*2 extracts, anti-FimA antibody and labelled anti-IgG revealed a tropism for mannosyl structures with strongest interactions with invertase. Yeast invertase contains 14 glycosylation sites and half of its mass consists of polymannans composed of 3 - 52 mannosyl residues, mainly α-(1→2/3/6)-linked (Zeng and Biemann 1999). Mannosyl residues could thus constitute the glycan ligand for the putative FimB adhesin from the *Pil2* cluster. Interestingly, a significant decrease (50 %) in adhesion was reported for *B. bifidum* PRL2010 cells when competed with mannose, suggesting again that it is a ligand of the bacteria (Serafini, Strati et al. 2013). *L. lactis* piliated clones also interacted strongly with ECM proteins such as laminin that contains oligomannans (Knibbs, Perini et al. 1989; Turroni, Serafini et al. 2013). Mannosyl residues are known microbial ligands of enterobacteria as described earlier for the type-1 fimbriae expressed by *E. coli.* A binding specificity for yeast oligomannosides by *Lactobacillus johnsonii* La1 has been reported, suggesting competition with enteropathogens such as *Salmonellae* for the same substrate (Neeser, Granato et al. 2000). In addition, 2 and 3% of the glycans present in the human GIT are respectively high mannose and hybrid *N*-linked structures according to the *in silico* analysis of CFG databases (Appendix).

Together with mannose, fucose has been previously described as the most abundant residue on glyco-conjugates from human intestinal epithelial Caco-2 cells (Youakim and Herscovics 1985). Fucose seems to be prominent on glycan ligands of *Pil*3 fimbriae since several bound glycans contained Lewis antigens (with some including sulfated structures). The best binder for *Pil*3 was the Le^b hexasaccharide Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc, an oligosaccharide found naturally in human milk, and with similar motifs found on mucins including MUC5AC (Van de Bovenkamp, Mahdavi et al. 2003; Bergstrom and Xia 2013). Le^b is also the major receptor for *Helicobacter pylori,* a pathogen which is rampant in the gastric mucosa (Van de Bovenkamp, Mahdavi et al. 2003). The secretor positive status is also associated strongly

with a higher colonisation of the human GIT by bifidobacterial strains including *B. bifidum* (Wacklin, Makivuokko et al. 2011).

In summary, these results suggest that *B. bifidum* PRL2010 expresses different pili to ensure adherence to different ecological niches. This might be an adaptation to stressful conditions within the human GIT, where the bacteria detect a pH change along the tract. *Pil*3 could thus be produced when *B. bifidum* senses an acidic environment causing the bacteria to aggregate and survive in the upper GIT. Several studies have reported the involvement of Gram positive pili in biofilm formation and aggregation such as in streptococci pathogens (Kline, Falker et al. 2009; Manetti, Köller et al. 2010; Becherelli, Manetti et al. 2012). Pili from *S. pneumoniae* contribute to cell aggregation when in saliva prior to colonisation of epithelial cells in the lungs (Barocchi, Ries et al. 2006). Correlation between auto-aggregation and colonisation of the human gut has likewise been shown in lactobacilli and bifidobacteria including *B. bifidum* (Del Re, Sgorbati et al. 2000; Collado, Meriluoto et al. 2008; Vlkova, Rada et al. 2008). Below pH 4.3, *B. bifidum* MIMBb75 displayed very strong auto-aggregation, inversely correlated with adhesion to Caco-2 cells (Guglielmetti, Tamagnini et al. 2009). Moreover, transformed *L. lactis* cells that express *Pil*2 and *Pil*3 were important contributors of bacterial autoaggregation (Turroni, Serafini et al. 2013) and these findings were supported by our TEM photographs. Environmental factors in the lower GIT may then trigger the expression of *Pil*2 by *B. bifidum* PRL2010 and help in retaining the bacteria and colonising its ecological niche such as the cecum. Recent findings in mice suggest the cecum to be the predominant niche for *B. bifidum* MIMBb75 (Singh, Arioli et al. 2013).

A spatio-temporal differential expression of pili could also be hypothesised in *B. bifidum* PRL2010. As well as along the tract, the pattern of glycosylation of GIT mucins indeed changes overtime under the influence of commensal microbes (Bergstrom and Xia 2013). In early life (foetal), sulfomucins predominate in the mucosal barrier and seem to exert a protective role (Deplancke and Gaskins 2001). Sialylation of *O*-glycans appear later after birth (Meslin, Fontaine et al. 1999) and an increase in sialomucins is also reported in mammals down the tract (Kilcoyne, Gerlach et al. 2012). *B. bifidum* PRL2010 colonise the new-born infant and is one of the strains reported to be monophyletic between the mother's and the infant's gut (Makino, Kushiro et al. 2013). Colonisation could thus start with the expression of *Pil3* that showed strong binding to sulfated ligands (3SuLe^a and 6SuLe^x) and exert immunogenic stimulations towards the host as previously reported for *Pil*3 (Turroni,

Serafini et al. 2013). Like other commensal strains, *B. bifidum* PRL2010 is able to thrive on host glycans including human milk oligosaccharides and mucins (Turroni, Bottacini et al. 2010). The action of the vast repertoire of glycosyl hydrolases could contribute to the modification of the mucosa in the infant's GIT and the establishment in the cecum would be secured by the expression of *Pil*2.

Chapter 5: Commensal exopolysaccharides

Key points:

- A lectin microarray platform was used to profile cell surface glycosylation of selected commensal bacteria
- This method was validated on a strain producing an exopolysaccharide (EPS) of known structure (*Lactobacillus rhamnosus* GG)
- The lectin binding signatures were unique for all species tested
- The lectin platform was able to discriminate isogenic clones from *Bifidobacterium breve* UCC 2003 producing or not EPS
- A novel method was used to label EPS for incubation on microarrays
- The lectin platform was able to differentiate commensal strains and their EPSs and can thus be used as a rapid screening tool

5.1 Introduction

The concept of communication between host and microbiota has developed following reports that bacteria-derived molecules, such as polysaccharide A in the commensal species, *Bacteroides fragilis*, were shown to be essential for maturation of the mouse immune system (Mazmanian, Round et al. 2008). As a result, commensal bacteria of the gastro-intestinal tract (GIT) have been the subject of increased research interest, with particular focus on their mode of adherence to the gut and the role of effector molecules they produce. Key players in these functions are the surface polysaccharides, which display extreme variation in terms of composition, linkage conformation, branching and the array of potential substituted groups they may carry. Surface polysaccharides present on both Gram positive and Gram negative bacteria, are often referred to as capsular polysaccharides (CPSs) or exopolysaccharides (EPSs). CPSs form a thick shell called the capsule or pellicle, permanently covering the bacterial cell whereas EPSs are secreted or loosely attached polymers bound to the cell envelop by electrostatic interactions. EPSs give the slimy coat found in some species and constitute the cement in biofilms (Christensen 1989; Laws, Gu et al. 2001; Sutherland 2001; Lee, Tomita et al.). Their non-covalent attachment makes them easily removable from the bacterial cell surface by chelating agents or, if secreted, by simple recovery from the bacterial growth media. It has been suggested that 17 % of human commensal strains from the genera *Lactobacillus* and *Bifidobacterium* are EPSs producers (Ruas-Madiedo, Moreno et al. 2007).

Several functions have been attributed to EPSs (Chapter 1, section 1.3.3). Some commensal EPSs are known to provide health benefits to their mammalian host. It has been reported, for example, that EPSs have a prebiotic function promoting the growth of other beneficial strains from the microbiota (Bello, Walter et al. 2001). Furthermore, roles for EPSs in lowering blood cholesterol levels, protecting against ulceration and anti-tumorigenic effects on malignant cells have been described. EPS have also been shown to affect the immune system of the host by promoting the production of immuno-active molecules such as cytokines, the proliferation of lymphocytes, and the activation of macrophages (Lebeer, Ceuppens et al. 2007; Lebeer, Vanderleyden et al. 2008; Ruas-Madiedo, Salazar et al. 2010; Şengül, Işık et al. 2011).

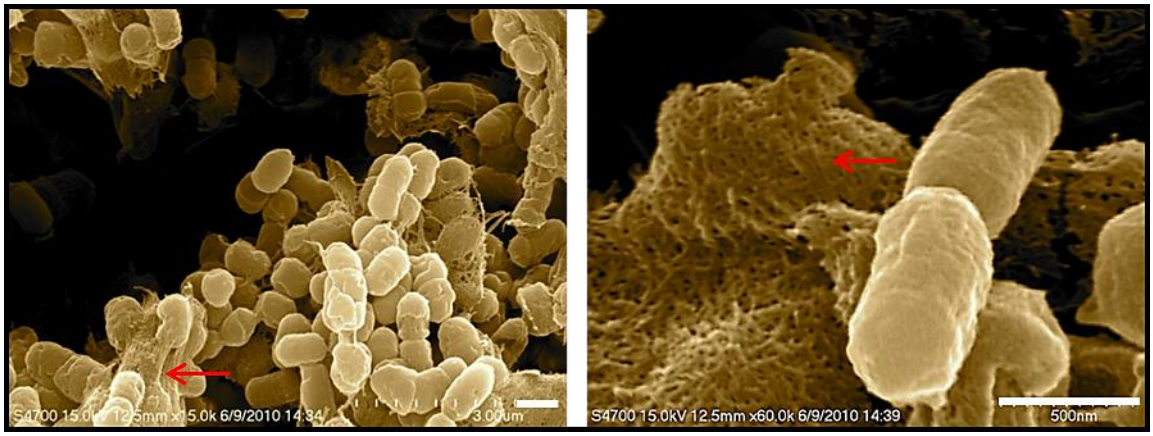

Figure 5.1: *Bifidobacterium longum* subsp. *infantis* ATCC15697
SEM photographs of *B. infantis* cells forming a biofilm (red arrows).
White scale bars represent 500 nm.

Despite the many functions attributed to EPS, little is known about the structure-function relationship of these molecules. This is primarily because of the physico-chemical complexity of these polymers. It is known that EPSs are usually organised in defined repeating units of the same few components and that the genes involved in polysaccharide synthesis are mostly found clustered together in genomes and are often located in plasmids (Sutherland 1994; Landersjo, Yang et al. 2002; Lebeer, Verhoeven et al. 2009). However, very few commensal EPS structures have been characterised as structural analysis is a lengthy and intricate process. The issue of EPSs structure and identification is complicated by the variety of procedures that have been reported in the literature for EPS isolation. Extraction can be accomplished chemically or physically from the cells, from the media or from the whole culture. In addition, several types of media and carbon source have been used to grow the micro-organisms that can influence EPSs composition (Ruas-Madiedo and de los Reyes-Gavilán 2005).

The putative EPS operon encoding a long galactose-rich EPS of the commensal strain *Lactobacillus rhamnosus* GG ATCC53103 (LGG) has been published (Lebeer, Verhoeven et al. 2009). It consists of a priming glycosyltransferase (pGT, termed WelE), which transfers a nucleotide sugar to a lipid carrier such as undecaprenyl phosphate, catalysing the first step of polysaccharide biosynthesis (see Chapter 1, Figure 1.7). This glycosyltransferase should therefore be present in all EPS clusters. Moreover, the operon features five other glycosyltransferases (GT, WelF to WelJ) that add monosaccharides to form the six-residue-EPS repetitive unit, which is translocated by a flippase (wzx) to the peptidoglycan layer. The EPS is then elongated by a polymerase (wzy) dependent on a

chain-length determinant (wzz). Various genes associated with EPS biosynthesis are present within the cluster, including phosphatases and regulatory elements. Another important aspect is the presence of transposases allowing acquisition of the EPS operon by horizontal gene transfer from strain to strain.

A bi-directional EPS operon of *Bifidobacterium breve* UCC2003 has also been proposed (Fanning, Hall et al. 2012). *In silico* analysis performed by our collaborators on *B. breve* UCC2003 genome lead to the identification of an EPS operon (termed 'eps') with an unusual *modus operandi*. This strain is able to produce two types of EPS, the synthesis of which is controlled by either half of a bidirectional gene cluster. Thus, alternative EPS products are possible by means of reorienting the promoter (Chapter 2, Figure 2.3). Various EPS strains, derived from *B. breve* UCC2003 wild type (WT) including the alternate EPS producer (INV) as well as a fortuitous non-EPS producer (DEL) were provided by Pr Douwe Van Sinderen, Alimentary Pharmabiotic Centre, UCC, Ireland (listed in Chapter 2, Table 2.2). Insertion mutants were also engineered to disrupt the transcription of genes encoding glycosyltransferases (such as Bbr_0440 and Bbr_0441 and the pGT Bbr_0430).

Glycosylation-related high throughput technologies that have emerged in recent years have been used as rapid glyco-profiling tools providing an alternative to detailed structural analysis (Hsu, Pilobello et al. 2006; Yasuda, Tateno et al. 2011). We present here the application of a microarray platform, consisting of specific glycan-binding proteins (lectins), for the generation of glycan profiles to compare bacterial strains and to obtain structural insights into their surface polysaccharides. Starting with the well-characterised LGG strain, we compared lectin-binding signatures of whole cells and their corresponding EPS, isolated by two different procedures, from a panel of commensal genera and species with potential probiotic activities and known to produce EPS, *Faecalibacterium prausnitzii, Lactobacillus johnsonii*, *Lactobacillus paracasei* subsp. *paracasei* (abbreviated from here on as *L. paracasei*) and *Bifidobacterium longum* subsp. *infantis* (abbreviated as *B. infantis*). Glycosignatures from *B. breve* UCC2003 isogenic mutants were also compared along with the EPSs produced by the WT and the INV strain. Lectin microarray binding data were validated by inhibition experiments with simple sugars and consolidated by EPS monosaccharide analysis.

5.2 Labelling and EPS isolation optimisations

5.2.1 EPS isolation quality control

The yields of EPS isolated from six commensal species covering three typical genera from the GIT are presented in Table 5.1. EPS amounts recovered from the cells using the chelation-based protocol (bEPS), as well as from ultrafiltration of the growth media (rEPS), differed significantly between strains (38 to 301 mg for bEPS; 106 to 541 for rEPS). The MRS media used to grow the Gram positive strains was processed as the rEPS samples.

Table 5.1: Commensal EPS yields

Sample	Wet biomass (mg)	EPS (mg)		[1]Yield (mg/L)		[2]Yield (%)		[3]Conversion yield	
		bEPS	rEPS	bEPS	rEPS	bEPS	rEPS	bEPS	rEPS
LGG[4]	14400	301	280	150	140	2.09	1.94	7.53	7.00
Lpp	14900	90	488	45	244	0.60	3.28	2.25	12.20
Ljns	5150	38	106	19	53	0.74	2.06	0.95	2.70
Bbr	12910	73	119	36	59	0.57	0.92	1.83	2.98
Bloni	14980	87	541	43	270	0.58	3.61	2.17	13.54
Fpr	3780	55	131	27	65	1.46	3.47	27.50	65.67
MRS	n/a	n/a	n/a	n/a	981	n/a	n/a	n/a	n/a

[1] Yield expressed as mg per litre of broth.
[2] Yield expressed as percentage of wet biomass (bacterial pellet weight).
[3] Glucose conversion yield (mg EPS/g of glucose in the media, with 5 and 20 g/L of glucose in WC and MRS media respectively).
[4] Abbreviations are as follows: *Lactobacillus rhamnosus* GG (LGG)*; L. paracasei* subsp. *paracasei* (Lpp); *L. johnsonii* (Ljns)*; Bifidobacterium breve* UCC2003 (Bbr); *B. longum* subsp. *infantis* (Bloni); *F. prausnitzii* (Fpr)*;* prefix b: (bound) chelation-based procedure; prefix r: (released) ultrafiltration-based procedure; n/a: not applicable.

In order to check the purity of the exopolysaccharides, approximately 100 µg of EPS originating from the two isolation procedures and from three strains were loaded on low percentage gels (3-8% SDS-PAGE). Four different stains were used to detect the presence of glycans or contaminants (proteins) (Figure 5.2). All proteins should have been removed following TCA precipitation and extensive proteinase K digests (Chapter 2, section 2.6). Silver-staining detects proteins but can also stain glycans as seen in the dark smearing

observed (Figure 5.2A). On the contrary, when Coomassie blue dye was used with the samples, no bands were observed (Figure 5.2B), suggesting no contamination was carried-over from the media or the cells following the EPS isolation procedures. The presence of carbohydrates as detected by two specific dyes, Alcian blue that stains charged polysaccharides (Figure 5.2C) and Periodic Acid Schiff (PAS), which colours neutral sugars in pink (Figure 5.2D).

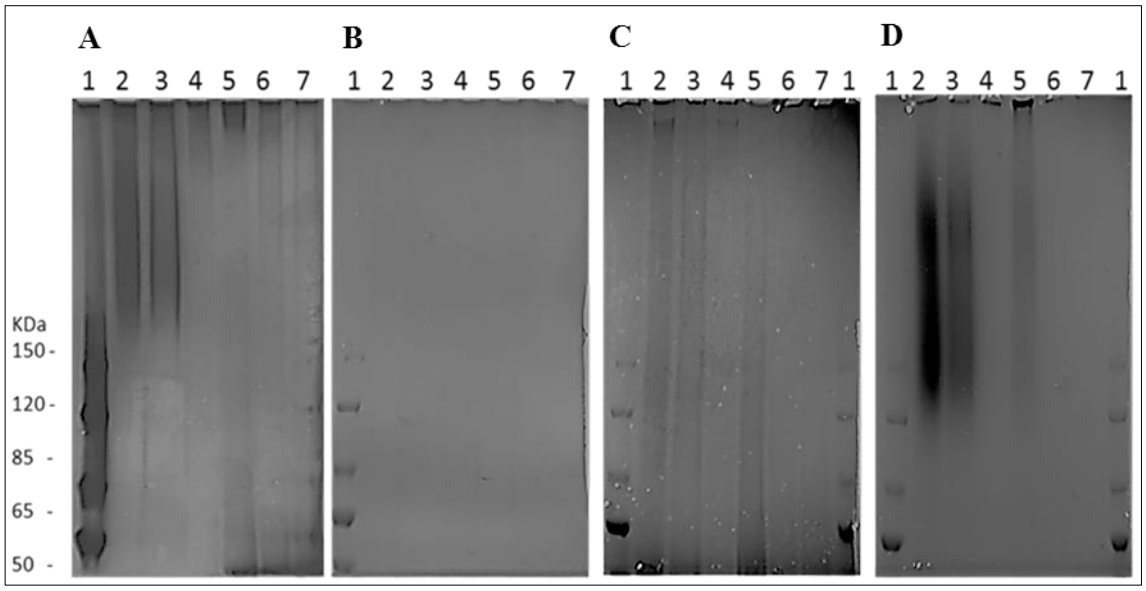

Figure 5.2: Quality control of EPS isolation

SDS-polyacrylamide gels (3-8%) electrophoresed in Tris-Acetate buffer and loaded with 100 µg of EPS from three commensal strains with lane 1: Fermentas PageRuler ladder; 2: bEPS from Bloni; 3: rEPS from Bloni 4: bEPS from Lpp; 5: rEPS from Lpp; 6: bEPS from Fpr; 7: rEPS from Fpr.
A: Silver-staining; B: Coomassie staining; C: Alcian blue staining; D: Periodic Acid Schiff staining. For abbreviations, refer to Table 5.1

5.2.3 Removal of free dye and collection of labelled EPS

EPS samples were labelled using Alexa Fluor® AF555 hydrazide molecule, an aldehyde-fixable dye, and purified post-labelling using gel filtration. Absorbance at 555nm was measured for each collected fraction to detect the presence of Alexa Fluor® AF555 dye. Monsigny assays (A_{450} nm) were also carried out to detect the presence of carbohydrates. Positive fractions (i.e. a potential AF555-EPS glyco-conjugate) were those where A_{450} nm and A_{555} nm peaks coincided. As seen in Figure 5.3, fraction 'Ljr5' consisting of rEPS from *Lactobacillus johnsonii* and 'Ljb5' corresponding to the bEPS from the same strain were both considered positives and subsequently used for the microarray analysis. They

respectively contained 1.24 mg/mL and 1.27 mg/mL sugars according to the standard curves.

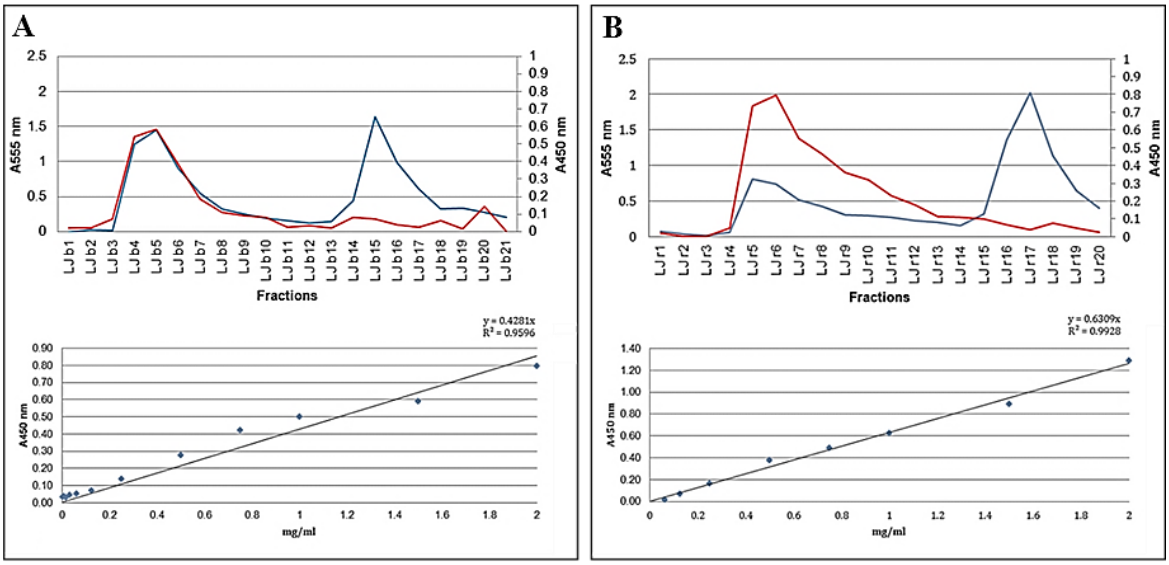

Figure 5.3: Purification of labelled EPS from free dye

Alexa Fluor® AF555 dye (A_{555} nm, blue curve) and total carbohydrate content (A_{450} nm, red curve) following Monsigny assay (top) of gel-purified EPS fractions from *Lactobacillus johnsonii*. Standard curve of unlabelled activated EPS from 0 to 2 mg/mL (A_{450} nm, bottom)
A: bEPS from *Lactobacillus johnsonii*; B: rEPS from *Lactobacillus johnsonii*

5.2.4 Glycoprotein controls on lectin microarrays

Glycoproteins of easy availability and known structure, fetuin and asialofetuin, were used as positive controls for slide quality and slide to slide comparison. The glycoproteins were routinely labelled with Alexa Fluor® AF555 dye, stored in aliquots at -20˚C, diluted to 1 µg/mL in TBSm-T and incubated on each slide. They produced two specific glyco-signatures. Fetuin is a complex glycoprotein found in blood serum that is widely used in glycoscience (Green, Adelt et al. 1988; Royle, Mattu et al. 2002; Iskratsch, Braun et al. 2009). The composition of bovine-sourced fetuin contains sialylated *N*-linked complex oligosaccharides and sialylated *O*-linked glycans (Figure 5.4).

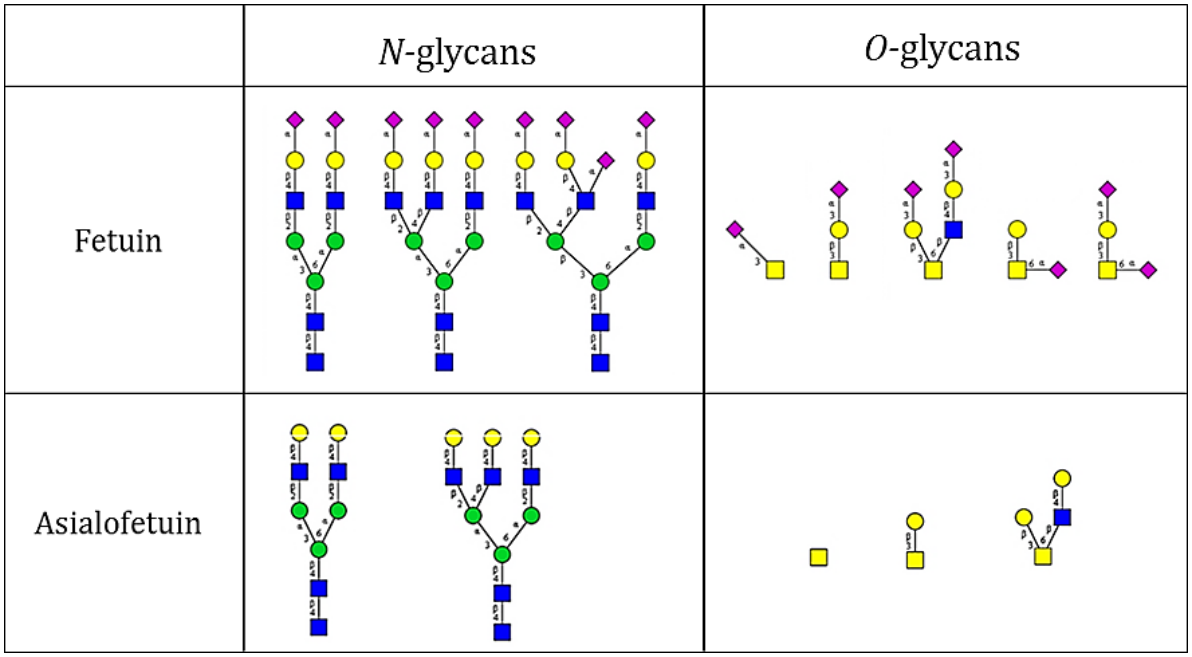

Figure 5.4: *O*- and *N*-glycosylation of fetuin and asialofetuin

Schematic using IUPAC glycan colour codes (Chapter1, Figure 1.3)

As expected, fetuin interacted with several lectins on the arrays (Figure 5.5). Primarily, labelled-fetuin bound strongly to lectins showing affinities for sialic-acid such as WGA, SNA-I and MAA. Binding to ABL, specific for Gal-β-(1→3)-GalNAc (i.e. Thomsen-Friedenreich antigen or T-antigen) and GlcNAc residues was also strong. ACA, another T-antigen-specific lectin that can also recognise sialylated galactosyl structures, presented some interactions with fetuin. In addition, fetuin bound to lectins specific for complex bi- and tri-antennary glycopeptides containing the structure Gal-β-(1→4)-GlcNAc such as RCA-I, PHA-E, PHA-L, CAA and PCA. Galactose binders AIA and MOA also displayed a strong affinity with fetuin. Finally, a last group of lectins binding mannosyl moieties including ConA, NPA, GNA and HHA show some interactions with the glycoprotein.

Asialofetuin corresponds to a remodelling of fetuin by the action of neuraminidase that removes sialic acid residues. The main difference with fetuin in terms of lectin profiles reside, firstly, in the strong interaction with SNA-II, which has affinity for galactose and GalNAc (Figure 5.6). Moreover, the *O*-glycans might be responsible for the binding to ABL, PNA and WFA lectins that recognise GalNAc moieties. In contrast to fetuin, asialofetuin does not bind the sialic acid-specific agglutinins MAA nor WGA on the arrays. However a strong interaction is still visible with SNA-I, the ligand of which is also sialic acid (attached to terminal galactose in α-(2→6) linkage). Complete digestion by

neuraminidase should result in bi- and tri-antennary complex *N*-glycans with terminal galactose residues and three *O*-linked glycans. This suggests incomplete desialylation with residual sialic acid (α-(2→6)-Neu5Ac) on asialofetuin but some interaction of SNA-I with galactose could also be possible. The reproducibility of the lectin microarray platform was assessed by looking at the coefficient of variation (% CV) to calculate intra-slide and inter-slide deviation. Averages for the % CV were 22 % for asialofetuin subarrays and 29 % for fetuin subarrays over 20 slides originating from two print batches.

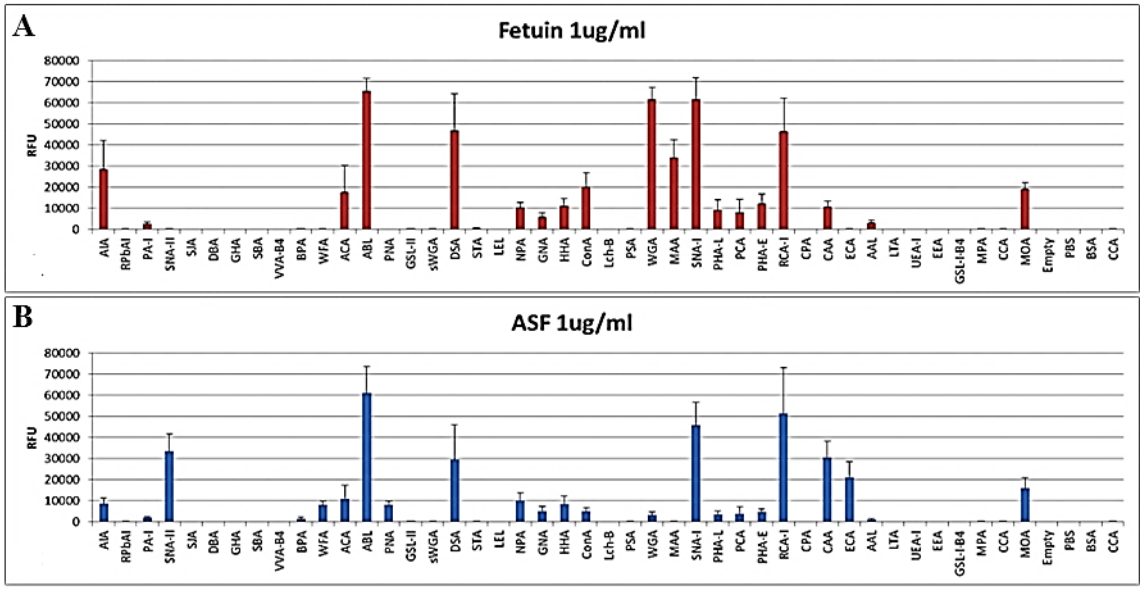

Figure 5.5: Lectin-binding profiles of control glycoproteins

Fetuin (A) and asialofetuin (B) signatures on lectin microarrays. Abbreviations for lectins are detailed in Table 2.13

5.3 Comparison of whole-cell and EPS lectin signatures

5.3.1 Proof of concept: LGG

LGG cells produce an EPS (Figure 5.6) for which a structure has been described: it is a long galactose-rich hetero-polymer with a repeating unit of six sugars - four galactose (Gal), *N*-acetyl-D-glucosamine (GlcNAc) and rhamnose (Rha) (Landersjo, Yang et al. 2002). The lectin-binding profiles of fluorescently-labelled LGG cells were examined together with labelled EPS, isolated from LGG cultures, to prove the applicability of the lectin microarray in studying bacterial glycans. In line with its reported EPS structure, LGG cells showed strong binding to the galactose-binding lectins AIA, RPbAI, SNA-II, MPA, VRA and MOA (Figure 5.7 A). Interactions were also observed with lectins having affinity for GlcNAc including DSA, GSL-II, LEL, sWGA and WGA, consistent with the reported EPS structure. No rhamnose-specific lectin was printed on the microarray. Binding was also noted to several lectins not directly associated with the known EPS, including ACA lectin, specific for the T-antigen (Gal-β-(1→3)-GalNAc), WFA and VVA-B4 (GalNAc-specific lectins) and some mannose-specific lectins (PSA, Lch-B, HHA, GNA and NPA). These unexpected binding events suggest interactions with other molecules decorating the cell wall. Either other polysaccharides are present, or certain lectins display broader sugar specificity on the microarray platform than currently described.

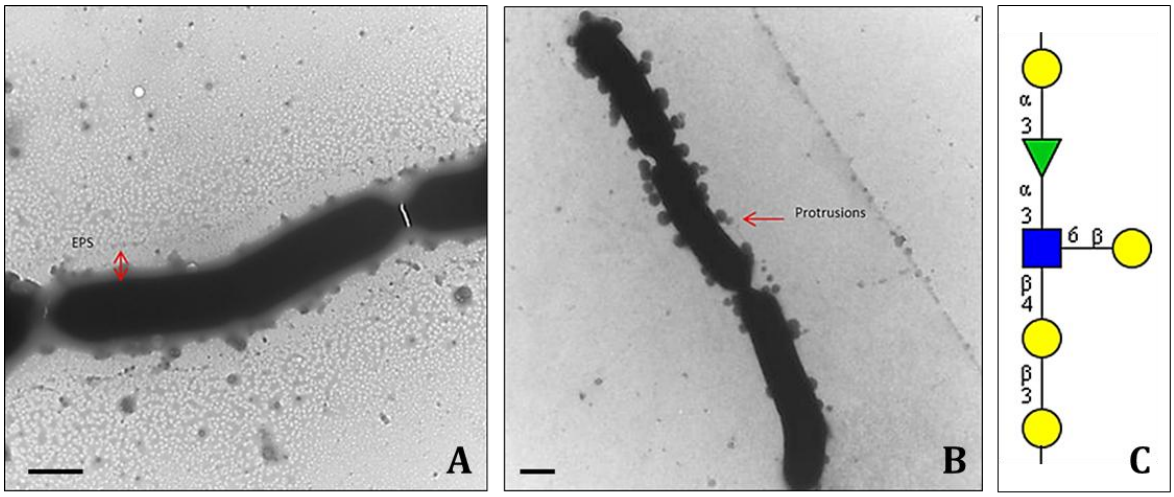

Figure 5.6: *Lactobacillus rhamnosus* GG whole cells and described EPS

EPS slime (A) and protrusions (B) visualised on 1% PTA-stained LGG bacteria by TEM with scale bar corresponding to 500nm. EPS known structure as described by Landersjö *et al* with yellow circle – galactose, green triangle - rhamnose and blue square – GlcNAc (C) (Landersjo, Yang et al. 2002)

Two forms of EPS were prepared from LGG cells. One form, referred to as bound EPS (bEPS), was isolated by treatment of intact cells with a chelating agent (EDTA). The second form, termed released EPS (rEPS), was isolated from the culture media by ultrafiltration (>100KDa). Both preparations were checked for purity on SDS-PAGE gels with coomassie staining and were shown to be protein-free. EPS samples were labelled and incubated on the microarray to compare their lectin signatures both with each other and with the profiles of whole cells (Fig. 5.7).

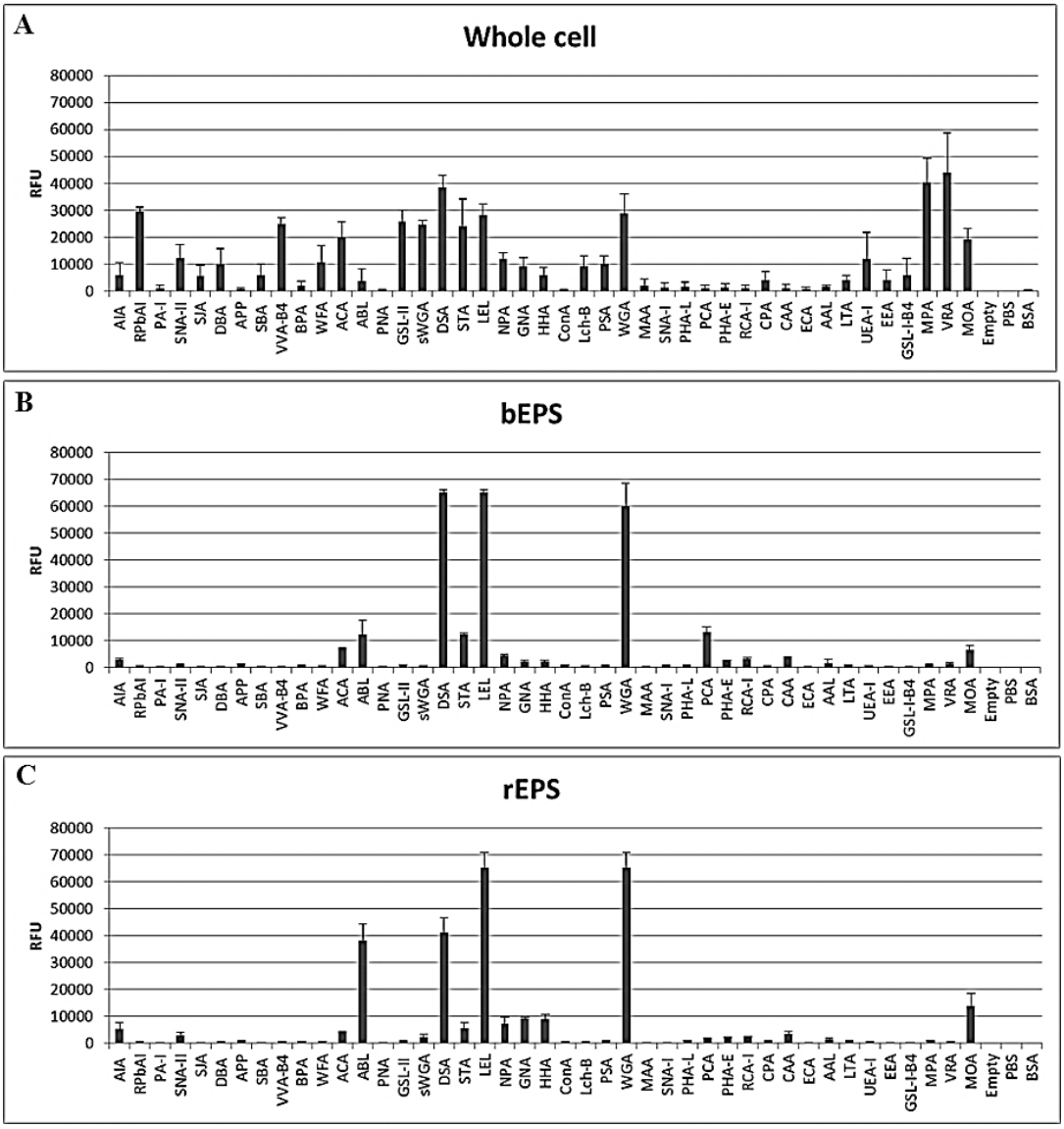

Figure 5.7: Lectin profiling of *Lactobacillus rhamnosus* GG

Normalised signals of LGG whole cells (A), LGG bEPS (B) and LGG rEPS (C) incubated on the lectin microarrays. Error bars correspond to the standard deviation of the mean for six microarrays (A) and three microarrays (B and C). 'Empty', 'PBS' and 'BSA' represent negative control spots. Abbreviations for lectins are detailed in Table 2.13

In general, similar lectins were bound by the two EPS preparations, although the magnitude of the signals varied. High signals were obtained, in both cases, when binding to lectins with affinity for GlcNAc (e.g. DSA, LEL and WGA) and both EPS fractions interacted with galactophilic lectins including AIA, SNA-II and MOA. The latter is specific for the Gal-α-(1→3)-Gal structure present on the LGG described EPS. Interactions were also observed with lectins specific for the T-antigen (Gal-β-(1→3)-GalNAc) such as ACA and ABL. Comparison of profiles given by the two EPS preparations (Table 5.2) showed higher signals for the rEPS binding to ABL (over 2-fold), as well as to mannose-recognising lectins NPA, GNA and HHA (differences ranging from 7- to 13-fold). In contrast, RCA-I, specific for Gal-β-(1→4)-GlcNAc, structure present in LGG EPS, showed significantly higher binding to bEPS compared to rEPS (by almost 10-fold).

Table 5.2: Significant binding differences between LGG EPS fractions

Lectin	[1]bEPS	SD	rEPS	SD	[2]*P*	[3]BF
AIA	6276.0	167.3	2839.0	1112.8	4.28E-03	2.6
MOA	2630.5	249.2	6771.2	2522.0	1.57E-02	-2.2
BPA	994.3	45.9	282.0	10.8	9.56E-07	3.5
ABL	7837.0	385.4	17960.1	3762.6	1.08E-03	-2.1
DSA	144711.7	1611.4	19139.8	3669.7	4.83E-06	7.8
LEL	74082.0	259.8	29109.5	1426.5	3.88E-06	2.5
NPA	435.0	30.8	3012.1	978.8	3.14E-05	-7.4
GNA	317.2	5.5	4125.7	159.6	2.34E-08	-13.2
HHA	365.9	20.5	3880.9	609.9	8.72E-07	-11.2
ConA	1701.2	116.8	348.2	7.4	4.71E-07	5.0
WGA	144746.4	1622.1	29105.6	1426.9	2.93E-07	4.9
PCA	343.4	22.4	798.3	19.8	1.00E-05	-2.3
PHA-E	357.0	16.3	958.6	91.2	1.48E-05	-2.6
RCA-I	10123.9	5932.3	1081.0	91.2	4.64E-04	9.9
CAA	223.9	27.6	1702.2	303.5	9.11E-06	-7.4
AAL	1793.5	130.3	687.1	91.8	5.86E-05	2.6

[1]Mean of scale normalised RFU signals from three microarrays with standard deviation (SD).
[2]Significance (*P*-value) following Student's *t*-test, between chelation-based isolation (bEPS) and the ultrafiltration-based methods (rEPS).
[3]Binding fold difference bEPS / rEPS.

Comparing the binding profiles of LGG cells and both EPS preparations, the cells bound a broader range of lectins than the EPS fractions, although many of the corresponding ligands (Gal, GlcNAc, GalNAc and Man), were common between the three samples.

To confirm sugar specificity LGG cell binding experiments were carried out in the presence of the relevant monosaccharides (100 mM). Binding to most of the lectins that interacted with the bacteria was inhibited by the corresponding ligand indicating that the binding events were sugar-mediated (Fig. 5.8 and Table 5.3). Exceptions were the absence of inhibitions when binding to PA-I, VRA and MPA (using Gal) and to ConA (using Man). Moreover, prior incubation of the LGG cells with their respective bEPS preparation caused most signals to be significantly decreased, supporting the involvement of bEPS in binding of the bacterial cells to the lectin microarray. Similar reductions were noted for the EPS preparations using free sugars (Table 5.4).

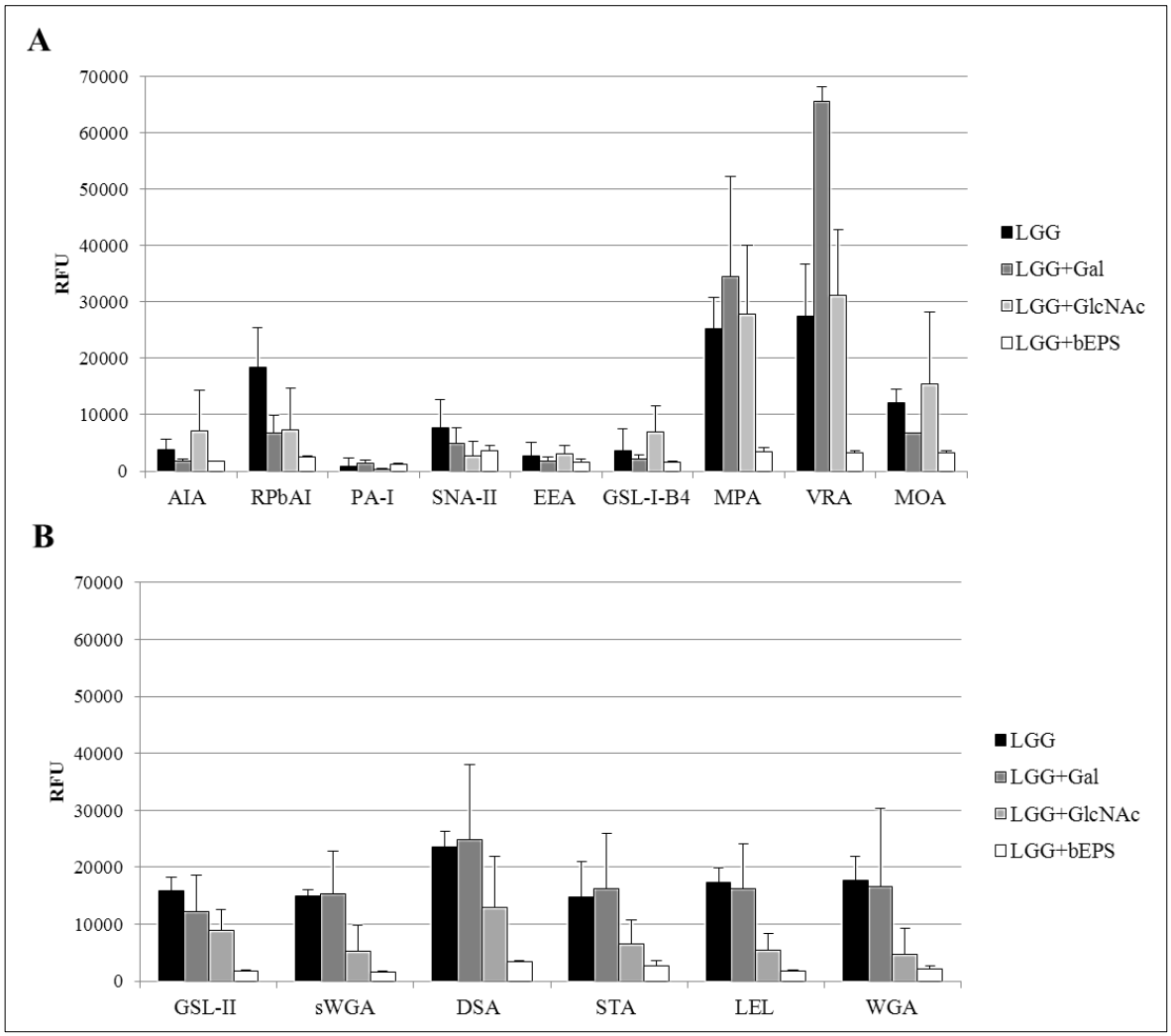

Figure 5.8: Effects of added sugars on the lectin-binding signature of whole LGG cells

Binding to Gal-specific (A) or GlcNAc-specific lectins (B) of LGG cells (black colour) co-incubated with 100 mM Gal (dark grey), 100 mM GlcNAc (light grey) or 1.25 mg/mL LGG bEPS (white). Normalised signals from six microarrays with error bars corresponding to the standard deviation of the mean. Abbreviations for lectins are detailed in Table 2.13

Table 5.3: Inhibition of lectin binding to LGG whole cells by haptens

Lectins	Known ligand	Haptenic sugar	Inhibition (%)	
			[1]Sugar	[2]bEPS
AIA	Gal	Gal	50.10	53.61
RPbAI	Gal	Gal	62.19	86.39
PA-I	Gal	Gal	0	0
SNA-II	Gal	Gal	34.35	52.51
EEA	α-Gal	Gal	35.21	39.67
GSL_I_B4	α-Gal	Gal	42.81	57.06
MPA	α-Gal	Gal	0	86.01
VRA	α-Gal	Gal	0	88.11
MOA	Gal-α-(1→3)-Gal	Gal	48.24	72.02
ACA	Gal-β-(1→3)-GalNAc	Gal	0	40.93
WFA	GalNAc	Gal	38.88	65.99
VVA-B4	GalNAc	Gal	52.29	88.26
DSA	GlcNAc oligomers	GlcNAc	49.8	85.87
GSL-II	GlcNAc oligomers	GlcNAc	49.02	88.52
LEL	GlcNAc-β-(1→4)-GlcNAc	GlcNAc	71.59	89.43
STA	GlcNAc oligomers	GlcNAc	59.19	81.44
sWGA	GlcNAc oligomers	GlcNAc	68.11	89.19
WGA	GlcNAc/Neu5Ac	GlcNAc	75.88	87.90
ConA	Man, Glc, GlcNAc	Man	0	0
GNA	Man-α-(1→3)-	Man	5.08	67.65
HHA	Man-α-(1→3)-Man-α-(1→6)	Man	0	39.03
NPA	α-(1→6)-Man	Man	62.91	83.01
PSA	Man, fucose dependent	Man	0	71.69
Lch-B	Man, fucose dependent	Man	0	91.93

[1]Percentages of signal inhibition on specific lectins by co-incubating LGG cells with 100mM haptenic sugars.

[2]Percentages of signal inhibition on specific lectins by co-incubating LGG cells with 1.25mg/mL bEPS.

Table 5.4: Competition experiments with LGG labelled EPS

Lectins	Known ligand	Sugar	[1]bEPS (%)	[1]rEPS (%)
AIA	Gal	Gal	87.62	70.29
SNA-II	Gal	Gal	86.52	91.39
MOA	Gal-α-(1→3)-Gal	Gal	91.05	95.13
ACA	Gal-β-(1→3)-GalNAc	Gal	0.00	0.00
ABL	Gal-β-(1→3)-GalNAc, GlcNAc	Gal	87.62	70.29
RCA-I	Gal-β-(1→4)-GlcNAc	Gal	59.97	59.23
DSA	GlcNAc	GlcNAc	62.01	34.16
LEL	GlcNAc	GlcNAc	63.02	25.1
sWGA	GlcNAc	GlcNAc	51.46	68.14
WGA	GlcNAc, Neu5Ac	GlcNAc	95.7	95.22
GNA	Man-α-(1→3)-	Man	93.18	79.28
HHA	Man-α-(1→3)-Man-α-(1→6)-	Man	96.11	82.47
NPA	α-(1→6)-Man	Man	86.02	74.17

[1]Percentages of signal inhibition on selected lectins when co-incubating LGG EPS with 100mM haptenic sugars.

Because EPS composition can be affected by strain differences and/or culture conditions, monosaccharide analysis was carried out to confirm the composition of the EPS preparations used in this study (Table 5.5). Gal was found to be the main component of the bEPS fraction and the presence of GlcNAc and rhamnose was also established as expected. Because of the low levels of rhamnose detected by monosaccharide analysis, its presence was also confirmed by ^{1}H-NMR analysis, showing the presence of the L-rhamnopyranose methyl group as a doublet at a chemical shift of 1.3 ppm (Figure 5.9). Gal and GlcNAc were also present in the rEPS fraction although in different proportions but rhamnose was not identified. Consistent with conclusions from analysis of their lectin binding profiles, other sugars including GalNAc and mannose were detected in both preparations of EPS and glucose in the rEPS fraction. Indeed, glucose, mannose and GalNAc were found at considerable levels in the rEPS fraction. This could be explained by contamination of the rEPS sample with polysaccharides from the MRS media or from dead cells.

Table 5.5: Composition of EPS fractions

	[1]Fuc	Rha	GalNAc	GlcNAc	Gal	Glc	Man
[2]**bLGG**	*nd*	**6.2** (0.5)	**43.1** (2.6)	**16.4** (1.0)	**171.5** (10.5)	*nd*	**7.5** (0.5)
rLGG	*nd*	*nd*	**120.0** (5.6)	**21.4** (1.0)	**28.8** (1.3)	**128.3** (6.0)	**116.6** (5.4)
bLjns	*nd*	*nd*	**98.8** (6.9)	**14.4** (1.0)	**52.7** (3.7)	**313.4** (21.8)	*nd*
rLjns	*nd*	*nd*	**136.0** (5.8)	**23.6** (1.0)	**42.3** (1.8)	**248.6** (10.5)	**14.8** (0.6)
bLpp	*nd*	*nd*	**58.7** (14.0)	**4.2** (1.0)	**17.8** (7.6)	**32.1** (8.8)	**37.1** (6.4)
rLpp	*nd*	*nd*	**117.0** (5.0)	**23.5** (1.0)	**38.3** (1.6)	**137.5** (5.9)	**143.9** (6.1)
bBloni	*nd*	*nd*	**10.0** (7.7)	**1.3** (1.0)	**87.9** (67.6)	**101.7** (78.2)	*nd*
rBloni	**6.1** (0.3)	*nd*	**131.7** (5.8)	**22.9** (1.0)	**38.8** (1.7)	**120.8** (5.3)	**735.1** (32.1)
bFpr	**19.4** (0.5)	*nd*	**10.6** (0.3)	**41.6** (1.0)	**56.2** (1.4)	**263.9** (6.3)	**431.7** (10.4)
rFpr	**19.2** (0.5)	*nd*	**11.0** (0.3)	**40.2** (1.0)	**77.7** (1.9)	**229.7** (5.7)	**383.8** (9.5)
[3]**MRS**	*nd*	*nd*	**68.0** (1.7)	**40.0** (1.0)	**97.0** (2.5)	**407.0** (10.2)	**237.0** (6.0)

[1]Concentration of monosaccharides (pmoles/µg, average of three samples) and the molar ratio (in brackets) calculated relative to GlcNAc. Abbreviations are as follows: *nd*, not detected.
[2]Prefix 'b' refers to the chelation-based isolation (bEPS), prefix 'r' to the ultrafiltration-based isolation (rEPS). EPS from *B. longum* subsp. *infantis* (Bloni), *F.prausnitzii* (Fpr), *L. paracasei* subsp. *paracasei* (Lpp), *L. rhamnosus GG* (LGG) and *L. johnsonii* (Ljns).
[3]MRS media control.

A comparison of lectin microarray binding, monosaccharide analysis and LGG bEPS structure is presented in Table 5.6.

Table 5.6: LGG analysis summary

Monosaccharide	[1]Lectin microarray	[2]Monosaccharide	[3]EPS
Gal	Present (6/9)	Present (95.3)	Present (4)
GlcNAc	Present (6/6)	Present (7.4)	Present (1)
GalNAc	Present (2/10)	Present (19.5)	Absent (0)
Man	Present (2/6)	Present (4.2)	Absent (0)
Rha	n/a (no lectin)	Present (3.8)	Present (1)

[1]Number of lectins bound to potential number on the array for each sugar in a glycan-mediated manner.
[2]Monosaccharide ratio
[3]Published EPS structure from Landersjö et al. (ratio of sugars) (Landersjo, Yang et al. 2002)

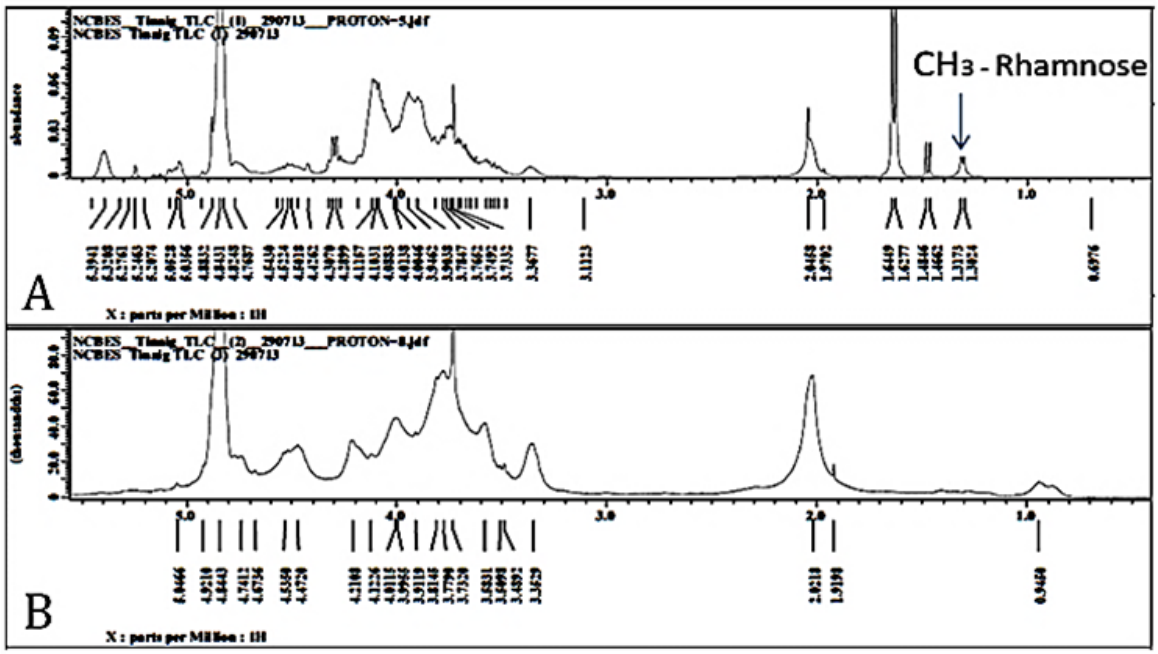

Figure 5.9: ^{1}H-NMR spectra

Chemical shifts of proton detecting the presence of L-rhamnopyranose methyl group as a doublet (1.3 ppm) in LGG bEPS (A, blue arrow) compared to MRS media (B)

5.3.2 Differential profiling of labelled commensal bacterial cells

Having confirmed that useful structural information on bacterial cell surface polysaccharides and EPS could be obtained from analysis of their lectin microarray-binding profiles, a panel of candidate probiotic strains were selected for determination of their cell and EPS lectin-binding profiles. To our knowledge, no structural information has yet been reported on the EPS from any of these strains other than two polysaccharides in a total cell wall extract of a similar *B. infantis* strain (YIT4025), which were shown to be galactans with glucose substitutions (Habu, Nagaoka et al. 1987) (http://csdb.glycoscience.ru/bacterial/; http://crdd.osdd.net/raghava/carbodb/). Due to the multimeric nature of the interactions between lectins and carbohydrates, the density and presentation of the lectins spotted on the glass slides as well as the possible physical hindrance of the ligands, all signals five times higher than the background fluorescence were taken into consideration. Each lectin profile was unique as depicted on the radar plot (Figure 5.10) and statistically significant differences in binding intensities were identified following ANOVA (Table 5.7).

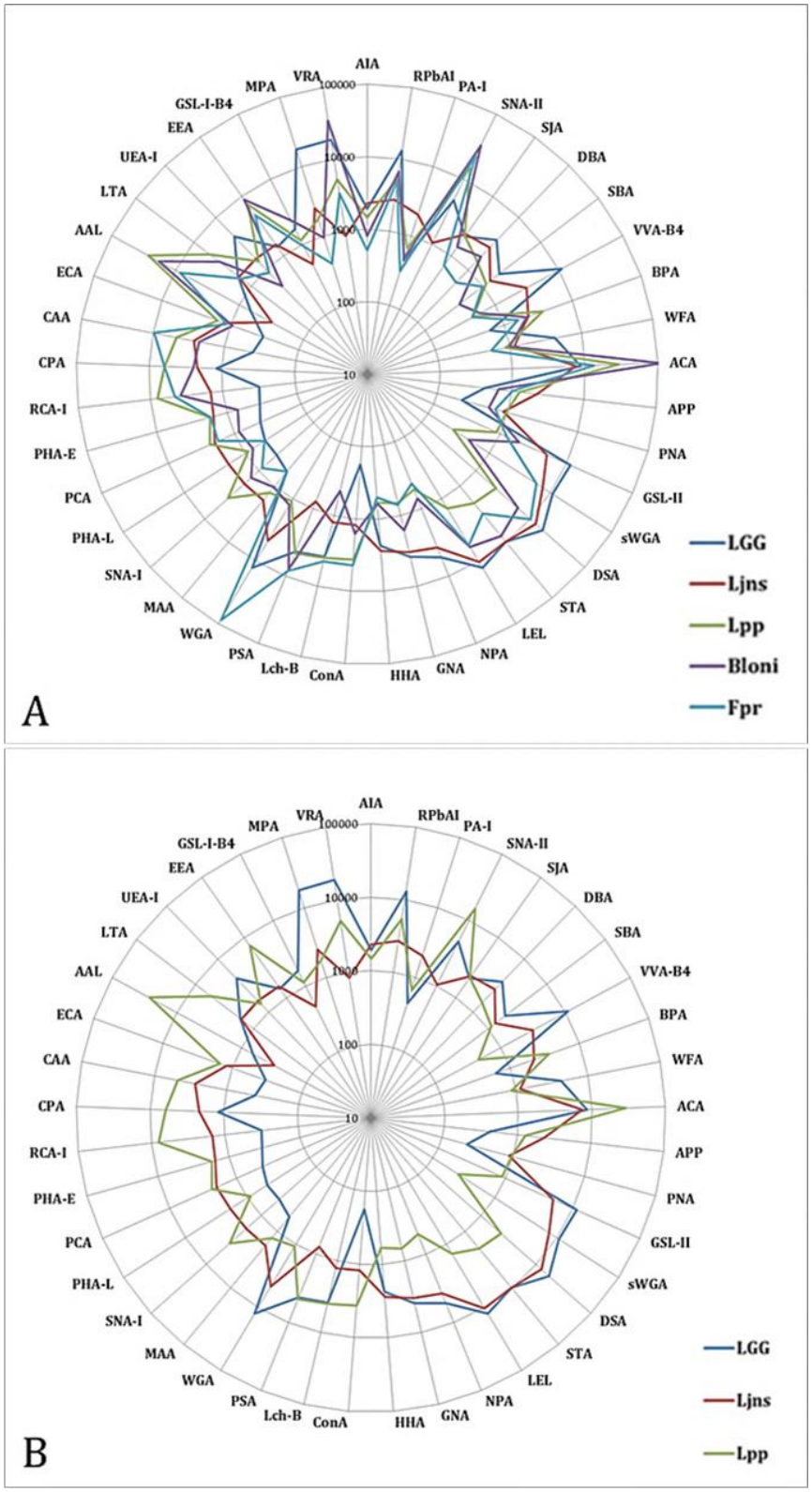

Figure 5.10: Lectin-binding signatures for whole cells of commensal species

Radar plots representing the lectin-binding signatures of five commensal species (A) and focussing on the three lactobacilli (B). Normalised RFU signals from three microarrays. Abbreviations are detailed in Table 5.5 and Table 2.13

Table 5.7: Differential lectin profiling of whole cells

ANOVA	Lpp	Fpr	Ljns	LGG
Bloni	[1]**14**	**15**	**23**	**31**
Lpp		**13**	**24**	**32**
Fpr			**26**	**31**
Ljns				**26**

[1]*Post hoc* test following ANOVA on lectin profiles from five commensal strains showing number of significant lectins ($P<0.05$, FDR 5%). Abbreviations are detailed in Table 5.5

When the profiles for the five species were subjected to hierarchical clustering, two major clusters were obtained (Figure 5.12). The first cluster included three species, *B. infantis*, *L. paracasei* and *F. prausnitzii*, with a subgroup differentiating the butyrate-producing strain (*F. prausnitzii*). The second cluster grouped two of the lactobacilli strains, *L. johnsonii* and LGG. Thus, the lectin-binding profiles of whole bacterial cells did not correlate with the phylogenetic relationships of the species examined.

In the first cluster, common high binders for *F. prausnitzii*, *B. infantis* and *L. paracasei* included ACA, AAL and SNA-II lectins. The most significant difference in binding observed for *F. prausnitzii* and the other two species was for the GlcNAc-specific lectin WGA (90-fold higher than *L. paracasei* and 67-fold higher than *B. infantis*). Binding to sWGA, CAA and CPA, which also have affinity for GlcNAc-containing oligomers, was also higher for *F. prausnitzii* compared to the *Bifidobacterium* and *Lactobacillus* species (24- and 13-fold, respectively, with sWGA).

In the second cluster, both strains showed strong binding to a group of GlcNAc-specific lectins including STA, LEL, DSA and sWGA. However, *L. johnsonii* bound preferentially to lectins with affinity for complex glycans such as ECA, CAA, PHA-E, PHA-L and RCA-I (3- to 7-fold). These glycans would also be expected to contain GlcNAc. In addition, *L. johnsonii* bound more strongly to GalNAc-specific lectins APP, PNA and BPA (3- to 5-fold), to sialic acid-specific agglutinins MAA and SNA-I (>3-fold) and to ConA (~7-fold).

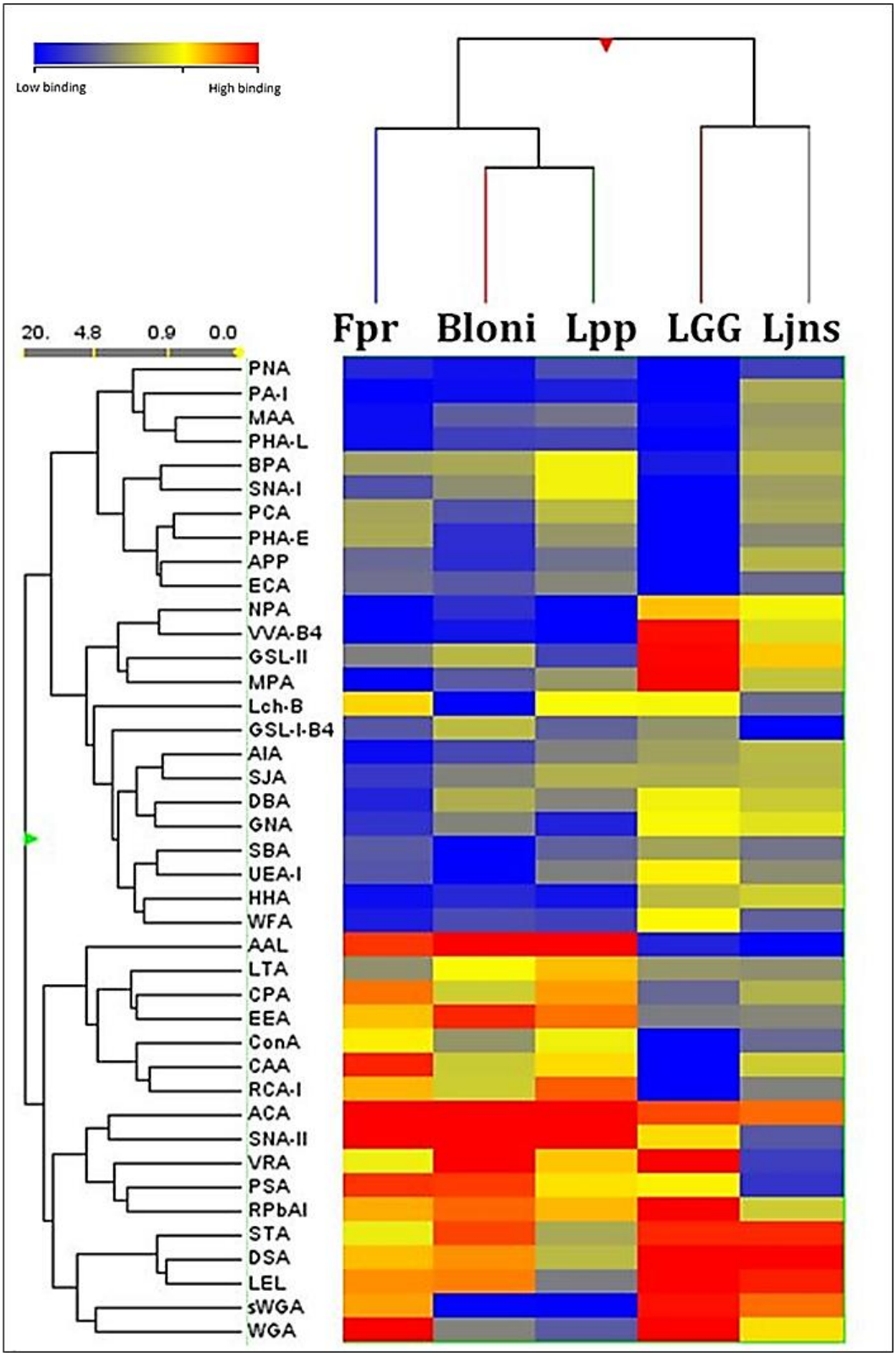

Figure 5.11: Hierarchical clustering of lectin microarray-binding profiles of whole cells

Clustering based on Euclidean distance of normalised lectin binding signals averaged for three microarrays. Abbreviations are detailed in Table 5.5 and Table 2.13

5.3.3 Lectin profiling of commensal exopolysaccharides

EPS were extracted from the five aforementioned commensal bacteria using two isolation methods: chelating agent (bEPS) or ultrafiltration (rEPS) of the growth media. EPS were labelled, incubated on the lectin microarrays in the absence or presence of competing monosaccharides and lectin signatures analysed. Lectin-binding profiles of the rEPS and bEPS preparations were generally simpler than the profiles of the corresponding whole cells with fewer lectins binding. The profiles of each of the bEPS and rEPS preparations examined were unique with binding intensities to at least four, and as many as 25 lectins, significantly different by ANOVA (Tables 5.8 and 5.9). Of particular interest were the large differences noted in the profiles of bEPS and rEPS from some species, suggesting different composition of each preparation and highlighting the importance of standardised procedures for EPS isolation. This was very evident when all ten EPS lectin-binding profiles were subjected to hierarchical clustering (Figure 5.12).

Table 5.8: Differential lectin profiles of EPS preparations between strains

	bEPS				rEPS			
	LGG	**Fpr**	**Bloni**	**Ljns**	**LGG**	**Fpr**	**Bloni**	**Ljns**
Lpp	16	20	17	15	16	16	5	17
LGG		25	19	20		23	14	16
Fpr			15	20			18	21
Bloni				19				21

Significant bound lectins (P<0.05, FDR 5%) following ANOVA between EPS lectin profiles of chelation-based isolation (bEPS) and the ultrafiltration-based isolation (rEPS fraction) from five species. Abbreviations are detailed in Table 5.5

Table 5.9: Differential lectin profiling of EPS preparations within strain

Fpr	**Bloni**	**LGG**	**Ljns**	**Lpp**
8	12	16	4	21

Significant bound lectins (P<0.05, FDR 5%) following *t*-test between the chelation-based isolation (bEPS) and the ultrafiltration-based isolation (rEPS) fractions. Abbreviations are detailed in Table 5.5

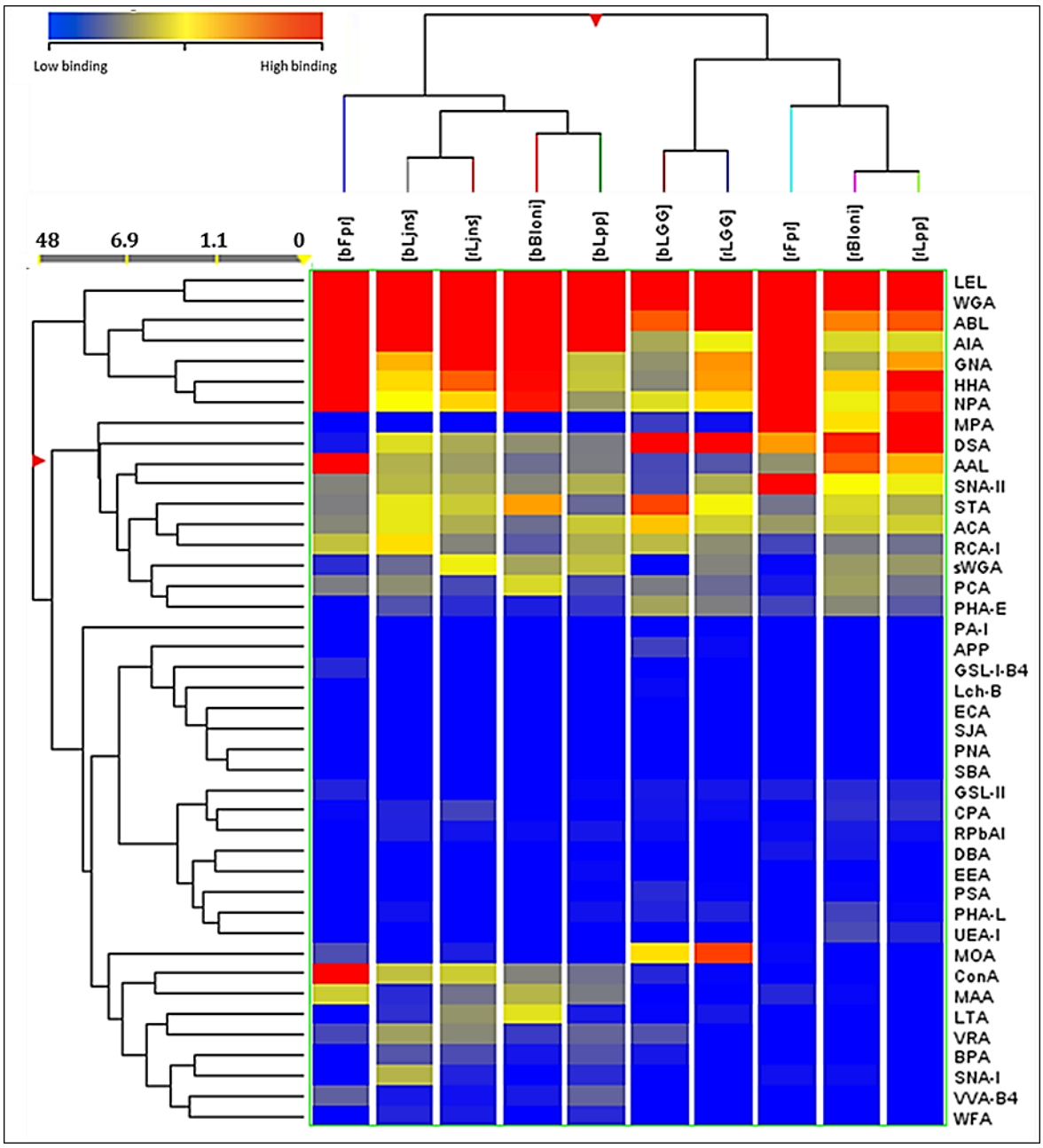

Figure 5.12: Hierarchical clustering of lectin microarray-binding signatures of EPS

Clustering of lectin binding signals from three microarrays of labelled EPS from five commensal strains using Euclidean distances

Prefixes ‘b’ and ‘r’ refer to bound and released EPS fractions respectively. Abbreviations are detailed in Table 5.5 and Table 2.13

Two major clusters were identified, one containing four of the five bEPS preparations and the other four of the five rEPS preparations. Lectin signatures of both preparations from LGG and *L. johnsonii* grouped together, the former with the other rEPS and the latter with the other bEPS preparations. Interestingly, within both the bEPS and rEPS clusters, *B. infantis* and *L. paracasei* grouped separately from *F. prausnitzii*, as did their whole cell lectin binding profiles.

Some lectins showed strong interactions with all EPS preparations, particularly LEL (specific for oligomers of β-(1→4)-linked GlcNAc), WGA (GlcNAc and sialic acid) and ABL (Gal-β-(1→3)-GalNAc, GlcNAc) and binding intensities were reduced in the presence of inhibiting sugar (Tables 5.3 and 5.10). These results suggested that GlcNAc is a component of all isolates.

Table 5.10: Signal inhibition following 100 mM GlcNAc challenge

Strain	Lectin	bEPS	[1]% bEPS	rEPS	[2]% rEPS
LGG	**ABL**	**12292**	78	**38045**	49
	LEL	**64926**	63	**65416**	25
	WGA	**60272**	95	**65409**	95
Fpr	**ABL**	**29067**	27	**45027**	10
	LEL	**71231**	83	**17192**	1
	WGA	**38508**	97	**38515**	94
Bloni	**ABL**	**21727**	0	**4824**	77
	LEL	**20281**	5	**34340**	49
	WGA	**21988**	95	**27753**	94
Ljns	**ABL**	**10438**	8	**14577**	0
	LEL	**62760**	69	**22630**	66
	WGA	**38153**	94	**30179**	95
Lpp	**ABL**	**21959**	47	**6828**	46
	LEL	**17389**	52	**55975**	30
	WGA	**10142**	94	**27411**	70

[1]Percentages of inhibition on specific lectins by co-incubating bEPS with 100mM GlcNAc.
[2]Percentages of inhibition on specific lectins by co-incubating rEPS with 100mM GlcNAc.

Interestingly, binding to DSA, another GlcNAc binder was more variable with strong binding observed to most rEPS preparations and to both LGG preparations, and no binding at all to bEPS from *F. prausnitzii*. This suggests differences in presentation of GlcNAc within the various EPS preparations. Other medium to strong binders to all EPS preparations were the mannose-binding lectins GNA, HHA and NPA. Con A, which also is

a general mannose binder, bound only to the EPS preparations in cluster 1 and with high intensity only to *F. prausnitzii*. The presence of mannose was confirmed by monosaccharide analysis in all but two of the preparations, bEPS from *L. johnsonii* and *B. longum* subsp. *infantis* (Table 5.5). The lectin array also suggested the presence of Gal/GalNAc in most EPS preparations with observed binding to ABL (mentioned above), AIA and SNA-II, although LGG bEPS bound very weakly to the latter two lectins. The rEPS fractions from *B. infantis*, *L. paracasei* and *F. prausnitzii* also bound strongly to the α-Gal-specific lectin MPA; however, except for the binding to MOA (α-Gal-specific) of both EPS preparations of LGG, no binding was observed with any EPS fraction to the other Gal- or GalNAc-binding lectins on the array (Table 5.11). Nevertheless, Gal and GalNAc were detected in all fractions on monosaccharide analysis. Hence, the mode of presentation of these sugars in the EPS molecules must be restricting accessibility to the lectins on the array. The presence of fucose in the bEPS from *F. prausnitzii* and rEPS from *B. longum* subsp. *infantis* was suggested by strong binding to the fucose-specific lectin, AAL. Weak binding was observed with more of the EPS preparations but, on monosaccharide analysis, fucose was only detected in the two strong binders and rEPS from *F. prausnitzii*. Interestingly, rhamnose was not detected in any EPS preparation other than the bEPS from LGG, where it was previously reported. Despite significant differences in the profiles of bEPS and rEPS from each species, no consistent pattern of interaction to specific lectins could be discerned. Dissimilarities between the two EPS preparations were species specific. This suggests that carryover from polysaccharides in the media were not responsible for the differences between bEPS and rEPS.

Since the MRS contained unknown sugars present in the yeast and meat extracts, it too underwent the same isolation procedure to determine its monosaccharide composition. In addition to the obvious high glucose content (probably resulting from hydrolysed glucans and glycogen), the MRS medium contained, in order of decreasing molar ratios, mannose, galactose, GalNAc and GlcNAc, with both fucose and rhamnose absent. The MRS media hydrolysate was compared initially to the hydrolysed rEPS from LGG by HPLC analysis and results showed little contamination of the fraction from the broth (Figure 5.13).

Table 5.11: Significant binding differences between EPS fractions

	Lectin	[1]*P*	[2]BF	[3]bEPS	[3]rEPS	Sugar	[4]% bEPS	[4]% rEPS
[5]Fpr	**ConA**	1.05E-03	71.6	14616.4	204.1	Man	97.3	0
	AAL	9.60E-02	9.7	13149.7	183.6	Fuc	83.3	33.5
	VRA	2.03E-02	5.7	979.7	172.8	Gal	42.5	0
	VVA-B4	1.46E-02	4.5	1203.1	268.2	Gal	80.6	22
	LTA	3.85E-02	3.7	457.6	124.6	Fuc	50.4	0
	DSA	1.46E-02	-8.7	639.6	5541.6	GlcNAc	0	53.4
	SNA-II	1.46E-02	-13.2	1599.5	21101.2	Gal	67.2	95.3
	MPA	5.48E-02	-153.7	171.5	26363.5	Gal	18.1	51.9
Bloni	**AAL**	3.45E-03	-5.2	1358.6	7028.2	Fuc	0	0
	AIA	3.74E-03	15.6	41165.1	2631.7	Gal	69.7	83.1
	ConA	4.71E-03	4.6	1617.5	350.7	Man	40.4	71.7
	DSA	6.58E-03	-5	1781.9	8939.4	GlcNAc	2.5	54.2
	GNA	1.77E-03	7.6	13802.4	1823.2	Man	21.3	77.2
	HHA	6.22E-03	2.7	12246.6	4462.7	Man	38.7	90.9
	LTA	2.49E-02	20.8	3505.8	168.7	Fuc	93.6	34.1
	MAA	3.74E-03	5.1	2428.3	474.1	na	na	na
	MPA	2.88E-03	-9.9	416.4	4102.8	Gal	40.8	81.1
	NPA	3.74E-03	3.7	11587.4	3169.2	Man	29.9	87.2
	SNA-II	1.59E-02	-2.1	1652	3495.1	Gal	55.2	80.4
	STA	4.71E-03	2.5	6579.8	2655.6	GlcNAc	0	55.8
Lpp	**ABL**	1.68E-02	6.5	45173.5	6948.6	Gal	32.2	74.3
	AIA	2.82E-03	29.4	74189.7	2526.3	Gal	82.1	74.5
	APP	4.43E-04	3.6	471.1	132.1	Gal	2.9	0
	BPA	4.43E-04	3.4	1382.9	406.9	Gal	7.8	0
	ConA	1.96E-04	6.2	1773.6	284.5	Man	56.9	0
	DSA	2.82E-03	-5.8	1919.3	11204.2	GlcNAc	25.4	27.8
	ECA	1.31E-02	4	652.2	164	GlcNAc	4.6	0
	GSL-I-B4	4.43E-04	2.3	463.7	204	Gal	6.3	15.2
	HHA	1.96E-04	-3.4	3536.5	12115.3	Man	70.9	92.2
	LTA	2.82E-03	4.8	866.1	178.8	Fuc	42.9	0
	MAA	6.62E-04	4.6	1914.9	419.7	na	na	na
	MOA	3.24E-03	2.8	518.9	184.1	Gal	7.6	4.1
	MPA	2.64E-05	-16.9	597.3	10114.8	Gal	19.7	78.3
	NPA	4.43E-04	-3.3	2464.6	8050	Man	64.5	87.7
	RCA-I	1.33E-02	2.7	2893.5	1084.4	GlcNAc	24.3	0
	SJA	2.96E-03	2.2	521.8	234.1	Gal	3.9	0
	SNA-I	8.67E-03	2.5	990.6	404	na	na	na
	sWGA	4.03E-03	2.3	3404.8	1504.6	GlcNAc	81	0

	Lectin	[1]*P*	[2]BF	[3]bEPS	[3]rEPS	Sugar	[4]% bEPS	[4]% rEPS
	VRA	2.96E-03	7.3	1596.2	219	Gal	0	8.7
	VVA-B4	2.82E-03	5.1	1564.1	309.6	Gal	67.4	12.5
	WFA	3.11E-04	4.3	974.8	224.3	Gal	45.3	0
Ljns	**sWGA**	4.11E-03	-3.6	987.5	3510.1	GlcNAc	18	80.7
	HHA	5.86E-03	-2	3849.9	7661.5	Man	54.2	60.7
	GNA	6.75E-03	-3.3	4502.8	14779.5	Man	55.7	65.3
	MAA	4.17E-02	-2.2	577.4	1256.7	na	na	na

[1]Significance (*P*-value) following Student's *t*-test, between chelation-based isolation (bEPS) and the ultrafiltration-based method (rEPS)
[2]Binding fold difference of bEPS/rEPS
[3]Mean of scale normalised RFU signals from three microarrays with bEPS and rEPS
[4]Percentages of signal inhibition by co-incubating bEPS or rEPS with 100mM haptenic sugars
[5]For abbreviations, refer to Table 5.5

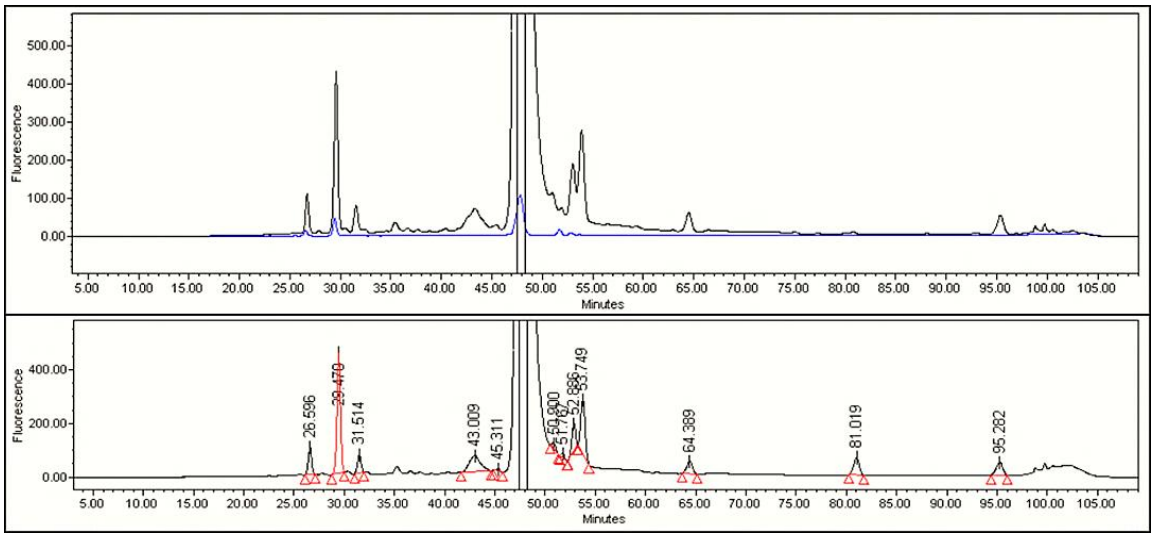

Figure 5.13: HPLC analysis of LGG rEPS compared to MRS media

Chromatograms of hydrolysed MRS media (blue curve, top graph) and rEPS from LGG (black curve, top graph) with associated retention times (red lines, bottom graph).

5.4 Putative commensal EPS operons

The sequenced genomes from *B. infantis, F. prausnitzii, L. paracasei* and *L. johnsonii* were investigated for the presence of priming glycosyltransferases (pGT) by blasting known pGT against individual genomes. For each strain, at least one pGT was identified. These include for instance a homologue of WelE in *L. paracasei* (protein accession EEI68377); a homologue of CpsD in *B. infantis* (protein accession BLIJ_2191); homologues of CpsE and AmsG in *F. prausnitzii* (respectively EEU97311 and EEU97300); EpsE in *L. johnsonii* (EEJ59758). *In silico* genome-walking and alignments to the described operons were then carried out to define additional genes involved in EPS clusters including glycosyltransferases (GT) and transcriptional regulatory elements (Figure 5.14). Very few GT were annotated in terms of sugar specificity in the NCBI database, but a summary is given in Table 5.12 and might suggest the presence of the corresponding sugars in the EPS.

Table 5.12: Specific glycosyltransferases identified in selected commensal genomes

[1]**Species**	**Name**	[2]**ID**
Bloni	Galactosyltransferase	BLIJ_2191
Fpr	β(1,6)GlcNAc transferase	EEU97308
Fpr	GlcNAc transferase	EEU97309
Fpr	Amylovoran biosynthesis glycosyltransferase AmsE	EEU97310
Ljns	UDP-*N*-acetylglucosamine:LPS *N* -acetylglucosamine transferase	EEJ59759
Lpp	UDP-*N*-acetylglucosamine:LPS *N*-acetylglucosamine transferase	EEI68380
Lpp	Mannosyltransferase	EEI68385
Lpp	Mannosyltransferase	EEI68078

[1]For abbreviations, refer to Table 5.5
[2]Sequence from NCBI database (http://www.ncbi.nlm.nih.gov/)

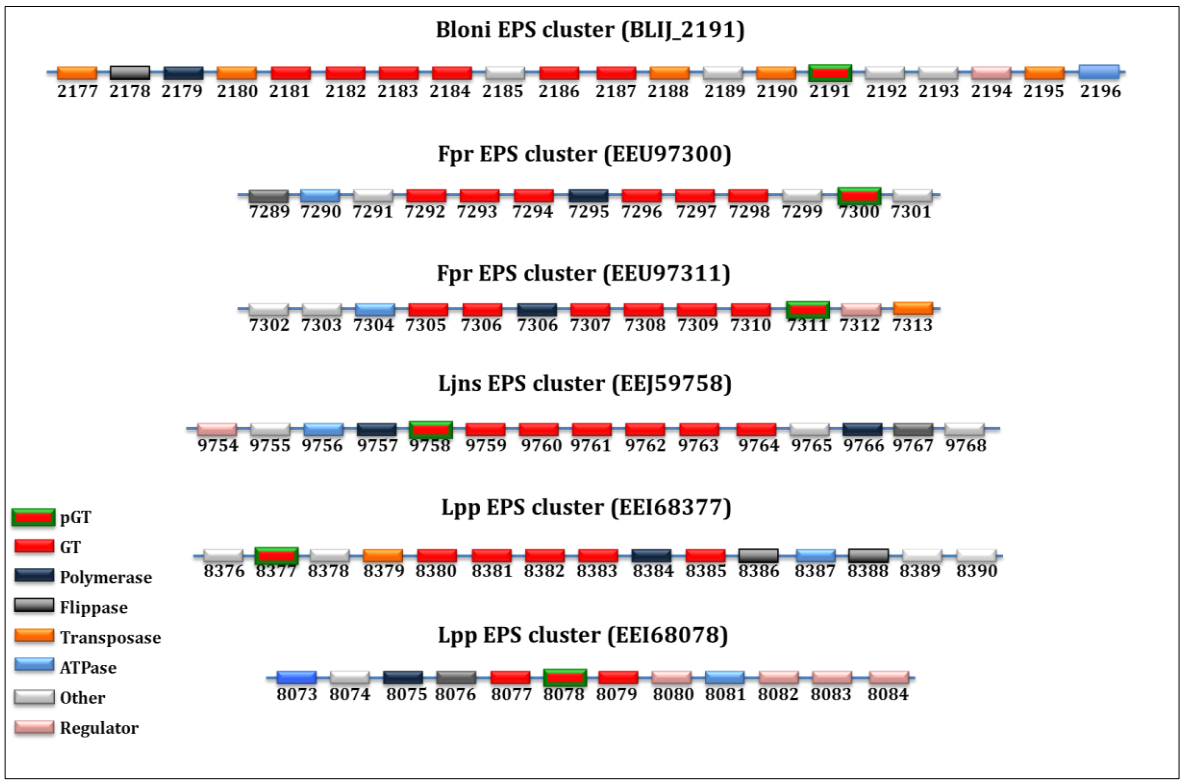

Figure 5.14: Putative EPS operons in four commensal strains from the GIT

Putative EPS gene clusters from investigated commensal species: *Bifidobacterium longum* subsp. *infantis* (Bloni), *Faecalibacterium prausnitzii* (Fpr), *Lactobacillus paracasei* subsp. *paracasei* (Lpp) and *Lactobacillus johnsonii* (Ljns).

The priming glycosyltransferase (pGT) has been used to name the operon (NCBI accession number between brackets).

Each rectangle relates to a gene from the cluster; ID numbers below each gene correspond to the last four digits of its operon's name (e.g. NCBI accession entry for gene ID 7289 in cluster EEU97300 is EEU97289).

5.5 Profiling of isogenic clones from *B. breve* UCC2003

Bifidobacterium breve UCC2003 (abbreviated from here on as *B. breve*) is able to produce two types of EPS, the synthesis of which is controlled by either half of a bidirectional gene cluster (Fanning, Hall et al. 2012). Thus, in addition to the wild type EPS (WT), the production of an alternate EPS (INV) is possible by means of promoter re-orientation. EPS-derivatives including non-EPS producers (DEL and SF430) and functional knock-outs (glycosyltransferase mutants SF440 and SF441) were provided by our collaborators for lectin profiling (see Chapter 2, section 2.2).

5.5.1 Comparison of whole-cell signatures

B. breve isogenic mutants displayed, as expected, more comparable lectin binding profiles than those of the commensal species aforementioned (Figure 5.16). The highest binding signals were obtained with ACA (specific for the T antigen Gal-β-(1→3)-GalNAc), MOA (α-Gal), AAL (Fuc-α-(1→6)-GlcNAc or Lewis x structures (Gal-β-(1→4)[Fuc-α-(1,3)]-GlcNAc) and PSA (high mannose *N*-linked glycans or Fuc-α-(1→6)-GlcNAc). A trend towards interaction with galactophilic lectins was also observed including the bacterial agglutinin PA-I (Gal), BPA (GalNAc/Gal), but especially to those specific for α-Gal such as EEA, MPA and VRA (Figure 5.15). Differential binding between the clones was apparent towards VRA, with three clones (INV, SF440 and SF441) binding with greater affinity. However, interaction with the wheat germ agglutinin WGA (specific for GlcNAc and sialic acid) constitutes the major difference between all strains, with the EPS mutant DEL and the ‘eps’ operon knock-out SF430, binding strongly to that lectin. Altogether, these profiles suggest diverse proportions of monosaccharides between isogenic clones (possibly in α-Gal, GlcNAc/sialic acid and fucose).

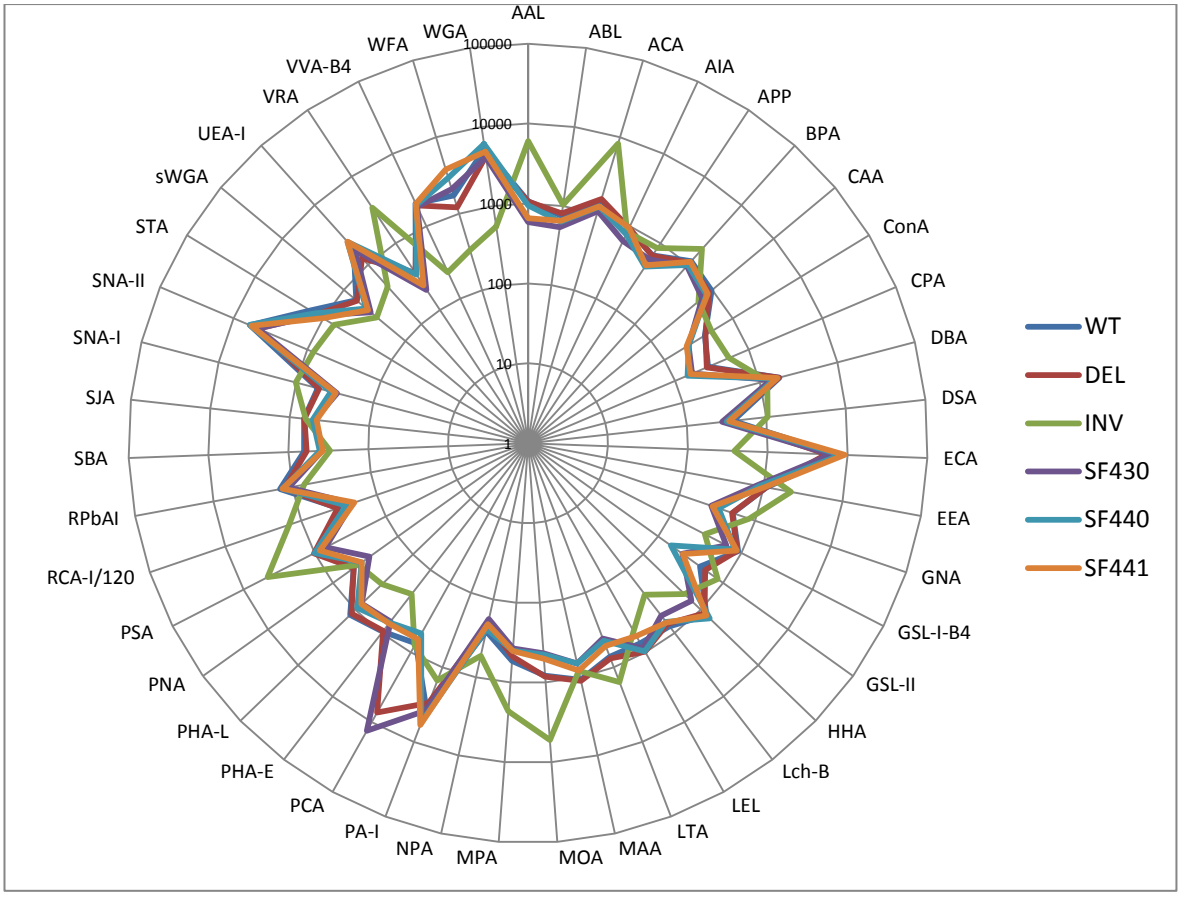

Figure 5.15: Lectin signatures of *B. breve* UCC2003 isogenic clones

B. breve UCC2003 wild type (WT) and mutants (INV, DEL, SF430, SF440 and SF441). Abbreviations for lectins are found in Table 2.13

Differential binding to WGA also clustered EPS producers separately from non-EPS producers (clones SF430 and DEL), which strongly interacted with the agglutinin (Fig. 5.16). WGA bound 9- to 23-fold more strongly to the two non-EPS producers compared to the rest (ANOVA, $P<0.001$, FDR 5%). Other lectins, such as MAA and SNA-I, also presenting specificity towards sialic acid, did not interact with the EPS negative mutants. This suggests that it is not sialic acid that is involved in the binding. WGA binds strongly to the chitobiose core of asparagine-linked trisaccharide Man-β-(1→4)-GlcNAc-β-(1→4)-GlcNAc (Yamamoto, Tsuji et al. 1981). A better explanation might be therefore that the absence of EPS exposes the bare surface of the mutants such as SF430 (TEM photograph Figure 5.17) making their cell wall more accessible to the lectins. The peptidoglycan layer surrounding Gram-positive bacteria is formed by a network of β-(1→4)-linked GlcNAc to MurNAc, cross-linked by amino-acids. It has been reported previously that WGA agglutinates strongly some bacteria such as *Micrococcus luteus*, *E. coli* and

Staphylococcus aureus and interactions are inhibited by competition with cell-wall polymers including soluble peptidoglycan (Lotan, Sharon et al. 1975).

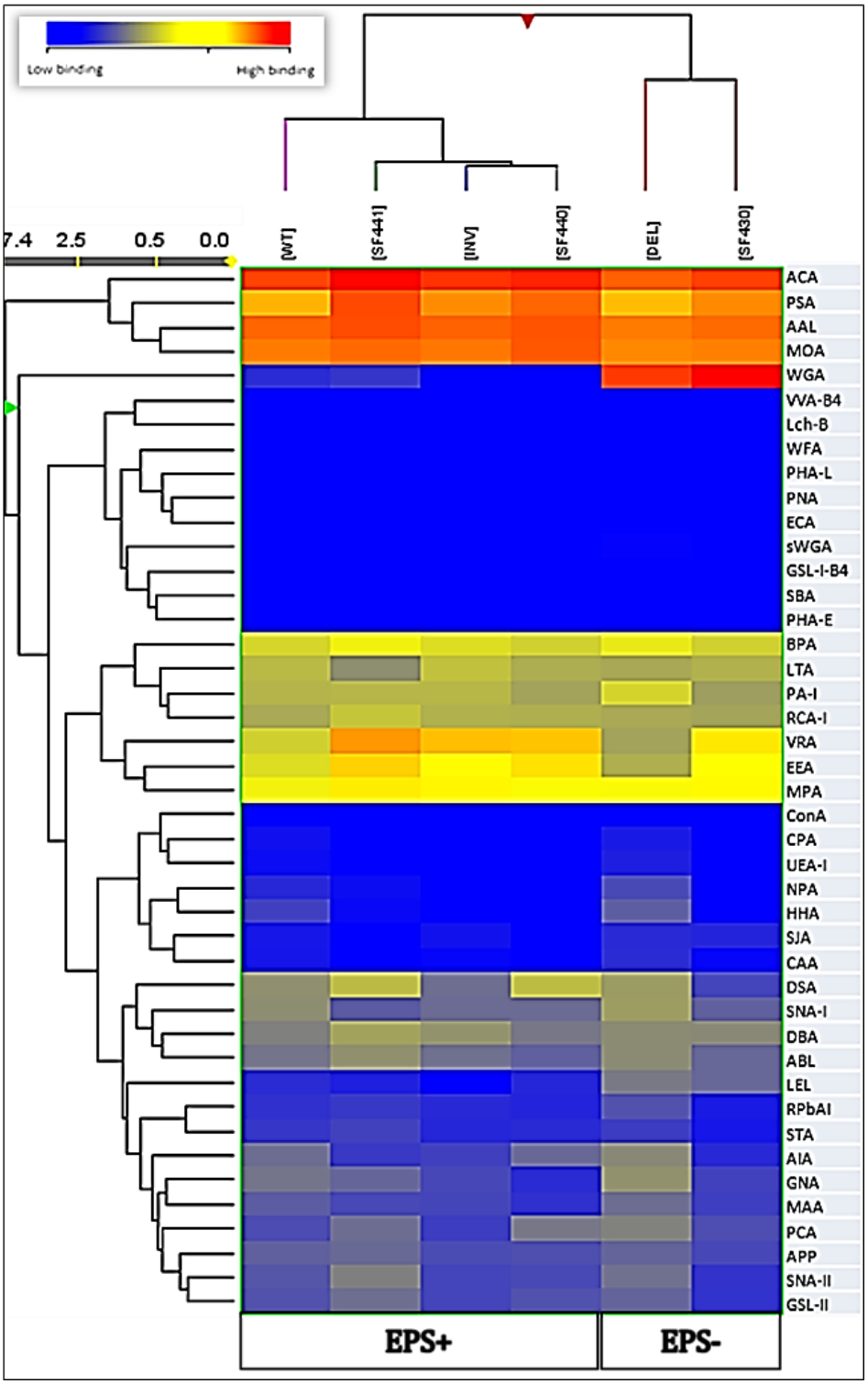

Figure 5.16: Hierarchical clustering of bound lectins by *B. breve* UCC2003 isogenic clones

Lectin profiles from *B. breve* UCC2003 EPS-producing clones (EPS+) with wild type (WT), alternate type (INV) and insertion mutants (SF440, SF441); non-EPS producers DEL and SF430 (EPS -). Abbreviations for lectins are found in Table 2.13

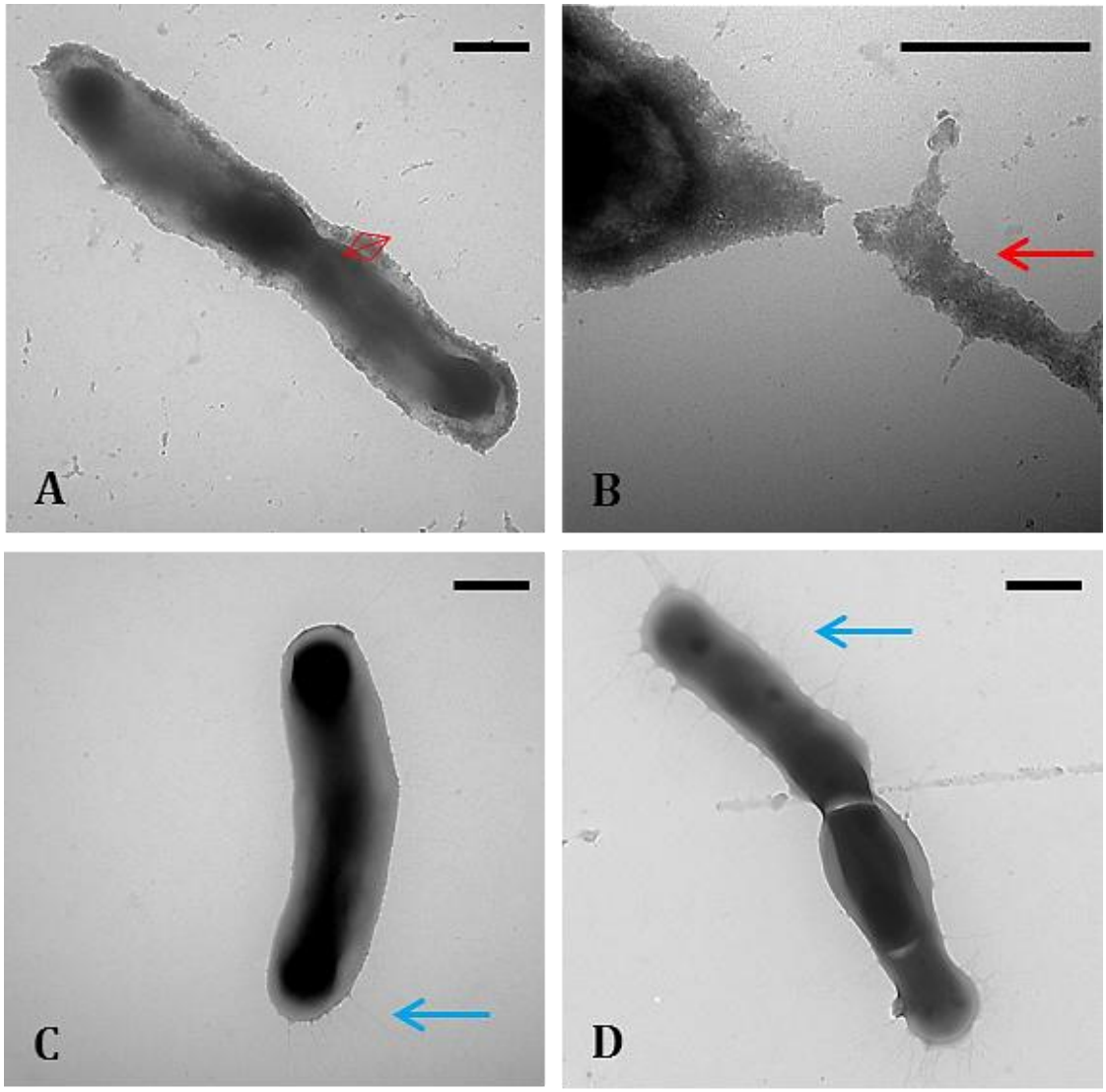

Figure 5.17: *B. breve* UCC2003 WT and EPS mutant

TEM photographs featuring wild type strain (A and B) and non EPS-producer isogenic mutant SF430 (C and D). Red arrows denotes loosely attached EPS forming a slime. Blue arrows point at hair-like appendages corresponding to possible fimbriae. Scale bar represents 500nm.

5.5.2 Profiling of *B. breve* UCC2003 EPS mutants

As described previously (Chapter 2, section 2.6), cell-bound EPS was isolated by a chelating agent (EDTA) and released EPS was obtained from ultrafiltration of the growth media (MRS). EPS was labelled, incubated on the lectin microarray and binding challenged with monosaccharides. Profiles are illustrated on the radar plots (Figure 5.18). The wild type bEPS interacted strongly with fucose-specific lectins AAL, LTA and UEA-I, mannose-specific agglutinins (ConA and HHA), galactose-specific lectins (SNA-II and RpbA-I), as well as with a group of lectins presenting affinity for complex glycans

possessing the disaccharide Gal-β-(1→4)-GlcNAc (ECA, EEA, PHA-E, PHA-L and RCA-I). The inverted bEPS (INV) showed stronger binding to GalNAc-specific ABL (T-antigen: Gal-β-(1→3)-GalNAc), AIA (Gal and T-antigen), MOA (α-Gal), and GlcNAc-specific lectins (DSA, LEL and WGA). Statistical analysis was carried out between the wild type (WT) and the EPS mutant (INV). Differential binding between the two bEPS preparations was significant for 33 lectins (Student *t*-test, $P < 0.05$, FDR 5%, Table 5.13). Comparison between the lectin signals, major ligands and the inhibition data allowed selection of a panel of lectins to act as a lectin 'bar-code' for each EPS, which might give structural information on each polymer (Figure 5.20).

A similar approach was used on the released EPS preparation from both clones. Fewer differences were obtained compared to the bEPS fractions (16 significant lectins following *t*-test, Table 5.14). Competition experiments, using monosaccharides, identified six lectins that accounted for the differences seen. Overall, the inverted type bound preferentially to mannose-specific lectins ConA, GNA and to LTA, specific for fucose and Lewis x structures, whereas the wild type rEPS interacted with higher affinity to galactophilic lectins (α-Gal-binding lectins MPA and MOA, as well as GHA, specific for GalNAc/Gal). As for the bEPS preparations, data were combined to generate a lectin 'bar-code' that might give structural hints for each strain (Figure 5.19).

The rEPS preparation from clone SF430 was also profiled on the lectin arrays (data not shown). Comparison to both WT and INV rEPS fractions revealed a preference of the functional knockout for a group of agglutinins specific for Gal-β-(1→4)-GlcNAc-containing oligomers (CAA, ECA, PCA, PHA-E, PHA-L, STA, sWGA). Moreover, highest signals using SF430 rEPS were obtained when interacting with WGA (specific to GlcNAc or sialic acid) as observed when labelling the whole cells. This suggests PG is in fact detected by the panel of lectins.

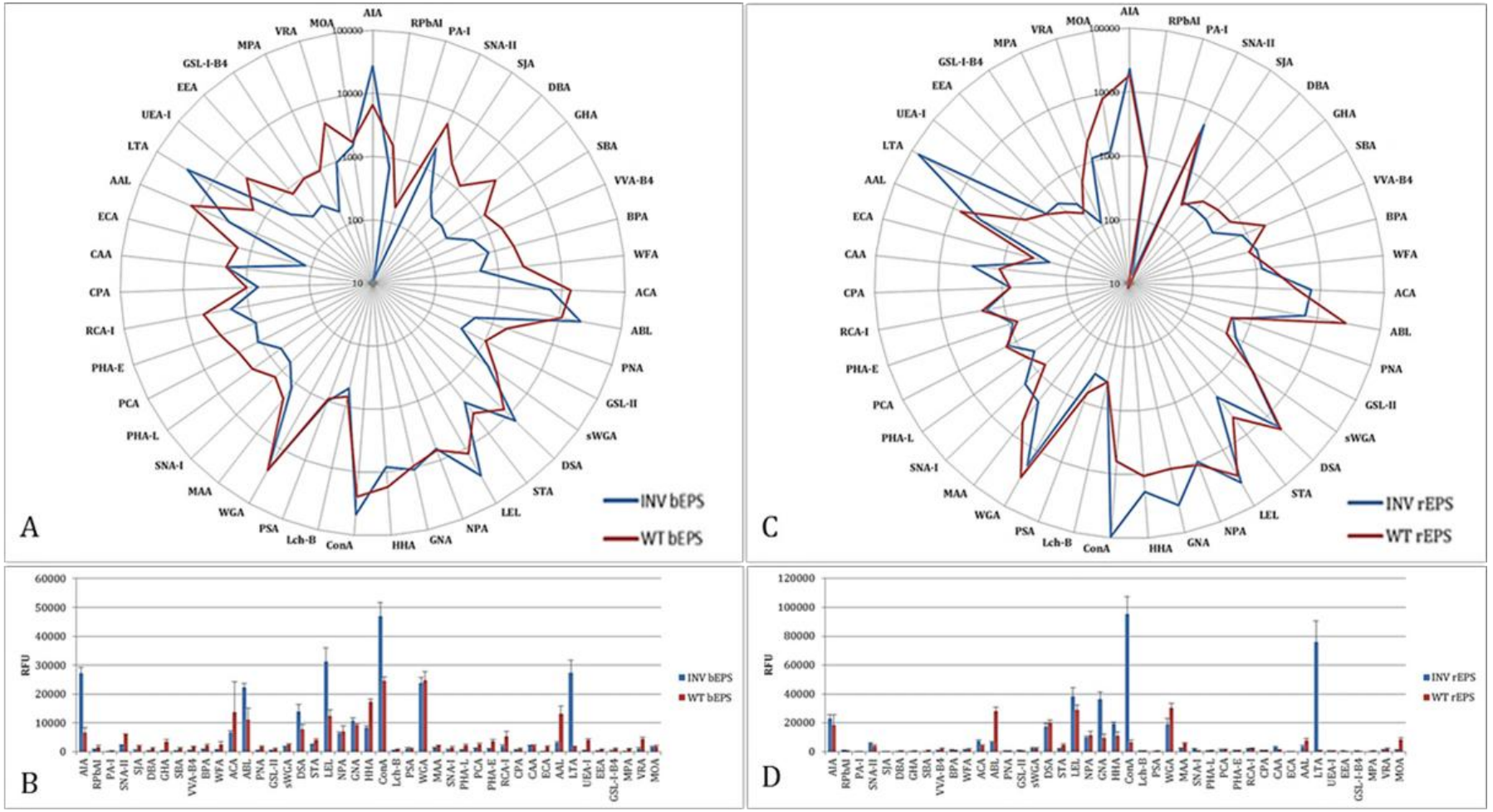

Figure 5.18: Lectin-binding profiling of EPS from *B. breve* UCC2003 wild type and inverted mutant

Profiling of bEPS (A, B) and rEPS (C, D) isolated from wild type strain (WT, red) and the inverted mutant (INV, blue).

Table 5.13: Significant interactions between lectins and bEPS

Lectin	Ligand	[1]*P*	[2]BF	[3]INV bEPS	[4]WT bEPS	[5]Sugar	[6]% INV	[7]% WT
AAL	Fuc	2.00E-03	-4.4	2977.6	13156.2	Fuc	17.3	20.9
UEA-I	Fuc	3.10E-04	-7.7	502.1	3898.2	Fuc	3.9	24.5
LTA	Fuc	3.00E-04	16.4	27391	1659.5	Fuc	26.5	35.3
VRA	α-Gal	5.90E-03	-4.6	975.1	4349.9	Gal	41.4	46.5
GSL-I-B4	α-Gal	2.80E-02	-3.3	295.6	940.7	Gal	0	4.9
MPA	α-Gal	9.50E-04	-5.1	177.4	918.2	Gal	0	92.4
EEA	α-Gal	1.90E-02	-3.2	262.2	786	Gal	0	0
RPbAI	Gal	3.30E-02	-2.2	694.4	1566.9	Gal	24.4	22.1
AIA	Gal	2.00E-03	4.2	27027	6591.1	Gal	0	74.6
SNA-II	Gal	1.70E-04	-2.7	2233.2	6041.5	Gal	87.5	51.8
SJA	β-GalNAc	6.90E-03	-4.2	457.9	1871.7	Gal	80.6	61.5
GHA	GalNAc	2.00E-03	-13.3	265.4	3414.4	Gal	0	61.5
ABL	GalNAc	3.00E-02	2.1	22238.7	11056.6	Gal	16.6	4.7
WFA	GalNAc	6.90E-03	-4.6	523.3	2507.3	Gal	73.8	31.8
SBA	GalNAc	2.00E-03	-4.9	233.7	1180.8	Gal	0	13.3
DBA	GalNAc	2.00E-03	-4.5	253.2	1168.1	Gal	0	32.8
BPA	GalNAc	6.50E-03	-2.6	778.1	2030.3	Gal	0	0
PNA	GalNAc	1.80E-03	-3.4	518.4	1748.8	Gal	0	7.9
VVA-B4	GalNAc	6.10E-04	-3.1	542.4	1669.4	Gal	0	36.7
LEL	GlcNAc	5.00E-03	2.5	31240.7	12410.2	GlcNAc	41.6	84.2
DSA	GlcNAc	3.40E-02	1.8	13842.3	7681.8	GlcNAc	46.3	69
STA	GlcNAc	5.90E-03	-1.7	2401.2	4037.9	GlcNAc	0	39.1
sWGA	GlcNAc	1.10E-02	-1.4	1769.1	2487.6	GlcNAc	0	78.6
PHA-E	GlcNAc	2.00E-03	-3.9	896.6	3552.6	Gal	7.9	67
HHA	Man	1.00E-03	-2.1	8288.4	17183	Man	88.4	98
ConA	Man/Glc	2.00E-03	1.9	46897.3	24534.9	Man	53	93.2
MAA	Sia	7.50E-03	-1.6	1268.5	2065.8	na	na	na
RCA-I	Complex	1.70E-02	-2.7	1908.4	5320.8	GlcNAc	96	37.5
PCA	Complex	1.40E-02	-2.2	1097.7	2442.5	GlcNAc	65.9	0
PHA-L	Complex	2.80E-03	-3.5	601.7	2152.3	GlcNAc	0	30.8
ECA	Complex	1.10E-03	-13.1	127.2	1637.9	GlcNAc	98.7	60.4
GSL-II	Complex	1.00E-02	-2.6	380.3	1029.9	GlcNAc	0	17.5
CPA	Complex	3.00E-02	-1.5	653.2	998.7	GlcNAc	0	0

[1] *P*-value following Student *t*-test

[2] Binding fold difference between EPS (INV/WT)

[3] Signal intensity for the inverted bEPS (INV)

[4] Signal intensity for the wild type bEPS (WT)

[5]Monosaccharide used for the inhibition. Abbreviations are as follows: Fuc, fucose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; Gal, galactose; Glc, glucose; Man, Mannose; Sia, sialic acid; na, not performed.

[6] Percentage of inhibition on INV bEPS

[7] Percentage of inhibition on WT bEPS

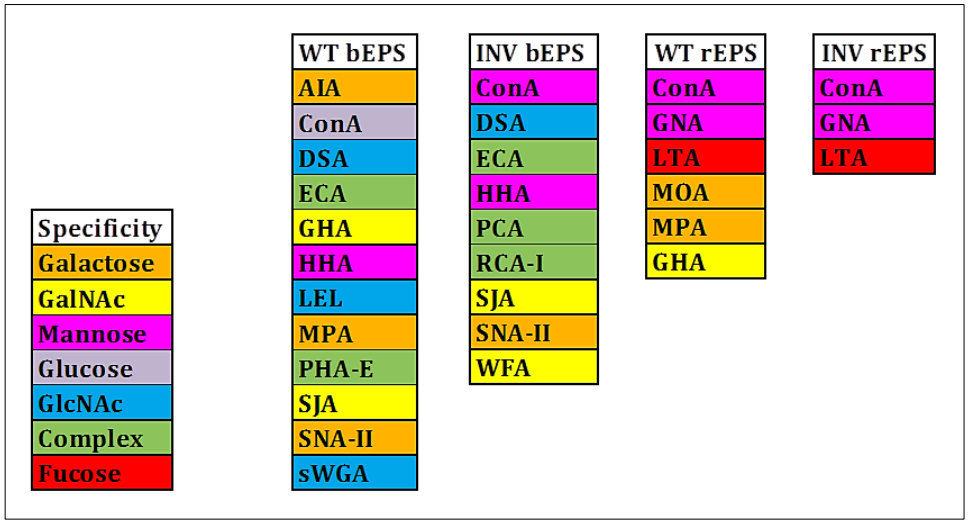

Figure 5.19: Lectin 'bar codes' of EPS from *B. breve* UCC2003

Lectin-binding signatures resulting from significant sugar-mediated interactions of EPS from *B. breve* UCC2003 wild type strain (WT) and the inverted (INV) mutant.

Table 5.14: Significant interactions between lectins and rEPS

Lectins	Ligand	[1]*P*	[2]BF	[3]INV rEPS	[4]WT rEPS	[5]Sugars	[6]% INV	[7]% WT
AAL	Fuc	3.24E-02	-2.2	3508.7	7564.9	Fuc	0	21
LTA	Fuc	1.83E-04	91.5	75842.8	820.7	Fuc	99	35
MOA	**α-Gal**	2.60E-03	-7	1200.1	8349.9	Gal	0	89.9
MPA	**α-Gal**	9.36E-03	-4.9	111.1	561.8	Gal	0	60.1
SNA-I	Gal	9.36E-03	2.7	1868.7	683.1	Gal	na	na
ABL	GalNAc	9.65E-04	-4.4	6382.5	28227.3	Gal	0	14
ACA	GalNAc	2.56E-02	1.7	7056.1	4207.7	Gal	0	0
GHA	GalNAc	1.82E-02	-1.7	363	633.1	Gal	0	65
SBA	GalNAc	2.35E-02	-2.1	333.9	706.9	Gal	0	26.9
VVA-B4	GalNAc	1.82E-02	-2.4	830.5	2035.4	Gal	0	26.9
CAA	GlcNAc	8.03E-03	2.7	3021.5	1130.2	GlcNAc	0	2.2
STA	GlcNAc	1.74E-02	-2.5	1772.7	4510.3	GlcNAc	0	16.3
ConA	Glc/Man	1.51E-03	15.5	95279.1	6255	Glc/Man	86.8/89.7	88/78.8
GNA	Man	9.36E-03	3.9	36406.6	9456.9	Man	86.2	83.8
PSA	Man	1.84E-02	-2.1	328.3	679.6	Man	0	0
MAA	Sia	9.28E-03	-2.5	2244.7	5609.6	n/a	na	na

[1] *P*-value following Student *t*-test
[2] Binding fold difference between EPS (INV/WT)
[3] Signal intensity for the inverted (INV) bEPS
[4] Signal intensity for the wild type (WT) bEPS
[5]Monosaccharide used for the inhibition. Abbreviations are as follows: Fuc: fucose, GalNAc: *N*-acetylgalactosamine, GlcNAc: *N*-acetylglucosamine, Gal: galactose, Glc: glucose, Man: Mannose
[6] Percentage of inhibition on INV bEPS
[7] Percentage of inhibition on WT bEPS

As suggested by the lectin profiles, the bEPS of the WT and the INV isogenic clones were composed of different ratios of monosaccharides (Table 5.15). Rhamnose was the only monosaccharide not detected in any of the samples. Fucose was detected in the WT and higher amounts of galactose (3-fold) and glucose (2-fold) were found when compared to the INV bEPS. A higher content of mannose (6-fold), GlcNAc (5-fold) and GalNAc (2-fold) contrasted the INV bEPS to the WT bEPS.

Regarding the released fractions (rEPS), fucose was not detected in the WT as opposed to the bEPS preparation. Closer ratios were obtained between clones, with galactose and GlcNAc content almost identical. The main difference was the mannose concentration, which was higher in the INV rEPS (4-fold). GalNAc and glucose contents were also higher (1.5-fold) compared to the wild type rEPS.

All sugars present in *B. breve* WT, with the exception of GlcNAc, were found in the 'non-EPS' functional knockout SF430 bEPS but in much lower concentration. This was also true for the rEPS fraction.

Table 5.15: Composition of UCC EPS fractions

	[1]Sample	Fuc	Rha	GalNAc	GlcNAc	Gal	Glc	Man
[2]**bEPS**	WT	**21**	*nd*	**20**	**3**	**1078**	**540**	**83**
		(1)		(1)	(0.2)	(54)	(27)	(4)
	INV	*nd*	*nd*	**46**	**19**	**365**	**235**	**464**
				(1)	(0.4)	(8)	(5)	(10)
	SF430	**10**	*nd*	**4**	*nd*	**165**	**174**	**20**
		(2)		(1)		(41)	(43)	(5)
[2]**rEPS**	WT	*nd*	*nd*	**223**	**71**	**213**	**103**	**161**
				(1)	(0.3)	(1)	(0.5)	(0.7)
	INV	*nd*	*nd*	**318**	**74**	**203**	**153**	**688**
				(1)	(0.2)	(0.6)	(0.5)	(2.2)
	SF430	*nd*	*nd*	**56**	**7**	**23**	**10**	**17**
				(1)	(0.1)	(0.4)	(0.2)	(0.3)
	MRS	*nd*	*nd*	**68**	**40**	**97**	**407**	**237**
				(1)	(0.6)	(1.4)	(6)	(3.5)

[1]Concentration of monosaccharides (pmoles/µg, average of three samples) with molar ratio calculated relative to GalNAc (between parentheses). Abbreviations are as follows: Fuc, fucose; GalNAc, *N*-acetyl-D-galactosamine; GlcNAc, *N*-acetyl-D-glucosamine; Gal, galactose; Glc, glucose; Man, Mannose and *nd*, not detected; EPS from *B. breve* UCC2003 isogenic clones, wild type :WT, alternate EPS: INV, non-EPS producer (SF430) and MRS media control.
[2]Prefix 'b' refers to the chelation-based isolation (bEPS), prefix 'r' to the ultrafiltration-based isolation (rEPS fraction).

5.6 Discussion

Exopolysaccharides are considered very promising for use as effector molecules in probiotic therapy (Lee, Tomita et al.). However, the current analytical approach, from isolation to structural determination, is complex and hinders the investigation of these important molecules (Laws, Gu et al. 2001). Lectin microarray profiling has the advantage of being a rapid high-throughput technology useful in discriminating cells and molecules with different glycosylation patterns (Hsu and Mahal 2006; Hsu, Pilobello et al. 2006; Yasuda, Tateno et al. 2011; Semchenko, Day et al. 2012; Hirabayashi, Yamada et al. 2013). The first application of lectin microarrays to profiling whole bacterial cells using a nucleic-acid-binding dye was described by Hsu *et al* (Hsu and Mahal 2006). This technique was shown to be able to discriminate pathogenic bacterial species, namely *E. coli*, *Salmonella* sp. and more recently, strains of *Campylobacter jejuni*, which are human pathogens but commensals in avian (Hsu and Mahal 2006; Hsu, Pilobello et al. 2006; Kilcoyne, Twomey et al.). Lectin binding specificities have been characterised for mammalian glycans, therefore they might not serve as structural predictors for prokaryotes (Gupta, Surolia et al. 2010). Bacterial polysaccharides could contain monosaccharides, linkages and substitutions not found in mammals and it is not known if or how these structures are recognized by the lectin array described herein. If bacteria do share similar monosaccharides with mammals including the most abundantly found D-glucose and D-Gal, they display a wider variety of them by ten-fold, with α-L-rhamnose, L-glycero-α-D-manno-heptose, α-D-galacturonic acid and α-Kdo being the most frequent monosaccharides unique to bacteria (Herget, Toukach et al. 2008). Thus, expanding the library of carbohydrate-binding proteins recognising bacterial sugars would be a major step forward in furthering bacterial polysaccharides structure prediction on lectin microarrays.

5.6.1 Lectin fingerprints of commensal strains and their EPSs

Lactobacillus casei is the only commensal species that has been subjected to lectin microarray profiling (Yasuda, Tateno et al. 2011). In this study, human commensal strains originating from faeces, blood or milk were selected based on their promising probiotic attributes. These included lactic acid bacteria species such as *Lactobacillus rhamnosus* GG (Morita 2008), *L. johnsonii* (Pridmore 2004), *L. paracasei* subsp. *paracasei* (Chiang and Pan 2012) and *Bifidobacterium* subsp. *infantis*, which is one of the first colonising strains of infants (Brenner and Chey 2009). In addition, a prominent Gram-negative strain found

in the healthy GIT of adults, *Faecalibacterium prausnitzii*, was included due to its anti-inflammatory properties described *in vitro* and in animal models (Sokol, Pigneur et al. 2008; Miquel, Martin et al. 2013). This study has shown that the lectin profiles of the selected panel of commensal species from the human GIT were unique.

As a proof of concept, we first used a strain of known EPS structure, LGG, producing a long galactose-rich EPS (Lebeer, Verhoeven et al. 2009), before applying to other species. Within the limitations of our lectin microarray, the LGG 'whole cell' profile correlated with the described EPS structure (Landersjo, Yang et al. 2002). Moreover, competition with LGG cell-bound EPS preparation (bEPS) inhibited most of the lectin signals, suggesting the surface-bound EPS was responsible for the whole cell profile in this strain. We then compared profiles of different strains of human commensals. Lectin signatures were unique and their clustering did not correlate with phylogeny. Profiles might instead reflect the ecological niche or the source of the strains as illustrated with the lectin signatures given by three *Lactobacillus* species originating from human milk (*L. paracasei*), blood (*L. johnsonii*) or faeces (LGG). Yasuda *et al*. also reported unique lectin signatures from four strains of *L. casei* suggesting the synthesis of EPS to be an adaptation trait to their host, possibly by altering their outer surface to escape from the host immune system (Yasuda, Tateno et al. 2011). It has been recently shown that strains such as LGG and *B. breve* UCC2003 avoid their host immune mechanisms by the synthesis and coating of their cell surface by EPS (Lebeer, Claes et al. 2011; Fanning, Hall et al. 2012). The human immune system is very well developed and new strategies are always needed to evade it, leading to differences in coating EPS produced by commensals. The lectin microarray profiling was able to identify differences in cell surface glycosylation.

This study has also shown that the lectin microarray platform gives unique binding signatures for different EPS preparations from the species studied, and that these can be correlated with differences in the composition of these molecules as determined by monosaccharide analysis. While profiling of isolated bacterial EPS on lectin microarrays has not been previously reported, lipooligosaccharides (LOS) from the pathogenic strain *Campylobacter jejuni* have been examined on an array of 15 lectins (Semchenko, Day et al. 2012). As with LOS, labelled EPS interacted with less features than whole cells, possibly because of the more complex surface of the cells and differences in presentation of the cell surface EPS/LOS molecules. Binding to a substantial number of lectins

discriminated each EPS preparation, indicating differences in composition of EPS produced by different commensals, as was suggested by the whole cell glycosignatures. This was further supported by the similarity of the clustering of bEPS profiles between species as for the 'whole cell' experiment. *L. johnsonii* bEPS was the one exception, clustering with bEPS from *B. infantis* and *L. paracasei* rather than LGG bEPS, as did whole *L. johnsonii* cells. Strong binding to AIA (Gal-specific) and only weak binding to DSA and STA (both specific for GlcNAc) seemed to differentiate it from LGG bEPS, although, like all the EPS preparations studied, they both showed strong binding to another two GlcNAc-specific lectins, LEL and WGA.

Of the species studied here, detailed structural information has only been reported for the EPS produced by LGG (Landersjo, Yang et al. 2002) and the lectin binding profiles obtained for both EPS preparations were shown to be substantially consistent with the reported structure, with a high level of binding to GlcNAc- and Gal-specific lectins. The presence of rhamnose could not be determined from our microarray binding profiles as no rhamnose binders were commercially available for use on our platform at the time of printing (Hirabayashi, Yamada et al. 2013). Rhamnose-specific agglutinins have to date only been isolated from fish (Watanabe, Tateno et al. 2009), e.g. CSL, extracted from salmon eggs that has been successfully immobilised on a lectin platform (Yasuda, Tateno et al. 2011). However, rhamnose was shown to be present by monosaccharide analysis and 1-D NMR. The bEPS and rEPS from LGG also showed binding consistent with the presence of GalNAc and mannose, and also glucose in the case of the rEPS, even though these were not described in the EPS study by Landersjö *et al.* (Landersjo, Yang et al. 2002). Fresh culture media was also shown to be rich in these monosaccharides but it is unlikely that residual mannans and glucans in the spent media contaminated the EPS preparations after 48h of culture, especially the bEPS fraction. Mannose and glucose-containing cell wall polysaccharides have actually been recently described in this strain (Francius, Lebeer et al. 2008; Lebeer, Claes et al. 2011). These may have been present in our EPS preparations due to co-elution in the case of bEPS or contamination of rEPS isolated from the media with debris from dead cells. Interestingly, these cell wall polysaccharides may also display important biological functions in their host. Lebeer *et al*. have suggested a role for the mannose-containing polysaccharides from LGG in triggering innate immune factors in the gut by inducing the mannose-binding lectins of the alternative complement pathway (Lebeer, Claes et al. 2011). Genes coding for a glucosyltransferase

and a mannosyltransferase are found amongst two defined operons in LGG in the bacterial operon database (id:612963 and id:613097 respectively http://csbl.bmb.uga.edu/DOOR), suggesting more than one EPS may be produced on the cell surface of LGG and released in the environment. Tallon *et al*. characterised two EPS from *L. plantarum* EP56 that were released in the culture media: an EPS containing GalNAc with a structure similar to the cell-bound polymer from the same species and a second EPS only detected in the media showing a different composition (Tallon, Bressollier et al. 2003).

Two neutral polysaccharides (PS1 and PS2) have been isolated using TCA and purified from cell wall extracts of *B. infantis* (strain ATCC 15697) and consist of branched galactans with glucose substitutions. The backbone structure of PS2, a disaccharide containing galactofuranose, has also been found in the lipopolysaccharide of pathogenic strain *Klebsiella pneumoniae* (Whitfield, Richards et al. 1991; Tone-Shimokawa, Toida et al. 1996). In addition, Habu *et al.* reported the structure of a branched polysaccharide of 100 KDa with an octasaccharide repeating unit of galactose and glucose from the same strain (Habu, Nagaoka et al. 1987). These polysaccharides could correlate with the bEPS fraction of *B. infantis* from our study, also consisting of large amounts of galactose and glucose. Furthermore, an *in silico* search for priming glycosyltransferases (GT) in the various species has suggested the presence of two EPS operons in *L. paracasei* subsp. *paracasei* and *F. prausnitzii* supporting the possible presence of several polymers. In addition, the rare specificities of GT annotated in the NCBI database matched the results from the microarray: a galactosyltransferase was identified in *B. infantis* as well as GlcNAc transferases in three species *F. prausnitzii*, *L. johnsonii* and *L. paracasei* subsp. *paracasei*. In *F. prausnitzii*, two GTs are actually specific for GlcNAc, one of which is homologous to the amylovoran biosynthesis glycosyltransferase AmsE from *Erwinia pyrifoliae*. Therefore, we could speculate that one polymer could resemble amylovoran, an EPS consisting of a repeating unit of five sugars with substitutions (Bugert and Geider 1995; Nimtz, Mort et al. 1996).

5.6.2 Lectin fingerprints of EPS mutants from *B. breve* UCC2003

EPS mutants derived from the same strain, *Bifidobacterium breve* UCC2003, were used to explore another level of lectin responses. Insertion mutants provided by collaborators were constructed by disrupting genes encoding glycosyltransferases (Bbr_0440, Bbr_0441, as well as the priming glycosyltransferase Bbr_0430). A fortuitous non-EPS producing

mutant was also provided (Fanning, Hall et al. 2012). Contrasting with the results obtained using various commensal species, lectin-binding profiles between *B. breve* isogenic clones were more similar. Nonetheless, the profiles of the two non-EPS producers (SF430 and DEL) were significantly different from all other isogenic mutants. SF430 and DEL showed very high affinity towards the wheat germ agglutinin WGA, specific to β-(1→4)-GlcNAc-containing oligomers or sialic acid residues (Yamamoto, Tsuji et al. 1981). Other lectins, though, that present specificity towards sialic acid, such as MAA and SNA-I, did not interact with the mutants, suggesting this structure was not involved. A better explanation might be that, in the absence of EPS, other surface-exposed molecules become accessible to the panel of lectins. These could be lipoteichoic acids (LTA), fimbriae that can be glycosylated (Smedley, Jewell et al. 2005), surface layer proteins or the cell wall itself (PG).

Despite some sugars detected by HPAEC, it was not possible to label the bEPS fraction from non-EPS-producer *B. breve* clone SF430 (as expected). However, the rEPS preparation from SF430 was analysed and, similarly to its whole cells, the labelled preparation interacted strongly with WGA, as well as with a group of eight agglutinins presenting affinity towards Gal-β-(1→4)-GlcNAc. The chitobiose core of asparagine-linked trisaccharide Man-β-(1→4)-GlcNAc-β-(1→4)-GlcNAc constitutes the strongest binder of WGA (Yamamoto, Tsuji et al. 1981). The PG layer surrounding Gram positive bacteria is formed by a network of β-(1→4)-linked GlcNAc to MurNAc cross-linked by oligopeptides. It has been reported previously that WGA agglutinates strongly some bacteria such as *Micrococcus luteus, E. coli* and *Staphylococcus aureus* and binding was inhibited by competition with cell-wall polymers including soluble PG (Lotan, Sharon et al. 1975). Fluorescent-WGA has also been used as an alternative of the Gram staining method as it binds only Gram positive species (Sizemore, Caldwell et al. 1990). The rEPS fractions thus likely contain remnants of cell walls.

The lectin-binding profiles of EPS from *B. breve* WT and the alternate (INV) EPS were also compared and signatures related to the bEPS differed especially. Following inhibition experiments, each EPS composition was presented in the form of a lectin 'bar code'. Suggested monosaccharides were largely confirmed by monosaccharide analysis except for the fucose-specific lectin LTA, which bound strongly to the INV EPS, although no fucose was detected following HPAEC. This is not unusual as lectins can be promiscuous and various parameters, including the type of immobilisation used, can affect the carbohydrate

recognition (Gerlach, Kilcoyne et al. 2014). The optimisation of the EPS labelling procedure was carried out using commercial bacterial EPS, namely dextran and gellan. The latter bound strongly another fucose-specific lectin (AAL) in a sugar-mediated manner, despite the absence of fucose in its structure (Houeix, Le Berre et al. 2012). Gellan is a heteropolysaccharide made of α-L-rhamnose, β-D-glucuronic acid and two β-D-glucose residues with glycerol and acyl substitutions (Noda, Funami et al. 2008). The differences in composition between the wild type and the alternate EPS confirmed the hypothesis formulated previously on the characterisation of the *eps* locus from *B. breve* (Fanning, Hall et al. 2012). However, we also detected EPS being released in the media from all clones as opposed to Fanning *et al.* Differences in culture conditions (in-house modified MRS media as opposed to a commercial source) could explain this. Another explanation might be the absence of a neutralising step in the isolation procedure used by Fanning *et al*., resulting in degradation of the EPS samples prior to analysis. Moreover, as for LGG, additional wall polysaccharides might be synthesised in *B. breve*. Other than the published operon, three glycosyltransferases involved in cell wall biogenesis are found in the genome of *B. breve* UCC2003. These include sequence Bbr_1796 that could constitute the priming glycosyltransferase of a second operon with neighbouring genes encoding a glycosyltransferase homologous to rhamnan synthesis protein F (Bbr_1795), phospho-glyceroltransferases (Bbr_1790 and Bbr_1791) as well as proteins of ABC transporter system for polysaccharides (ATP-binding protein, Bbr_1793 and permease, Bbr_1794). Rhamnose was not detected by HPAEC from our experiments. However, as performed on LGG bEPS, it could be of interest to perform NMR analysis on the samples to confirm this.

This study has shown that the lectin profiles of a selected panel of commensal species from the human gastro-intestinal tract were unique. Lectin binding signatures from the EPS produced by the panel studied were strain-specific but the binding patterns suggested they shared similar components. The profiles of EPS isolated using two different methods also differed, highlighting the importance of the isolation procedure when comparing various EPS.

Chapter 6: Conclusion and perspectives

The main goal of this project was to advance knowledge on commensal glycobiology, especially the molecules involved in the interactions between gut commensal bacteria and their host. Studies so far have focused on adhesion or biofilm formation of pathogenic strains (Chapter 1). Little information is available on mechanisms of adhesion of commensal bacteria to the human GIT. Very few commensal adhesins and even fewer ligands have been identified. Similarly, effector molecules such as exopolysaccharides produced by commensals have been poorly described to date.

Developments in genomics, proteomics and glycomics have, in recent years, generated a large amount of publicly available data that can deliver useful insights into a wide variety of biological processes with focused probing and exploration and thereby facilitate the generation of hypotheses for experimental testing. This study was unique in that it applied available *in silico* tools for the identification of potential commensal adhesins and their ligands and also demonstrated that glycosylation-related high-throughput microarray technologies could be used to discriminate between molecules involved in bacterial adherence to the mucosa of the GIT and understand how they interact (Chapter 2). The first objective was to test a novel approach using *in silico* tools to study host-microbes interaction with a focus on bacterial adherence to the mucosa of the GIT (Chapter 3). The current approaches relying on *in silico* identification of putative adhesion molecules permit the rapid screening of whole proteomes.

Predictive computer modelling of bacterial adhesion has been used in vaccine development, and was described as a potentially useful tool to select surface-exposed proteins (Barinov, Loux et al. 2009). In this work, the UNIX-based Java program SurfG+ was used to investigate the surfaceome, as defined by Cullen, of species from three main commensal genera (*Bifidobacterium*, *Lactobacillus* and *Faecalibacterium*) (Cullen, Xu et al. 2005; Barinov, Loux et al. 2009). This software was designed to predict the localisation of proteins within the Gram positive bacterial cell, with an emphasis on cell-wall binding domains to detect surface-exposed proteins for the purpose of vaccine engineering. Our results showed that a substantial number of SEPs could be identified in all three species. Using SurfG+, Barinov et al. predicted a higher proportion of SEPs in human commensal lactobacilli compared to dairy strains, suggesting that expression of SEPs represent an adaptive trait to the residence in the GIT environment (Barinov 2009; Ghoddusi 2012). Our results were consistent with this hypothesis. The SEPs identified by SurfG+ were further investigated for homology to conserved domains from described adhesins. This led to the

identification of several potential adhesins in selected commensal strains, representative of the microbiota of healthy individuals.

Our next step was to evaluate expression of the selected adhesin genes under different experimental conditions. Expression of several of the putative adhesin genes was successfully demonstrated in *L. paracasei* subsp. *paracasei* and adhesive structures visualised, particularly when cultured in conditions resembling the human gut, providing support for the utility of the approach taken. Indeed, reanalysis of the genome led to four models of pili clusters being proposed in this strain. Expression analysis did not, however, provide any confirmation of the adhesin list drawn up for the other two species. *In vitro* expression of adhesins or pili is reported to be very difficult in some species (Turroni, Serafini et al. 2014), which may explain these results. It was surprising that pili structures were clearly visible in *Faecalibacterium prausnitzii* in regular media, yet no expression of any of the identified adhesins was detected. An alternative approach might have been to continue the analysis of the predicted adhesins by *in silico* molecular docking using putative ligands, as described for pathogenic lectins BambL and LecB (Topin, Arnaud et al. 2013), to gain greater assurance of the ability of the selected proteins to bind to glycans. In either case, the next steps would be, as has already been done for other strains such as LGG (Reunanen, von Ossowski et al. 2012), to engineer adhesin gene knockout clones and perform adhesion assays in parallel with the wild type strain grown in the presence of mucin and salt. The pilin subunits and putative adhesins could also be expressed constitutively in bacterial systems, such as *E. coli* or *Lactococcus lactis*, the proteins purified and used in binding assays or on glycan arrays for ligand identification.

Glycan arrays are increasingly being used to characterise the binding specificities of glycan-binding proteins. These arrays have been developed in a number of formats, with either glycans, glycolipids or glycoproteins spotted on array and are still being validated, even for abundant and well-studied glycan-binding proteins (Wang, Cummings et al. 2014). In this study, we used in-house-developed NGC arrays containing 76 features, comprising mainly glycans found on mammalian glycoproteins, to discover and identify the glycan epitopes of selected bacterial adhesins (Chapter 4). Bespoke glycan arrays, comprising the 95 GIT glycan structures compiled from the CFG database as putative ligands for commensal bacteria (sections 2.1.1 and 3.1), would probably have been more useful in this regard. With the on-going progress in glycan synthesis, an adhesin screening platform could be available in the future for identification of ligands of commensal lectins.

Initially, as a proof of concept, two strains of *E. coli*, which produce two types of pili *in vitro* with known carbohydrate ligands were incubated on NGC microarrays. type 1- and P fimbriae-producing strains bind, respectively, mannosyl structures and galabiose-containing glycans. Analysis of the binding data highlighted the expected glycan-binding pattern of both pili-producing strains and specificity was supported by inhibition experiments with free monosaccharides. This gave us confidence to profile the binding pattern of selected commensal strains, including the ones studied *in silico*, and identify glycan ligands of their lectins. To our knowledge, the adhesion mechanisms of these commensals have not yet been addressed.

Whole cells, as well as isolated pili, from each of the strains were incubated on NGC microarrays. Specific ligands were identified for each strain by challenging with free monosaccharides. Whole cells generally bound several NGCs; most probably representing the many ligands of the many surface exposed proteins discovered *in silico*. Incubation of directly labelled pili extracts gave clearer binding profiles, but there were still a number of binding targets identified for each preparation. As mentioned above, recombinant expression of putative adhesins is necessary for their full characterisation. These could be incubated directly on the NGC microarray to identify corresponding ligands and could also be used for antibody production. Nevertheless, clear differences in profiles were obtained with each of the sample types, indicating that the commensal strains examined expressed several glycan-binders on their surface and much further work is needed to decipher their roles in commensal biology. Mucins are among the first substrates the bacteria adhere to when colonising a particular biotope in the GIT. Therefore, we also profiled the binding of the commensal strains on natural mucins from various body sites immobilised on a microarray. This platform is considered a better representation of the glycan environment as sensed by the bacteria *in vivo* (Kilcoyne, Gerlach et al. 2012). The commensals showed greatest binding to the mucins from GIT origin as expected. Improving the platform by using human mucins from various section of the GIT would be of interest to detect commensal particular niches. Moreover, the analysis of the glycans composing these mucins would be very useful to narrow down the ligands bound by commensals.

The bacterial cell has to expend a considerable metabolic effort for the synthesis of multimeric appendages such as pili (Müller, Åberg et al. 2009). As a result, regulation by phase variation could be a selective advantage for a particular strain. As described for the Type-1 fimbriae of *E. coli*, the ‘ON’ phase is suggested to cater for initial attachment to

mucins and an ‘OFF’ phase then follows with close adherence to its host maintained using other surface exposed molecules, such as MSCRAMM. Thus, the complex profiles of single binding snapshots might be due to the presence of a mixture of cells in different phases in the sample. This emphasises the need for recombinant versions of the adhesins to enable full characterisation of binding activities. Alternatively, the NGC and/or mucin profiles could be used to generate affinity columns for isolation of individual adhesins from surface-exposed proteins for further characterisation.

The third phase of this study involved an evaluation of the potential of another type of glycosylation-related microarray, the lectin microarray, to fingerprint commensal species in order to gain insights into cell surface glycosylation and, also, to test if this platform could be used as a screening tool for novel exopolysaccharides (EPSs) (Chapter 5). The lectin platform was first tested using a *Lactobacillus* strain, LGG, producing a long galactose-rich EPS the structure of which is known (Lebeer, Verhoeven et al. 2009). The lectin-binding profile of whole LGG cells correlated with the described EPS structure and inhibition experiments suggested the surface-bound EPS was responsible for the whole cell profile. Analysis of the lectin-binding profiles of the selected commensal strains showed that each profile was unique and, although some clustering of profiles was observed, it did not correlate with phylogeny. The profiles may instead reflect the niche or the source of the strains, as illustrated by the distinct lectin signatures given by three *Lactobacillus* strains, originating from human milk (*L. paracasei*), blood (*L. johnsonii*) or faeces (LGG). Further studies using strains of the same species originating from the human GIT and from dairy sources such as *L. rhamnosus* strains would be of interest to confirm this hypothesis. That the differences in lectin-binding profiles of the *Lactobacillus* strains indicate differences in surface glycosylation is supported by the similarity of profiles given by isogenic EPS mutants, derived from the same strain of *Bifidobacterium breve* UCC2003, whereas the profiles of the two non-EPS producers were significantly different. These displayed very high affinity towards wheat germ agglutinin (WGA), suggesting WGA was in fact binding the ‘bare’ cell wall (peptidoglycan layer).

There is increasing evidence that the EPSs produced by commensal species play an important role in commensal biology and in the benefits commensals exert on the host. Progress in the study of EPSs is, however, very slow as current analytical methods are very complex and labour-intensive. Thus, it was of interest to explore whether the lectin microarray platform could be used to discriminate between EPS species and thus provide a

useful screening tool for the study of EPSs. Using two different methods for isolation of EPS, chelation-based (bEPS) or ultrafiltration of the media (rEPS), and a novel labelling procedure, developed in-house, unique binding profiles were obtained for different EPS preparations from the species studied, which could be correlated with differences in the composition of each EPS preparation, as determined by monosaccharide analysis. Interestingly, the lectin-binding profiles of EPS isolated from the wild-type *B. breve* and the 'alternate' EPS mutant (INV) showed several differences, especially the bEPS preparation, supporting the clustering given by the corresponding whole cells.

While we have shown that lectin-binding profiles of bacterial cells or bacterial-derived carbohydrates can discriminate between species and molecules, it must be remembered that the specificities and affinities of most lectins have been studied with mammalian glycans. Bacterial polysaccharides contain monosaccharides, linkages and substituted groups not found in mammals. Thus, expanding the library of carbohydrate-binding proteins recognising bacterial sugars would be a major step forward in furthering bacterial polysaccharide structure prediction on lectin microarrays. In particular, binders for α-L-rhamnose, L-glycero-α-D-manno-heptose, α-D-galacturonic acid and α-Kdo, the most frequently-occurring residues unique to bacteria (Marino, Bones et al. 2010), would be expected to improve the performance of the lectin microarray platform for bacterial profiling. Nonetheless, we have shown that the mostly plant lectin-based microarray platform used in this study could be of value as a rapid screening tool to differentiate bacterial strains or EPS types or in the identification of novel bacterial strains with desirable properties in strain selection/mutation studies. This would in turn facilitate studies investigating the beneficial effects of various EPS structures on the host, including their prebiotic or immunomodulatory roles.

The results presented in this thesis have furthered our understanding of commensal glycobiology and uncovered possible new effector molecules involved in the interplay between prominent commensal bacterial strains and the gut mucosa. Along with the increasing technological advances in *omics*, this thesis could form the basis of future collaborative projects that could lead to commercial applications. These could include the development of probiotics or nutraceuticals for prophylaxis or treatment of inflammatory diseases such as IBD.

Bibliography

Abbad Andaloussi, S., H. Talbaoui, et al. (1995). "Isolation and characterization of exocellular polysaccharides produced by *Bifidobacterium longum*." Appl Microbiol Biotechnol **43**(6): 995-1000.

Adamczyk, B., W. B. Struwe, et al. (2013). "Characterization of fibrinogen glycosylation and its importance for serum/plasma N-glycome analysis." J Proteome Res **12**(1): 444-454.

Adams, R. A., C. Schachtrup, et al. (2007). "Fibrinogen signal transduction as a mediator and therapeutic target in inflammation: lessons from multiple sclerosis." Curr Med Chem **14**(27): 2925-2936.

Agris, P. F. (2004). "Decoding the genome: a modified view." Nucleic Acids Research **32**(1): 223-238.

Alemka, A., M. Clyne, et al. (2010). "Probiotic Colonization of the Adherent Mucus Layer of HT29MTXE12 Cells Attenuates *Campylobacter jejuni* Virulence Properties." Infection and Immunity **78**(6): 2812-2822.

Alemka, A., S. Whelan, et al. (2010). "Purified chicken intestinal mucin attenuates *Campylobacter jejuni* pathogenicity in vitro." Journal of Medical Microbiology **59**(8): 898-903.

Alp, G., B. Aslim, et al. (2010). "The role of hemagglutination and effect of exopolysaccharide production on bifidobacteria adhesion to Caco-2 cells in vitro." Microbiology and Immunology **54**(11): 658-665.

Altermann, E. (2005). "Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM." Proc. Natl Acad. Sci. USA **102**: 3906-3912.

Amieva, M. and E. El-Omar (2008). "Host-bacterial interactions in *Helicobacter pylori* infection." Gastroenterology **134**: 306 - 323.

Andrews, R. K., L. J. Bendall, et al. (1995). "Inhibition of Binding of von Willebrand Factor to the Platelet Glycoprotein Ib-IX Complex, Heparin and Sulfatides by Polyanionic Compounds. The Mechanism of Modulation of the Adhesive Function of von Willebrand Factor." Platelets **6**(5): 252-258.

Archambaud, M., P. Courcoux, et al. (1988). "Detection by molecular hybridization of pap, afa, and sfa adherence systems in *Escherichia coli* strains associated with urinary and enteral infections." Ann Inst Pasteur Microbiol **139**(5): 575-588.

Arumugam, M., J. Raes, et al. (2011). "Enterotypes of the human gut microbiome." Nature **473**(7346): 174-180.

Arvidson, C. G. and M. So (1995). "The *Neisseria* Transcriptional Regulator PilA Has a GTPase Activity." Journal of Biological Chemistry **270**(44): 26045-26048.

Atuma, C., V. Strugala, et al. (2001). "The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo." Am J Physiol Gastrointest Liver Physiol **280**(5): G922-929.

Audfray, A., J. Claudinon, et al. (2012). "Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes." J Biol Chem **287**(6): 4335-4347.

Backert, S., M. Clyne, et al. (2011). "Molecular mechanisms of gastric epithelial cell adhesion and injection of CagA by *Helicobacter pylori*." Cell Communication and Signaling **9**(1): 28.

Backhed, F., R. E. Ley, et al. (2005). "Host-bacterial mutualism in the human intestine." Science **307**: 1915-1920.

Badea, L., S. Doughty, et al. (2003). "Contribution of Efa1/LifA to the adherence of enteropathogenic *Escherichia coli* to epithelial cells." Microb Pathog **34**(5): 205-215.

Barinov, A. (2009). An in Silico Approach to the Identification of Proteins Involved in Bacteria-host Interactions: A Case-study of *Lactobacillus delbrueckii* ssp. *bulgaricus* and related lactobacilli. Thesis.

Barinov, A., V. Loux, et al. (2009). "Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria." PROTEOMICS **9**(1): 61-73.

Barocchi, M. A., J. Ries, et al. (2006). "A pneumococcal pilus influences virulence and host inflammatory responses." Proc Natl Acad Sci U S A **103**(8): 2857-2862.
Barre, A., Y. Bourne, et al. (2001). "Mannose-binding plant lectins: different structural scaffolds for a common sugar-recognition process." Biochimie **83**(7): 645-651.
Batchelor, M., S. Prasannan, et al. (2000). "Structural basis for recognition of the translocated intimin receptor (Tir) by intimin from enteropathogenic *Escherichia coli*." EMBO J **19**(11): 2452-2464.
Becherelli, M., A. G. O. Manetti, et al. (2012). "The ancillary protein 1 of *Streptococcus pyogenes* FCT-1 pili mediates cell adhesion and biofilm formation through heterophilic as well as homophilic interactions." Molecular Microbiology **83**(5): 1035-1047.
Behrens, I., P. Stenberg, et al. (2001). "Transport of lipophilic drug molecules in a new mucus-secreting cell culture model based on HT29-MTX cells." Pharm Res **18**(8): 1138-1145.
Belley, A., K. Keller, et al. (1999). "Intestinal mucins in colonization and host defense against pathogens." Am J Trop Med Hyg **60**(90040): 10-15.
Bello, F. D., J. Walter, et al. (2001). "In vitro study of Prebiotic Properties of Levan-type Exopolysaccharides from Lactobacilli and Non-digestible Carbohydrates Using Denaturing Gradient Gel Electrophoresis." Systematic and Applied Microbiology **24**(2): 232-237.
Belzer, C. and W. M. de Vos (2012). "Microbes inside-from diversity to function: the case of *Akkermansia*." Isme J **6**(8): 1449-1458.
Bendali, F., N. Madi, et al. (2011). "Beneficial effects of a strain of *Lactobacillus paracasei* subsp. *paracasei* in *Staphylococcus aureus*-induced intestinal and colonic injury." International Journal of Infectious Diseases **15**(11): e787-e794.
Bendtsen, J. D., H. Nielsen, et al. (2004). "Improved prediction of signal peptides: SignalP 3.0." J Mol Biol **340**(4): 783-795.
Berg, R. D. (1999). "Bacterial translocation from the gastrointestinal tract." Adv Exp Med Biol **473**: 11-30.
Bergstrom, K. S. and L. Xia (2013). "Mucin-type *O*-glycans and their roles in intestinal homeostasis." Glycobiology **23**(9): 1026-1037.
Biagi, E., L. Nylund, et al. (2010). "Through Ageing, and Beyond: Gut Microbiota and Inflammatory Status in Seniors and Centenarians." PLoS One **5**(5): e10667.
Bielaszewska, M., R. Prager, et al. (2009). "Detection and characterization of the fimbrial sfp cluster in enterohemorrhagic *Escherichia coli* O165:H25/NM isolates from humans and cattle." Appl Environ Microbiol **75**(1): 64-71.
Bilge, S. S., C. R. Clausen, et al. (1989). "Molecular characterization of a fimbrial adhesin, F1845, mediating diffuse adherence of diarrhea-associated *Escherichia coli* to HEp-2 cells." J Bacteriol **171**(8): 4281-4289.
Biol-N'Garagba, M. C., S. Greco, et al. (2002). "Polyamine participation in the maturation of glycoprotein fucosylation, but not sialylation, in rat small intestine." Pediatric Research **51**(5): 625-634.
Bjerke, G. A., R. Wilson, et al. (2011). "Mother-to-Child Transmission of and Multiple-Strain Colonization by *Bacteroides fragilis* in a Cohort of Mothers and Their Children." Applied and Environmental Microbiology **77**(23): 8318-8324.
Blackburn, D., A. Husband, et al. (2009). "Distribution of the *Escherichia coli* Common Pilus among Diverse Strains of Human Enterotoxigenic *E. coli*." Journal of Clinical Microbiology **47**(6): 1781-1784.
Blyn, L. B., B. A. Braaten, et al. (1990). "Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states." Embo J **9**(12): 4045-4054.
Boehm, G., J. Jelinek, et al. (2004). "Prebiotics in infant formulas." J Clin Gastroenterol **38**(6 Suppl): S76-79.
Boekhorst, J., M. W. de Been, et al. (2005). "Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs." J Bacteriol **187**(14): 4928-4934.
Boisen, N., C. Struve, et al. (2008). "New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family." Infect Immun **76**(7): 3281-3292.
Bonazzi, M., M. Lecuit, et al. (2009). "*Listeria monocytogenes* internalin and E-cadherin: from structure to pathogenesis." Cellular Microbiology **11**(5): 693-702.

Booth, W. J., F. H. Furby, et al. (1984). "Factor VIII/von Willebrand factor has potent lectin activity." Biochem Biophys Res Commun **118**(2): 495-501.
Bos, M. P., V. Robert, et al. (2007). "Biogenesis of the gram-negative bacterial outer membrane." Annu Rev Microbiol **61**: 191-214.
Brandtzaeg, P. (2013). "Secretory IgA: Designed for anti-microbial defense." Frontiers in Immunology **4**(AUG).
Bravo, J. A., P. Forsythe, et al. (2011). "Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve." Proc Natl Acad Sci U S A **108**(38): 16050-16055.
Brenner, D. M. and W. D. Chey (2009). "*Bifidobacterium infantis* 35624: a novel probiotic for the treatment of irritable bowel syndrome." Rev Gastroenterol Disord **9**(1): 7-15.
Brockhausen I, S. H., Stanley P. (1999). "O-GalNAc Glycans. Chapter 9." In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd edition.
Buck, B. L., E. Altermann, et al. (2005). "Functional Analysis of Putative Adhesion Factors in *Lactobacillus acidophilus* NCFM." Applied and Environmental Microbiology **71**(12): 8344-8351.
Bugert, P. and K. Geider (1995). "Molecular analysis of the ams operon required for exopolysaccharide synthesis of *Erwinia amylovora*." Molecular Microbiology **15**(5): 917-933.
Burrows, L. L. (2012). "*Pseudomonas aeruginosa* twitching motility: type IV pili in action." Annu Rev Microbiol **66**: 493-520.
Busch, A. and G. Waksman (2012). "Chaperone-usher pathways: diversity and pilus assembly mechanism." Philosophical transactions of the Royal Society of London. Series B, Biological sciences **367**(1592): 1112-1122.
Buts, L., J. Bouckaert, et al. (2003). "The fimbrial adhesin F17-G of enterotoxigenic *Escherichia coli* has an immunoglobulin-like lectin domain that binds *N*-acetylglucosamine." Mol Microbiol **49**(3): 705-715.
Byrd, J. and R. Bresalier (2004). "Mucins and mucin binding proteins in colorectal cancer." Cancer and Metastasis Reviews **23**(1-2): 77-99.
Caballero-Franco, C., K. Keller, et al. (2007). "The VSL#3 probiotic formula induces mucin gene expression and secretion in colonic epithelial cells." Am J Physiol Gastrointest Liver Physiol **292**(1): 14.
Cahenzli, J., M. L. Balmer, et al. (2013). "Microbial-immune cross-talk and regulation of the immune system." Immunology **138**(1): 12-22.
Caimano, M. J., R. Iyer, et al. (2007). "Analysis of the RpoS regulon in *Borrelia burgdorferi* in response to mammalian host signals provides insight into RpoS function during the enzootic cycle." Mol Microbiol **65**(5): 1193-1217.
Candela, M., F. Perna, et al. (2008). "Interaction of probiotic *Lactobacillus* and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production." Int J Food Microbiol **125**(3): 286-292.
Carlow, D. A., K. Gossens, et al. (2009). "PSGL-1 function in immunity and steady state homeostasis." Immunol Rev **230**(1): 75-96.
Carrington, S. D., M. Clyne, et al. (2010). Chapter 33 - Microbial interaction with mucus and mucins. Microbial Glycobiology. O. Holst, P. J. Brennan, M. v. Itzstein and A. P. Moran. San Diego, Academic Press**:** 655-671.
Carvalho, E., A. T. Ching Ching, et al. (2012). "Breaking the bond: recent patents on bacterial adhesins." Recent patents on DNA & gene sequences **6**(2): 160-171.
Casadesús, J. and R. D'Ari (2002). "Memory in bacteria and phage." BioEssays **24**(6): 512-518.
Celej, M. S., S. A. Dassie, et al. (2005). "Ligand-induced thermostability in proteins: thermodynamic analysis of ANS-albumin interaction." Biochim Biophys Acta **30**(2): 122-133.
Chagnot, C., A. Listrat, et al. (2012). "Bacterial adhesion to animal tissues: protein determinants for recognition of extracellular matrix components." Cellular Microbiology **14**(11): 1687-1696.

Chiang, S.-S. and T.-M. Pan (2012). "Beneficial effects of *Lactobacillus paracasei* subsp. *paracasei* NTU 101 and its fermented products." Applied Microbiology and Biotechnology **93**(3): 903-916.

Chow, J. and S. K. Mazmanian (2010). "A Pathobiont of the Microbiota Balances Host Colonization and Intestinal Inflammation." Cell Host & Microbe **7**(4): 265-276.

Christensen, B. E. (1989). "The role of extracellular polysaccharides in biofilms." Journal of Biotechnology **10**(3–4): 181-202.

Claesson, M. J. (2006). "Multireplicon genome architecture of *Lactobacillus salivarius*." Proc. Natl Acad. Sci. USA **103**: 6718-6723.

Claesson, M. J., S. Cusack, et al. (2011). "Composition, variability, and temporal stability of the intestinal microbiota of the elderly." Proceedings of the National Academy of Sciences **108**(Supplement 1): 4586-4591.

Clarke, S. C., R. D. Haigh, et al. (2003). "Virulence of enteropathogenic *Escherichia coli*, a global pathogen." Clin Microbiol Rev **16**(3): 365-378.

Cleary, J., L. C. Lai, et al. (2004). "Enteropathogenic *Escherichia coli* (EPEC) adhesion to intestinal epithelial cells: role of bundle-forming pili (BFP), EspA filaments and intimin." Microbiology **150**(Pt 3): 527-538.

Clegg, S., J. Wilson, et al. (2011). "More than One Way To Control Hair Growth: Regulatory Mechanisms in Enterobacteria That Affect Fimbriae Assembled by the Chaperone/Usher Pathway." Journal of Bacteriology **193**(9): 2081-2088.

Collado, M., J. Meriluoto, et al. (2008). "Adhesion and aggregation properties of probiotic and pathogen strains." European Food Research and Technology **226**(5): 1065-1073.

Collins, M. D., B. A. Phillips, et al. (1989). "Deoxyribonucleic-Acid Homology Studies of *Lactobacillus casei*, *Lactobacillus paracasei* sp-Nov, subsp *paracasei* and subsp *tolerans*, and *Lactobacillus rhamnosus* sp-Nov, Comb-Nov." International Journal of Systematic Bacteriology **39**(2): 105-108.

Collins, S. M. and P. Bercik (2009). "The relationship between intestinal microbiota and the central nervous system in normal gastrointestinal function and disease." Gastroenterology **136**(6): 2003-2014.

Collinson, S. K., S. C. Clouthier, et al. (1996). "*Salmonella enteritidis* agfBAC operon encoding thin, aggregative fimbriae." J Bacteriol **178**(3): 662-667.

Corthésy, B., H. R. Gaskins, et al. (2007). "Cross-Talk between Probiotic Bacteria and the Host Immune System." The Journal of Nutrition **137**(3): 781S-790S.

Cryan, J. F. and T. G. Dinan (2012). "Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour." Nat Rev Neurosci **13**(10): 701-712.

Cuesta, G., N. Suarez, et al. (2003). "Quantitative determination of pneumococcal capsular polysaccharide serotype 14 using a modification of phenol-sulfuric acid method." Journal of Microbiological Methods **52**(1): 69-73.

Cullen, P. A., X. Xu, et al. (2005). "Surfaceome of *Leptospira* spp." Infect Immun **73**(8): 4853-4863.

Culligan, E., C. Hill, et al. (2009). "Probiotics and gastrointestinal disease: successes, problems and future prospects." Gut Pathogens **1**(1): 19.

Dalziel, M., M. Crispin, et al. (2014). "Emerging Principles for the Therapeutic Exploitation of Glycosylation." Science **343**(6166).

Davalos, D. and K. Akassoglou (2012). "Fibrinogen as a key regulator of inflammation in disease." Seminars in Immunopathology **34**(1): 43-62.

Davicino, R. C., R. J. Elicabe, et al. (2011). "Coupling pathogen recognition to innate immunity through glycan-dependent mechanisms." Int Immunopharmacol **11**(10): 1457-1463.

De Angelis, M., M. Piccolo, et al. (2013). "Fecal Microbiota and Metabolome of Children with Autism and Pervasive Developmental Disorder Not Otherwise Specified." PLoS One **8**(10): e76993.

de Jong, P., M. M. M. Vissers, et al. (2007). "*In Silico* Model as a Tool for Interpretation of Intestinal Infection Studies." Applied and Environmental Microbiology **73**(2): 508-515.

De Vos, W. M. (1987). "Gene cloning and expression in lactic streptococci." Fems Microbiology Letters **46**(3): 281-295.

De Vuyst, L. and B. Degeest (1999). "Heteropolysaccharides from lactic acid bacteria." FEMS Microbiology Reviews **23**(2): 153-177.
Dekker, J., J. W. Rossen, et al. (2002). "The MUC family: an obituary." Trends Biochem Sci **27**(3): 126-131.
Del Re, B., B. Sgorbati, et al. (2000). "Adhesion, autoaggregation and hydrophobicity of 13 strains of *Bifidobacterium longum*." Lett Appl Microbiol **31**(6): 438-442.
Deplancke, B. and H. R. Gaskins (2001). "Microbial modulation of innate defense: goblet cells and the intestinal mucus layer." The American Journal of Clinical Nutrition **73**(6): 1131S-1141S.
Dertli, E., I. J. Colquhoun, et al. (2013). "Structure and Biosynthesis of Two Exopolysaccharides Produced by *Lactobacillus johnsonii* FI9785." The Journal of biological chemistry **288**(44): 31938-31951.
Desvaux, M., E. Dumas, et al. (2006). "Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure." Fems Microbiology Letters **256**(1): 1-15.
Dethlefsen, L., P. B. Eckburg, et al. (2006). "Assembly of the human intestinal microbiota." Trends in Ecology & Evolution **21**(9): 517-523.
Dhakal, B. K., R. R. Kulesus, et al. (2008). "Mechanisms and consequences of bladder cell invasion by uropathogenic *Escherichia coli*." European Journal of Clinical Investigation **38**: 2-11.
Diaz Heijtz, R., S. Wang, et al. (2011). "Normal gut microbiota modulates brain development and behavior." Proc Natl Acad Sci U S A **108**(7): 3047-3052.
Dore, J. and G. Corthier (2010). "The human intestinal microbiota." Gastroenterol Clin Biol **34**(1): 70015-70014.
Doron, S., D. R. Snydman, et al. (2005). "*Lactobacillus* GG: Bacteriology and Clinical Applications." Gastroenterology Clinics of North America **34**(3): 483-498.
Dorsey, C. W., M. C. Laarakker, et al. (2005). "*Salmonella enterica* serotype Typhimurium MisL is an intestinal colonization factor that binds fibronectin." Mol Microbiol **57**(1): 196-211.
Duncan, S., G. Hold, et al. (2002). "Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov." Int J Syst Evol Microbiol **52**: 2141 - 2146.
Dunne, W. M. (2002). "Bacterial Adhesion: Seen Any Good Biofilms Lately?" Clinical Microbiology Reviews **15**(2): 155-166.
Eddy, S. R. (1998). "Profile hidden Markov models." Bioinformatics **14**(9): 755-763.
Eisenstein, B. I. (1981). "Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control." Science **214**(4518): 337-339.
Estrada-Garcia, T., K. Hodges, et al. (2013). Chapter 8 - *Escherichia coli*. Foodborne Infections and Intoxications (Fourth Edition). J. M. Glenn and P. Morris. San Diego, Academic Press**:** 129-164.
Etzold, S., D. A. MacKenzie, et al. (2014). "Structural and molecular insights into novel surface-exposed mucus adhesins from Lactobacillus reuteri human strains." Mol Microbiol **92**(3): 543-556.
Faherty, C. S., J. C. Redman, et al. (2012). "*Shigella flexneri* effectors OspE1 and OspE2 mediate induced adherence to the colonic epithelium following bile salts exposure." Mol Microbiol **85**(1): 107-121.
Fanning, S., L. J. Hall, et al. (2012). "Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection." Proceedings of the National Academy of Sciences.
Feng, T. and C. O. Elson (2011). "Adaptive immunity in the host-microbiota dialog." Mucosal Immunol **4**(1): 15-21.
Fitzhenry, R. J., D. J. Pickard, et al. (2002). "Intimin type influences the site of human intestinal mucosal colonisation by enterohaemorrhagic *Escherichia coli* O157:H7." Gut **50**(2): 180-185.
Flugel, A., H. Schulze-Koops, et al. (1994). "Interaction of enteropathogenic *Yersinia enterocolitica* with complex basement membranes and the extracellular matrix proteins

collagen type IV, laminin-1 and -2, and nidogen/entactin." J Biol Chem **269**(47): 29732-29738.
Folch, J. L., M. and Stanley, G.H.S. (1957). "A simple method for the isolation and purification of total lipides from animal tissues." J. Biol. Chem. **226**: (1957).
Foroni, E., F. Serafini, et al. (2011). "Genetic analysis and morphological identification of pilus-like structures in members of the genus *Bifidobacterium*." Microb Cell Fact **10 Suppl 1**: S16.
Francius, G. g., S. Lebeer, et al. (2008). "Detection, Localization, and Conformational Analysis of Single Polysaccharide Molecules on Live Bacteria." ACS Nano **2**(9): 1921-1929.
Franks, A. H., H. J. M. Harmsen, et al. (1998). "Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes." Applied and Environmental Microbiology **64**(9): 3336-3345.
Fujimoto, T., H. Imaeda, et al. (2013). "Decreased abundance of *Faecalibacterium prausnitzii* in the gut microbiota of Crohn's disease." J Gastroenterol Hepatol **28**(4): 613-619.
Fujisawa, T., Y. Benno, et al. (1992). "Taxonomic study of the *Lactobacillus acidophilus* group, with recognition of *Lactobacillus gallinarum* sp. nov. and *Lactobacillus johnsonii* sp. nov. and synonymy of *Lactobacillus acidophilus* group A3 (Johnson et al. 1980) with the type strain of *Lactobacillus amylovorus* (Nakamura 1981)." International Journal of Systematic Bacteriology **42**(3): 487-491.
Fukuda, S., H. Toh, et al. (2011). "Bifidobacteria can protect from enteropathogenic infection through production of acetate." Nature **469**: 543 - U791.
Funkhouser, L. J. and S. R. Bordenstein (2013). "Mom knows best: the universality of maternal microbial transmission." PLoS Biol **11**(8): 20.
Gaastra, W. and F. K. de Graaf (1982). "Host-specific fimbrial adhesins of noninvasive enterotoxigenic *Escherichia coli strains*." Microbiol Rev **46**(2): 129-161.
Gaastra, W. and A. M. Svennerholm (1996). "Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC)." Trends Microbiol **4**(11): 444-452.
Gabius, H. J., S. Andre, et al. (2011). "From lectin structure to functional glycomics: principles of the sugar code." Trends Biochem Sci **36**(6): 298-313.
Galdeano, C. M., A. de Moreno de LeBlanc, et al. (2007). "Proposed Model: Mechanisms of Immunomodulation Induced by Probiotic Bacteria." Clinical and Vaccine Immunology **14**(5): 485-492.
Galecka, M., P. Szachta, et al. (2013). "*Faecalibacterium prausnitzii* and Crohn's disease - is there any connection?" Pol J Microbiol **62**(1): 91-95.
Garcia, M. I., A. Labigne, et al. (1994). "Nucleotide sequence of the afimbrial-adhesin-encoding afa-3 gene cluster and its translocation via flanking IS1 insertion sequences." J Bacteriol **176**(24): 7601-7613.
Gardy, J. L., M. R. Laird, et al. (2005). "PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis." Bioinformatics **21**(5): 617-623.
Garrido, D., J. H. Kim, et al. (2011). "Oligosaccharide Binding Proteins from *Bifidobacterium longum* subsp. *infantis* Reveal a Preference for Host Glycans." PLoS One **6**(3): e17315.
Gasson, M. J. (1983). "Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing." J Bacteriol **154**(1): 1-9.
Gerlach, J. Q., M. Kilcoyne, et al. (2014). "Microarray evaluation of the effects of lectin and glycoprotein orientation and data filtering on glycoform discrimination." Analytical Methods **6**(2): 440-449.
Gerlach, J. Q., A. Krüger, et al. (2013). "Surface Glycosylation Profiles of Urine Extracellular Vesicles." PLoS One **8**(9): e74801.
Gerlach, R. G. and M. Hensel (2007). "Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens." Int J Med Microbiol **297**(6): 401-415.
Ghoddusi, H. B. (2012). "Lactic Acid Bacteria and Bifidobacteria: Current Progress in Advanced Research (2011)." International Journal of Dairy Technology **65**(3): 462-464.

Górska-Frączek, S., C. Sandström, et al. (2011). "Structural studies of the exopolysaccharide consisting of a nonasaccharide repeating unit isolated from Lactobacillus rhamnosus KL37B." Carbohydrate Research **346**(18): 2926-2932.
Gorska, S., P. Grycko, et al. (2007). "Exopolysaccharides of lactic acid bacteria: structure and biosynthesis." Postepy Hig Med Dosw **61**: 805-818.
Gorska, S., W. Jachymek, et al. (2010). "Structural and immunochemical studies of neutral exopolysaccharide produced by Lactobacillus johnsonii 142." Carbohydr Res **345**(1): 108-114.
Gophna, U., M. Barlev, et al. (2001). "Curli fibers mediate internalization of *Escherichia coli* by eukaryotic cells." Infect Immun **69**(4): 2659-2665.
Granato, D., G. E. Bergonzelli, et al. (2004). "Cell surface-associated elongation factor Tu mediates the attachment of *Lactobacillus johnsonii* NCC533 (La1) to human intestinal cells and mucins." Infect Immun **72**(4): 2160-2169.
Grangette, C. (2005). "Enhanced antiinflammatory capacity of a *Lactobacillus plantarum* mutant synthesizing modified teichoic acids." Proc. Natl Acad. Sci. USA **102**: 10321-10326.
Green, E. D., G. Adelt, et al. (1988). "The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of *N*-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy." Journal of Biological Chemistry **263**(34): 18253-18268.
Greene, J. D. and T. R. Klaenhammer (1994). "Factors involved in adherence of lactobacilli to human Caco-2 cells." Appl Environ Microbiol **60**(12): 4487-4494.
Gross, G., J. van der Meulen, et al. (2008). "Mannose-specific interaction of *Lactobacillus plantarum* with porcine jejunal epithelium." FEMS Immunol Med Microbiol **54**(2): 215-223.
Gudina, E. J., J. A. Teixeira, et al. (2010). "Isolation and functional characterization of a biosurfactant produced by *Lactobacillus paracasei.*" Colloids Surf B Biointerfaces **76**(1): 298-304.
Guglielmetti, S., I. Tamagnini, et al. (2009). "Study of the adhesion of *Bifidobacterium bifidum* MIMBb75 to human intestinal cell lines." Curr Microbiol **59**(2): 167-172.
Gupta, G., A. Surolia, et al. (2010). "Lectin microarrays for glycomic analysis." Omics **14**(4): 419-436.
Habu, Y., M. Nagaoka, et al. (1987). "Structural studies of cell wall polysaccharides from *Bifidobacterium breve* YIT 4010 and related *Bifidobacterium* species." J Biochem **102**(6): 1423-1432.
Hakomori, S. and Y. Igarashi (1995). "Functional role of glycosphingolipids in cell recognition and signaling." J Biochem **118**(6): 1091-1103.
Hamer, H. M., D. Jonkers, et al. (2008). "Review article: the role of butyrate on colonic function." Alimentary Pharmacology & Therapeutics **27**(2): 104-119.
Hanson, M. S., J. Hempel, et al. (1988). "Purification of the *Escherichia coli* type 1 pilin and minor pilus proteins and partial characterization of the adhesin protein." Journal of Bacteriology **170**(8): 3350-3358.
Harazono, A., N. Kawasaki, et al. (2006). "Site-specific *N*-glycosylation analysis of human plasma ceruloplasmin using liquid chromatography with electrospray ionization tandem mass spectrometry." Anal Biochem **348**(2): 259-268.
Hauck, C. R. and K. Ohlsen (2006). "Sticky connections: extracellular matrix protein recognition and integrin-mediated cellular invasion by *Staphylococcus aureus*." Current Opinion in Microbiology **9**(1): 5-11.
Hawrelak, J. A. and S. P. Myers (2004). "The causes of intestinal dysbiosis: a review." Altern Med Rev **9**(2): 180-197.
He, B., W. Xu, et al. (2007). "Intestinal Bacteria Trigger T Cell-Independent Immunoglobulin A2 Class Switching by Inducing Epithelial-Cell Secretion of the Cytokine APRIL." Immunity **26**(6): 812-826.
Hendrickx, A. P., J. M. Budzik, et al. (2011). "Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds." Nat Rev Microbiol **9**(3): 166-176.
Henry, S., R. Oriol, et al. (1995). "Lewis histo-blood group system and associated secretory phenotypes." Vox Sang **69**(3): 166-182.

Herget, S., P. V. Toukach, et al. (2008). "Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans." BMC Struct Biol **8**(35): 1472-6807.

Hirabayashi, J., M. Yamada, et al. (2013). "Lectin microarrays: concept, principle and applications." Chem Soc Rev **42**(10): 4443-4458.

Holden, N. J., M. Totsika, et al. (2006). "Demonstration of regulatory cross-talk between P fimbriae and type 1 fimbriae in uropathogenic Escherichia coli." Microbiology **152**(Pt 4): 1143-1153.

Holmes, A. R., R. McNab, et al. (2001). "The pavA gene of Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence." Mol Microbiol **41**(6): 1395-1408.

Holst, O., A. P. Moran, et al. (2010). Chapter 1 - Overview of the glycosylated components of the bacterial cell envelope. Microbial Glycobiology. O. Holst, P. J. Brennan, M. v. Itzstein and A. P. Moran. San Diego, Academic Press**:** 1-13.

Hooper, L. V. and J. I. Gordon (2001). "Commensal host-bacterial relationships in the gut." Science **292**: 1115-1118.

Hooper, L. V., T. Midtvedt, et al. (2002). "How host-microbial interactions shape the nutrient environment of the mammalian intestine." Annu. Rev. Nutr. **22**: 283-307.

Houeix, B., M. Le Berre, et al. (2012). Lectin profiling of bacterial exopolysaccharides using microarray technology, Increasing the Impact of Glycoscience through New Tools and Technologies. International Glycomics Symposium, San Sebastian, Spain, cicbiomagune.

Houston, S., G. W. Blakely, et al. (2010). "Binding and degradation of fibrinogen by *Bacteroides fragilis* and characterization of a 54 kDa fibrinogen-binding protein." Microbiology **156**(Pt 8): 2516-2526.

Hsiao, E. Y., S. W. McBride, et al. (2013). "Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders." Cell **155**(7): 1451-1463.

Hsu, K. L. and L. K. Mahal (2006). "A lectin microarray approach for the rapid analysis of bacterial glycans." Nat Protoc **1**(2): 543-549.

Hsu, K. L., K. T. Pilobello, et al. (2006). "Analyzing the dynamic bacterial glycome with a lectin microarray approach." Nat Chem Biol **2**(3): 153-157.

Hull, R. A., W. H. Donovan, et al. (2002). "Role of Type 1 Fimbria- and P Fimbria-Specific Adherence in Colonization of the Neurogenic Human Bladder by *Escherichia coli*." Infection and Immunity **70**(11): 6481-6484.

Hyland, R. M., J. Sun, et al. (2008). "The bundlin pilin protein of enteropathogenic *Escherichia coli* is an *N*-acetyllactosamine-specific lectin." Cellular Microbiology **10**(1): 177-187.

Ikezawa, H. (2002). "Glycosylphosphatidylinositol (GPI)-anchored proteins." Biol Pharm Bull **25**(4): 409-417.

Imberty, A., M. Wimmerova, et al. (2004). "Structures of the lectins from *Pseudomonas aeruginosa*: insights into the molecular basis for host glycan recognition." Microbes and Infection **6**(2): 221-228.

Inoue, T., I. Tanimoto, et al. (1998). "Molecular Characterization of Low-Molecular-Weight Component Protein, Flp, in *Actinobacillus actinomycetemcomitans* Fimbriae." Microbiology and Immunology **42**(4): 253-258.

Iskratsch, T., A. Braun, et al. (2009). "Specificity analysis of lectins and antibodies using remodeled glycoproteins." Analytical Biochemistry **386**(2): 133-146.

Jacobson, S. H. (1986). "P-fimbriated *Escherichia coli* in adults with renal scarring and pyelonephritis." Acta Medica Scandinavica **220**(S713): 1-64.

Jakava-Viljanen, M., S. Åvall-Jääskeläinen, et al. (2002). "Isolation of Three New Surface Layer Protein Genes (slp) from *Lactobacillus brevis* ATCC 14869 and Characterization of the Change in Their Expression under Aerated and Anaerobic Conditions." Journal of Bacteriology **184**(24): 6786-6795.

Jarczak, J., E. M. Kościuczuk, et al. (2013). "Defensins: Natural component of human innate immunity." Human Immunology **74**(9): 1069-1079.

Joh, D., E. R. Wann, et al. (1999). "Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells." Matrix Biol **18**(3): 211-223.
Johansson, M. E., M. Phillipson, et al. (2008). "The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria." Proc Natl Acad Sci U S A **105**(39): 15064-15069.
Johnson-Henry, K. C., K. E. Hagen, et al. (2007). "Surface-layer protein extracts from *Lactobacillus helveticus* inhibit enterohaemorrhagic *Escherichia coli* O157:H7 adhesion to epithelial cells." Cellular Microbiology **9**(2): 356-367.
Johnson, J. R., I. Orskov, et al. (1994). "O, K, and H Antigens Predict Virulence Factors, Carboxylesterase B Pattern, Antimicrobial Resistance, and Host Compromise among Escherichia coli Strains Causing Urosepsis." Journal of Infectious Diseases **169**(1): 119-126.
Jordan, D. M., N. Cornick, et al. (2004). "Long polar fimbriae contribute to colonization by Escherichia coli O157:H7 in vivo." Infect Immun **72**(10): 6168-6171.
Jost, T., C. Lacroix, et al. (2012). "New Insights in Gut Microbiota Establishment in Healthy Breast Fed Neonates." PLoS One **7**(8): e44595.
Jost Wingender, T. R. N., Hans-Curt Flemming (1999). "Microbial Extracellular Polymeric Substances: Characterization, Structure and Function." Springer: 171-200.
Juge, N. (2012). "Microbial adhesins to gastrointestinal mucus." Trends Microbiol **20**(1): 30-39.
Juncker, A. S., H. Willenbrock, et al. (2003). "Prediction of lipoprotein signal peptides in Gram-negative bacteria." Protein Sci **12**(8): 1652-1662.
Kahsay, R. Y., G. Gao, et al. (2005). "An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes." Bioinformatics **21**(9): 1853-1858.
Kajava, A. V., N. Cheng, et al. (2001). "Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins." Mol Microbiol **42**(2): 279-292.
Kalliomaki, M., S. Salminen, et al. (2001). "Probiotics in primary prevention of atopic disease: a randomised placebo-controlled trial." Lancet **357**(9262): 1076-1079.
Kalliomaki, M., S. Salminen, et al. (2003). "Probiotics and prevention of atopic disease: 4-year follow-up of a randomised placebo-controlled trial." Lancet **361**(9372): 1869-1871.
Kandori, H., K. Hirayama, et al. (1996). "Histochemical, lectin-histochemical and morphometrical characteristics of intestinal goblet cells of germfree and conventional mice." Exp Anim **45**: 155 - 160.
Kankainen, M., L. Paulin, et al. (2009). "Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein." Proc Natl Acad Sci U S A **106**: 17193 - 17198.
Kaper, J. B., J. P. Nataro, et al. (2004). "Pathogenic *Escherichia coli*." Nat Rev Microbiol **2**(2): 123-140.
Kaper, J. B. and V. Sperandio (2005). "Bacterial cell-to-cell signaling in the gastrointestinal tract." Infect Immun **73**(6): 3197-3209.
Karatzas, K.-A. G., O. Brennan, et al. (2010). "Intracellular Accumulation of High Levels of γ-Aminobutyrate by *Listeria monocytogenes* 10403S in Response to Low pH: Uncoupling of γ-Aminobutyrate Synthesis from Efflux in a Chemically Defined Medium." Applied and Environmental Microbiology **76**(11): 3529-3537.
Kasai, K.-i. and J. Hirabayashi (1996). "Galectins: A Family of Animal Lectins That Decipher Glycocodes." Journal of Biochemistry **119**(1): 1-8.
Kau, A. L., P. P. Ahern, et al. (2011). "Human nutrition, the gut microbiome and the immune system." Nature **474**(7351): 327-336.
Kazmierczak, M. J., M. Wiedmann, et al. (2005). "Alternative Sigma Factors and Their Roles in Bacterial Virulence." Microbiology and Molecular Biology Reviews **69**(4): 527-543.
Kilcoyne, M., J. Q. Gerlach, et al. (2012). "Construction of a Natural Mucin Microarray and Interrogation for Biologically Relevant Glyco-Epitopes." Analytical Chemistry **84**(7): 3330-3338.
Kilcoyne, M., J. Q. Gerlach, et al. (2012). "Surface chemistry and linker effects on lectin-carbohydrate recognition for glycan microarrays." Analytical Methods **4**(9): 2721-2728.

Kilcoyne, M., M. E. Twomey, et al. (2014). "*Campylobacter jejuni* strain discrimination and temperature-dependent glycome expression profiling by lectin microarray." Carbohydr Res **389**(0): 123-133.
Kim, H., E. Goo, et al. (2012). "Regulation of universal stress protein genes by quorum sensing and RpoS in *Burkholderia glumae*." J Bacteriol **194**(5): 982-992.
Kingsley, R. A., A. D. Humphries, et al. (2003). "Molecular and phenotypic analysis of the CS54 island of *Salmonella enterica* serotype Typhimurium: identification of intestinal colonization and persistence determinants." Infect Immun **71**(2): 629-640.
Kitaoka, M. (2012). "Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides." Adv Nutr **3**(3): 001420.
Klemm, P. and M. A. Schembri (2000). "Fimbrial surface display systems in bacteria: from vaccines to random libraries." Microbiology **146**(12): 3025-3032.
Kline, K. A., K. W. Dodson, et al. (2010). "A tale of two pili: assembly and function of pili in bacteria." Trends Microbiol **18**(5): 224-232.
Kline, K. A., S. Falker, et al. (2009). "Bacterial adhesins in host-microbe interactions." Cell Host Microbe **5**(6): 580-592.
Knibbs, R. N., F. Perini, et al. (1989). "Structure of the major concanavalin A reactive oligosaccharides of the extracellular matrix component laminin." Biochemistry **28**(15): 6379-6392.
Knight, S. D., J. Berglund, et al. (2000). "Bacterial adhesins: structural studies reveal chaperone function and pilus biogenesis." Current Opinion in Chemical Biology **4**(6): 653-660.
Knutton, S., R. K. Shaw, et al. (1999). "The type IV bundle-forming pilus of enteropathogenic *Escherichia coli* undergoes dramatic alterations in structure associated with bacterial adherence, aggregation and dispersal." Mol Microbiol **33**(3): 499-509.
Korea, C. G., J. M. Ghigo, et al. (2011). "The sweet connection: Solving the riddle of multiple sugar-binding fimbrial adhesins in *Escherichia coli*: Multiple *E. coli* fimbriae form a versatile arsenal of sugar-binding lectins potentially involved in surface-colonisation and tissue tropism." BioEssays **33**(4): 300-311.
Korakli, M., M. G. Gänzle, et al. (2002). "Metabolism by bifidobacteria and lactic acid bacteria of polysaccharides from wheat and rye, and exopolysaccharides produced by Lactobacillus sanfranciscensis." Journal of Applied Microbiology **92**(5): 958-965.
Kopitz, J., M. Bergmann, et al. (2010). "How adhesion/growth-regulatory galectins-1 and -3 attain cell specificity: case study defining their target on neuroblastoma cells (SK-N-MC) and marked affinity regulation by affecting microdomain organization of the membrane." IUBMB Life **62**(8): 624-628.
Korhonen, T. K., E. L. Nurmiaho, et al. (1980). "New Method for isolation of immunologically pure pili from *Escherichia coli*." Infect Immun **27**(2): 569-575.
Kornfeld, R. and S. Kornfeld (1985). "Assembly of asparagine-linked oligosaccharides." Annu Rev Biochem **54**: 631-664.
Koropatkin, N., E. Cameron, et al. (2012). "How glycan metabolism shapes the human gut microbiota." Nat Rev Microbiol **10**: 323 - 335.
Kovacs-Simon, A., R. W. Titball, et al. (2011). "Lipoproteins of Bacterial Pathogens." Infection and Immunity **79**(2): 548-561.
Kuehn, M. J., J. Heuser, et al. (1992). "P pili in uropathogenic *E. coli are* composite fibres with distinct fibrillar adhesive tips." Nature **356**(6366): 252-255.
Lalioui, L., M. Jouve, et al. (1999). "Molecular cloning and characterization of the afa-7 and afa-8 gene clusters encoding afimbrial adhesins in *Escherichia coli* strains associated with diarrhea or septicemia in calves." Infect Immun **67**(10): 5048-5059.
Landersjo, C., Z. Yang, et al. (2002). "Structural studies of the exopolysaccharide produced by *Lactobacillus rhamnosus* strain GG (ATCC 53103)." Biomacromolecules **3**(4): 880-884.
Lantz, M. S., R. D. Allen, et al. (1990). "*Bacteroides gingivalis* and *Bacteroides intermedius* recognize different sites on human fibrinogen." J Bacteriol **172**(2): 716-726.
Lantz, M. S., R. D. Allen, et al. (1991). "Specific cell components of *Bacteroides gingivalis* mediate binding and degradation of human fibrinogen." Journal of Bacteriology **173**(2): 495-504.

Laws, A., Y. Gu, et al. (2001). "Biosynthesis, characterisation, and design of bacterial exopolysaccharides from lactic acid bacteria." Biotechnology Advances **19**(8): 597-625.
Lebeer, S., J. Ceuppens, et al. (2007). "Mechanisms of probiotic-host interaction with IBD as a case study: a role for exopolysaccharides?" Commun Agric Appl Biol Sci **72**(1): 41-45.
Lebeer, S., I. J. J. Claes, et al. (2011). "Exopolysaccharides of *Lactobacillus rhamnosus* GG form a protective shield against innate immune factors in the intestine." Microbial Biotechnology **4**(3): 368-374.
Lebeer, S., J. Vanderleyden, et al. (2010). "Adaptation factors of the probiotic *Lactobacillus rhamnosus* GG." Beneficial Microbes **1**(4): 335-342.
Lebeer, S., J. Vanderleyden, et al. (2008). "Genes and Molecules of lactobacilli Supporting Probiotic Action." Microbiol. Mol. Biol. Rev. **72**(4): 728-764.
Lebeer, S., J. Vanderleyden, et al. (2010). "Host interactions of probiotic bacterial surface molecules: Comparison with commensals and pathogens." Nature Reviews Microbiology **8**(3): 171-184.
Lebeer, S., J. Vanderleyden, et al. (2010). "Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens." Nat Rev Micro **8**(3): 171-184.
Lebeer, S., T. L. Verhoeven, et al. (2009). "Identification of a Gene Cluster for the Biosynthesis of a Long, Galactose-Rich Exopolysaccharide in *Lactobacillus rhamnosus* GG and Functional Analysis of the Priming Glycosyltransferase." Appl Environ Microbiol **75**(11): 3554-3563.
LeBlanc, J. G., C. Milani, et al. (2013). "Bacteria as vitamin suppliers to their host: a gut microbiota perspective." Curr Opin Biotechnol **24**(2): 160-168.
Lee, I. C., S. Tomita, et al. (2013). "The quest for probiotic effector molecules--unraveling strain specificity at the molecular level." Pharmacol Res **69**(1): 61-74.
Lee, Y. K., C. Y. Lim, et al. (2000). "Quantitative approach in the study of adhesion of lactic acid bacteria to intestinal cells and their competition with enterobacteria." Appl Environ Microbiol **66**(9): 3692-3697.
Leivers, S., C. Hidalgo-Cantabrana, et al. (2011). "Structure of the high molecular weight exopolysaccharide produced by Bifidobacterium animalis subsp. lactis IPLA-R1 and sequence analysis of its putative eps cluster." Carbohydr Res **346**(17): 2710-2717.
Ley, K. and G. S. Kansas (2004). "Selectins in T-cell recruitment to non-lymphoid tissues and sites of inflammation." Nat Rev Immunol **4**(5): 325-335.
Ley, R., P. Turnbaugh, et al. (2006). "Microbial ecology - human gut microbes associated with obesity." Nature **444**: 1022 - 1023.
Ley, R. E., D. A. Peterson, et al. (2006). "Ecological and evolutionary forces shaping microbial diversity in the human intestine." Cell **124**: 837-848.
Li, Y.-F., S. Poole, et al. (2009). "Structure of CFA/I fimbriae from enterotoxigenic *Escherichia coli*." Proceedings of the National Academy of Sciences **106**(26): 10793-10798.
Lindahl, M. and T. Wadstrom (1984). "K99 surface haemagglutinin of enterotoxigenic *E. coli* recognize terminal N-acetylgalactosamine and sialic acid residues of glycophorin and other complex glycoconjugates." Vet Microbiol **9**(3): 249-257.
Lindberg, A. (1977). Bacterial Surface Carbohydrates and Bacteriophage Adsorption. Surface Carbohydrates of the Prokaryotic Cell. I. W. Sutherland. London; United Kingdom, Academic Press Inc. (London) Ltd.**:** -356.
Linden, S., J. Mahdavi, et al. (2008). "Role of ABO secretor status in mucosal innate immunity and H. pylori infection." PLoS Pathog **4**(1): 0040002.
Linke, D. and A. Goldman (2011). Bacterial adhesion. Advances in Experimental Medicine and Biology, Springer. **715**.
Lipinski, T., C. Jones, et al. (2003). "Structural analysis of the Lactobacillus rhamnosus strain KL37C exopolysaccharide." Carbohydr Res **338**(7): 605-609.
Littman, D. R. and E. G. Pamer (2011). "Role of the commensal microbiota in normal and pathogenic host immune responses." Cell Host Microbe **10**(4): 311-323.
Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Ljungh, Å., A. P. Moran, et al. (1996). "Interactions of bacterial adhesins with extracellular matrix and plasma proteins: pathogenic implications and therapeutic possibilities." FEMS Immunology and Medical Microbiology **16**(2): 117-126.

Lo, C. Y., A. Antonopoulos, et al. (2013). "Competition between core-2 GlcNAc-transferase and ST6GalNAc-transferase regulates the synthesis of the leukocyte selectin ligand on human P-selectin glycoprotein ligand-1." J Biol Chem **288**(20): 13974-13987.

LoCascio, R. G., M. R. Ninonuevo, et al. (2007). "Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation." J Agric Food Chem **55**(22): 8914-8919.

Lotan, R., N. Sharon, et al. (1975). "Interaction of wheat-germ agglutinin with bacterial cells and cell-wall polymers." Eur J Biochem **55**(1): 257-262.

Lupp, C. and B. B. Finlay (2005). "Intestinal microbiota." Current Biology **15**(7): R235-R236.

Ma, Y., M. O. Lassiter, et al. (1996). "Multiple glucan-binding proteins of *Streptococcus sobrinus*." J Bacteriol **178**(6): 1572-1577.

MacKenzie, D. A., L. E. Tailford, et al. (2009). "Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity." J Biol Chem **284**(47): 32444-32453.

Macpherson, A. J., M. B. Geuking, et al. (2012). "Homeland Security: IgA immunity at the frontiers of the body." Trends Immunol **33**(4): 160-167.

Macpherson, A. J. and E. Slack (2007). "The functional interactions of commensal bacteria with intestinal secretory IgA." Curr Opin Gastroenterol **23**(6): 673-678.

Mainil, J. (2013). "Escherichia coli virulence factors." Veterinary Immunology and Immunopathology **152**(1–2): 2-12.

Makino, H., A. Kushiro, et al. (2013). "Mother-to-Infant Transmission of Intestinal Bifidobacterial Strains Has an Impact on the Early Development of Vaginally Delivered Infant's Microbiota." PLoS One **8**(11): e78331.

Makras, L., G. Van Acker, et al. (2005). "*Lactobacillus paracasei* subsp. *paracasei* 8700:2 Degrades Inulin-Type Fructans Exhibiting Different Degrees of Polymerization." Applied and Environmental Microbiology **71**(11): 6531-6537.

Mandlik, A., A. Swierczynski, et al. (2008). "Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development." Trends Microbiol **16**(1): 33-40.

Manetti, A. G. O., T. Köller, et al. (2010). "Environmental Acidification Drives *S pyogenes* Pilus Expression and Microcolony Formation on Epithelial Cells in a FCT-Dependent Manner." PLoS One **5**(11): e13864.

Marchler-Bauer, A., S. Lu, et al. (2011). "CDD: a Conserved Domain Database for the functional annotation of proteins." Nucleic Acids Res **39**(Database issue): 24.

Marchler-Bauer, A., C. Zheng, et al. (2013). "CDD: conserved domains and protein three-dimensional structure." Nucleic Acids Res **41**(Database issue): 28.

Marco, M., S. Pavan, et al. (2006). "Towards understanding molecular modes of probiotic action." Curr Opin Biotechnol **17**: 204 - 210.

Marcobal, A., A. M. Southwick, et al. (2013). "A refined palate: bacterial consumption of host glycans in the gut." Glycobiology **23**(9): 1038-1046.

Mariat, D., O. Firmesse, et al. (2009). "The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age." BMC Microbiology **9**(1): 123.

Marraffini, L., A. Dedent, et al. (2006). "Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria." Microbiol Mol Biol Rev **70**: 192 - 221.

Martín, R., E. Jiménez, et al. (2009). "Isolation of Bifidobacteria from Breast Milk and Assessment of the Bifidobacterial Population by PCR-Denaturing Gradient Gel Electrophoresis and Quantitative Real-Time PCR." Applied and Environmental Microbiology **75**(4): 965-969.

Martin, R., S. Miquel, et al. (2013). "Role of commensal and probiotic bacteria in human health: a focus on inflammatory bowel disease." Microb Cell Fact **12**(71): 1475-2859.

Martín, R. o., S. Langa, et al. (2004). "The commensal microflora of human milk: new perspectives for food bacteriotherapy and probiotics." Trends in Food Science & Technology **15**(3–4): 121-127.

Marzotto, M., C. Maffeis, et al. (2006). "*Lactobacillus paracasei* A survives gastrointestinal passage and affects the fecal microbiota of healthy infants." Res Microbiol **157**(9): 857-866.

Matamoros, S., C. Gras-Leguen, et al. (2013). "Development of intestinal microbiota in infants and its impact on health." Trends in Microbiology **21**(4): 167-173.

Mattar, A. F., D. H. Teitelbaum, et al. (2002). "Probiotics up-regulate MUC-2 mucin gene expression in a Caco-2 cell-culture model." Pediatr Surg Int **18**(7): 586-590.

Mattarelli, P., C. Bonaparte, et al. (2008). "Proposal to reclassify the three biotypes of Bifidobacterium longum as three subspecies: *Bifidobacterium longum* subsp. *longum* subsp. nov., *Bifidobacterium longum* subsp. *infantis* comb. nov. and *Bifidobacterium longum* subsp. *suis* comb. nov." Int J Syst Evol Microbiol **58**(Pt 4): 767-772.

Maukonen, J., J. Matto, et al. (2008). "Intra-individual diversity and similarity of salivary and faecal microbiota." J Med Microbiol **57**(Pt 12): 1560-1568.

Mayer, E. A., T. Savidge, et al. (2014). "Brain-gut microbiome interactions and functional bowel disorders." Gastroenterology **146**(6): 1500-1512.

Mazmanian, S. K., J. L. Round, et al. (2008). "A microbial symbiosis factor prevents intestinal inflammatory disease." Nature **453**(7195): 620-625.

McGuckin, M. A., S. K. Linden, et al. (2011). "Mucin dynamics and enteric pathogens." Nat Rev Microbiol **9**(4): 265-278.

McMichael, J. C. and J. T. Ou (1979). "Structure of common pili from *Escherichia coli*." J Bacteriol **138**(3): 969-975.

Melican, K., R. M. Sandoval, et al. (2011). "Uropathogenic *Escherichia coli* P and Type 1 Fimbriae Act in Synergy in a Living Host to Facilitate Renal Colonization Leading to Nephron Obstruction." PLoS Pathog **7**(2): e1001298.

Meslin, J. C., N. Fontaine, et al. (1999). "Variation of mucin distribution in the rat intestine, caecum and colon: effect of the bacterial flora." Comp Biochem Physiol A Mol Integr Physiol **123**(3): 235-239.

Metchnikoff, E. (1903). Etudes sur la nature humaine: essai de philosophie optimiste, Masson.

Mierau, I. and M. Kleerebezem (2005). "10 years of the nisin-controlled gene expression system (NICE) in Lactococcus lactis." Appl Microbiol Biotechnol **68**(6): 705-717.

Miller, M. B. and B. L. Bassler (2001). "Quorum sensing in bacteria." Annu Rev Microbiol **55**: 165-199.

Miquel, S., R. Martin, et al. (2013). "*Faecalibacterium prausnitzii* and human intestinal health." Curr Opin Microbiol.

Mishra, A., A. Das, et al. (2007). "Sortase-catalyzed assembly of distinct heteromeric fimbriae in *Actinomyces naeslundii*." J Bacteriol **189**: 3156 - 3165.

Miyoshi, Y., S. Okada, et al. (2006). "A mucus adhesion promoting protein, MapA, mediates the adhesion of *Lactobacillus reuteri* to Caco-2 human intestinal epithelial cells." Biosci Biotechnol Biochem **70**(7): 1622-1628.

Mobili, P., L. Serradell Mde, et al. (2009). "Heterogeneity of S-layer proteins from aggregating and non-aggregating *Lactobacillus kefir* strains." Antonie Van Leeuwenhoek **95**(4): 363-372.

Molloy, Michael J., John R. Grainger, et al. (2013). "Intraluminal Containment of Commensal Outgrowth in the Gut during Infection-Induced Dysbiosis." Cell Host & Microbe **14**(3): 318-328.

Monsan, P., S. Bozonnet, et al. (2001). "Homopolysaccharides from lactic acid bacteria." International Dairy Journal **11**(9): 675-685.

Monsigny, M., C. Petit, et al. (1988). "Colorimetric determination of neutral sugars by a resorcinol sulfuric acid micromethod." Analytical Biochemistry **175**(2): 525-530.

Monteville, M. R., J. E. Yoon, et al. (2003). "Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the CadF outer-membrane protein and microfilament reorganization." Microbiology **149**(Pt 1): 153-165.

Moody, R. G. and M. P. Williamson (2013). "Structure and function of a bacterial Fasciclin I Domain Protein elucidates function of related cell adhesion proteins such as TGFBIp and periostin." FEBS Open Bio **3**(0): 71-77.

Morita, H. (2008). "Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production." DNA Res. **15**: 151-161.

Mueller, C. and A. J. Macpherson (2006). "Layers of mutualism with commensal bacteria protect us from intestinal inflammation." Gut **55**(2): 276-284.

Mukherjee, S., S. Vaishnava, et al. (2008). "Multi-layered regulation of intestinal antimicrobial defense." Cellular and Molecular Life Sciences **65**(19): 3019-3027.

Müller, C. M., A. Åberg, et al. (2009). "Type 1 Fimbriae, a Colonization Factor of Uropathogenic *Escherichia coli*, Are Controlled by the Metabolic Sensor CRP-cAMP." PLoS Pathog **5**(2): e1000303.

Muscas, P., G. M. Rossolini, et al. (1994). "Purification and characterization of type 1 fimbriae of Salmonella typhi." Microbiol Immunol **38**(5): 353-358.

Nachin, L., U. Nannmark, et al. (2005). "Differential roles of the universal stress proteins of Escherichia coli in oxidative stress resistance, adhesion, and motility." J Bacteriol **187**(18): 6265-6272.

Nagaoka, M., S. Hashimoto, et al. (1996). "Structure of a galactan from cell walls of Bifidobacterium catenulatum YIT4016." Carbohydr Res **281**(2): 285-291.

Nagaoka, M., M. Muto, et al. (1990). "Structure of polysaccharide-peptidoglycan complex from the cell wall of Lactobacillus casei YIT9018." J Biochem **108**(4): 568-571.

Nagaoka, M., M. Muto, et al. (1988). "Structure of 6-deoxytalose-containing polysaccharide from the cell wall of Bifidobacterium adolescentis." J Biochem **103**(4): 618-621.

Nagler-Anderson, C. (2001). "Man the barrier! Strategic defences in the intestinal mucosa." Nat Rev Immunol **1**(1): 59-67.

Nallapareddy, S., K. Singh, et al. (2006). "Endocarditis and biofilm-associated pili of *Enterococcus faecalis*." J Clin Invest **116**: 2799 - 2807.

Nallapareddy, S. R., X. Qin, et al. (2000). "*Enterococcus faecalis* adhesin, ace, mediates attachment to extracellular matrix proteins collagen type IV and laminin as well as collagen type I." Infect Immun **68**(9): 5218-5224.

Navarre, W. W. and O. Schneewind (1999). "Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope." Microbiol Mol Biol Rev **63**(1): 174-229.

Neeser, J. R., D. Granato, et al. (2000). "*Lactobacillus johnsonii* La1 shares carbohydrate-binding specificities with several enteropathogenic bacteria." Glycobiology **10**(11): 1193-1199.

Neutra, M. and P. Kozlowski (2006). "Mucosal vaccines: the promise and the challenge." Nat Rev Immunol **6**: 148 - 158.

Neves, B. C., R. K. Shaw, et al. (2003). "Polymorphisms within EspA filaments of enteropathogenic and enterohemorrhagic *Escherichia coli*." Infect Immun **71**(4): 2262-2265.

Nicholson, B. and D. Low (2000). "DNA methylation-dependent regulation of pef expression in Salmonella typhimurium." Mol Microbiol **35**(4): 728-742.

Niemann, H. H., W.-D. Schubert, et al. (2004). "Adhesins and invasins of pathogenic bacteria: a structural view." Microbes and Infection **6**(1): 101-112.

Nimtz, M., A. Mort, et al. (1996). "Structure of amylovoran, the capsular exopolysaccharide from the fire blight pathogen Erwinia amylovora." Carbohydr Res **287**(1): 59-76.

Nishiyama, K., A. Ochiai, et al. (2013). "Identification and Characterization of Sulfated Carbohydrate-Binding Protein from *Lactobacillus reuteri*." PLoS One **8**(12): e83703.

Noda, S., T. Funami, et al. (2008). "Molecular structures of gellan gum imaged with atomic force microscopy in relation to the rheological behavior in aqueous systems. 1. Gellan gum with various acyl contents in the presence and absence of potassium." Food Hydrocolloids **22**(6): 1148-1159.

Noel, C., N. Diaz, et al. (2010). "*Trichomonas vaginalis* vast BspA-like gene family: evidence for functional diversity from structural organisation and transcriptomics." BMC Genomics **11**(1): 99.

Nou, X., B. Skinner, et al. (1993). "Regulation of pyelonephritis-associated pili phase-variation in *Escherichia coli*: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation." Molecular Microbiology **7**(4): 545-553.

O'Connell Motherway, M., J. O'Driscoll, et al. (2009). "Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003." Microbial Biotechnology **2**(3): 321-332.

O'Connell Motherway, M., A. Zomer, et al. (2011). "Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor." Proceedings of the National Academy of Sciences **108**(27): 11217-11222.

O'Flaherty, S. and T. R. Klaenhammer (2010). "The role and potential of probiotic bacteria in the gut, and the communication between gut microflora and gut/host." International Dairy Journal **20**(4): 262-268.

O'Hara, A. M. and F. Shanahan (2007). "Mechanisms of action of probiotics in intestinal diseases." Scientific World J. **7**: 31-46.

O'Hara, A. M. S. F. (2007). "The gut flora as a forgotten organ." EMBO Reports **7**(7): 688-693.

O'Toole, P. W. and M. J. Claesson (2010). "Gut microbiota: Changes throughout the lifespan from infancy to elderly." International Dairy Journal **20**(4): 281-291.

Odenbreit, S., K. Swoboda, et al. (2009). "Outer membrane protein expression profile in Helicobacter pylori clinical isolates." Infect Immun **77**: 3782 - 3790.

Ohtsubo, K. and J. D. Marth (2006). "Glycosylation in Cellular Mechanisms of Health and Disease." Cell **126**(5): 855-867.

Okamoto-Shibayama, K., Y. Sato, et al. (2006). "Identification of a glucan-binding protein C gene homologue in Streptococcus macacae." Oral Microbiol Immunol **21**(1): 32-41.

Otte, J.-M. and D. K. Podolsky (2004). Functional modulation of enterocytes by gram-positive and gram-negative microorganisms.

Ouwerkerk, J. P., W. M. de Vos, et al. (2013). "Glycobiome: Bacteria and mucus at the epithelial interface." Best Practice & Research Clinical Gastroenterology **27**(1): 25-38.

Packey, C. D. and R. B. Sartor (2009). "Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases." Curr Opin Infect Dis **22**(3): 292-301.

Palmer, C., E. M. Bik, et al. (2007). "Development of the human infant intestinal microbiota." PLoS Biol **5**(7): 26.

Palmer, J., S. Flint, et al. (2007). "Bacterial cell attachment, the beginning of a biofilm." J Ind Microbiol Biotechnol **34**(9): 577-588.

Parracho, H. M., M. O. Bingham, et al. (2005). "Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children." Journal of Medical Microbiology **54**(10): 987-991.

Paton, A. W., P. Srimanote, et al. (2001). "Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans." Infect Immun **69**(11): 6999-7009.

Patti, J. M., B. L. Allen, et al. (1994). "MSCRAMM-mediated adherence of microorganisms to host tissues." Annu Rev Microbiol **48**: 585-617.

Paulson, J. C., O. Blixt, et al. (2006). "Sweet spots in functional glycomics." Nat Chem Biol **2**(5): 238-248.

Perez-Vilar, J. (2007). "Mucin granule intraluminal organization." Am J Respir Cell Mol Biol **36**(2): 183-190.

Peter-Katalinic, J. (2005). "Methods in enzymology: O-glycosylation of proteins." Methods Enzymol **405**: 139-171.

Petersen, T. N., S. Brunak, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat Methods. 2011 Sep 29;8(10):785-6. doi: 10.1038/nmeth.1701.

Pizarro-Cerda, J. and P. Cossart (2006). "Bacterial adhesion and entry into host cells." Cell **124**: 715 - 727.

Pooley, N., L. Ghosh, et al. (1995). "Up-regulation of E-selectin and intercellular adhesion molecule-1 differs between Crohn's disease and ulcerative colitis." Digestive Diseases and Sciences **40**(1): 219-225.

Pretzer, G. (2005). "Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*." J. Bacteriol. **187**: 6128-6136.

Pridmore, R. D. (2004). "The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533." Proc. Natl Acad. Sci. USA **101**: 2512-2517.

Prockop, J. D. (1995). "Collagens: Molecular Biology, Diseases, and Potentials for Therapy." Annual Review of Biochemistry **64**(1): 403-434.

Proft, T. and E. N. Baker (2009). "Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease." Cell Mol Life Sci **66**(4): 613-635.

Puente, J. L., D. Bieber, et al. (1996). "The bundle-forming pili of enteropathogenic *Escherichia coli*: transcriptional regulation by environmental signals." Mol Microbiol **20**(1): 87-100.

Qin, J., R. Li, et al. (2010). "A human gut microbial gene catalogue established by metagenomic sequencing." Nature **464**: 59 - U70.

Quigley, B. R., D. Zähner, et al. (2009). "Linkage of T3 and Cpa pilins in the *Streptococcus pyogenes* M3 pilus." Molecular Microbiology **72**(6): 1379-1394.

Rademacher, T. W., R. B. Parekh, et al. (1988). "Glycobiology." Annu Rev Biochem **57**: 785-838.

Raghu, H. and M. J. Flick (2011). "Targeting the coagulation factor fibrinogen for arthritis therapy." Curr Pharm Biotechnol **12**(9): 1497-1506.

Rajilic-Stojanovic, M., H. Smidt, et al. (2007). "Diversity of the human gastrointestinal tract microbiota revisited." Environ. Microbiol. **9**: 2125-2136.

Rakoff-Nahoum, S., J. Paglino, et al. (2004). "Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis." Cell **118**(2): 229-241.

Ramiah, K., C. A. van Reenen, et al. (2007). "Expression of the mucus adhesion genes Mub and MapA, adhesion-like factor EF-Tu and bacteriocin gene plaA of *Lactobacillus plantarum* 423, monitored with real-time PCR." Int J Food Microbiol **116**(3): 405-409.

Rausch, P., A. Rehman, et al. (2011). "Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype." Proc Natl Acad Sci U S A **108**(47): 19030-19035.

Reiff, C. and D. Kelly (2010). "Inflammatory bowel disease, gut bacteria and probiotic therapy." Int J Med Microbiol **300**(1): 25-33.

Rendon, M. A., Z. Saldana, et al. (2007). "Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization." Proceedings of the National Academy of Sciences of the United States of America **104**(25): 10637-10642.

Reunanen, J., I. von Ossowski, et al. (2012). "Characterization of the SpaCBA pilus fibers in the probiotic *Lactobacillus rhamnosus* GG." Appl Environ Microbiol **78**(7): 2337-2344.

Rhee, K.-J., P. Sethupathi, et al. (2004). "Role of Commensal Bacteria in Development of Gut-Associated Lymphoid Tissues and Preimmune Antibody Repertoire." The Journal of Immunology **172**(2): 1118-1124.

Richardson, E. J. and M. Watson (2012). "The automatic annotation of bacterial genomes." Briefings in Bioinformatics.

Rintahaka, J., X. Yu, et al. (2014). "Phenotypical Analysis of the Lactobacillus rhamnosus GG Fimbrial spaFED Operon: Surface Expression and Functional Characterization of Recombinant SpaFED Pili in Lactococcus lactis." PLoS One **9**(11).

Robijn, G. W., H. L. J. Wienk, et al. (1996). "Structural studies of the exopolysaccharide produced by Lactobacillus paracasei 34-1." Carbohydrate Research **285**(0): 129-139.

Robles Alonso, V. and F. Guarner (2013). "Linking the gut microbiota to human health." Br J Nutr **109**(2).

Rogers, E. A., A. Das, et al. (2011). "Adhesion by pathogenic corynebacteria." Adv Exp Med Biol **715**: 91-103.

Rojas, M., F. Ascencio, et al. (2002). "Purification and characterization of a surface protein from *Lactobacillus fermentum* 104R that binds to porcine small intestinal mucus and gastric mucin." Appl Environ Microbiol **68**(5): 2330-2336.

Romer, C. and E. L. (2011). "Galectins: Structures, Binding Properties and Function in Cell Adhesion." Biomaterials - Physics and Chemistry.

Roos, S. and H. Jonsson (2002). "A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components." Microbiology **148**(2): 433-442.

Round, J. L. and S. K. Mazmanian (2009). "The gut microbiota shapes intestinal immune responses during health and disease." Nat Rev Immunol **9**(5): 313-323.

Rousseaux, C., X. Thuru, et al. (2007). "*Lactobacillus acidophilus* modulates intestinal pain and induces opioid and cannabinoid receptors." Nature Medicine **13**: 35 - 37.

Roussel, P. (2005). Airway Glycoconjugates Secreted in Cystic Fibrosis and Severe Chronic Airway Inflammation Relationship with Pseudomonas aeruginosa. Defects of Secretion in Cystic Fibrosis. C. Schultz, Springer US. **558:** 145-167.

Royle, L., T. S. Mattu, et al. (2002). "An analytical and structural database provides a strategy for sequencing O-glycans from microgram quantities of glycoproteins." Analytical Biochemistry **304**(1): 70-90.

Ruas-Madiedo, P. and C. G. de los Reyes-Gavilán (2005). "Invited Review: Methods for the Screening, Isolation, and Characterization of Exopolysaccharides Produced by Lactic Acid Bacteria." Journal of Dairy Science **88**(3): 843-856.

Ruas-Madiedo, P., J. A. Moreno, et al. (2007). "Screening of Exopolysaccharide-Producing *Lactobacillus* and *Bifidobacterium* Strains Isolated from the Human Intestinal Microbiota." Applied and Environmental Microbiology **73**(13): 4385-4388.

Ruas-Madiedo, P., N. Salazar, et al. (2010). Chapter 45 - Exopolysaccharides produced by lactic acid bacteria in food and probiotic applications. Microbial Glycobiology. H. Otto, J. B. Patrick, P. J. B. Mark von ItzsteinA2 - Otto Holst and I. Mark von. San Diego, Academic Press**:** 885-902.

Sabui, S., A. Ghosal, et al. (2010). "Allelic variation in colonization factor CS6 of enterotoxigenic *Escherichia coli* isolated from patients with acute diarrhoea and controls." J Med Microbiol **59**(Pt 7): 770-779.

Sachdeva, G., K. Kumar, et al. (2005). "SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks." Bioinformatics **21**(4): 483-491.

Sartor, R. B. (2008). "Microbial Influences in Inflammatory Bowel Diseases." Gastroenterology **134**(2): 577-594.

Sato, S., C. St-Pierre, et al. (2009). "Galectins in innate immunity: dual functions of host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs)." Immunol Rev **230**(1): 172-187.

Satokari, R., T. Gronroos, et al. (2009). "*Bifidobacterium* and *Lactobacillus* DNA in the human placenta." Lett Appl Microbiol **48**(1): 8-12.

Schubert, A., K. Zakikhany, et al. (2002). "A fibrinogen receptor from group B Streptococcus interacts with fibrinogen by repetitive units with novel ligand binding sites." Mol Microbiol **46**(2): 557-569.

Schulze-Koops, H., H. Burkhardt, et al. (1992). "Plasmid-encoded outer membrane protein YadA mediates specific binding of enteropathogenic yersiniae to various types of collagen." Infect Immun **60**(6): 2153-2159.

Schwan, W. R. (2011). "Regulation of genes in uropathogenic." World J Clin Infect Dis **1**(1): 17-25.

Schwertz, H., G. A. Zimmerman, et al. (2009). "Fibrinogen selects selectins." Blood **114**(2): 234.

Scott, J. R. and D. Zahner (2006). "Pili with strong attachments: Gram-positive bacteria do it differently." Mol Microbiol **62**(2): 320-330.

Sebaihia, M., B. Wren, et al. (2006). "The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome." Nat Genet **38**: 779 - 786.

Sela, D. A., D. Garrido, et al. (2012). "*Bifidobacterium longum* subsp. *infantis* ATCC 15697 α-Fucosidases Are Active on Fucosylated Human Milk Oligosaccharides." Applied and Environmental Microbiology **78**(3): 795-803.

Sela, D. A., Y. Li, et al. (2011). "An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides." J Biol Chem **286**(14): 11909-11918.

Semchenko, E. A., C. J. Day, et al. (2012). "Structural Heterogeneity of Terminal Glycans in *Campylobacter jejuni* Lipooligosaccharides." PLoS One **7**(7): e40920.
Şengül, N., S. Işık, et al. (2011). "The Effect of Exopolysaccharide-Producing Probiotic Strains on Gut Oxidative Damage in Experimental Colitis." Digestive Diseases and Sciences **56**(3): 707-714.
Serafini, F., F. Strati, et al. (2013). "Evaluation of adhesion properties and antibacterial activities of the infant gut commensal *Bifidobacterium bifidum* PRL2010." Anaerobe **21**: 9-17.
Servin, A. L. (2005). "Pathogenesis of Afa/Dr Diffusely Adhering *Escherichia coli*." Clinical Microbiology Reviews **18**(2): 264-292.
Servin, A. L. and M.-H. Coconnier (2003). "Adhesion of probiotic strains to the intestinal mucosa and interaction with pathogens." Best Practice & Research Clinical Gastroenterology **17**(5): 741-754.
Shakhsheer, B., M. Anderson, et al. (2013). "SugarBind database (SugarBindDB): a resource of pathogen lectins and corresponding glycan targets." J Mol Recognit **26**(9): 426-431.
Sharma, A., H. Sojar, et al. (1998). "Cloning, expression, and sequencing of a cell surface antigen containing a leucine-rich repeat motif from Bacteroides forsythus ATCC 43037." Infect Immun **66**: 5703 - 5710.
Shiner, E. K., K. P. Rumbaugh, et al. (2005). "Inter-kingdom signaling: deciphering the language of acyl homoserine lactones." FEMS Microbiol Rev **29**(5): 935-947.
Sillanpaa, J., B. Martinez, et al. (2000). "Characterization of the collagen-binding S-layer protein CbsA of Lactobacillus crispatus." J Bacteriol **182**(22): 6440-6450.
Sillanpaa, J., S. R. Nallapareddy, et al. (2008). "Identification and phenotypic characterization of a second collagen adhesin, Scm, and genome-based identification and analysis of 13 other predicted MSCRAMMs, including four distinct pilus loci, in Enterococcus faecium." Microbiology **154**(Pt 10): 3199-3211.
Silva, M., N. V. Jacobus, et al. (1987). "Antimicrobial substance from a human *Lactobacillus* strain." Antimicrob Agents Chemother **31**(8): 1231-1233.
Singh, N., S. Arioli, et al. (2013). "Impact of *Bifidobacterium bifidum* MIMBb75 on mouse intestinal microorganisms." FEMS Microbiol Ecol **85**(2): 369-375.
Sinkiewicz, G. and E. A. Nordström (2005). "353 Occurrence of *Lactobacillus reuteri*, Lactobacilli and Bifidobacteria in Human Breast Milk." Pediatric Research **58**(2): 415-415.
Sizemore, R. K., J. J. Caldwell, et al. (1990). "Alternate gram staining technique using a fluorescent lectin." Applied and Environmental Microbiology **56**(7): 2245-2247.
Smedley, J. G., 3rd, E. Jewell, et al. (2005). "Influence of pilin glycosylation on Pseudomonas aeruginosa 1244 pilus function." Infect Immun **73**(12): 7922-7931.
Sohel, I., J. L. Puente, et al. (1993). "Cloning and characterization of the bundle-forming pilin gene of enteropathogenic *Escherichia coli* and its distribution in *Salmonella* serotypes." Mol Microbiol **7**(4): 563-575.
Sokol, H., B. Pigneur, et al. (2008). "*Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients." Proc Natl Acad Sci USA **105**: 16731 - 16736.
Sokol, H., P. Seksik, et al. (2009). "Low counts of *Faecalibacterium prausnitzii* in colitis microbiota." Inflamm Bowel Dis **15**: 1183 - 1189.
Sonnenburg, J. L., J. Xu, et al. (2005). "Glycan foraging in vivo by an intestine-adapted bacterial symbiont." Science **307**(5717): 1955-1959.
Spellerberg, B., E. Rozdzinski, et al. (1999). "Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of *Streptococcus agalactiae* to human laminin." Infect Immun **67**(2): 871-878.
Spiro, R. G. (1967). "The Structure of the Disaccharide Unit of the Renal Glomerular Basement Membrane." Journal of Biological Chemistry **242**(20): 4813-4823.
Spraggon, G., E. Koesema, et al. (2010). "Supramolecular Organization of the Repetitive Backbone Unit of the *Streptococcus pneumoniae* Pilus." PLoS One **5**(6): e10919.
Steinbrecher, K. A., N. A. Horowitz, et al. (2010). "Colitis-associated cancer is dependent on the interplay between the hemostatic and inflammatory systems and supported by integrin alpha(M)beta(2) engagement of fibrinogen." Cancer Res **70**(7): 2634-2643.

Styriak, I. and R. Nemcova (2003). "Lectin-like binding of lactobacilli considered for their use in probiotical preparations for animal use." Berl Munch Tierarztl Wochenschr **116**(3-4): 96-101.

Sultana, R., A. J. McBain, et al. (2013). "Strain-Dependent Augmentation of Tight-Junction Barrier Function in Human Primary Epidermal Keratinocytes by *Lactobacillus* and *Bifidobacterium* Lysates." Applied and Environmental Microbiology **79**(16): 4887-4894.

Sutak, R., E. Lesuisse, et al. (2008). "Crusade for iron: iron uptake in unicellular eukaryotes and its significance for virulence." Trends Microbiol **16**: 261 - 268.

Sutherland, I. (2001). "Biofilm exopolysaccharides: a strong and sticky framework." Microbiology **147**(Pt 1): 3-9.

Sutherland, I. W. (1994). "Structure-function relationships in microbial exopolysaccharides." Biotechnology Advances **12**(2): 393-448.

Svenson, S. B., G. Källenius, et al. (1983). "P-fimbriae of pyelonephritogenic *Escherichia coli*: Identification and chemical characterization of receptors." Infection **11**(1): 61-67.

Szeri, I., P. Anderlik, et al. (1976). "Decreased cellular immune response of germ-free mice." Acta Microbiol Acad Sci Hung **23**(3): 231-234.

Tallon, R., P. Bressollier, et al. (2003). "Isolation and characterization of two exopolysaccharides produced by *Lactobacillus plantarum* EP56." Research in Microbiology **154**(10): 705-712.

Tarr, P. I., S. S. Bilge, et al. (2000). "Iha: a novel *Escherichia coli* O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure." Infect Immun **68**(3): 1400-1407.

Tarsi, R. and C. Pruzzo (1999). "Role of surface proteins in *Vibrio cholerae* attachment to chitin." Appl Environ Microbiol **65**(3): 1348-1351.

Telford, J., M. Barocchi, et al. (2006). "Pili in gram-positive pathogens." Nat Rev Microbiol **4**: 509 - 519.

Tiihonen, K., A. C. Ouwehand, et al. (2010). "Human intestinal microbiota and healthy ageing." Ageing Res Rev **9**(2): 107-116.

Tlaskalova-Hogenova, H., R. Stepankova, et al. (2011). "The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases." Cell Mol Immunol **8**(2): 110-120.

Tom, B. H., L. P. Rutzky, et al. (1976). "Human colonic adenocarcinoma cells. I. Establishment and description of a new line." In Vitro **12**(3): 180-191.

Tomaras, A. P., M. J. Flagler, et al. (2008). "Characterization of a two-component regulatory system from *Acinetobacter baumannii* that controls biofilm formation and cellular morphology." Microbiology **154**(Pt 11): 3398-3409.

Ton-That, H., L. A. Marraffini, et al. (2004). "Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*." Molecular Microbiology **53**(1): 251-261.

Tone-Shimokawa, Y., T. Toida, et al. (1996). "Isolation and structural analysis of polysaccharide containing galactofuranose from the cell walls of Bifidobacterium infantis." Journal of Bacteriology **178**(1): 317-320.

Topin, J., J. Arnaud, et al. (2013). "Deciphering the Glycan Preference of Bacterial Lectins by Glycan Array and Molecular Docking with Validation by Microcalorimetry and Crystallography." PLoS One **8**(8): e71149.

Tsai, J. C., M.-R. Yen, et al. (2010). "The Bacterial Intimins and Invasins: A Large and Novel Family of Secreted Proteins." PLoS One **5**(12): e14403.

Tschop, M. H., P. Hugenholtz, et al. (2009). "Getting to the core of the gut microbiome." Nat Biotechnol **27**(4): 344-346.

Turnbaugh, P., M. Hamady, et al. (2009). "A core gut microbiome in obese and lean twins." Nature **457**: 480 - 484.

Turnbaugh, P., R. Ley, et al. (2007). "The human microbiome project." Nature **449**: 804 - 810.

Turner, A., S.-N. Chen, et al. (2005). "Inhibition of Uropathogenic *Escherichia coli* by Cranberry Juice: A New Antiadherence Assay." Journal of Agricultural and Food Chemistry **53**(23): 8940-8947.

Turroni, F., F. Bottacini, et al. (2010). "Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging." Proc Natl Acad Sci U S A **107**: 19514 - 19519.

Turroni, F., A. Ribbera, et al. (2008). "Human gut microbiota and bifidobacteria: from composition to functionality." Antonie Van Leeuwenhoek **94**: 35 - 50.

Turroni, F., F. Serafini, et al. (2013). "Role of sortase-dependent pili of *Bifidobacterium bifidum* PRL2010 in modulating bacterium-host interactions." Proc Natl Acad Sci U S A **110**(27): 11151-11156.

Uchida, H., H. Kinoshita, et al. (2006). "Lactobacilli binding human A-antigen expressed in intestinal mucosa." Research in Microbiology **157**(7): 659-665.

Umesaki, Y., H. Setoyama, et al. (1993). "Expansion of alpha beta T-cell receptor-bearing intestinal intraepithelial lymphocytes after microbial colonization in germ-free mice and its independence from thymus." Immunology **79**(1): 32-37.

Uribe, A., M. Alam, et al. (1994). "Microflora modulates endocrine cells in the gastrointestinal mucosa of the rat." Gastroenterology **107**(5): 1259-1269.

Vaishnava, S., M. Yamamoto, et al. (2011). "The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine." Science **334**(6053): 255-258.

Van Calsteren, M. R., C. Pau-Roblot, et al. (2002). "Structure determination of the exopolysaccharide produced by Lactobacillus rhamnosus strains RW-9595M and R." Biochem J **363**(Pt 1): 7-17.

Van de Bovenkamp, J. H., J. Mahdavi, et al. (2003). "The MUC5AC glycoprotein is the primary receptor for Helicobacter pylori in the human stomach." Helicobacter **8**(5): 521-532.

Van der Meer, J. R., J. Polman, et al. (1993). "Characterization of the Lactococcus lactis nisin A operon genes nisP, encoding a subtilisin-like serine protease involved in precursor processing, and nisR, encoding a regulatory protein involved in nisin biosynthesis." Journal of Bacteriology **175**(9): 2578-2588.

Van der Waaij, D., J. M. Berghuis-de Vries, et al. (1971). "Colonization resistance of the digestive tract in conventional and antibiotic-treated mice." Epidemiology & Infection **69**(03): 405-411.

Van Leeuwen, S. S., S. Kralj, et al. (2008). "Structural analysis of the alpha-D-glucan (EPS35-5) produced by the Lactobacillus reuteri strain 35-5 glucansucrase GTFA enzyme." Carbohydrate Research **343**(7): 1251-1265.

Van Leeuwen, S. S., S. Kralj, et al. (2008). "Structural analysis of the α-d-glucan (EPS180) produced by the Lactobacillus reuteri strain 180 glucansucrase GTF180 enzyme." Carbohydrate Research **343**(7): 1237-1250.

Van Passel, M. W., R. Kant, et al. (2011). "The genome of Akkermansia muciniphila, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes." PLoS One **6**(3): 0016876.

Van Tassell, M. L. and M. J. Miller (2011). "*Lactobacillus* adhesion to mucus." Nutrients **3**(5): 613-636.

Varki A, C. R., Esko J, et al., editors. (1999). " N-Glycans, Chapter;." Essentials of Glycobiology **2**d edition.

Ventura, M. (2007). "Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum." Microbiol. Mol. Biol. Rev. **71**: 495-548.

Ventura, M., F. Turroni, et al. (2012). "Host-microbe interactions that facilitate gut colonization by commensal bifidobacteria." Trends Microbiol **20**(10): 467-476.

Vlkova, E., V. Rada, et al. (2008). "Auto-aggregation and co-aggregation ability in bifidobacteria and clostridia." Folia Microbiol **53**(3): 263-269.

Vollaard, E. J. and H. A. Clasener (1994). "Colonization resistance." Antimicrobial Agents and Chemotherapy **38**(3): 409-414.

von Ossowski, I., J. Reunanen, et al. (2010). "Mucosal Adhesion Properties of the Probiotic *Lactobacillus rhamnosus* GG SpaCBA and SpaFED Pilin Subunits." Applied and Environmental Microbiology **76**(7): 2049-2057.

von Ossowski, I., R. Satokari, et al. (2011). "Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic *Lactobacillus rhamnosus* GG." Appl Environ Microbiol **77**(13): 4465-4472.
Vrieze, A., F. Holleman, et al. (2010). "The environment within: how gut microbiota may influence metabolism and body composition." Diabetologia **53**(4): 606-613.
Wacklin, P., H. Makivuokko, et al. (2011). "Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine." PLoS One **6**(5): 19.
Wacklin, P., J. Tuimala, et al. (2014). "Faecal Microbiota Composition in Adults Is Associated with the FUT2 Gene Determining the Secretor Status." PLoS One **9**(4).
Waksman, G. and S. J. Hultgren (2009). "Structural biology of the chaperone-usher pathway of pilus biogenesis." Nat Rev Microbiol **7**(11): 765-774.
Waligora, A.-J., M.-C. Barc, et al. (1999). "*Clostridium difficile* Cell Attachment Is Modified by Environmental Factors." Applied and Environmental Microbiology **65**(9): 4234-4238.
Wang, J. F., Y. H. Zhu, et al. (2004). "In vitro fermentation of various fiber and starch sources by pig fecal inocula." Journal of Animal Science **82**(9): 2615-2622.
Wang, J. Y. and M. H. Roehrl (2002). "Glycosaminoglycans are a potential cause of rheumatoid arthritis." Proceedings of the National Academy of Sciences **99**(22): 14362-14367.
Wang, K.-C., Y.-H. Hsu, et al. (2012). "A previously uncharacterized gene stm0551 plays a repressive role in the regulation of type 1 fimbriae in *Salmonella enterica* serotype Typhimurium." BMC Microbiology **12**(1): 111.
Ward, R. E., M. Ninonuevo, et al. (2006). "In vitro fermentation of breast milk oligosaccharides by *Bifidobacterium infantis* and *Lactobacillus gasseri*." Appl Environ Microbiol **72**(6): 4497-4499.
Watanabe, Y., H. Tateno, et al. (2009). "The function of rhamnose-binding lectin in innate immunity by restricted binding to Gb3." Developmental & Comparative Immunology **33**(2): 187-197.
Weintraub, A. (2007). "Enteroaggregative *Escherichia coli*: epidemiology, virulence and detection." Journal of Medical Microbiology **56**(1): 4-8.
Wells, C. L., E. A. Moore, et al. (2000). "Inducible expression of *Enterococcus faecalis* aggregation substance surface protein facilitates bacterial internalization by cultured enterocytes." Infect Immun **68**(12): 7190-7194.
Wells, J. M., O. Rossi, et al. (2011). "Epithelial crosstalk at the microbiota–mucosal interface." Proceedings of the National Academy of Sciences **108**(Supplement 1): 4607-4614.
Whitfield, C., J. C. Richards, et al. (1991). "Expression of two structurally distinct D-galactan O antigens in the lipopolysaccharide of *Klebsiella pneumoniae* serotype O1." J Bacteriol **173**(4): 1420-1431.
Wiles, T. J., R. R. Kulesus, et al. (2008). "Origins and virulence mechanisms of uropathogenic *Escherichia coli*." Experimental and Molecular Pathology **85**(1): 11-19.
Wilson, I. G. (1997). "Inhibition and facilitation of nucleic acid amplification." Appl Environ Microbiol **63**(10): 3741-3751.
Wold, A. E., M. Thorssen, et al. (1988). "Attachment of *Escherichia coli* via mannose- or Gal alpha 1----4Gal beta-containing receptors to human colonic epithelial cells." Infect Immun **56**(10): 2531-2537.
Wolf, D. M. and A. P. Arkin (2002). "Fifteen minutes of fim: control of type 1 pili expression in *E. coli*." Omics **6**(1): 91-114.
Wong, J. M., R. de Souza, et al. (2006). "Colonic health: fermentation and short chain fatty acids." J Clin Gastroenterol **40**(3): 235-243.
Wright, K. J., P. C. Seed, et al. (2007). "Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili." Cellular Microbiology **9**(9): 2230-2241.
Wrzosek, L., S. Miquel, et al. (2013). "*Bacteroides thetaiotaomicron* and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent." BMC Biology **11**(1): 61.
Wurpel, D. J., S. A. Beatson, et al. (2013). "Chaperone-Usher Fimbriae of *Escherichia coli*." PLoS One **8**(1): e52835.

Xicohtencatl-Cortes, J., V. Monteiro-Neto, et al. (2009). "The Type 4 Pili of Enterohemorrhagic *Escherichia coli* O157:H7 Are Multipurpose Structures with Pathogenic Attributes." Journal of Bacteriology **191**(1): 411-421.

Xu, J., M. Bjursell, et al. (2003). "A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis." Science **299**: 2074 - 2076.

Xu, Q., B. Christen, et al. (2012). "Structure of the pilus assembly protein TadZ from *Eubacterium rectale*: Implications for polar localization." Molecular Microbiology **83**(4): 712-727.

Yamamoto, K., T. Tsuji, et al. (1981). "Structural requirements for the binding of oligosaccharides and glycopeptides to immobilized wheat germ agglutinin." Biochemistry **20**(20): 5894-5899.

Yang, H., S. Lang, et al. (2009). "Fibrinogen is required for maintenance of platelet intracellular and cell-surface P-selectin expression." Blood **114**(2): 425-436.

Yasuda, E., H. Tateno, et al. (2011). "Lectin microarray reveals binding profiles of *Lactobacillus casei* strains in a comprehensive analysis of bacterial cell wall polysaccharides." Appl Environ Microbiol **77**(13): 4539-4546.

Yoshida, E., H. Sakurama, et al. (2012). "*Bifidobacterium longum* subsp. *infantis* uses two different beta-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides." Glycobiology **22**(3): 361-368.

Youakim, A. and A. Herscovics (1985). "Cell surface glycopeptides from human intestinal epithelial cell lines derived from normal colon and colon adenocarcinomas." Cancer Res **45**(11 Pt 1): 5505-5511.

Yousef, F. and M. Espinosa-Urgel (2007). "In silico analysis of large microbial surface proteins." Res Microbiol **158**(6): 545-550.

Zarbock, A., K. Ley, et al. (2011). "Leukocyte ligands for endothelial selectins: specialized glycoconjugates that mediate rolling and signaling under flow." Blood **118**(26): 6743-6751.

Zeng, C. and K. Biemann (1999). "Determination of N-linked glycosylation of yeast external invertase by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." J Mass Spectrom **34**(4): 311-329.

Zhang, H., Y. Wang, et al. (2013). "Safety evaluation of *Lactobacillus paracasei* subsp. *paracasei* LC-01, a probiotic bacterium." J Microbiol **51**(5): 633-638.

Zhao, M., J. Sillanpaa, et al. (2009). "Adherence to host extracellular matrix and serum components by *Enterococcus faecium* isolates of diverse origin." Fems Microbiology Letters **301**(1): 77-83.

Zhu, J., M. B. Miller, et al. (2002). "Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*." Proc Natl Acad Sci U S A **99**(5): 3129-3134.

Zivkovic, A. M., J. B. German, et al. (2010). "Human milk glycobiome and its impact on the infant gastrointestinal microbiota." Proceedings of the National Academy of Sciences.

Zdorovenko, E. L., V. V. Kachala, et al. (2009). "Structure of the cell wall polysaccharides of probiotic bifidobacteria Bifidobacteriumbifidum BIM B-465." Carbohydr Res **344**(17): 2417-2420.

Appendix

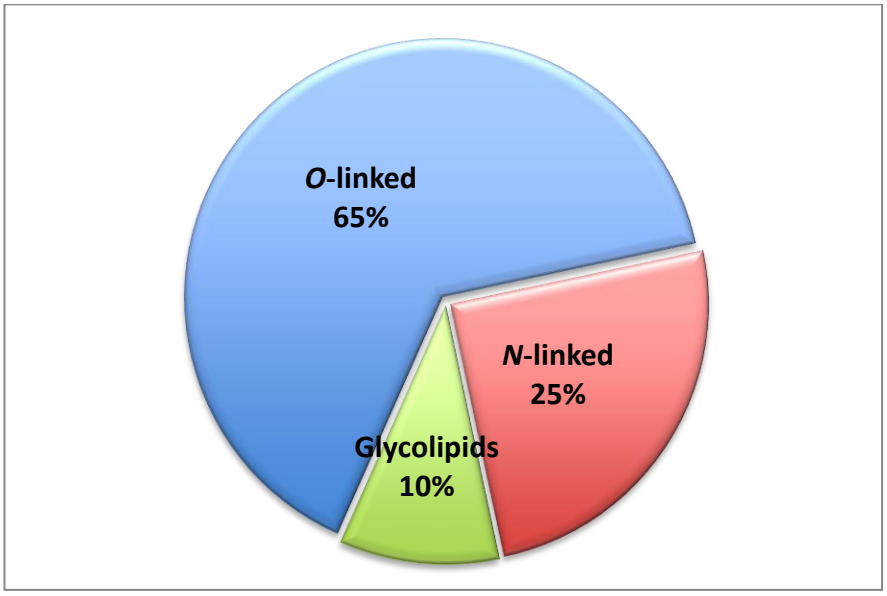

Figure 1A: Families of glycans found in the human GIT

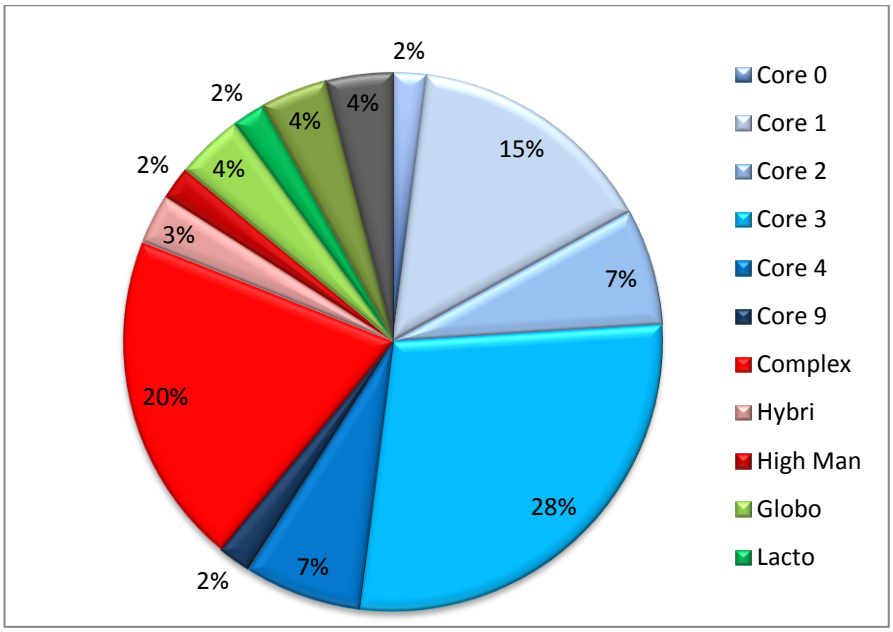

Figure 2A: Sub-families of glycans found in the human GIT

Table 1A: List of glycans found in the GIT

Glycans from the GIT
Fucα(1→2)Galβ(1→3)[Galβ(1→6)]GalNAc
Fucα(1→2)Galβ(1→3/4)GlcNAcβ(1→3)Galβ(1→3)GalNAc
Fucα(1→2)Galβ(1→3)[NeuAcα(2→6)]GalNAc
Fucα(1→2)Galβ(1→3/4)[Fucα(1→4/3)]GlcNAcβ(1→3/6)[Fucα(1→2)Galβ(1→3/4) GlcNAcβ(1→6/3)]Galβ(1→3)[Galβ(1→4)GlcNAcβ(1→6)]GalNAc
Fucα(1→2)Galβ(1→3/4)[Fucα(1→4/3)]GlcNAcβ(1→3)Galβ(1→3)[GlcNacβ(1→6)] GalNAc
Fucα(1→2)Galβ(1→3/4)GlcNAcβ(1→3/6)[Fucα(1→2)Galβ(1→3/4)[Fucα(1→4/3)] GlcNAcβ(1→6/3)]Galβ(1→3/4)GlcNAcβ(1→3)Galβ(1→3)[Galβ(1→6)]GalNAc
Fucα(1→2)Galβ(1→3/4)GlcNAcβ(1→3/6)[Fucα(1→2)Galβ(1→3/4)GlcNacβ(1→6/3)] Galβ(1→3)[Galβ(1→4)GlcNAcβ(1→6)]GalNAc
Fucα(1→2)Galβ(1→3/4)GlcNAcβ(1→3/6)[Fucα(1→2)Galβ(1→3/4)GlcNacβ(1→6/3)] Galβ(1→3)[GlcNAcβ(1→6)]GalNAc
Fucα(1→2)Galβ(1→3/4)GlcNAcβ(1→3/6)[Fucα(1→2)Galβ(1→3/4)GlcNacβ(1→6/3)] Galβ(1→3/4)GlcNAcβ(1→3)Galβ(1→3)GalNAc
Fucα(1→2)Galβ(1→3/4)GlcNAcβ(1→3/6)[Fucα(1→2)Galβ(1→3/4)GlcNacβ(1→6/3)] Galβ(1→3)GalNAc
Fucα(1→2)Galβ(1→3/4)GlcNAcβ(1→3)Galβ(1→3)[GlcNAcβ(1→6)]GalNAc
Galβ(1→3)GalNAcα
Galβ(1→3)GlcNAcβ(1→3)Galβ(1→3)[Fucα(1→4)]GlcNAcβ(1→3)Galβ(1→4)Glcβ
Galβ(1→3)GlcNAcβ(1→3)Galβ(1→3)GlcNAcβ(1→3)Galβ(1→4)Glcβ
Galβ(1→4)GlcNAcβ(1→2)[Galβ(1→4)GlcNAcβ(1→4)]Manα(1→3)[Galβ(1→ 4)GlcNAcβ(1→2) Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)[Fucα(1→6)]GlcNAc
Galβ(1→4GlcNAcβ(1→3)[GlcNAcβ(1→6]Galβ(1→4)GlcNAcβ(1→3)[Neu5Acα(2→6)]GalNAc
Galβ(1→4)GlcNAcβ(1→3)Galβ(1→4)GlcNAcβ(1→3)GalNAc
Galα(1→3)Galβ(1→4)GlcNAcβ(1→2Manα(1→6)Man
Galβ(1→3)[Fucα(1→4)]GlcNAcβ(1→3)Galβ(1→3)GalNAc
Galβ(1→3)[Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→6)]GalNAc
Galβ(1→3)[GlcNAcβ(1→6)]GalNAc
Galβ(1→3)[NeuAcα(2→6)]GalNAc
Galβ(1→3)[NeuAcα(2→6)]GalNAcα
Galβ(1→3)GalNAc
Galβ(1→3)GlcNAcβ(1→3)[Galβ(1→4)GlcNAcβ(1→6)]GalNAc
Galβ(1→3)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
Galβ(1→3)GlcNAcβ(1→3)Galβ(1→4)Glc
Galβ(1→3)GlcNAcβ(1→3)GalNAc
Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→2Man
Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→3)Galβ(1→3)GalNAc
Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→3)Galβ(1→4)Glc
Galβ(1→4)[NeuAcα(2→6)]GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
Galβ(1→4)[Galβ(1→4)GlcNAcβ(1→2Manα(1→3)[GlcNAcβ(1→2)GlcNAcβ(1→6)Manα(1→6)] Manβ(1→4)GlcNAcβ(1→4)[Fucα(1→6)]GlcNAc
Galβ(1→4)GlcNAcβ(1→2)[Galβ(1→4)GlcNAcβ(1→4)]Manα(1→3)[Galβ(1→ 4)GlcNAcβ(1→2) [Galβ(1→4)GlcNAcβ(1→6)]Manα(1→6)]Manβ(1→4)GlcNacβ(1→4)[Fucα(1→6)]GlcNAc
Galβ(1→4)GlcNAcβ(1→2)Manα(1→3)[Galβ(1→4)GlcNAcβ(1→2)[Galβ(1→4) GlcNAcβ(1→4)] Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)GlcNAc

Galβ(1→4)GlcNAcβ(1→2)Manα(1→3)[Galβ(1→4)GlcNAcβ(1→2)Manα(1→6)] Manβ(1→4)GlcNAcβ(1→4)[Fucα(1→6)]GlcNAc
Galβ(1→4)GlcNAcβ(1→2)Manα(1→3)[Galβ(1→4)GlcNAcβ(1→2)Manα(1→6)]Manβ(1→4) GlcNAcβ(1→4)GlcNAc
Galβ(1→4)GlcNAcβ(1→2)Manα(1→3)[Manα(1→3)[Manα(1→6)]Manα(1→6)]Manβ(1→4) GlcNAcβ(1→4)GlcNAc
Galβ(1→4)GlcNAcβ(1→2)Manα(1→3)[Manα(1→3)Manα(1→6)]Manβ(1→4)GlcNacβ(1→4) GlcNAc
Galβ(1→4)GlcNAcβ(1→2)Manα(1→3)[Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)[Fucα(1→6)] GlcNAc
Galβ(1→4)GlcNAcβ(1→3)[Galβ(1→4)GlcNAcβ(1→6)]Galβ(1→4)GlcNAcβ(1→3)GalNAc
Galβ(1→4)GlcNAcβ(1→3)[Galβ(1→4)GlcNAcβ(1→6)]GalNAc
Galβ(1→4)GlcNAcβ(1→3)[GlcNAcβ(1→6)]Galβ(1→4)GlcNAcβ(1→3)GalNAc
Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
Galβ(1→4)GlcNAcβ(1→3)Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
Galβ(1→4)GlcNAcβ(1→3)GalNAc
GalNAc
GalNAc[3S]β(1→4)Galβ(1→4)Glcβ
GalNAc[4S]β(1→4)GlcNAcβ(1→2)Manα
GalNAcα(1→3)[Fucα(1→2)]Galβ(1→3)[Fucα(1→4)]GlcNAcβ(1→3)Galβ(1 →4)Glcβ
GalNAcα(1→3)[Fucα(1→2)]Galβ(1→3)/4)GlcNAcβ(1→3)[NeuAcα(2→6)]Galβ(1→4) GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
GalNAcα(1→3)[Fucα(1→2)]Galβ(1→3)GlcNAcβ(1→3)[GalNAcα(1→3)[Neu5Acα(2→6)] Galβ(1→3)GlcNAcβ(1→6)]Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
GalNAcα(1→3)[Fucα(1→2]Galβ(1→4)GlcNAcα(1→3)Galβ(1→4)GlcNAcβ(1→3)GalNAc
GalNAcα(1→3)[Fucα(1→2)]Galβ(1→4)GlcNAcβ(1→3)[Galβ(1→4)GlcNAc β(1→6)] Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
GalNAcα(1→3)[Fucα(1→2)]Galβ(1→4)GlcNAcβ(1→3)[GalNAcα(1→3)Gal β(1→4) GlcNAcβ(1→6)]Galβ(1→4)GlcNAcβ(1→3)GalNAc
GalNAcα(1→3)[Fucα(1→2)]Galβ(1→4)GlcNAcβ(1→3)Galβ(1→4)GlcNAcβ(1→3) Galβ(1→4)GlcNAcβ(1→3)GalNAc
GalNAcα(1→3)Galβ(1→3)GlcNAcβ(1→3)[NeuAcα(2→6)Galβ(1→4)GlcNAc β(1→6)] Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
GalNAcα(1→3)Galβ(1→4)GlcNAcβ(1→3)[GalNAcα(1→3)Galβ(1→4) GlcNα(cβ(1→6)]Galβ(1→4)GlcNAcβ(1→3)GalNAc
GalNAcα(1→3)Galβ(1→4)GlcNAcβ(1→3)[GlcNAcβ(1→6)]Galβ(1→4)GlcNacβ(1→3)GalNAc
GalNAcα(1→6)Galβ(1→3)GlcNAcβ(1→3)[Galβ(1→4)GlcNAcβ(1→6)]Galβ(1→4) GlcNAcβ(1→3)GalNAc
GalNAcβ(1→3)Galα(1→4)Galβ(1→4)Glc
GlcNAc
GlcNAcβ(1→2[GlcNAcβ(1→6)]Manα(1→3)[GlcNAcβ(1→2Manα(1→6)]Manβ(1→4) GlcNAcβ(1→4)[Fucα(1→6)]GlcNAc
GlcNAcβ(1→3)[Fucα(1→2Galβ(1→4)GlcNAcβ(1→6)]GalNAc
GlcNAcβ(1→3)[Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→6)]GalNAc
GlcNAcβ(1→3)[Galβ(1→4)GlcNAcβ(1→6)]GalNAc
GlcNAcβ(1→3)[GlcNAcβ(1→6)]GalNAc
GlcNAcβ(1→3)[GlcNAcβ(1→6)]GalNAc
GlcNAcβ(1→3)[NeuAcα(2→6)]Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc

GlcNAcβ(1→3)Galα(1→4)Galβ(1→4)Glcβ
GlcNAcβ(1→3)Galβ(1→3)[NeuAcα(2→6)]GalNAc
GlcNAcβ(1→3)Galβ(1→3)GalNAc
GlcNAcβ(1→3)Galβ(1→4)GlcNAcβ(1→3)GalNAc
GlcNAcβ(1→3)GalNAc
GlcNAcβ(1→4)[NeuAcα(2→6)Galβ(1→4)GlcNAcβ(1→2Manα(1→3)]Manα(1→3) [Manα(1→6)]Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)[Fucα(1→6)]GlcNAc
GlcNAcβ(1→6)Manα(1→6)Man
Manα(1→3)[GlcNAcβ(1→6)Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)GlcNAc
Manα(1→3)[Manα(1→3)[Manα(1→6)]Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)GlcNAc
Manα(1→3)[Manα(1→3)Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)GlcNAc
Manα(1→3)[Manα(1→6)]Manβ(1→4)GlcNAcβ(1→4)GlcNAc
Manα(1→3)Manβ(1→4)GlcNAcβ(1→4)[Fucα(1→6)]GlcNAc
NeuAcα(2→3)Galβ(1→3)[NeuAcα(2→6)]GalNAc
NeuAcα(2→3)Galβ(1→3)GalNAcβ(1→4)[NeuAcα(2→3)]Galβ(1→4)Glcβ
NeuAcα(2→3)Galβ(1→3)GalNAcβ(1→4)[NeuAcα(2→8)][NeuAcα(2→3)]Galβ(1→4)Glcβ
NeuAcα(2→3)Galβ(1→4)GlcNAcβ(1→3)Gal
NeuAcα(2→6)Galβ(1→3)GlcNAcβ(1→3)[NeuAcα(2→6)]Galβ(1→4)GlcNAc β(1→3) [NeuAcα(2→6)]GalNAc
NeuAcα(2→6)Galβ(1→3)GlcNAcβ(1→3)[NeuAcα(2→6)Galβ(1→3)GlcNAcβ(1→6)] Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
NeuAcα(2→6)Galβ(1→4)GlcNAcβ(1→3)[NeuAcα(2→6)]GalNAc
NeuAcα(2→6)Galβ(1→4)GlcNAcβ(1→3)Gal
NeuAcα(2→6)GalNAc
NeuAcα(2→6)GalNAcβ(1→4)GlcNAcβ(?)Manα(1→3)
NeuAcα(2→3)Galβ(1→3)GalNAcα
NeuAcα(2→3)Galβ(1→4)Glcβ

Data were compiled from databases of the Consortium for functional glycomics (www.functionalglycomics.org)